

Overconfidence and Health Insurance Participation among the Elderly*

Wei Huang and Mi Luo[†]

Abstract

People may have imperfect information about their health status and thus make suboptimal decisions in insurance participation. Using national representative samples of the elderly in US and China, we find that people with lower socio-economic status and poorer health are relatively less likely to realize how unhealthy they are and this overconfidence is associated with no insurance participation. Accurate health information provided through physical examinations induces relatively higher participation among the overconfident people afterwards. These findings contribute a new explanation for the insufficient participation and advantageous selection in health insurance, and provide new insights on the insurance market and policy suggestions. (*JEL* codes: I12, I13, J14)

Keywords: Overconfidence, Health, Health Insurance Participation

*We term “overconfidence” as the observation of individuals perceiving their health better than it actually is. We introduce the overconfidence as individuals having imperfect information about their own health risk type. We follow this name in Sandroni and Squintani (2007). Overconfidence here may be just due to inattention or unawareness and not “too confident” in real life.

[†]Huang: (E-mail: weihuang@fas.harvard.edu) Department of Economics, Harvard University, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138. Luo: (E-mail: mi.luo@nyu.edu) Department of Economics, New York University, 19 West 4th Street, 6 FL, New York, NY 10012. We thank Amitabh Chandra, John Cawley, David Cutler, Hanming Fang, Amy Finkelstein, Edward Glaeser, Christopher Jencks, Lawrence Katz, Kathleen McGarry, Joseph Newhouse, Matthew D. Shapiro, Mark Shepard, Theda Skocpol and John Strauss for their helpful suggestions and encouragement. We are grateful to the participants in Proseminar on Inequality and Social Policy at Harvard Kennedy School, NBER Health and Aging lunch, and HCP Health Economics seminar at Harvard Medical School. Wei Huang gratefully acknowledges financial support from NBER Pre-doctoral Fellowship and Harvard University Jeanne Humphrey Block Dissertation Award. All errors are ours.

I. Introduction

The textbook model of health insurance and subsequent literature have long assumed that people have perfect private information about their own health and have emphasized the asymmetric information between the demand and supply side of insurance (Rothschild and Stiglitz, 1976; Wilson, 1977). Although the individual subjective bias or inconsistency between perceived and actual health has been known by many researchers,¹ few have studied the possible bias in the perceived health and the corresponding consequences.²

This paper fills the gap by investigating whether the perceived health is biased from the actual health status and, if yes, how this subjective bias affects individual insurance participation.³ We build a simple theoretical framework to incorporate the subjective bias. Assuming that individual behaviors are based on the perceived rather than actual health status, we find that 1) people may insufficiently participate in health insurance if perceived health are better than actual, and 2) the insurance participation may be either positively or negatively correlated with actual health status. If those with poorer health are relatively less likely to realize how unhealthy they are, the important prediction of the classic model - adverse selection - may be weakened or even reversed to the opposite direction, i.e. advantageous selection.

Measuring the subjective bias is a prerequisite for our ends but a difficult one. An ideal measure would be derived from comparison between a subjective measure for health and its

¹For example, Cawley and Philipson (1999) found a weak correlation of self-perceived risk of death with the *ex post* actual mortality for the US seniors; Barseghyan et al. (2013) documented the probability distortions (i.e. substantial overweighting of small probabilities and only mild insensitivity to probability changes) are important to explain the risk aversion in deductible choices; Spinnewijn (2015) analyzed how the willingness to pay for insurance to be biased in presence of heterogeneous demand frictions caused by differences between perceived and actual risk.

²There are some exceptions. For example, Sandroni and Squintani (2007), studied the insurance market equilibrium consequences with overconfidence in perceived health in a theoretical framework but did not provide empirical evidence; Abaluck and Gruber (2011) documented the inconsistency of the individual insurance options under the Medicare Part D Prescription Drug plan with prescription drug claims but it is unclear whether and how the irrationality is relevant to the difference between perceived and actual health.

³Different from the demand frictions in previous literature (e.g. Spinnewijn, 2015) which usually assumes that the frictions are independently distributed with zero mean, we study the systematic or mean difference between perceived and actual health.

genuinely objective counterpart: the difference between the two would define the subjective bias.⁴ This paper uses two independent measures. The first one is the unrecognized height shrinkage.⁵ We use this measure for seniors aged over 55 in the US from the National Health and Nutrition Examination Survey (NHANES) dataset by the Centers for Disease Control (CDC). This survey is unique in providing the necessary information to calculate both the subjective and the biological height shrinkage. The second measure is under-diagnosed hypertension. We use this measure mainly for a national representative sample of people aged over 45 from the China Health and Retirement Longitudinal Studies (CHARLS). This survey contains comprehensive measures for both subjective and objective hypertension status. More importantly, the respondents were informed of their blood pressure from the physical examinations during the first wave and their insurance status were also collected in the following wave, giving us a unique opportunity to identify the effects of better health information on insurance participation. Apart from the fact that height shrinkage and hypertension are more prevalent among seniors, we use the elderly sample in both countries because they are the most intensive consumers of health care in both countries. In the US, for example, individuals over 65 consume 36 percent of health care (Centers for Medicaid and Medicare Services, 2005; Chandra et al., 2010).

We term “overconfidence” as the observation of individuals perceiving their health better than it actually is.⁶ Our results show that the majority of the elderly population are overconfident about their health. Seniors in the US realize merely half of their height shrinkage and over one quarter of Chinese elderly with hypertension fail to recognize their health is-

⁴Because the subjective bias is unobserved, it is rarely measured or collected solely from survey questionnaires or physical examinations. Most of the popularly used health measures may not work for our end. For example, it is almost impossible to define and measure an objective counterpart for self-reported health. For the mortality, it is also difficult to measure an *objective* mortality probability for a particular person.

⁵Height shrinks mainly due to osteoporosis (bones losing density and becoming more fragile) as people age. Medical literature showed that osteoporosis – a “silent disease” strongly associated with bone fracture and disability in the future - is commonly unrecognized, under-diagnosed and under-treated in both developed and developing countries (Ali et al., 2009; Bilezikian, 1999; Feldstein et al., 2006). More details can be found in the Appendix.

⁶We follow this name in Sandroni and Squintani (2007). Overconfidence here may be just due to inattention or unawareness and not “too confident” in real life.

sues. Our estimates show that those with higher socio-economic status (SES) are less likely to ignore their health problems.⁷ Specifically, those with some college or above shrink 0.9 centimeters (cm) less in height but report 0.1 cm more shrinkage than those with fewer than 12-year schooling do. In China, those with some education or urban *hukou* are also 6 percentage points more likely to realize their hypertension condition if they have it. In addition, those with actual worse health are even less likely to recognize their health issues, indicating that they are relatively more overconfident. Specifically, those with scarcer bone mineral density, hypertension or lower cognition ability are more likely to underestimate their own height shrinkage and thus are more overconfident as defined.

We further explore how overconfidence affects the insurance enrollment decision. Estimates using different measures from both countries consistently suggest that, conditional on the actual health status, people with more overconfidence are less likely to participate in health insurance programs. Among those aged 55-65 in the US, those with under-recognition about shrinkage by 1 cm are associated with 1.3 percentage points less likelihood of health insurance participation; among Chinese over 45 with hypertension, those not realizing it are 3.3 percentage points less likely to have insurance. Back-of-the-envelope calculations based on these estimates suggest that insurance participation rate would increase by 1 percentage point among those with fewer than 12-years schooling in the US and those with hypertension in China, had they fully realized their health conditions.

Yet the overconfidence measures we use above have potential endogeneity problems. First potential issue is that the overconfidence measure may reflect unobserved lower SES that is correlated with less insurance participation. For one thing, using NHANES data, we find that the overconfidence is associated with better *subjective* health but with worse *objective* health while the higher SES is associated with better health for both subjective and objective measures; for the other, in the US, Medicaid is mainly for those with lower income and the Medicare is eligible for those with disability when aged below 65; if the overconfidence

⁷This finding is also consistent with Cawley and Choi (2015), who found that the better educated individuals report their health more accurately.

only reflects lower SES, it should be positively correlated with participation in these insurances. In contrast, our results find that that the overconfidence is negatively correlated with participation in both of them, no matter for those aged under 65 or for the full sample.

It is also possible that those with insurance may visit doctors more frequently so they could have more accurate health information. Following Card et al. (2009), we test this by using the Medicare eligibility as an exogenous shock for insurance participation. The results show that insurance participation rate increases by 15 percentage points from those aged 64 to those aged 65 but there is almost no change in our overconfidence measures, suggesting that the reverse causality may not be a first order issue in this analysis.

The panel structure and the hypertension condition feedback during the physical examination of CHARLS help to provide additional insights on causality because the combination creates an “experimental” setting to identify the effects of accurate health information. In the first 2011 wave of the survey, after the subjective health information collection, the respondents were asked to take the physical examinations and their blood pressures were measured by professional nurses. More importantly, the respondents would be notified by the nurse about their blood pressure and hypertension condition. The following 2013 wave collected the insurance status of the respondents again. We show that the information from the nurses makes those with under-diagnosed hypertension more likely to take up health insurance. Compared to others, those with under-diagnosed hypertension in 2011 are 2 percentage points more likely to enroll in health insurance programs if they have not previously, and 1 percentage point less likely to quit the original program in 2013 if they have already enrolled.

This study contributes to several strands of literature. First, we provide a new explanation to potentially reconcile the mixed empirical findings regarding the selection issue in health insurance markets.⁸ Since individuals with actual higher risk are relatively more

⁸Although the “adverse selection” is predicted by the theory and has been supported by an established strand of empirical literature, the phenomenon of “advantageous selection” has been widely documented in both large insurance markets such as life insurance (Cawley and Philipson, 1999, McCarthy and Mitchell, 2010) and Medigap insurance (Hurd and McGarry, 1997, Ettner, 1997, Fang et al., 2008), and in extremely

overconfident in their health status and thus less likely to enroll in insurance, the subjective bias is an important factor contributing to the advantageous selection. In addition, this paper also contributes to the behavior economics literature by documenting the prevalence of overconfidence among the elderly and establishing the negative associations between overconfidence and health insurance participation.⁹ The overconfidence in health could arise from several possibilities, including inattention, lack of symptom salience, information avoidance and many other psychological factors. It potentially explains why there are a large number of individuals who do not take up health insurance even when it is available at a very low cost (Baicker et al., 2012). Last but not least, the findings provide new insights in theoretical analysis of health insurance, as well as consequent welfare implications and policy suggestions (Sandroni and Squintani, 2007; Einav et al., 2010a; Chetty and Finkelstein, 2013). For example, since willingness to pay and the slope of the insurance demand curve is shaped by the perceived risk, our findings suggest that both the level and slope of health demand curve would be altered by accurate health information provided, thus the welfare implications may change correspondingly.

The paper is organized as follows. Section II presents a simple model with subjective bias. Section III introduces the data used in this paper. Sections IV and V provide empirical results from NHANES and CHARLS, respectively. Section VI extends the model in Einav et al. (2010b) with the presence of overconfidence and Section VII concludes.

thin markets such as long-term care insurance (Finkelstein and McGarry, 2006) and reverse mortgages (Davidoff and Welke, 2007).

⁹For example, there are several field experiments that tried to look at the effectiveness of behavioral interventions in health insurance markets and found that such nudges are effective. Volpp et al. (2008) studied the lottery incentives in increasing people's adherence to anti-stroke medication. Loewenstein et al. (2007) studied how to use decision errors to help people improve their health. Chandra et al. (2010) found that increases in cost sharing by a supplemental insurer can exert financial externalities in the Medicare program. Moreover, Baicker et al. (2015) tried to provide a foundation for the behavioral moral hazard prevalent in health insurance.

II. Theoretical Framework

We start by showing how the subjective bias in actual health status will influence the insurance decision at optimization. Individuals have the perceived risk π^* and the actual risk π . Individual subjective bias on his/her own risk, B , is defined as the difference between perceived and actual risk, $B = \pi^* - \pi$. If individuals have perfect information about their health status, and $B = 0$. Imperfect information about own health status means $\pi^* \neq \pi$. Specifically, if $\pi^* > \pi$ and $B < 0$, perceived risk being larger than the actual, individuals are over-concerned; if $\pi^* < \pi$ and $B > 0$, perceived risk smaller than actual, individuals are over-confident. Follow the framework in Cawley and Philipson (1999), we assume that the insurance is provided by the government with a given price p . Individuals choose how much insurance D to purchase to maximize their expected utility based on their perceived health status π^* :

$$D(\pi^*) = \arg \max_D \{ \pi^* U(Y + D - pD) + (1 - \pi^*) U(\bar{Y} - pD) \}$$

where Y and \bar{Y} is the income of the individual when he/she is ill or healthy, respectively, with $\bar{Y} \geq Y + D$. Therefore, the first order condition is

$$\pi^*(1 - p)U'(Y + D(\pi^*) - pD(\pi^*)) = (1 - \pi^*)pU'(\bar{Y} - pD(\pi^*))$$

thus we will have $D'(\pi^*) > 0$, which means that those perceived higher risk will purchase more. If the overconfidence is prevalent and the price is given and fixed, $D'(\pi^*) > 0$ also suggests that the insurance participation rate would be higher if all the individuals had realized how unhealthy they are in actual.

However, the correlation between D and actual risk π is provided by the following:

$$\frac{dD(\pi^*)}{d\pi} = \frac{\partial D(\pi^*)}{\partial \pi^*} \frac{d\pi^*}{d\pi} = D'(\pi^*) \left(1 + \frac{dB}{d\pi}\right)$$

Therefore, if B is a constant or distribute uniformly among the population, the adverse selection will still hold since $\frac{dB}{d\pi} = 0$ and $D'(\pi^*) > 0$. However, if $\frac{dB}{d\pi} > 0$, i.e. those with

higher risk will overestimate the health risk, the adverse selection would be strengthened by the subjective bias. To the contrary, if $\frac{dB}{d\pi} < 0$, i.e. those who are unhealthier are relatively more likely to fail to realize how unhealthy they are, the sign of $\frac{dD(\pi^*)}{d\pi}$ will be undetermined, implying that whether the classic “adverse selection” holds or not actually relies on how the subjective bias is distributed in the population.

