IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12524

Do Pension Benefits Accelerate Cognitive Decline? Evidence from Rural China

Plamen Nikolov
Alan Adelman

AUGUST 2019

DISCUSSION PAPER SERIES

IZA DP No. 12524

Do Pension Benefits Accelerate Cognitive Decline? Evidence from Rural China

Plamen Nikolov
State University of New York (Binghamton), Harvard Institute for Quantitative Social Science and IZA

Alan Adelman
State University of New York (Binghamton)

AUGUST 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany
Phone: +49-228-3894-0
Email: publications@iza.org
www.iza.org

ABSTRACT

Do Pension Benefits Accelerate Cognitive Decline? Evidence from Rural China*

Higher life expectancy and rapidly aging populations have led to the introduction of pension programs in developing countries in the last two decades. Using the introduction of a new public policy in China, we estimate the effects of pension benefits on individual cognition, measured by episodic memory and intact mental status, among individuals ages 60 and above. We find large and significant negative effects of the provision of pension benefits on cognitive functioning among the elderly. We find the largest effect of the program on delayed recall, a measure implicated in neurobiological research as an important predictor of the onset of dementia. We show that the program leads to more negative impacts among the female sample. Our findings support the mental retirement hypothesis that decreased mental activity results in atrophy of cognitive skills. We show that retirement plays a significant role in explaining cognitive decline at older ages.

JEL Classification: J14, H55, H75, J26, J24, D91, O12, N35, O10

Keywords: life-cycle, cognitive functioning, cognition, aging, health, mental retirement, middle-income countries, developing countries, China

Corresponding author:
Plamen Nikolov
Department of Economics
State University of New York
4400 Vestal Parkway East
Binghamton, NY 13902
USA
E-mail: pnikolov@binghamton.edu

* We thank Matthew Bonci for outstanding research support during this project. We thank Subal Kumbhakar, Eric Edmonds, Wei Huang, Livia Montana, Susan Wolcott, Evan Riehl, Nusrat Jimi, Wei Xiao, Xu Wang, Solomon Polachek, Petra Moser, James MacKinnon, and Morten Nielsen for constructive feedback and helpful comments.

I. Introduction

Life expectancy in the developed world has dramatically improved in the last century. Developing countries have also recently benefited from this pattern of rising life expectancy. Over the last 50 years, life expectancy at birth has increased globally by 20 years (WHO 2003). However, the greatest longevity gains have been in developing countries; between 1950 and 2002, longevity gain in developing countries alone has been 26 years (in low-income developing countries). The early 1970s witnessed world population growth plateau at 2 percent annually. As the decade's conclusion grew nearer, the growth rate decreased to 1.1 percent, largely due to declining fertility rates (UN 2019, World Bank 2017). Declining fertility coupled with higher longevity has shifted the population structure toward the elderly, whose share of the population has increased dramatically. The increase of the elderly as a fraction of total population has been especially pronounced in developing countries; the fraction of individuals aged 65 to 85 years increased globally from 13 percent to 33 percent between 1950 and 2010 (World Bank 2017). The same number for this period in developing countries increased from 8 percent to 29 percent (UN 2013). Globally, the number of persons aged 80 or over is projected to triple between 2017 and 2050, from 137 million to 425 million; the same population group is projected to reach 909 million, seven times its value in 2017 (UN 2017). The largest demographic source will be countries in Asia and Latin America, where it is predicted that this age group will quadruple in number between 2017 and 2050 (US Census 2016). This rapidly increasing demographic trend has generated an urgent need for policymakers to introduce new and sustainable pension systems.

The problem is particularly exacerbated in China, as the country is aging more rapidly than almost any country in recent history. For example, in the latter part of the twentieth century, life expectancy at birth in China increased from 40 to 70 years. China's dependency ratio, the difference between those not in the labor force and those who are working, for retirees could rise as high as 44 percent by 2050 (UN 2017). To tackle these demographic challenges, the country introduced the New Rural Pension Scheme (NRPS) in 2009. The defined contribution program was made available to all rural residents over the age of 16 years. A rapidly increasing aging population, erosion of traditional sources of old-age security, and a large urban-rural income gap have been important factors in the development of China's NRPS, first introduced in 2009. The program was introduced in a recent push for more social pension programs in many developing countries due to demographic pressures and concerns about old-age poverty in the last decade (Holzman, Robalino, and Takayama 2009).[1,2]