Take the extreme case, $\pi^* = 0$, as an example, which means the individuals completely fail to realize their risk type. The model predicts $D^* = 0$, indicating that individuals would not participate the insurance when they maximize their utility based on their perceived risk. Then it is obvious that the insurance participation is not relevant to the actual risk type and that adverse selection does not hold any more.

III. Data and Measures of Overconfidence

3.1 National Health and Nutrition Examination Survey (NHANES)

The first data we use are from the NHANES, a program of studies designed to assess the health and nutritional status of adults and children in the US, which collected demographics, socioeconomic status, health conditions and health behavior patterns.

We mainly use the unrecognized height shrinkage as the overconfidence measure for NHANES. Height shrinks mainly due to osteoporosis (bones losing density and becoming more fragile) as people age, and height shrinkage is strongly associated with bone fracture and disability in the future. To our best knowledge, this is the only existing survey that provides the full set of subjective and objective measures for the height shrinkage. On the one hand, the respondents were asked about their height at age 25 and their current height, enabling us to calculate the *perceived* height shrinkage. On the other hand, professional nurses were instructed to measure the current height and lengths of upper arms and upper legs in millimeters in the physical examination. Following Huang et al. (2013), we use the limb lengths to estimate the *actual or biological* pre-shrinkage height for the seniors

and Appendix A presents the details. In addition, nurses conducted X-ray examinations and recorded respondents’ bone mineral density (BMD).¹⁰ We restrict the sample to those with valid measured biomarkers. Since NHANES oversampled those with lower SES in the US, we use the sampling weights throughout this analysis so that the results are national representative.¹¹

Appendix A also provides the biological background and evidence that biological shrinkage is strongly and consistently associated with bad health measured by other popular health measures (both subjective and objective), including self-reported health, Difficulties in Activities of daily living (ADLs) and Instrumental Activities of Daily Living (IADLs), hypertension and cognitive ability. Recall that the subjective bias defined in Section II is $B = \pi^* - \pi$. The perceived risk, π^* , is measured by reported shrinkage and the actual risk, c , measured by the biological shrinkage in this scenario. We use the difference between reported and biological shrinkage as a measure for overconfidence in NHANES.¹² “Overconfidence” refers to when the subjective shrinkage is smaller than the actual shrinkage, and “over-concern” happens when people think they shrink more than actually do. Panel A of Table 1 reports means and standard deviations for the key variables used in this study.

[Table 1 about here]

For those aged over 55 in the US, the average biological height is 167.1 cm and estimated pre-shrinkage height 168.7 cm, implying the average biological shrinkage is 1.6 cm. However, the reported current height and the reported height at 25 year-old are 168.8 and 169.8 cm, which yield the seniors only realize 0.96 cm, about 60% of the actual value. If height shrinkage is a good health measure, this finding suggests that the seniors do not fully realize how much they actually shrink and are overconfident about their health in general according

¹⁰The examined position for BMD may change over years. So we calculate the yearly specific z-score for it. BMD provides another benchmark for biological and reported height shrinkage, and serves as another objective health measure as well.

¹¹<http://www.cdc.gov/nchs/tutorials/Nhanes/SurveyDesign/Weighting/OverviewKey.htm>

¹²We use the absolute difference in the main text. We also use the ratio of reported and biological shrinkage as well as another (i.e. under-diagnosed hypertension) to measure the overconfidence, and find consistent results.

to our definition. Next section will show how this measure is correlated with subjective and objective health measures as well as other behaviors to shed some light on what is measured by this constructed variable.

The measure of unrecognized shrinkage has some virtues in practice. First, shrinkage is easily to be ignored since osteoporosis is notoriously a “silent disease,” and, unlike other diseases, people cannot know exact shrinkage directly by just taking physical examinations. Second, biological shrinkage is continuous and accurately reflects the actual situation.¹³ Third, the subjective shrinkage is derived through difference in reported heights and thus individual-invariant bias could be cancelled by doing so.

3.2 China Health and Longitudinal Study

CHARLS is a nationally representative longitudinal survey of the middle-aged and elderly population (45+) in China along with their spouse, which includes an assessment of the social, economic, and health circumstances of community-residents.¹⁴ It is a twin study of the Health and Retirement Study (HRS) in the US. The main questionnaire includes information on basic demographics, family, health status, health care, employment, and household economy (income, consumption and wealth).

The survey collected information for both realized and actual high blood pressure. For one thing, respondents were asked about detailed information about their hypertension. The two questions we use to measure realized hypertension are “Have you been diagnosed with hypertension” and “Do you know if you have hypertension?” If the answer to either question is “Yes,” then we categorize them as people who realized their hypertension. For the other, CHARLS also provided a free physical examination to all respondents and blood pressures were measured then. In practice, the professional nurses measured the blood systolic and

¹³Compare height shrinkage with mortality, for instance. In the HRS, people were asked to provide their subjective death likelihood. The corresponding ideal objective morality measure should be a probability *ex ante* but the *ex post* measure is a binary variable.

¹⁴The national baseline survey of CHARLS was conducted in 2011 on 17,692 respondents. CHARLS baseline data includes detailed information of respondents and their living spouses. For a detailed description of the CHARLS survey, see Zhao et al. (2012).

diastolic pressures for the respondents three times, respectively. We assign a hypertension condition equal to 1 if minimum systolic is 140 or greater, or if minimum diastolic is 90 or greater.¹⁵ Noting that who realized their hypertensive status may have taken some measures to control their blood pressure, we thus define actual hypertension as having measured high blood pressure or realized hypertension even when the measured blood pressure is normal. There would only be under-realized hypertension according to our definition, but no over-realized hypertension.

The sample is restricted to those aged over 45 with valid measures of blood pressures and subjective hypertensive status. Panel B of Table 1 presents the means and standard deviations of key variables used in the study. Noticeably the under-diagnosed hypertension rate for the seniors is really high in China; for those 9350 respondents in the sample, 3169 have hypertension (i.e. 34% of all the respondents) but 863 of them (i.e. 27% of those having hypertension and 9.2% of all the respondents) did not realize it; while the under-diagnosed hypertension in the US is 12% according to the NHANES data.¹⁶

As mentioned, the panel structure of CHARLS provides additional insights on causality. During the 2011 survey, the nurses collected the blood pressures and were required to inform the respondents about their hypertension condition directly. This information is confidential and will not be released to any one else, including their family, friends, local hospitals, etc. The “consent form” sent to the households before the survey reads:¹⁷

“If you agree to participate in the survey, the physical examination and test results related to your health will feed back to you directly ... All your personal records including fingerprint, questionnaires, and physical examination and test results are confidential; we will not tell others, include your family, friends, local hospitals, etc... ”

¹⁵Following Huang et al. (2013), we use these two thresholds to define hypertension.

¹⁶We use under-diagnosed hypertension as the overconfidence measure in CHARLS and provide additional and consistent evidence from China by examining whether it is associated with SES and insurance status. Its correlation with unhealthy status is obvious because it is only those suffering hypertension who are possibly ignore it.

¹⁷Please see details at http://charls.ccer.edu.cn/en/page/documentation/2011_national_baseline

The following 2013 wave collected the insurance status of the respondents again, enabling us to track the changes in insurance status after telling them their actual hypertensive status in the first wave. This “experimental” setting enables us to examine whether those with under-diagnosed hypertension are more likely to take up health insurance in the 2013 wave than others after having more accurate information about their health status. The design helps to rule out the influence from third party since the health information is only released to the respondents themselves rather than any one else.

The next two sections present the results from NHANES and CHARLS, respectively.

IV. Empirical Results from NHANES

4.1 What is about the “Overconfidence”?

Because overconfidence itself is unobserved, the first concern is whether our measures successfully capture the concept of overconfidence as we hope. We are concerned that the involved construction makes our measures a pure “black box” and only pick up other relevant factors but not overconfidence. Before investigating the relationship of overconfidence with SES, health and insurance, we start by providing some consistent evidence on what the constructed overconfidence variable measures.

Since the overconfidence is defined as perceived health better than actual, we verify this by investigating its correlation with a series of other health measures. We expect that people with more overconfidence tend to report better subjective health status but do not score better for the objective health status. In practice, we use the following equation (1) to test this hypothesis:

$$Y_i = \alpha_0 + \alpha_1 \text{Overconf}_i + \alpha_2 \text{Shrink}_i + \alpha X_i + \epsilon_i \quad (1)$$

The dependent variable, Y_i , is other health measures than height shrinkage, which may be

self-reported health (a subjective measure) or hypertensive status (an objective measure).¹⁸ $Overconf_i$ refers to the overconfidence measure constructed above by differencing the subjective shrinkage and the actual shrinkage. $Shrink_i$ is the biological height shrinkage. We need to control for this variable first because we construct our overconfidence measure from height shrinkage, and second it also reflects the actual individual health status. The coefficient α_1 is of central interest here because it captures the association between overconfidence and the outcomes, conditional on the actual health. X_i denotes a set of control variables. It includes indicators for education levels in three categories (i.e. fewer than 12 years schooling, high school graduates and some college or above), gender, ethnicities, country of birth, marital status, survey year and a full set of dummies for age in years.

The first column of Table 2 reports the OLS estimates for α_1 and α_2 for two different dependent variables. Consistent with Appendix A, the biological shrinkage is associated with bad health, for both subjective and objective measure.¹⁹ More importantly, the signs on overconfidence are different for the two outcomes. The coefficient in the first column suggests that people with more overconfidence tend to report better health. However, the coefficient on it becomes positive when dependent variable is hypertension, implying that overconfidence is correlated with higher likelihood of hypertension. These results are fairly consistent with our hypothesis. Appendix B provides additional consistent evidence by examining other health measures.²⁰

[Table 2 about here]

To further verify what is captured by the overconfidence measure, we further examine a set of outcomes about healthiness of one’s diet. In NHANES, respondents were asked about

¹⁸In NHANES, the individuals were asked about their evaluation on their health status and the answer ranges from 1 “Excellent” to 5 “Poor.” We use this variable directly here. The hypertensive status was defined by reported diagnosed hypertension and actual high blood pressure measured in physical examination. It has the same definition as that in CHARLS.

¹⁹It is worthwhile to note that either of the coefficients in column 2 is not significant according to the t -statistics, but they are jointly significant in a F -test, with a p -value of 0.006.

²⁰The subjective measures include self-reported health and reported ADLs and IADLs, while the objective ones are hypertension and cognition.

their evaluation on their diet “How healthy do you think your diet is?” and the answer ranges from 1 “very healthy” to 5 “very unhealthy.” The respondents were also asked about some detailed questions about the frequency that the respondents check the information of certain ingredients on the food labels.²¹ If a person is overconfident, it is likely that he/she would report better healthiness of diet but would not care about the information on the food labels. Using the econometric framework same as above, we provide consistent evidence in columns 3 and 4 of Table 2: Those with higher overconfidence are more likely to report having a healthy diet but they check notice the health information on the food labels less frequently.

Previous studies suggest that risk aversion is also an important factor that affects on insurance participation and could be easily confounded with overconfidence. Cutler et al. (2008) used the risky health behaviors to proxy the risk aversion attitude and provides evidence that people with higher risk tolerance are less likely to participate in insurance programs. Although risk aversion and overconfidence are theoretically different,²² some might argue that the two mechanisms could be observationally equivalent. We examine here whether we can tell them apart empirically. In this study, we also use smoking and drinking as proxies for risk tolerance. The last two columns of Table 2 conduct the same regressions of equation (1) but replace the dependent variable to indicators for smokers and heavy drinkers. The estimates show that overconfidence measure has only weak correlation with these behaviors, suggesting that the overconfidence and risk aversion should be two distinct factors that affect individual behavior differently.

4.2 Overconfidence, Education and Health

We can now start our analysis by examining the association of overconfidence with education and actual health status. If overconfidence is universally constant among the population,

²¹Actually, NHANES asked four questions about different ingredients, respectively, i.e. calories, fat, cholesterol and carbohydrates. The answers varies from 1 “Very often” to 4 “never”. In practice, we just conduct the principle component analysis and use the constructed principle component.

²²The risk aversion refers to the curvature of the utility function but the overconfidence means the difference between a subjective value and an objective one.

the implications of the classical model still holds and essentially nothing changes. The interesting case is $\frac{dB}{d\pi} \neq 0$ when the overconfidence is correlated with health status. In our case, $\frac{dB}{d\pi} > 0$ means that those with more shrinkage relatively overestimate it, and $\frac{dB}{d\pi} < 0$ means the same group of people relatively underestimate it. As Section II suggests, the latter case may qualify the degree of adverse selection or even reverse its direction to advantageous selection. In this part, we examine how the overconfidence is correlated with the health status. We use the following equation to examine whether the difference between biological and subjective shrinkage correlates with health and education:

$$Overconf_i = \beta_0 + \beta_1 Hlth_i + \beta X_i + \epsilon_i \quad (2)$$

where the dependent variable $Overconf_i$ is the overconfidence of individual i , i.e. the unrecognized height shrinkage (i.e. reported shrinkage - biological shrinkage). $Hlth_i$ denotes the actual health status of individual i , which could be bone mineral density, hypertension or cognition.²³ X_i denotes a set of control variables same as in equation (1). Table 3 reports the OLS estimates for equation (2).

[Table 3 about here]

The three health measures are constructed such that higher means unhealthier. All estimates for β_1 are positive and significant, suggesting that worse health is correlated with more overconfidence. One standard deviation increase of BMD scarcity and hypertension predict 0.5 cm and 0.4 cm more in unrecognized shrinkages, respectively. The third column examines the association of overconfidence with cognition and finds that people with cognition impairment have more unrecognized shrinkage.²⁴ Therefore, unhealthier people shrink more but unfortunately fail to fully realize it. Consistent with the results above, estimates

²³We do not use the height shrinkage here because the shrinkage has been used for constructing the dependent variable.

²⁴The results of cognition is consistent with those in Fang et al. (2008), who find that cognition impairment is an important source for advantageous selection.

for the coefficient before education present that higher education is significantly associated with less overconfidence: compared to those with some college or above, people with fewer than 12 years schooling have 1.2 cm more unrecognized shrinkage on average. All the results here suggest that better health is associated with lower overconfidence.

4.3 Overconfidence, Insurance Participation and Healthcare Usage

This part further examines how overconfidence in health affects the insurance participation and healthcare usage by conducting similar regressions of equation (1) and switching the dependent variables.

We report the associations of overconfidence with insurance status by types of health insurance in different columns of Table 4. Since the age profile of insurance participation is greatly influenced by Medicare program’s availability for those aged 65 and above, we further divide the sample by the 65-year-old threshold and report the results in Panel B and Panel C, respectively.²⁵ With the actual health status measured by biological shrinkage controlled for, the coefficient on overconfidence indicates how much insurance participation is associated with an additional one-centimeter unrecognized shrinkage. Estimates in the first column show that the overconfidence is significantly associated with lower general health insurance participation. The magnitudes suggest that unrecognized shrinkage increased by one centimeter predicts 0.4 percentage points decrease in insurance participation rate for the full sample, and 1.3 and 0.1 for those aged below and above 65, respectively. The magnitudes are not trivial when comparing to the uninsured rates for the three groups, which are 6.9, 14 and 1.3 percent, respectively. Due to the large measurement error in the overconfidence variable, we argue the magnitudes could be much larger.²⁶

[Table 4 about here]

²⁵The participation rate increases from 86 percent among those aged below 65 to 99 percent for those aged above 65.

²⁶Appendix C uses under-diagnosed hypertension as another measure for overconfidence in NHANES. We find that the magnitudes are much larger. Specifically, the insurance participation rate for those with under-diagnosed hypertension is 3.2 percentage points lower in the full sample.

The next three columns show consistent results for different health insurance products, including Medicare, Medicaid and private insurance.²⁷ Estimates in column 2 shows the results for Medicare participation. The large difference in participation rates of Medicare for the two age groups reflects the policy implementation: for those below 65, it is mainly those with disability who are eligible for Medicare; but for those above 65, almost everyone is eligible. In line with it, we find larger and more significant associations of objective poorer health (i.e. biological height shrinkage) with higher Medicare participation; and more importantly, those with more overconfidence are associated with lower participation rate, especially for those aged below 65. One possible explanation is that those with more overconfidence are less likely to realize or report their disability.

The third column reports the results when the dependent variable is Medicaid participation (Yes = 1). Medicaid is generally a social health care program for families and individuals with low income and limited resources. The Health Insurance Association of America describes Medicaid as a “government insurance program for persons of all ages whose income and resources are insufficient to pay for health care.” Therefore, there is almost no monetary cost of participating in Medicaid when respondents meet the requirement. If the overconfidence only reflects lower SES, it should be probably positively correlated with insurance participation. The participation rates across different age groups are almost constant, about 4 percent. Estimates in column 3 also show people with overconfidence tend to less participate in Medicaid.

The final column also shows the consistent results when dependent variable is private insurance participation (Yes = 1), and the magnitudes of the coefficients the overconfidence are similar in the two age groups. But It is noteworthy that the three insurances may have some overlaps due to different policies for the programs. For example, “dual eligibles” have both Medicare and Medicaid, and those having Medicare may further purchase private

²⁷NHANES does not provide information on whether the insurance is employer-provided or not. We also conduct the regressions controlling for the work status and find robust results, suggesting that the work status may not be a large problem for this analysis (results available upon request).

insurance like Medigap and Medicare Advantage etc. Therefore, the coefficients should be interpreted carefully and we only conclude that the overconfidence is negatively correlated with insurance participation.

Table 5 further examines the associations between overconfidence and healthcare usage measured by times of doctor visits for people with different insurance status. NHANES asked the respondents the frequency of doctor visits in the last year and the answer ranged from 0 to 5, with higher value for higher frequency.²⁸ Participation is an “extensive” margin while healthcare usage could be interpreted as an “intensive” margin. Consistent with our hypothesis, the coefficients on overconfidence are all negative across all the columns, indicating that conditional on insurance status people with more overconfidence are less likely to visit doctors. All of the coefficients are significant except for those without insurance. Recall that the classical insurance model without subjective bias predicts the “positive correlation” of *ex post* risk with both insurance participation and healthcare usage, and this positive correlation with healthcare usage would be reinforced due to moral hazard. However, the empirical results here suggest that overconfidence among those unhealthy people plays an opposite role against the classical positive correlation, as overconfidence is stronger among those unhealthy people and predicts lower insurance participation and less healthcare usage.

[Table 5 about here]

Appendix C shows that the results are robust using different measures of overconfidence in NHANES. The first measure is the shrinkage in terms of ratios of heights. We examine this because the perception about shrinkage may be biased in proportion rather than absolute differences. The second measure we examine is the under-diagnosed hypertension. Parallel analysis for both measures provides consistent evidence.

²⁸Specifically, 0 for none, 1 for once, 2 for 2-3 times, 3 for 4-9 times, 4 for 10-12 times and 5 for 13 or more.

4.4 Higher Insurance, Less Overconfidence?

The associations of overconfidence with insurance and healthcare usage are not causal. It is likely that those with insurance visit doctors more frequently and receive better information about their health. It is also possible that insurance may directly affect health status and thus is correlated with overconfidence. Due to the possible endogeneity we should be careful to interpret the results above. To shed some light on the issue, we follow the strategy of Card et al. (2009) and test this by comparing the overconfidence of the cohorts right before and immediately after the Medicare eligibility cutoff at age 65. In practice, we keep those between 60 and 70 and control for dummies for gender, marital status, birth place, ethnicity, education and survey year. We then plot the coefficients against age dummies for different dependent variables and normalize the coefficient on age 65 to zero.

Figure 1a first presents the coefficients and 90 percent confidence intervals for the age dummies when the dependent variable is either insurance participation or times of doctor visits. The insurance participation rate significantly jumps by over 10 percentage points from age 64 to 65 and the times of doctor visiting also increases significantly. Figure 1b plots the estimates when dependent variable is either insurance participation or overconfidence measured by unrecognized height shrinkage. However, there is no significant change in overconfidence at age 65, implying that the correlation between insurance and overconfidence should not be driven by the higher probability of receiving accurate health information among those with insurance. Appendix C also provides consistent evidence when using under-diagnosed hypertension as the overconfidence measure.

[Figure 1a and 1b about here]

V. Further Evidence from CHARLS

The above analysis explores NHANES and shows that the overconfidence is correlated with bad health and the uninsured status in the US. In this section we further investigate the

CHARLS data and use under-diagnosed hypertension as an alternative overconfidence measure to provide additional evidence from China. More importantly, the panel structure of the data combined with the physical examination in the first wave provides an natural experimental setting to examine the impact of the exogenous shock of health information on insurance participation.

5.1 Consistent Evidence from China

We start by using the cross-sectional data in 2011, the first wave of the survey, to conduct similar regressions as equation (2) in the US case, and report the results in Appendix D. Consistent with previous analysis, those with higher SES measured by education and urban *hukou* have lower overconfidence and the magnitude is not small. Among those with hypertension, 27 percent of them do not know their hypertensive status. However, the under-diagnosed hypertension rate of those with some education or urban *hukou* is about 5 percentage points lower.

Before investigating the impact of exact information provided on insurance participation later we establish the association between overconfidence and insurance participation in the first wave by estimating the following equation:

$$Y_i = \gamma_0 + \gamma_1 \text{Overconf}_i + \gamma_2 \text{Hyper}_{ij} + \gamma Z_{ij} + \text{Cty}_j + \epsilon_i \quad (3)$$

The overconfidence measure, Overconf_{ij} , is an indicator whether the individual i in county j had under-diagnosed hypertension and the objective measure for health, Hyper_{ij} , is dummy for actually suffering hypertension. The covariates, Z_i , are the indicators for gender, age in years, type of *hukou* (urban/rural) and levels of education (i.e. no formal education or illiterate, primary school and junior high school or above). We control for the *hukou* dummy because of the different health insurance systems for urban and rural residents.²⁹ We also control for county fixed effects, Cty_j , because the insurance programs

²⁹In 2011, the insurance coverage rate for rural people is 95% and that for urban people is 88%. The

like NCMS for rural and URBMI for urban areas are implemented at the county level and the indicators capture the heterogeneity across different counties.

First two columns of Table 6 report the estimates. Panel A and Panel B examine the correlations in the full sample and among those with hypertension, respectively. Consistent with what we found in NHANES, the estimates in column 1 show that those with under-diagnosed hypertension are correlated with 2.8 and 3.8 percentage points lower insurance participation rate than that in the corresponding control groups, respectively.³⁰

[Table 6 about here]

According to the seriousness of the hypertension condition, we further categorize the overconfidence into minor and serious accordingly with the consideration that those with under-diagnosed serious hypertension (i.e. more overconfident) should be even less likely to participate in health insurance in 2011 compared to those with only under-diagnosed minor hypertension.³¹ The estimates in column 2 provide sound evidence for it. Specifically, the insurance participation rate for those with under-diagnosed serious hypertension is 3.7 percentage points lower in the full sample, while, for those with under-diagnosed minor hypertension, the number is 2.4.

insurance coverage rate is higher for rural areas due to the New Cooperative Medical Scheme (NCMS) initiated in 2003. In 2011, the coverage rate of NCMS is 94% for those with rural *hukou*. Among those who have health insurance, the NCMS coverage rate is 98% and the rest are mainly covered by private insurance. The major two insurance products in the urban regions are the Urban Employee Basic Medical Insurance (UEBMI) and the Urban Resident Basic Medical Insurance (URBMI). The coverage rate for the former is about 56% and the latter 24%. Then is the government medical insurance (about 12%) mainly for government employees and state personnels, which may be viewed as an upgraded version of UEBMI. The last 8% is private insurance and “urban and rural resident medical insurance,” which is a new insurance product as a combination of URBMI and NCMS.

³⁰Note that the control groups for Panel A and Panel B are those who have accurate information about their hypertensive status. But that in Panel A includes both those with no hypertension and believing so, and those having hypertension and fully aware of it, while the control group in Panel B only include those having hypertension and fully aware of it.

³¹We define serious hypertension if the minimum systolic pressure is larger than 170 or the minimum diastolic pressure is larger than 100.

5.2 The Effects of Information in Physical Examination on Insurance Participation Later

We still use the equation (3) to analyze the impact of the physical examination in the first wave that every respondent took on the insurance participation choice later. We use insurance status change of individual i in county j between the two years as our new dependent variable, denoted by ΔIns_i . There are three categories: the first one is “quit”, which means the individual i was covered by insurance in 2011 but not in 2013; the second one is “unchanged”, meaning the individual i did not change the insurance status; and the third one is “newly enrolled,” standing for the individual i was not covered by insurance in 2011 but was in 2013. The first variable we use is a categorical variable ranging from -1 to 1, denoting the three cases above, respectively. The second is an indicator for “newly enrolled,” which equals to one for those not covered by insurance in 2011 but was in 2013 and zero if otherwise. The third is an indicator for “not quit”, which equals to one for those covered by insurance in both 2011 and 2013.

The remaining columns of Table 6 reports the OLS estimates for the impacts of physical examination on later insurance participation. Since the physical examination was conducted for all individuals, the effects would only bite for individuals who received new information. We predict that, compared to those without hypertension and those who have but know they have hypertension, those with under-diagnosed hypertension are more likely to enroll in or not quit the insurance programs in 2013 after the physical examination. The estimates in column 3 provide evidence for this, as the coefficients on overconfidence are positive. After decomposing the overconfidence into the two levels according to the seriousness of the actual hypertension, we find that the effects on taking up insurance are even larger for those with under-diagnosed serious hypertension. The remaining two columns are for the other two different dependent variables - newly enrollees and those who do not quit. Specifically, after the physical examination, those with under-diagnosed minor hypertension have over 2 percentage points higher likelihood to become new enrollees compared to those

who have already recognized their conditions, and this statistic is 4.2 for those with serious under-diagnosed hypertension. The estimates in the last column also consistently suggest that those with under-diagnosed hypertension are more likely to stay insured, though the evidence is not as strong.

The health insurance programs in China generally have very different implementations.³² Due to work-related and non-voluntary health insurance, the individuals in urban areas may not choose to participate or opt out freely. In addition, the under-diagnosed hypertension rate is much lower for urban people (i.e. the rate is 6.6 percentage points lower for those with urban *hukou*), we further conduct the *hukou* type specific analysis in Appendix D with expectation that the results above should hold in particular for the people with rural *hukou* because the NCMS is generally voluntary and the policy allows individuals to either enroll or opt out at any time. The results are consistent with this hypothesis.

A potential issue for the analysis is lack of a perfect control group. The physical examination was conducted for all survey participants and the health information was provided to all individuals. An ideal control group would be those taking the physical examination but not notified of health information. Without the control group, it may be difficult to rule out, for example, a trend of insurance expansion as an alternative explanation for increased insurance participation. We argue this may not be a first order concern. First, the insurance coverage rate was already 94.7 percent in 2011 and that slowly climbed to 96.3 percent in 2013, implying that the insurance program was almost universal in earlier wave and there was not a large regional expansion; second, our analysis controls for the county level dummies, at which level the insurance program was implemented, thus it may be reasonable to assume that the institutional implementation would be the same for the residents within the same county; third, we do not only examine the behavior of new enrollees but also those quitting and find consistent evidence for both, while insurance expansion may only explain

³²NCMS and URBMI are voluntary but the basic medical care insurance for urban employees is supplemented by the employers and are required by the policies. The retirees will continue to enjoy the employee insurance if they retired in their original firm.

the behavior of new enrollment but not for no quitting.

VI. Discussion about Overconfidence and Insurance Market

Our findings also provide some new insights on the theoretical analysis, welfare implications and relevant policy suggestions about the health insurance market.³³ Incorporating the overconfidence into the analytical framework of Einav et al. (2010b),³⁴ we analyze qualitatively in this section how overconfidence provides a new explanation to the insufficient participation and sheds some light on the mixed empirical evidence for adverse or advantageous selection in the insurance market. We further discuss some welfare consequences and relevant policy interventions.

6.1 Stylized Model with Overconfidence: Explanation to Insufficient Insurance and Advantageous Selection

For simplicity, we follow the setup in Einav et al. (2010b) to consider a market for a single insurance contract with exogenous characteristics as in Akerlof (1970) and extend it to incorporate overconfidence.

Model Setup Individuals decide whether to purchase the uniform insurance offered by risk-neutral insurers. An individual i 's *actual* risk type determines the *ex post* cost to the insurer, denoted by π_i ; the individual i 's *perceived* risk type captures the *ex ante* expected cost, which we denote by π_i^* . The *true* value of insurance v_i equals the *actual* cost π_i plus the *ex post* risk premium r_i , which is defined as the risk premium if i had perfect information about her risk type. However, the *perceived* value of insurance v_i^* equals the perceived cost

³³We are really grateful to Professor Amitabh Chandra, Professor David Cutler, Professor Amy Finkelstein and Professor Hanming Fang for their great insights and suggestions for this part. All errors are ours.