In this paper, we estimate the causal effect of the NRPS program on cognition among individuals ages 60 and above. The expansion of the program affected an easily identifiable group, as the policy was introduced only in select areas. Our identification employs a triple difference (DDD) strategy. We exploit the staggered policy implementation between 2009 and 2013 and compare the cognitive outcomes of individuals 60 years and older who live in areas that implemented the NRPS program to the cognitive outcomes of individuals of the same age group who live in areas that did not implement the NRPS program. Given the similarity across countries regarding the age for individuals' retirement, our quasi-experiment closely simulates the case of the introduction of retirement in other countries. Our analysis relies on a new data source—the Chinese Health and Retirement Longitudinal Survey (CHARLS)—that is nationally representative of individuals ages 45 and above within the Chinese population. The survey, a sister survey of the US Health and Retirement Survey, directly tests cognition with a focus on two important cognitive domains: *episodic memory*, and components of *intact mental status*. Examining the effect on cognition for the older population in a country such as China, may be especially important given the country's lack of any intermediary market institutions to assist with financial decisions related to income security or health care provision. The triple difference

[1] Feldstein and Liebman (2002) and Cutler and Johnson (2004) overview social insurance programs in developed countries.

[2] The primary factors that precipitated the introduction of the program was demographic and economic challenges: population aging (Bloom and McKinnon 2014), a large rural fraction of the population, and rising income inequality (Sabates-Wheeler and Koettl 2010), and weak local institutions to support social protection on their own (Musalem and Ortiz 2011). The program was financed from comingled funds: the first source of the program cost was financed by the local and federal funds; the second source came from individual contributions. The central government subsidizes 100 percent of the program cost in provinces with low fiscal capacity, whereas the federal subsidies constitute only 50 percent of the total funding in wealthier provinces with high fiscal capacity.

empirical design allows us to attribute any changes in cognition in the treatment regions to the presence of the NRPS program.

Previous empirical research documents a causal link between pension programs and early cognitive decline but these studies rely on data from high-income countries (Adam et al. 2007, Rohwedder and Willis 2010, Bonsang et al. 2012, Coe at al. 2012, Mazzonna and Peracchi 2012, Bingley and Martinello 2013, de Grip et al. 2015). Using comparable survey data from the United States, England, and Europe, Rohwedder and Willis (2010) demonstrate that early retirement has a significant negative impact on the cognitive ability of people in their early 60s, and document a phenomenon that they call *mental retirement*.[34] This phenomenon relates to an early onset of retirement with accelerated cognitive decline among the elderly.

Our analysis yields several interesting results. First, the NRPS program has a significantly negative effect on several cognitive ability outcomes among individuals aged 60 or above. We find that the provision of pension benefits negatively influences immediate recall, delayed recall, and total word recall. For total word recall, the intent-to-treat effect for individuals aged 60 and over is a significant decline of 0.40 points, from a baseline average of 6.6 (on a 20-point scale). Second, we find that the provision of pension benefits leads to a much more considerable impact, compared to other cognition measures, on delayed recall. The "delayed recall" test records the number of words that the respondent remembers approximately five minutes after having heard a list of 10 nouns read aloud. This test is one of the most sensitive tests to distinguish the effects of normal aging from the symptoms of Alzheimer's disease (Laakso et al. 2000). Neurological research demonstrates that proxy measures of delayed recall memory are highly accurate detectors of dementia (Welsh et al. 1991). Third, we show that the program leads to faster and more negative impacts among the female sample. This finding may have alarming implications. The average performance on cognition tests for Chinese women is much lower than the performance of Chinese men; the gender difference is particularly

[3] Using data from the US, England, Canada and 11 European countries, Rohwedder and Willis (2010), Bonsang et al. (2012) and Adam et al. (2007) examine how retirement rates influence cognitive functioning and find a significant negative effect between retirement and cognitive functioning.[3] Conversely, Coe et al. (2012) find no conclusive evidence with data from the US Health and Retirement Survey (HRS).

[4] Other recent empirical studies also examine the effect of the NRPS on other individual or household-related outcomes. For example, Nikolov and Adelman (2009ab) examine the effects of the NRPS program on intergenerational transfers and health behaviors.

pronounced among older Chinese cohorts (Lei et al. 2012). Coupled with the fact that women have a longer life expectancy (Liu et al. 2009), a faster cognitive decline due to an earlier onset of retirement could be an additional contributor to a gender-based expansion of morbidity in older age and greater life expectancy with a disability among women (Wang 1993).

We contribute to the existing literature in at least four major ways. First, this study is the first, to the best of our knowledge, to examine the effects of pension participation on individual cognitive functioning in the context of a developing country.[5,6] China is a particularly suitable context given the size of its overall population and the size of its elderly population. China's population is aging rapidly. In 2007, approximately 11 percent of China's population was age 60 or over making up 21 percent of the world's elderly population (UN 2007). Our analysis focuses on China, the country with the largest population in the world, home to 1.4 billion people. Therefore, the implications of this study's findings are likely to affect a lot of people, which additionally underscores the importance of the findings from a welfare standpoint. The study setting is unique because we analyze data from China's rural areas, whose demographic and economic dynamics resemble the economies of low-income countries. Therefore, our findings have important implications for other low-income countries.