³⁴Einav et al. (2010b) used this framework to show that the relative slope of the cost curve, depicting the average cost of participants at different prices, to the demand curve captures the degree of adverse selection and is sufficient to estimate the corresponding welfare cost in the insurance pricing. This classical framework has been used in the following literature including Einav and Finkelstein (2011) and Spinnewijn (2015). Spinnewijn (2015) additionally considered the impact of the demand friction on the welfare implications of the insurance market.

π_i^* plus the perceived risk premium r_i^* based on the *perceived* risk type π_i^* . In the presence of the overconfidence, $\pi_i > \pi_i^*$, $r_i > r_i^*$, thus $v_i > v_i^*$.³⁵ The overconfidence is measured by the difference between π_i and π_i^* , $\pi_i - \pi_i^*$. The empirical analysis suggests that unhealthier people tend to be more overconfident in health, implying that overconfidence is positively correlated with π_i .

Without health information intervention, individual i chooses to purchase the insurance contract given the insurance contract price p if and only if the perceived value is higher than the price, $v_i^* > p$. Therefore, the sorting of individual based on their perceived risk types determines the cost to the insurers. The average and marginal cost at price p equal $AC(p) = E(\pi|v_i^* \geq p)$ and $MC(p) = E(\pi|v_i^* = p)$, respectively. We consider a competitive equilibrium where the equilibrium price p^c equals to the average *ex post* cost of providing insurance,

$$AC(p^c) = E(\pi|v_i^* \geq p^c) = p^c.$$

Figure 2a plots a linear demand curve together with the corresponding marginal and average cost curves in absence of overconfidence. The market equilibrium is depicted by (p^c, q^c) , the intersection between the demand curve D and the average cost curve AC . The efficient price p^* is the price for which the demand curve and the marginal cost curve intersect. The welfare cost equals the area of shaded triangular area between the demand curve and the marginal cost curve and between the competitive and the efficient level of insurance coverage.

[Figure 2 about here]

Insufficient Participation Figure 2b plots the case with overconfidence in absence of information intervention in the *ex ante* insurance market. Due to overconfidence, the perceived marginal cost MC^* is below the actual marginal cost MC , and the gap between them

³⁵For simplicity, we do not consider the demand friction Spinnewijn (2015) and heterogeneity in risk preferences analyzed in Finkelstein and McGarry (2006) and Cutler et al. (2008).

is narrower when MC is lower. There are two demand lines $D^{ex-ante}$ and $D^{ex-post}$ based on the two marginal cost lines MC^* and MC , respectively. Individuals decide to purchase the insurance based on their perceived marginal cost, and firms have zero profit. Therefore, the market equilibrium is depicted by (p^c, q^c) , the intersection between *ex ante* demand curve and *ex post* cost curve. However, if the individuals had fully realized how unhealthy they are, the actual demand curve would be the dashed $D^{ex-post}$ line, shifted up due to the existence of overconfidence and rotated clockwise because the unhealthier people tend to be relatively more overconfident. The new market equilibrium would be the intersection of $D^{ex-post}$ and AC , which yields a higher insurance participation since $q^{cc} > q^c$. Therefore, our analysis suggests that overconfidence could be an explanation to the insufficient insurance participation: the demand based on perceived health, which determines the actual market equilibrium, is below that according to the actual risk type because of overconfidence.

Advantageous Selection In the previous two subplots we assumed that overconfidence does not change the sign of the slope of MC curve.³⁶ But what will happen if the correlation between overconfidence and health is large enough to change the signs of slopes of MC and MC^* curves? Figure 2c plots this case, when unhealthier people feel they are even healthier than those with healthier status. Again because of the sorting of the perceived marginal cost on the horizontal axis is from high to low and the demand curve is determined by the perceived risk type, the demand curve D and the perceived marginal cost MC^* goes downward. However, the actual MC curve and average cost curve go upward. The market equilibrium is still the intersection between D and AC . In consequence, those effectively healthier are more likely to participate in insurance, i.e. advantageous selection emerges. Note that this is consistent with the theoretical framework in section II which indicates $\frac{dB}{d\pi} < -1$ leading to advantageous selection. The analysis also suggests that the extent of overconfidence in different insurance markets should be a promising explanation to the

³⁶However, the case is similar when the overconfidence changes the sign of the slope: because of the demand curve is higher on average, the insurance participation would increase if individuals had realized how unhealthy they actually are. But the slope of MC and AC curves may differ due to different sorting.

inconsistent findings across markets.

6.2 Welfare Implications in the Presence of Overconfidence

Above all, the presence of overconfidence complicates the welfare analysis by introducing two different demand curves (i.e. $D^{ex-post}$ and $D^{ex-ante}$) derived from the actual and perceived marginal costs (i.e. MC and MC^*) shown in Figure 2d.³⁷ The social optimal insurance levels are different: the intersections of MC line with the two demand curves are (p^*, q^*) and (p^{**}, q^{**}) , respectively. The two different optimal level reflect individual inconsistency in *ex ante* and *ex post* markets, which implies the possibility that some people without insurance *ex post* may wish they had purchased the insurance *ex ante*. Therefore, this inconsistency yields two welfare measures: *ex ante* welfare and *ex post* welfare. This study will not take a stand on which one is better and we discuss both measures.

Second, the presence of overconfidence leads to different quantitative estimates for social welfare loss. As Einav et al. (2010b) and Chetty and Finkelstein (2013) point out, the slope of demand curve and that of the marginal cost curve are sufficient statistics to estimate the welfare analysis of equilibrium and non-equilibrium pricing of existing contracts. The welfare loss calculated by Einav et al. (2010b) is represented by the triangle between the *ex ante* demand curve and the marginal cost curve and between the competitive and the *ex ante* efficient level of insurance coverage (i.e. q^c and q^*). The estimates in Einav et al. (2010b) suggest a small welfare loss associated with the inefficient pricing under adverse selection. However, because the dashed $D^{ex-post}$ line is shifted up and rotated clockwise due to the presence of overconfidence and its association with bad health, the welfare loss based on the *ex post* demand curve may be larger, shown by the triangular area between *ex ante* demand curve and the marginal cost curve and between the competitive and the *ex post* efficient level of insurance coverage (i.e. q^c and q^{**}). However, since the observed demand curve would be the *ex ante* one, it is an empirical yet difficult question to estimate the slope of *ex post*

³⁷Since “adverse selection” is the main direction for the health insurance market suggested by Chetty and Finkelstein (2013), we mainly consider a baseline market with adverse selection for welfare analysis.

demand curve.

Finally, the findings also raise some empirical questions about health insurance policies. The policy intervention considered here is whether to provide exact health information through universal physical examinations. We ignore the cost of physical examinations and assume all people now have perfect information about their health after the intervention. Figure 4d suggests that the new market equilibrium is (p^{cc}, q^{cc}) , the intersection between *ex post* demand curve and the average cost curve AC , and the welfare loss (for both *ex post* and *ex ante* cases) is the triangle between MC and $D^{ex-post}$ curves and between the new competitive and the *ex post* efficient level of insurance coverage (i.e. q^{cc} and q^{**}). Based on the *ex ante* welfare standard, it is an empirical question whether the loss is smaller after the information intervention; but it is certain that the intervention will reduce the loss based on the *ex post* welfare if $q^{cc} > q^c$.³⁸

VII. Conclusions

The classic insurance model assumes that individuals have perfect information about their health status and make optimal decision for health insurance participation based on perfect information. This induces a positive correlation between their health insurance participation and *ex post* risk (i.e. adverse selection). However, empirical papers find that individuals may not participate in health insurance even under a really low price (Baicker et al., 2015) and suggest mixed findings as for the presence of adverse selection (Chetty and Finkelstein, 2013). Although many scholars have acknowledged the possibility of individuals having imperfect information about their own health, our paper is the first study providing empirical evidence for the potential impact of overconfidence on the health insurance market, which provides

³⁸The case is similar when the overconfidence causes advantageous selection in the baseline market. The information provision will change the market from Figure 2c to Figure 2a. Based on the *ex ante* welfare standard, the advantageous selection will reverse to adverse selection, and it is undetermined whether the welfare loss becomes larger or smaller; based on the *ex post* welfare standard, the welfare loss would be reduced after the intervention: we suppose $MC > MC^*$ holds for people on the right-hand side of q^* in Figure 2b since overconfidence is prevalent. then there is a large loss based on the *ex post* demand among these people.

novel explanations potentially to reconcile the two puzzling phenomena.

A simple theoretical framework allowing for subject bias predicts that the individuals may make suboptimal choices in health insurance participation and that adverse selection may not hold in the presence of overconfidence (i.e. perceived health is better than actual health) and its. Using national representative samples of the seniors in US and China, we construct two overconfidence measures based on height shrinkage and hypertension, and find consistent evidence that individuals generally do not fully recognize their health issues. Since people make decisions based on their perceived information rather than actual risks, overconfidence may be a barrier for insurance participation even when the cost is low.

Furthermore, overconfidence does not distribute uniformly across the population. People with higher education or actual better health are more likely to have accurate health information. We further find that the overconfidence is associated with less insurance coverage and healthcare usage. The reverse causality may not be the first order issue as we find that the exogenous increase in the insurance participation at age 65 in the US does not influence the overconfidence. These findings imply that the overconfidence is an important factor contributing to the advantageous selection in the insurance market.

Investigation using data from China does not only provide consistent evidence but also sheds some new light on the causality issue. Using the panel structure of the data and the physical examination conducted in the first wave, we find that those with under-diagnosed hypertension at first were more likely to participate in health insurance in the next wave after nurses provided the accurate information about hypertensive status to the respondents.

Finally, our consistent findings in both US and China provide brand new insights on the health insurance markets, relevant welfare analysis and the corresponding policies. In the presence of overconfidence, some implications from the classical model need to be revisited. Our findings thus put the research in agenda to analyze the potential outcomes and social consequences from the overconfidence in health.

There are some concerns about this study. The first is about the overconfidence measure

we used. Because of the difficulty to find an ideal measure based on comparison between a subjective measure and its genuinely objective counterpart, we solely use the height shrinkage and hypertension to construct our overconfidence measures. Although analysis based on the two different measures yields the consistent results, it is still possible that the overconfidence measures may miss some part of the actual overconfidence in health in general or contain some other unobservable factors, because it is not crystal clear what these two constructed measures capture, exactly as the overconfidence concept itself cannot be observed directly. So the magnitude of the coefficients should be interpreted with caution when generalizing the results. In addition, the pre-shrinkage heights estimation may bring large measurement errors, thus it is likely the coefficients on overconfidence measured by unrecognized height shrinkage may be biased towards zero. If this is the case, the associations may be underestimated.

The second one is lack of analysis for different types of insurance. The previous literature suggest the “adverse selection” is the main direction for the health insurance market with some exceptions such as Medigap, life insurance or long-term care. Although this study provides the evidence that overconfidence could be a force that pushes the market away from adverse selection to advantageous selection for general health insurance participation, it may be more interesting to find whether and how much the overconfidence explains any specific insurance markets. Due to data limitation, we hope the research in the future may shed more light on these questions.

References

- Abaluck, Jason and Jonathan Gruber**, “Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program,” *American Economic Review*, 2011, *101* (4), 1180–1210.
- Akerlof, George A.**, “The Market for “Lemons”: Quality Uncertainty and the Market Mechanism,” *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.
- Ali, Tauseef, David Lam, Michael S. Bronze, and Mary Beth Humphrey**, “Osteoporosis in Inflammatory Bowel Disease,” *The American Journal of Medicine*, 2009, *122* (7), 599–604.
- Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein**, “Behavioral Hazard in Health Insurance,” *The Quarterly Journal of Economics*, 2015, *130* (4), 1623–1667.
- , **William J Congdon, and Sendhil Mullainathan**, “Health Insurance Coverage and Take-Up: Lessons from Behavioral Economics,” *Milbank Quarterly*, 2012, *90* (1), 107–134.
- Barseghyan, Levon, Francesca Molinari, Ted O’Donoghue, and Joshua C Teitelbaum**, “The nature of risk preferences: Evidence from insurance choices,” *American Economic Review*, 2013, *103* (6), 2499–2529.
- Bilezikian, J. P.**, “Osteoporosis in Men,” *The Journal of Clinical Endocrinology and Metabolism*, 1999, *84* (10), 3431–3434.
- Card, David, Carlos Dobkin, and Nicole Maestas**, “Does Medicare Save Lives?,” *The Quarterly Journal of Economics*, May 2009, *124* (2), 597–636.
- Cawley, John and Anna Choi**, “Health Disparities Across Education: The Role of Differential Reporting Error,” NBER Working Papers 21317, National Bureau of Economic Research, Inc July 2015.

- and **Tomas Philipson**, “An Empirical Examination of Information Barriers to Trade in Insurance,” *American Economic Review*, September 1999, *89* (4), 827–846.
- Centers for Medicaid and Medicare Services**, “National Health Expenditures,” 2005.
- Chandra, Amitabh, Jonathan Gruber, and Robin McKnight**, “Patient Cost-Sharing and Hospitalization Offsets in the Elderly,” *American Economic Review*, March 2010, *100* (1), 193–213.
- Chetty, Raj and Amy Finkelstein**, “Social Insurance: Connecting Theory to Data,” *Handbook of Public Economics*, 2013, *5*, 111.
- Cutler, David M., Amy Finkelstein, and Kathleen McGarry**, “Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance,” *American Economic Review*, May 2008, *98* (2), 157–62.
- Davidoff, Thomas and Gerd Welke**, “Selection and Moral Hazard in the Reverse Mortgage Market,” 2007.
- Einav, Liran, Amy Finkelstein, and Jonathan Levin**, “Beyond Testing: Empirical Models of Insurance Markets,” *Annual Review of Economics*, 09 2010, *2* (1), 311–336.
- , – , and **Mark R Cullen**, “Estimating Welfare in Insurance Markets Using Variation in Prices,” *The Quarterly Journal of Economics*, 2010, *125* (3), 877–921.
- and – , “Selection in Insurance Markets: Theory and Empirics in Pictures,” *The Journal of Economic Perspectives*, 2011, *25* (1), 115.
- Ettner, Susan L.**, “Adverse selection and the purchase of Medigap insurance by the elderly,” *Journal of Health Economics*, October 1997, *16* (5), 543–562.
- Fang, Hanming, Michael P. Keane, and Dan Silverman**, “Sources of Advantageous Selection: Evidence from the Medigap Insurance Market,” *Journal of Political Economy*, 04 2008, *116* (2), 303–350.

- Feldstein, Adrienne, Patricia J. Elmer, David H. Smith, Michael Herson, Eric Orwoll, Chuhe Chen, Mikel Aickin, and Martha C. Swain**, “Electronic medical record reminder improves osteoporosis management after a fracture: a randomized, controlled trial.,” *Journal of the American Geriatrics Society*, 2006, *54* (3), 450–457.
- Finkelstein, Amy and Kathleen McGarry**, “Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market,” *American Economic Review*, September 2006, *96* (4), 938–958.
- Huang, Wei, Xiaoyan Lei, Geert Ridder, John Strauss, and Yaohui Zhao**, “Health, Height, Height Shrinkage, and SES at Older Ages: Evidence from China,” *American Economic Journal: Applied Economics*, April 2013, *5* (2), 86–121.
- Hurd, Michael D. and Kathleen McGarry**, “Medical insurance and the use of health care services by the elderly,” *Journal of Health Economics*, April 1997, *16* (2), 129–154.
- Loewenstein, George, Troyen Brennan, and Kevin G. Volpp**, “Asymmetric Paternalism to Improve Health Behaviors,” *Journal of the American Medical Association*, 2007, *298* (20), 2415–2417.
- McCarthy, David and Olivia S Mitchell**, *International adverse selection in life insurance and annuities*, Springer, 2010.
- Rothschild, Michael and Joseph E Stiglitz**, “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,” *The Quarterly Journal of Economics*, November 1976, *90* (4), 630–49.
- Sandroni, Alvaro and Francesco Squintani**, “Overconfidence, Insurance, and Paternalism,” *American Economic Review*, December 2007, *97* (5), 1994–2004.
- Spinnewijn, Johannes**, “Heterogeneity, Demand for Insurance and Adverse Selection,” 2015.

Volpp, Kevin G, George Loewenstein, Andrea B Troxel, Jalpa Doshi, Maureen Price, Mitchell Laskin, and Stephen E Kimmel, “A test of financial incentives to improve warfarin adherence,” *BMC Health Services Research*, 2008, 8 (272).

Wilson, Charles, “A model of insurance markets with incomplete information,” *Journal of Economic Theory*, December 1977, 16 (2), 167–207.

Zhao, Yaohui, Yisong Hu, James P Smith, John Strauss, and Gonghuan Yang, “Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS),” *International Journal of Epidemiology*, 2012, 43 (1), 61–68.

Table 1: Summary statistics

Variables	(1) Obs.	(2) Mean	(3) Std. Dev.
<i>Panel A. NHANES data</i>			
Measured current height (cm)	9,864	167.08	9.97
Estimated pre-shrinkage height (cm)	9,864	168.71	8.68
Actual height shrinkage (cm)	9,864	1.64	4.76
Reported current height (cm)	9,864	168.80	10.32
Reported height when age 25 (cm)	9,864	169.77	10.12
Reported height shrinkage (cm)	9,864	0.97	2.09
Overconfidence (Unrecognized shrinkage)	9,864	0.67	4.91
Male	9,864	0.47	0.50
Married (Yes = 1)	9,864	0.64	0.48
Age	9,864	67.18	8.44
<i>Education levels</i>			
Schooling fewer than 12 years	9,864	0.23	0.42
High School Grad (= 12 years)	9,864	0.27	0.44
Some College or above (> 12 years)	9,864	0.50	0.50
<i>Panel B. CHARLS data</i>			
Hypertension (Yes = 1)	11,902	0.34	0.47
Under-diagnosed hypertension (Yes =1)	11,902	0.09	0.29
Male (Yes =1)	11,902	0.47	0.50
Age	11,902	59.41	9.47
Urban hukou (Yes = 1)	11,902	0.18	0.39
<i>Education levels</i>			
No formal education	11,902	0.29	0.45
Primary school	11,902	0.41	0.49
Junior high or above	11,902	0.30	0.46

Notes: Data source are NHANES and CHARLS for Panel A and Panel B, respectively.

Table 2: Overconfidence, Diet Habit and Health Behaviors

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Health Measures		Diet Healthiness		Risky Behaviors	
	Self-reported Health (1-5, Higher for unhealthier)	Hypertension (Yes = 1)	Subject evaluation on diet (1-5, Higher for unhealthier)	Not check food label info. (Higher for less checking)	Smoker (Yes = 1)	Heavy drinker (Yes = 1)
Mean of Dep. Var.	2.66	0.55	2.64	0.00	0.135	0.0193
Overconfidence	-0.0194*** (0.00615)	0.00224 (0.00268)	-0.0208*** (0.00719)	0.0147* (0.00772)	0.00166 (0.00158)	0.000484 (0.000536)
Objective poor health (Biological Shrinkage)	0.0299*** (0.00643)	0.00175 (0.00285)	0.0297*** (0.00756)	-0.0181** (0.00818)	-0.00466*** (0.00168)	-0.00104** (0.000526)
Observations	9,845	9,607	5,718	4,793	10,652	11,906
<i>R</i> -squared	0.168	0.069	0.084	0.149	0.086	0.039

Notes: Data source is NHANES 1999-2011. Covariates controlled for in all columns include dummies for gender, levels of education, age in years, country of birth, ethnicity and survey years. Robust standard errors are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: Overconfidence, SES and Health

Sample VARIABLES	(1) Full sample	(2) BMD not missing	(3) Cognition not missing
Mean of Dep. Var.	Overconfidence (Higher for more overconfident)		
<i>Objective Health Measures</i>			
Hypertension (Yes = 1)	0.370*** (0.120)	0.459*** (0.130)	0.536* (0.279)
BMD scarce <i>z</i> -score		0.506*** (0.0958)	0.544** (0.215)
Cognitive impairment <i>z</i> -score			0.544*** (0.166)
<i>Education level: Reference group is fewer than 12 years</i>			
High School Grad	-0.509*** (0.157)	-0.458*** (0.170)	-0.0839 (0.367)
Some College or above	-1.401*** (0.138)	-1.166*** (0.152)	-0.919*** (0.348)
Observations	12,433	9,863	2,396
<i>R</i> -squared	0.042	0.046	0.067

Notes: Data source is NHANES 1999-2011. Covariates controlled for in all columns include dummies for gender, age in years, country of birth, ethnicity and survey years. Robust standard errors are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Overconfidence and insurance participation, by age and type of insurances

Sample	(1) Any Health Insurance Participation (Yes = 1)	(2) Having Medicare (Yes = 1)	(3) Having Medicaid (Yes = 1)	(4) Having Private Health Insurance (Yes = 1)
<i>Panel A: Full sample (N = 12,385)</i>				
Mean of Dep. Var.	0.93	0.49	0.05	0.66
Overconfidence	-0.00368*** (0.00105)	-0.00437*** (0.00157)	-0.00191* (0.00109)	-0.00392* (0.00236)
Objective poor health (Biological Shrinkage)	0.00259** (0.00115)	0.00441*** (0.00166)	0.00186* (0.00110)	0.00132 (0.00251)
<i>Panel B: Those aged below 65 (N= 4,836)</i>				
Mean of Dep. Var.	0.86	0.07	0.04	0.73
Overconfidence	-0.0132*** (0.00396)	-0.00836** (0.00352)	-0.00395* (0.00234)	-0.00344 (0.00456)
Objective poor health (Biological Shrinkage)	0.0114*** (0.00423)	0.0107*** (0.00370)	0.00361 (0.00236)	0.00337 (0.00490)
<i>Panel C: Those aged above 65 (N = 7,549)</i>				
Mean of Dep. Var.	0.99	0.87	0.05	0.59
Overconfidence	-0.00121** (0.000514)	-0.00209 (0.00175)	-0.00130 (0.00122)	-0.00449 (0.00277)
Objective poor health (Biological Shrinkage)	0.000842* (0.000501)	0.000223 (0.00183)	0.00163 (0.00124)	-0.000128 (0.00293)

Notes: Data source is NHANES 1999-2011. Covariates controlled for in all regressions are the same with those in Table 2. Robust standard errors are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Overconfidence and times of visiting doctors, by age and insurance status

	(1)	(2)	(3)	(4)	(5)
Sample	People with Insurance	Age < 65 & Insured	Age >= 65 & Covered by Medicare /Medicaid	Age >= 65 & No Medicare/Medicaid but having other insurance	No insurance
Dependent variable	Times of doctor visits (0-5, higher for more frequent)				
Mean of Dep. Var.	2.55	2.35	2.72	2.52	1.58
Overconfidence	-0.0376*** (0.00727)	-0.0344* (0.0201)	-0.0371*** (0.00794)	-0.0553* (0.0282)	-0.0160 (0.0514)
Objective poor health (Biological Shrinkage)	0.0361*** (0.00790)	0.0323 (0.0217)	0.0369*** (0.00854)	0.0641** (0.0287)	0.0188 (0.0524)
Observations	8,971	3,017	5,320	634	837
R-squared	0.040	0.042	0.022	0.094	0.140

Notes: Data source is NHANES 1999-2011. Covariates controlled for in all regressions are the same with those in Table 2. Robust standard errors are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6: Overconfidence, Insurance participation and the effects of provided health information

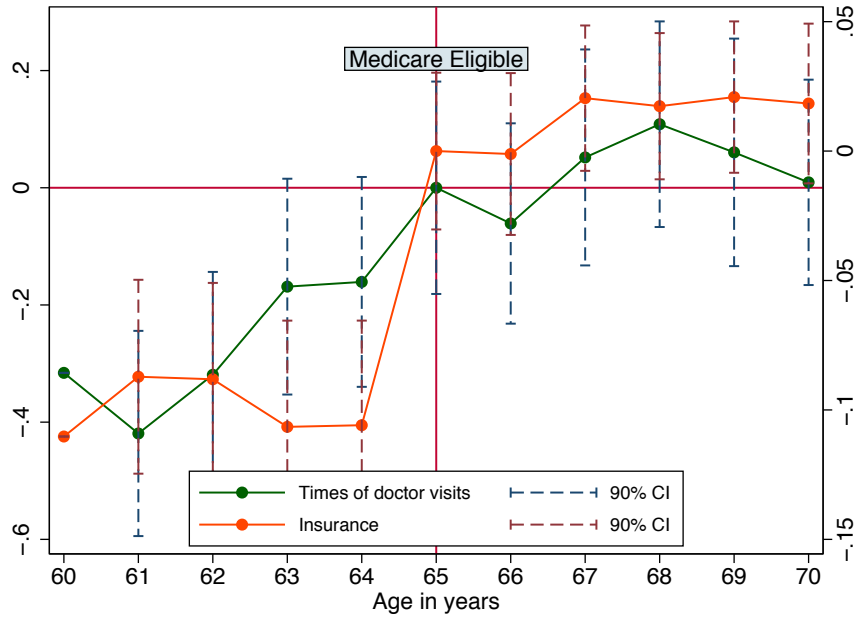
Variables	(1) Any Insurance Participation in 2011 (Yes = 1)	(2)	(3) Δ Insurance Enrollment between 2011 and 2013 (-1, quit; 0, same; 1, enroll)	(4)	(5) Insurance newly enroll (Yes = 1)	(6) Insurance no quit (Yes = 1)
<i>Panel A. All the respondents (N = 11,902)</i>						
Mean of Dep. Var.	0.94		0.03		0.05	0.97
Objective poor health in 2011 (Hypertension, Yes = 1)	0.00627 (0.00520)	0.00623 (0.00520)	-0.00161 (0.00606)	-0.00150 (0.00606)	-0.00418 (0.00471)	0.00268 (0.00361)
Overconfidence in 2011	-0.0281*** (0.00962)		0.0262** (0.0106)			
Minor overconfidence in 2011		-0.0239** (0.0109)		0.0159 (0.0122)	0.0181* (0.00982)	-0.00220 (0.00675)
Serious overconfidence in 2011		-0.0369** (0.0162)		0.0482*** (0.0172)	0.0318** (0.0148)	0.0165** (0.00791)
<i>Panel B. Those who have hypertension (N = 4,068)</i>						
Mean of Dep. Var.	0.93		0.03		0.05	0.97
Overconfidence in 2011	-0.0333*** (0.0102)		0.0314*** (0.0112)			
Minor overconfidence in 2011		-0.0268** (0.0113)		0.0191 (0.0124)	0.0207** (0.0101)	-0.00161 (0.00678)
Serious overconfidence in 2011		-0.0476*** (0.0170)		0.0586*** (0.0184)	0.0427*** (0.0155)	0.0159* (0.00897)

Notes: Data source is CHARLS 2011-2013. The overconfidence measure here is an indicator for under-diagnosed hypertension (Yes = 1). The minor and serious overconfidence measures are indicators for under-diagnosed minor and serious hypertension, respectively. Covariates in all regressions include dummies for gender, age in years, type of *hukou* (urban/rural), and counties. Robust standard errors in parentheses.

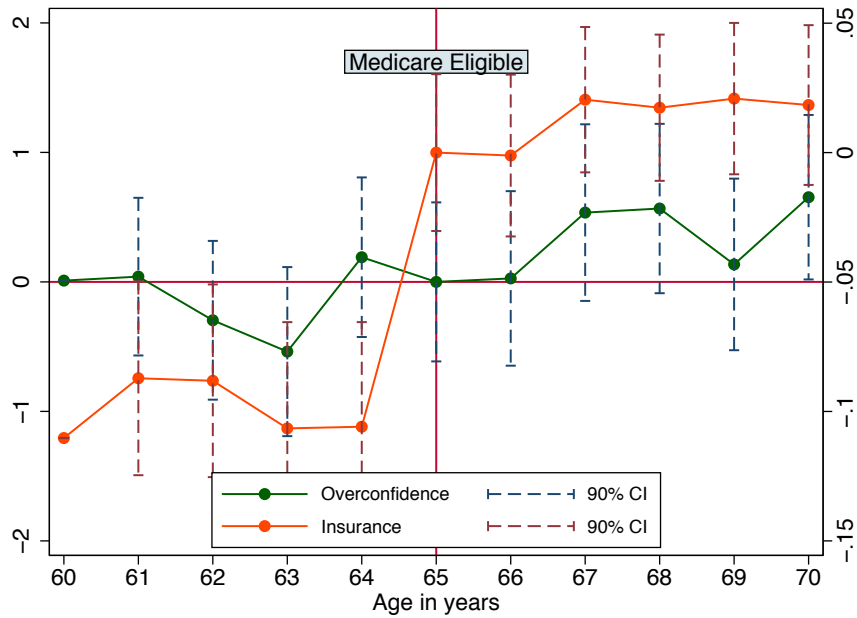
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 1: Effects of Medicare-eligibility on times of doctor visits and overconfidence

(a) Effect of Medicare-eligibility on times of doctor visits



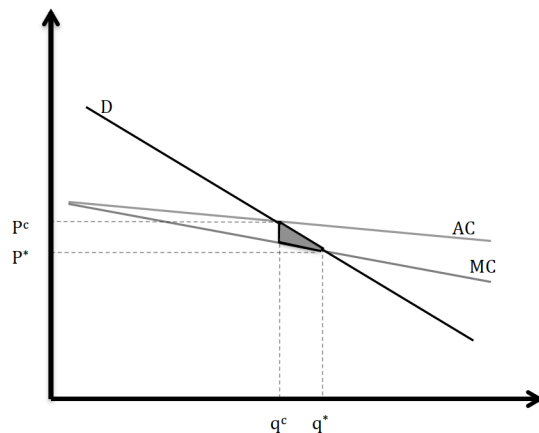
(b) Effect of Medicare-eligibility on overconfidence



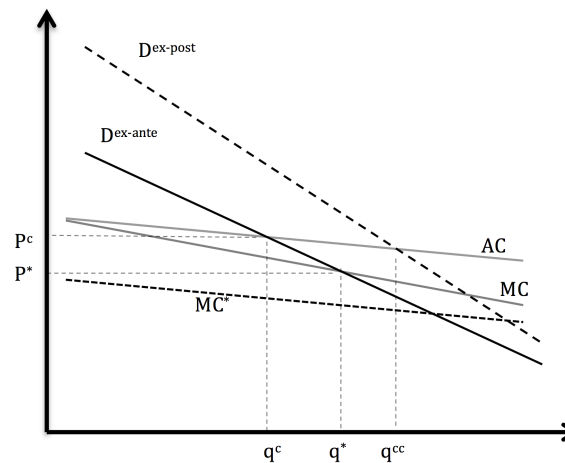
Notes: Data source is NHANES 1999-2011. The sample is restricted for only those aged between 60 and 70. Coefficients and 90% CI are reported. Reference group is those aged 65. Covariates include dummies for gender, education level, country of birth, ethnicity and survey years.

Figure 2: Equilibrium and welfare loss in Insurance market with Overconfidence

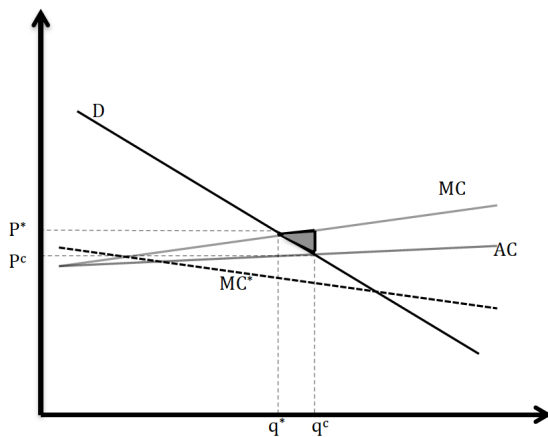
(a) The original model: Inefficient pricing and welfare loss



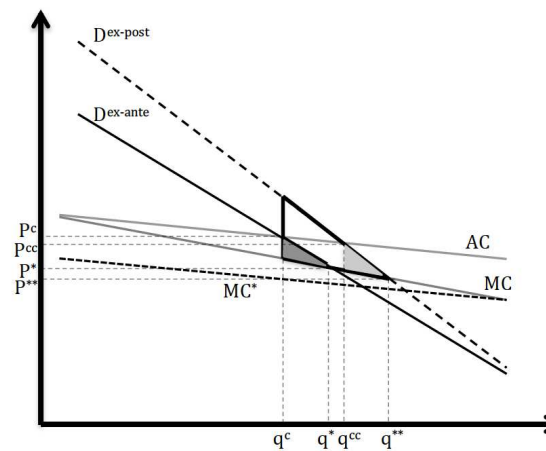
(b) Model with overconfidence: Explanation to insufficient insurance participation



(c) Model with overconfidence: Explanation to advantageous selection



(d) Model with overconfidence: Welfare Implications



Online Appendix

Appendix A: Height Shrinkage, SES and Health

A.1 Height Shrinkage Estimation

The first step is to estimate biological height shrinkage and compare it to reported height shrinkage. Following the methodology in Huang et al. (2013), we choose those aged between 30 and 40 the in NHANES and regress their measured height on limb length(s) and use the prediction as a measure of pre-shrinkage height for those aged over 60. This is a good measure because the limbs generally do not shrink as people age.¹ The results are reported in Appendix Table A1. Note that the R-squares in columns 1 and 3 are over 0.6, indicating fairly precise estimation. The age terms and the interactions between age and limb lengths are also jointly insignificant (not reported), indicating the ratio of limb lengths to height does not change by cohorts.

[Table A1 about here]

A.2 Associations of Shrinkage with SES and Health

We then use the coefficients combined with the limb lengths of those aged over 60 to predict the pre-shrinkage height. The biological height shrinkage is defined as the difference between pre-shrinkage height and measured height in the survey. Then we establish the relationship of biological height shrinkage with a series of health measures in Table A2, including self-reported health, reported ADLs and IADLs, hypertension and cognitive abilities. Note that the ADLs, IADLs and cognitive abilities are standardized to zero mean and unit standard deviation. The covariates include indicators for gender, marital status, working status, age in years, ethnicity, birth country and survey years, the same with Table 1 and Table 2. Robust

¹Lower leg length (Chumlea et al., 1985; Chumlea et al., 1998), arm span from roughly the shoulder to the wrist (Kwok and Whitelaw, 1991; Kwok et al., 2002), total arm length (Mitchell and Lipschitz, 1982; Auyeung et al., 2001), upper arm or humeral length, tibia length (Haboubi et al., 1990) and bula length (Auyeung et al. (2001)) have all been used to estimate pre-shrinkage stature.

standard errors are in parentheses. For all the health measures, the higher value shows unhealthier status. Therefore, the negative coefficients on education indicate that higher education are significantly correlated with better health and all the positive coefficients on biological height shrinkage imply that more shrinkage is consistently associated with worse health for all measures.

[Table A2 about here]

A.3 Comparison between Reported and Biological Height Shrinkages

NHANES asked respondents about their height when they were 25 (*Subjective* Pre-shrinkage height) and height as of now (*Subjective* height) respectively. We take the difference to derive the reported height shrinkage. Because the overconfidence in the paper is measured by the difference between reported and biological height shrinkage, this section compares the reported shrinkage and biological shrinkage estimated above to shed some light on the overconfidence measure.

Shrinkage over Age Figure A1 plots the different heights over age, by gender and whether it is subjective (reported) or objective (biological) measure. Top two plot the heights for men and the bottom two do those for women; the left two plot those for subjective measures and the right two for objective ones. In each figure, the diamond points are the mean values of the pre-shrinkage height for each age in the corresponding sample; and the circle points are those of current height. The smoothed lines show the aging trends for each particular measure. One important phenomenon is that the gap between pre-shrinkage height and the current height is widening as aging for all the figures, indicating that height is shrinking in the aging process. Another important point is that the magnitude of the gap is much larger for the biological measures, indicating that people may not fully recognize their height shrinkage. In particular, people overreport both their pre-shrinkage height and current height. Assuming the report bias is fixed for a particular person in the two measures, we are able to cancel

the invariant subjective reported bias out by taking the difference between reported current height and reported pre-shrinkage height. That's the reason we argue that unrecognized height shrinkage may be a good measure for overconfidence.

[Figure A1 about here]

Shrinkage by Education Since Huang et al. (2013) found that the higher education is negatively associated with biological shrinkage, we also investigate the shrinkage-education gradient here to check this consistency and also make a comparison with the results from reported height shrinkage. In practice, we use the shrinkage (biological or reported) as the dependent variable and education level indicators (i.e. fewer than 12 years, 12 years and more than 12 years) as the independent variables, with controlling for dummies for gender, marital status, age in years, ethnicity, birth country and survey years. Figure A2 report the education gradients in biological and reported shrinkage, respectively.² First, the reported height shrinkage is lower than the biological one for each education group, suggesting prevalent ignorance on the height shrinkage among the population. For those with less than 12-year education, the estimated biological shrinkage is 2.5 cm but they only report 0.9 cm. Second, people with higher education report more shrinkage but they actually shrink less. The difference between reported and biological shrinkages is 1.6 cm for the lowest education group but only 0.3 cm for the highest education group. This finding suggests people of higher education recognize more health problems but they are actually in better health status.

[Figure A2 about here]

Associations with Bone Mineral Density (BMD) Since height shrinks due to osteoporosis, the estimated height shrinkage should be strongly correlated with the bone mineral density (BMD). NHANES collected the BMD information through the Dual-energy x-ray

²Specifically, we report the summation of the constant term and coefficient on the particular education dummy.

absorptiometry from 1999 to 2009.³ Since the position examined and the methodology are different, we use the BMD-scarce z-score for each year in our analysis. In practice, we first regress the shrinkage (biological and reported, respectively) on a set of covariates including education and those in Figure A2, and keep the residuals. Then we regress the BMD-scarce z-score on the same covariates and also keep the corresponding residuals. We repeat the process for biological and reported shrinkages for each gender sample, respectively. In the end, we use kernel-weighted local polynomial smoothing to plot the corresponding residuals of height shrinkage against those of BMD, by gender and whether the shrinkage is reported or biological, and show them in Figures A3a-A3d. The scale of y-axis is consistent for all figures. Figure A3a and A3b show that biological height shrinkage is strongly correlated with BMD scarce for both genders. In contrast, Figure A3c and Figure A3d show that the reported height shrinkage is not as strongly correlated with BMD. These results suggest that the height shrinkage is a valid biomarker for osteoporosis, and that people may fail to realize how much they shrink and are overconfident as we defined.

[Figures A3a-A3d about here]

Appendix B: Overconfidence and Health Measures

This section investigates the correlation of overconfidence constructed in the paper with a series of health outcomes. We use three subjective health measures (i.e. self-reported health, reported ADLs and reported IADLs) and two objective health measures (i.e. hypertension and cognitive function), conduct the regressions of equation (1), and report the OLS estimates in Table B1. The higher value means worse health for all the dependent variables. We find that the overconfidence measure (unrecognized height shrinkage) is correlated with better perceived/subjective health but worse actual/objective health measures. Note the P-values for the two variables are small enough to reject the null hypothesis; the t-statistics

³Details can be found here: For the data before 2005, <http://www.cdc.gov/Nchs/Nhanes/dxx/dxa.htm>; for the data afterwards, please refer to: http://www.cdc.gov/Nchs/Nhanes/2007-2008/DXXFEM_E.htm.

across the columns may be underestimated due to correlation between overconfidence and biological shrinkage.

[Table B1 about here]

Appendix C: Other Overconfidence Measures and Robust Results in NHANES

This section investigates the associations of other overconfidence measures in NHANES with SES, health and insurance participation, and further checks whether the Medicare-eligible at age 65 has any effects on the overconfidence measures examined here.

C.1 Associations of Other Overconfidence measure with Health and SES

Table C1 reports the OLS estimates when the overconfidence measures are unrecognized height shrinkage (in ratios) and under-diagnosed hypertension.⁴ Consistent with the results in the main text, the bad health measured by BMD scarce and hypertension are positively correlated with the overconfidence measure, suggesting that the unhealthier people do not recognize their unhealthiness more. Education level is negative correlated with overconfidence when the measure is unrecognized height shrinkage in ratios but has no significant correlation with under-diagnosed hypertension. One possible reason is that individual overconfidence may be differential for different health measures; and the other explanation could be that those with hypertension are more “selected” among those with high education in the NHANES data.⁵

[Table C1 about here]

⁴The unrecognized height shrinkage is defined as $\frac{height_{bio}^{pre}}{height_{bio}^{cur}} / \frac{height_{rep}^{pre}}{height_{rep}^{cur}}$, where the superscripts *pre* and *cur* are short for “pre-shrinkage” and “current”, respectively; and the subscripts *bio* and *rep* are short for “biological” and “reported”, respectively.

⁵We also conduct another regression without health controls and find that higher education is negatively correlated with under-diagnosed hypertension. The coefficients are not significant as well. These results are available upon request.

C.2 Associations of Other Overconfidence Measures with Insurance Participation

Table C2 further investigates the associations between insurance participation and the overconfidence measures. The columns report the results for different types of insurance and panels for different overconfidence measures or samples. Note that all the coefficients are negative, which is fairly consistent with the Table 3 in the main text. For another, the coefficient on under-diagnosed hypertension is large in magnitude.

[Table C2 about here]

C.3 Does Medicare-eligibility affect the under-diagnosed hypertension?

Figure C1 follows the strategy in the main text and plots the overconfidence measured by under-diagnosed hypertension changing over age. Consistent with the unrecognized height shrinkage, we do not find a significant change in overconfidence at age 65, though the insurance participation rate increased by over 15 percentage points.

[Figures C1 about here]

Appendix D: Consistent Results from CHARLS

D.1 Associations of Overconfidence with Lower SES

We use the cross-sectional data in 2011, the first wave of the survey, to conduct the following equation:

$$Overconf_{ij} = \beta_0 + \beta_1 Hyper_{ij} + Z_{ij} + Cty_j + \epsilon_i$$

where the dependent variable, $Overconf_{ij}$, is an indicator for the individual i in county j has under-diagnosed hypertension or not. This is the overconfidence measure we use for CHARLS data. All the other covariates are consistent with the equation (3) in the main text. Table D1 report the OLS estimates for education, type of *hukou* and hypertensive status. The

positive correlation between under-diagnosed hypertension and actual hypertension is obvious,⁶ and thus we only keep those with hypertension, conduct the same analysis and report the results in column 2. Both columns show that higher education or urban *hukou* is correlated with less under-diagnosed hypertension. Particularly, among those with hypertension, urban *hukou* is correlated with 6.6 percentage points lower under-diagnosed hypertension.

[Table D1 about here]

D.2 Effects of Exact Health Information, by Type of Hukou

The health insurance programs in China generally have very different implementations. In 2011, the insurance coverage rate for rural people is 95% and that for urban people is 88%. The insurance coverage rate is higher for rural areas due to the New Cooperative Medical Scheme (NCMS) initiated in 2003. In 2011, the coverage rate of NCMS is 94% for those with rural *hukou*. Among those who have health insurance, the NCMS coverage rate is 98% and the rest are mainly covered by private insurance. The major two insurance products in the urban regions are the Urban Employee Basic Medical Insurance (UEBMI) and the Urban Resident Basic Medical Insurance (URBMI). The coverage rate for the former is about 56% and the latter 24%. Then is the government medical insurance (about 12%) mainly for government employees and state personnels, which may be viewed as an upgraded version of UEBMI. The last 8% is private insurance and “urban and rural resident medical insurance,” which is a new insurance product as a combination of URBMI and NCMS.

Therefore, we further conduct the *hukou* type specific analysis with expectation that the results above should hold in particular for the people with rural *hukou* because the NCMS is generally voluntary and the policy allows individuals to either enroll or opt out at any time. Table D2 reports the estimates. In general, the results show that those with under-diagnosed hypertension in the first wave is associated with lower insurance participation in

⁶Because all with under-diagnosed hypertension suffer from hypertension in actual and that it is impossible for those who do not suffer hypertension in actual to report hypertension.

2011, especially for those with rural *hukou*. Consistent with the expectation, the information provided in the first wave increases the insurance participation rate in the following wave for those with rural *hukou*, and most of the take-up can be explained by the newly enrolling.

[Table D2 about here]

References

- Auyeung, Tung Wai, Jenny Shun Wah Lee et al.**, “Estimation of height in older Chinese adults by measuring limb length,” *Journal of the American Geriatrics society*, 2001, 49 (5), 684–685.
- Chumlea, William Cameron, Alex F Roche, and Maria L Steinbaugh**, “Estimating stature from knee height for persons 60 to 90 years of age,” *Journal of the American Geriatrics Society*, 1985, 33 (2), 116–120.
- Chumlea, WM Cameron, Shumei S Guo, Kevin Wholihan, David Cockram, Robert J Kuczmarski, and Clifford L Johnson**, “Stature prediction equations for elderly non-Hispanic white, non-Hispanic black, and Mexican-American persons developed from NHANES III data,” *Journal of the American Dietetic Association*, 1998, 98 (2), 137–142.
- Haboubi, NY, PR Hudson, and MS Pathy**, “Measurement of height in the elderly,” *Journal of the American Geriatrics society*, 1990, 38 (9), 1008–1010.
- Huang, Wei, Xiaoyan Lei, Geert Ridder, John Strauss, and Yaohui Zhao**, “Health, height, height shrinkage, and SES at older ages: evidence from China,” *American Economic Journal: Applied Economics*, 2013, 5 (2), 86–121.
- Kwok, Terry, E Lau, and J Woo**, “The prediction of height by armspan in older Chinese people,” *Annals of human biology*, 2002, 29 (6), 649–656.
- Kwok, Timothy and Michael N Whitelaw**, “The use of armspan in nutritional assessment of the elderly,” *Journal of the American Geriatrics Society*, 1991, 39 (5), 492–496.
- Mitchell, CO and David A Lipschitz**, “Arm length measurement as an alternative to height in nutritional assessment of the elderly,” *Journal of parenteral and enteral nutrition*, 1982, 6 (3), 226–229.

Table A1.Pre-shrinkage Height Estimation using Limb Lengths, Age 30-40, By Gender

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Measure Height (cm)					
Sample	Male sample			Female sample		
(Upper) Arm length	-1.043 (1.052)	-1.078 (1.052)	-0.825 (1.133)	3.284*** (0.839)	3.285*** (0.840)	3.348*** (0.919)
(Upper) Leg length	2.238*** (0.652)	2.232*** (0.652)	2.645*** (0.742)	0.942* (0.570)	0.949* (0.570)	0.603 (0.657)
Arm length square	4.759*** (1.820)	4.780*** (1.822)	5.088*** (1.830)	0.244 (1.629)	0.259 (1.634)	0.336 (1.685)
Leg length square	-0.284 (0.874)	-0.285 (0.873)	-0.410 (0.881)	2.317*** (0.687)	2.322*** (0.687)	2.392*** (0.694)
Arm*leg	-2.732 (1.940)	-2.703 (1.941)	-2.699 (1.957)	-5.215*** (1.828)	-5.245*** (1.830)	-5.326*** (1.899)
Age		0.267 (0.722)	1.131 (0.858)		0.357 (0.622)	0.0417 (0.770)
Age square		-0.324 (1.015)	-0.265 (1.015)		-0.510 (0.879)	-0.432 (0.885)
Age * Arm length			-0.0139 (0.0145)			-0.00246 (0.0140)
Age * Leg length			-0.00873 (0.0103)			0.00906 (0.00934)
Observations	2,575	2,575	2,575	2,947	2,947	2,947
R-squared	0.668	0.668	0.669	0.619	0.619	0.619
F-statistic tests						
P-values for limb length	0.000	0.000	0.000	0.000	0.000	0.000
P-values for limb square	0.052	0.051	0.029	0.000	0.000	0.000
P-values for age		0.413	0.146		0.834	0.737
P-values for age*limb			0.167			0.581

Notes: Data source are NHANES and CHARLS for Panel A and Panel B, respectively.

Table A2. Associations between Height Shrinkage and Health, Age 55+

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Subjective Measures (Higher for unhealthier)			Objective Measures (Higher for unhealthier)	
	Self-reported Health (1-5)	Reported ADLs	Reported IADLs	Hypertension (Yes = 1)	Cognitive impairment z-score
Height shrinkage (cm)	0.0121*** (0.00283)	0.0115*** (0.00281)	0.00966*** (0.00300)	0.00450*** (0.00137)	0.0156*** (0.00406)
<i>Education level: Reference group is fewer than 12 years</i>					
High School Grad	-0.317*** (0.0378)	-0.175*** (0.0411)	-0.290*** (0.0438)	-0.0122 (0.0173)	-0.572*** (0.0546)
Some College or above	-0.702*** (0.0341)	-0.306*** (0.0369)	-0.413*** (0.0394)	-0.0574*** (0.0156)	-0.879*** (0.0496)
Observations	9,879	9,883	9,883	9,640	2,403
R-squared	0.128	0.061	0.087	0.069	0.418

Notes: Data source is NHANES 1999 - 2011. Those aged over 55 (included) are kept. Covariates include indicators for gender, marital status, age in years, ethnicity, birth country and survey years. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table B1. Overconfidence and Health Measures

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Subjective Measures (Higher for unhealthier)			Objective Measures (Higher for unhealthier)	
	Self-reported Health (1-5)	Reported ADLs	Reported IADLs	Hypertension (Yes = 1)	Cognition impairment z-score
Overconfidence	-0.0154** (0.00627)	-0.0305*** (0.00754)	-0.0248*** (0.00784)	0.00805 (0.00971)	0.00145 (0.00293)
Objective poor health (Biological Shrinkage)	0.0268*** (0.00661)	0.0409*** (0.00775)	0.0337*** (0.00797)	0.00808 (0.0100)	0.00324 (0.00311)
Observations	9,845	9,849	9,849	9,607	2,393
R-squared	0.168	0.093	0.111	0.069	0.416
P-values of F-tests	0.000	0.000	0.000	0.00301	0.000442

Notes: Data source is NHANES 1999-2011. The overconfidence measure is unrecognized height shrinkage in centimeters. Covariates controlled for in all columns are the same with those in Table 2. Robust standard errors are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table C1. Overconfidence Measure, SES and Health

Dependent variable	(1) Unrecognized height shrinkage (in ratios)	(2) Under-diagnosed Hypertension (Yes =1)
<i>Objective Health Measures</i>		
BMD scarce z-score (Higher for unhealthier)	0.331*** (0.0588)	0.0199*** (0.00409)
Hypertension (Yes = 1)	0.288*** (0.0804)	0.107*** (0.00561)
<i>Education level: Reference group is fewer than 12 years</i>		
High School Grad	-0.277*** (0.106)	0.00398 (0.00846)
Some College or above	-0.731*** (0.0956)	0.00423 (0.00731)
Observations	9,874	9,633
R-squared	0.046	0.029

Notes: Data source is NHANES 1999-2011. Covariates controlled for in all regressions are the same with those in Table 2.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table C2. Overconfidence and Insurance Participation

VARIABLES	(1) Any Health Insurance Participation (Yes = 1)	(2) Having Medicare (Yes = 1)	(3) Having Medicaid (Yes = 1)	(4) Having Private Health Insurance (Yes = 1)
<i>Panel A: Unrealized shrinkage in ratios, full sample</i>				
Mean of Dep. var.	0.929	0.492	0.0459	0.655
Overconfidence	-0.589*** (0.158)	-0.714*** (0.264)	-0.239 (0.182)	-0.681* (0.389)
Observations	12,385	12,385	12,385	12,433
R-squared	0.115	0.653	0.070	0.126
<i>Panel B: Under-diagnosed hypertension, full sample</i>				
Overconfidence	-0.0319** (0.0130)	-0.0241** (0.0115)	-0.0237*** (0.00788)	-0.0181 (0.0223)
Observations	12,051	12,051	12,051	12,094
R-squared	0.117	0.653	0.071	0.126
<i>Panel C: Under-diagnosed hypertension, hypertension sufferers</i>				
Mean of Dep. var.	0.943	0.569	0.0556	0.625
Overconfidence	-0.0348*** (0.0130)	-0.0225* (0.0116)	-0.0240*** (0.00790)	-0.0207 (0.0223)
Observations	7,297	7,297	7,297	7,320
R-squared	0.102	0.603	0.088	0.115

Notes: Data source is NHANES 1999-2011. Covariates controlled for in all regressions are the same with those in Table 2. Robust standard errors are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table D1. Overconfidence and SES in China

	(1)	(2)
Sample	Full sample	Those with Hypertension
Dependent variable	Overconfidence (Under-diagnosed Hypertension)	
Mean of Dep. Var.	0.0948	0.277
<i>Education level: No formal education is reference group</i>		
Primary school	-0.0191*** (0.00665)	-0.0540*** (0.0188)
Junior high or above	-0.0117 (0.00783)	-0.0446* (0.0231)
Urban Hukou	-0.0256*** (0.00822)	-0.0658*** (0.0217)
Hypertension (Yes = 1)	0.280*** (0.00708)	
Observations	11,902	4,068
R-squared	0.245	0.135

Notes: Data source is CHARLS 2011-2013. The overconfidence measure here is an indicator for under-diagnosed hypertension (Yes = 1). Covariates in both regressions include dummies for gender, age in years, and counties. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

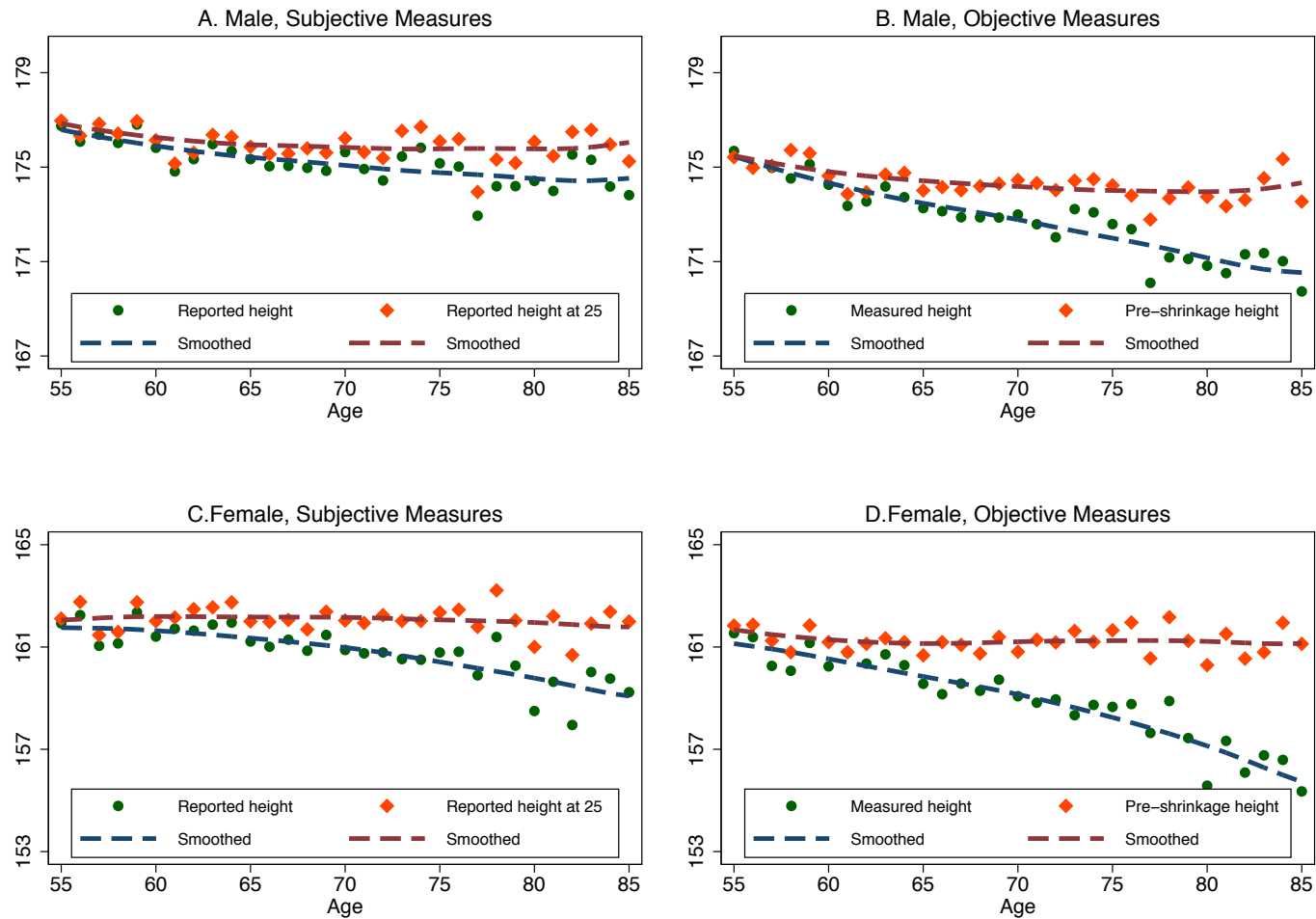
Table D2. Under-diagnosed Hypertension, Exact Health Information Provided and Insurance Participation, by Type of hukou

VARIABLES	(1) Insurance Participation in 2011 (Yes = 1)	(2) Insurance Enrollment Change (-1, quit; 0, same; 1, enroll)	(3)	(4) Insurance newly enroll (Yes = 1)	(5) Insurance no quit (Yes = 1)
<i>Panel A: Those with Urban hukou (N=879)</i>					
Mean of Dep. Var.	0.891		0.0466	0.0785	0.968
Overconfidence in 2011	-0.0175 (0.0297)	-0.0124 (0.0310)			
Minor overconfidence in 2011			-0.0323 (0.0341)	-0.0141 (0.0265)	-0.0181 (0.0197)
Serious overconfidence in 2011			0.0355 (0.0560)	0.00525 (0.0454)	0.0302 (0.0316)
<i>Panel B: Those with Rural hukou (N=3,189)</i>					
Mean of dependent var.	0.945		0.0210	0.0464	0.975
Overconfidence in 2011	-0.0364*** (0.0107)	0.0433*** (0.0119)			
Minor overconfidence in 2011			0.0347*** (0.0132)	0.0326*** (0.0110)	0.00203 (0.00695)
Serious overconfidence in 2011			0.0618*** (0.0194)	0.0479*** (0.0164)	0.0139 (0.00933)

Notes: Data source is CHARLS 2011-2013. The overconfidence measure here is an indicator for under-diagnosed hypertension (Yes = 1). The minor and serious overconfidence measures are indicators for under-diagnosed minor and serious hypertension, respectively. Covariates in all regressions include dummies for gender, age in years, and counties. Robust standard errors in parentheses.

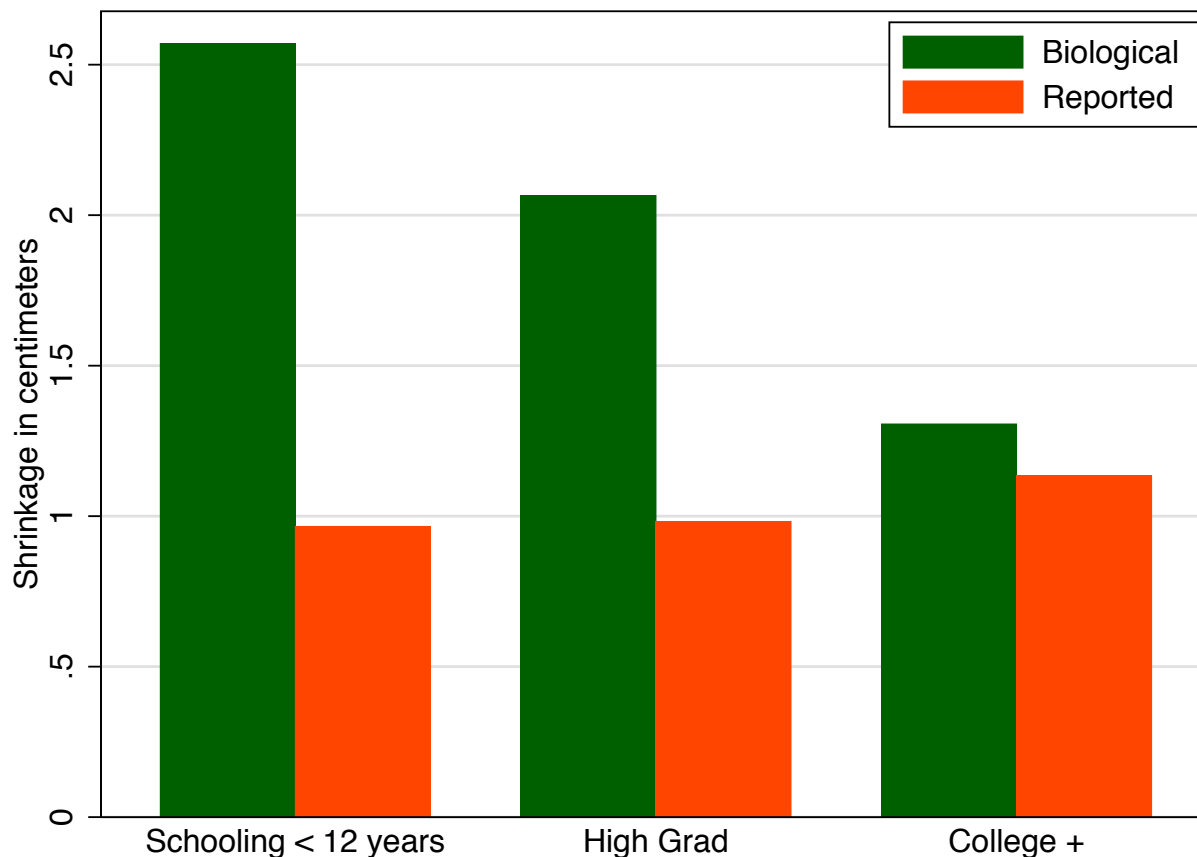
*** p<0.01, ** p<0.05, * p<0.1

Figure A1. Pre-shrinkage Height and Current Height over Age, by gender and objective/subjective measure



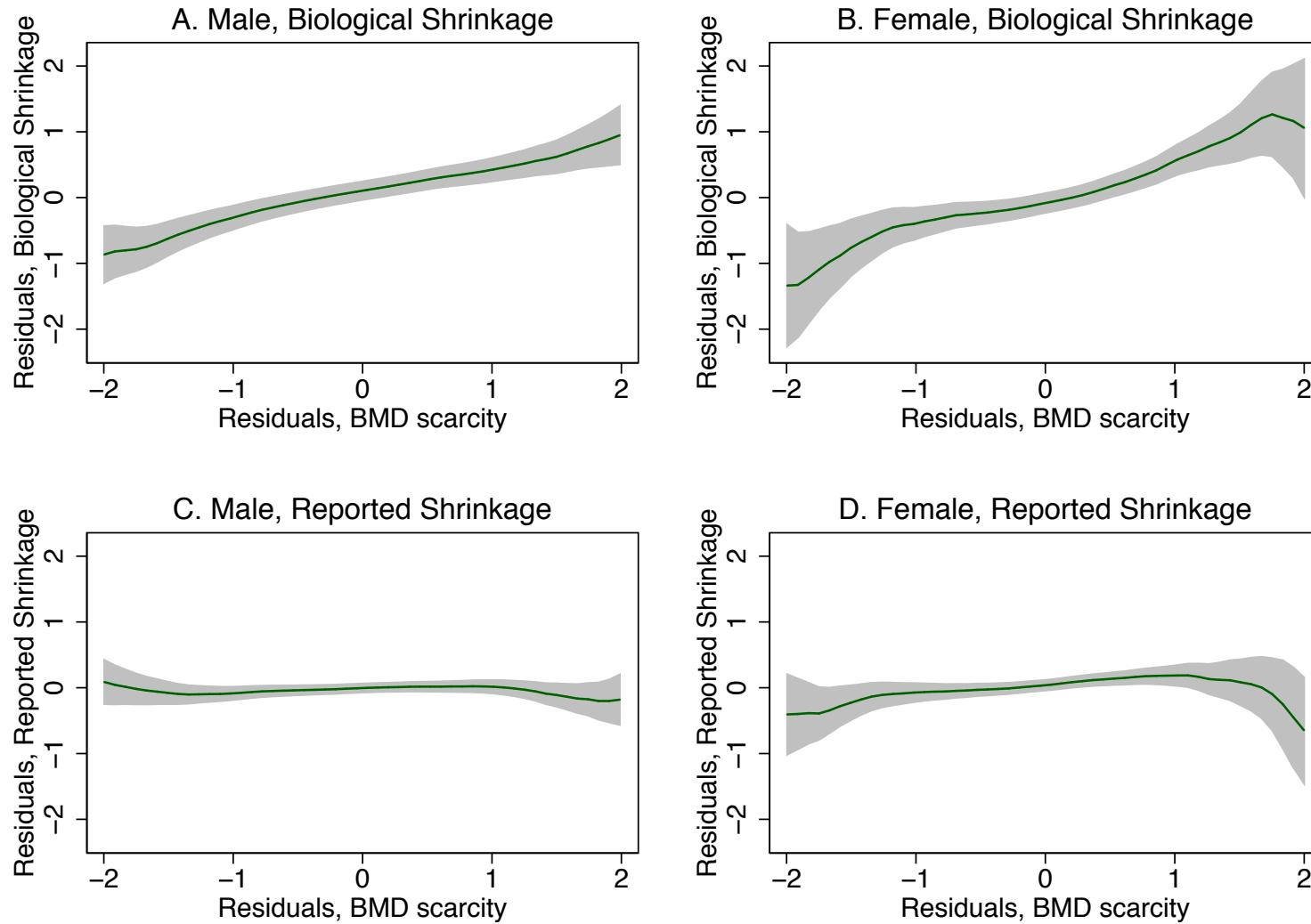
Notes: Data source is NHANES 1999-2011. The reported height at 25, reported current height and biological height are provided directly by the data. The biological pre-shrinkage height is estimated by the coefficients in columns 1 and 4 of Table A1.

Figure A2. Height Shrinkage-Education Gradients, by objective/subjective measure



Notes: Data source is NHANES 1999-2011. We use the shrinkage (biological or reported) as the dependent variable and education level indicators (i.e. fewer than 12 years, 12 years and more than 12 years) as the independent variables, with controlling for dummies for gender, marital status, age in years, ethnicity, birth country and survey years. The values reported are derived by the coefficients on education plus the constant term in each regression.

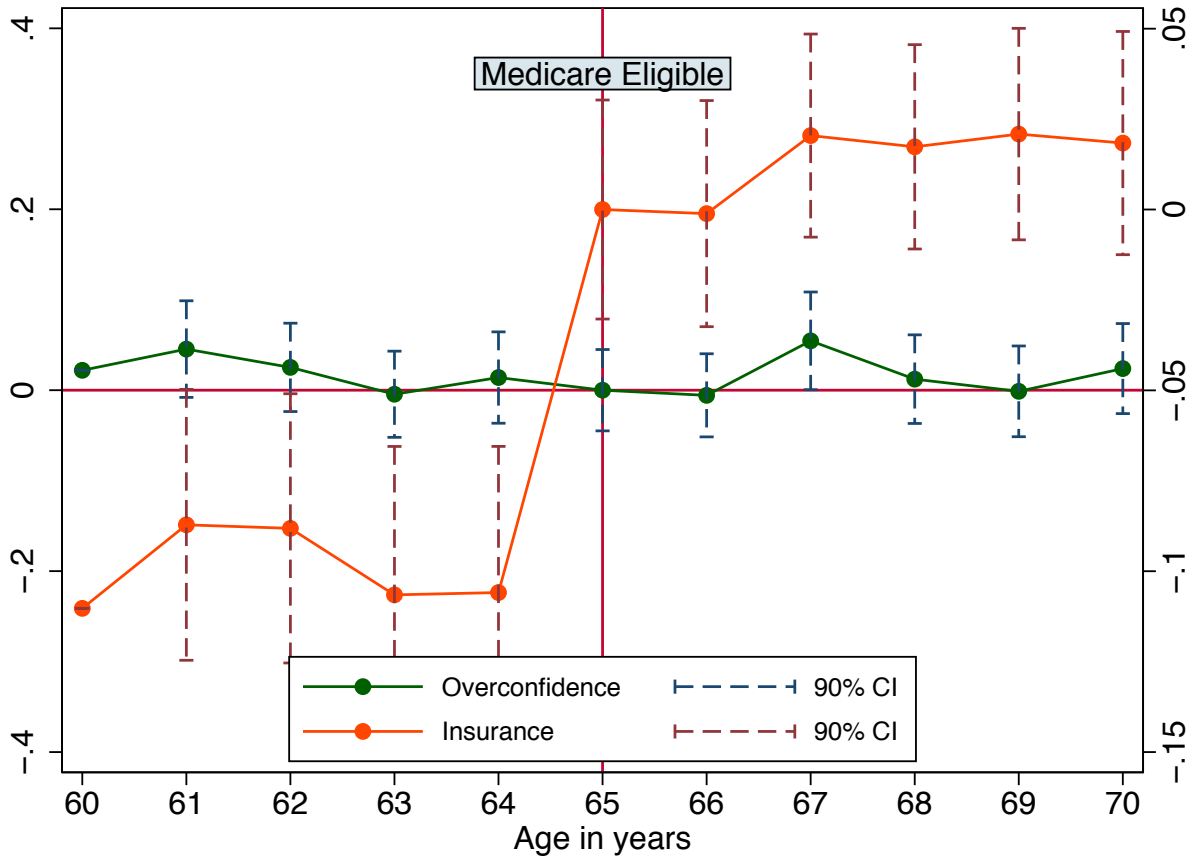
Figure A3. Height Shrinkage over Bone Mineral Density (BMD) Scarcity, by objective/subjective measure



61

Notes: Data source is NHANES 1999-2011. We use the shrinkage (biological or reported) and BMD scarcity as the dependent variable and regress them on dummies for gender, marital status, age in years, ethnicity, birth country and survey years. Keep residuals for each dependent variable then plot the residuals of height shrinkages against those of BMD, by gender and whether shrinkage is objective/subjective measure.

Figure C1. Effect of Medicare-eligibility on overconfidence measured by under-diagnosed hypertension



Notes: Data source is NHANES 1999-2011. The overconfidence measure is under-diagnosed hypertension (Yes = 1). The sample is restricted for only those aged between 60 and 70. Coefficients and 90% CI are reported. Reference group is those aged 65. Covariates include dummies for gender, education level, country of birth, ethnicity and survey years.