Second, we illuminate how program participation affects several cognitive domains, and we examine impacts on a wider set of cognitive domains than previously examined. Although some cognitive decline appears to be an inevitable byproduct of aging, faster onset of cognitive decline can have profound adverse consequences on various aspects of one's life—from financial planning for retirement (Banks and Oldfield 2007) and medical treatment adherence, to the planning of sequential activities (Fillenbaum et al. 1988). In this paper, we focus on proxies of cognition, specifically episodic memory, which neurobiology research documents to be particularly sensitive to the aging process. Several studies highlight that this domain is the first to exhibit setbacks as aging sets in (Souchay et al. 2000; Tulving and Craik 2000; Prull et al. 2000). The second reason relates to its provision of high individual variation across individuals as

[5] Recent studies examine the effect of retirement policies on health behaviors in the context of high-income countries (Eibich 2015) or developing economies (Nikolov and Adelman 2019b).
[6] Cheng et al. (2016) examine the health implications of the NRPS using a different survey, and do not explicitly test for cognitive functioning.

opposed to other cognitive measures. For example, the word learning and recall tasks do not exhibit floor or ceiling effects (excess of maximum or minimum values) and the individual distribution of the score does not exhibit extreme observation bunching around minimum and maximum values. Related to this, the CHARLS includes several cognitive measurements. We combine this into an aggregate cognition index. The use of an index of outcomes, a method based on the approach adopted by Kling, Ludwig and Liebman (2004) and Kling, Liebman and Katz (2007), addresses the possibility that the results are an artifact of multiple hypothesis testing and provides robust evidence of the global impact of the program.

Furthermore, this study uses data from the CHARLS, a survey harmonized with the US Health and Retirement Study (HRS) and other sister health surveys in high- and middle-income countries.[7] The survey harmonization of cognition measures across surveys can enable additional analyses and international comparisons across different country settings based on data from these retirement surveys. Our fourth contribution relates to the study's robust identification strategy exploiting the staggered implementation of a unique policy experiment. We estimate the causal effects of retirement program enrollment on cognitive functioning and we exploit plausibly exogenous variation in the district-level offering of a voluntary pension program.

The rest of the paper is organized as follows. Section II provides background for the study and the rural pension scheme, and summarizes the data. Section III presents the identification strategy. Section IV presents the results, while Section V reports additional robustness checks and bolsters the validity of the empirical approach. Section VI concludes.

---

[7] Started in 1992, the Health and Retirement Study is a biennial longitudinal survey. The main objective of the survey is to facilitate the interdisciplinary study of aging and retirement. The core component of the survey collects that on data on a wide array of topics, including current health, cognition, current labor market participation, employment history and subjective expectations about future events. Over the last three decades, it has collected information on more than 43,000 individuals in the U.S. Because of its successful implementation, the survey has become a model for similar surveys that specifically examine issues of health and retirement around the world. Currently, harmonized constructs on health and demographic information exists across eighteen longitudinal aging sister studies (e.g., Ageing and Retirement in Europe, the English Longitudinal Study of Ageing, Longitudinal Aging Study in India, Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community in South Africa, Costa Rican Longevity and Health Aging Study, etc.) around the world. Because data in these surveys is calibrated based on the U.S. HRS, they allow for analysis of data that is harmonized for cross-national comparisons. More information on this data project is available at https://hrs.isr.umich.edu/about/international-sister-studies.

II. Background and Data

A. China's New Rural Pension Scheme

Prior to the 1980s, China's public policies regarding its elderly population were largely decentralized. Although pension programs existed, they were initiated on an ad hoc basis, and managed and financed at the provincial level; there was a lack of a centralized policy. Vilela (2013) reviews the history and the evolution of China's pension policy since the establishment of the new People's Republic of China in 1949 up to 2013. The study posits that the country's policies toward its old-age segment have been moving away from its historical focus on formal-sector workers to a stronger emphasis on universal coverage of formal and informal workers alike. Vilela (2013) highlights three distinct historical phases of the country with regard to retirement policies: the "Iron Rice Bowl" (1949–1978), formal-sector pension reform and rural pension piloting (1978–2001), and a gear change in pension expansion (2003 to the present).

Weaknesses in the old-age pension system established in the 1950s began to surface as the country moved more aggressively towards market reforms and a market-oriented economy in the early 1980s. It became increasingly difficult for the old enterprise-based system to meet its financial obligations as the demographics of the Chinese population began to change. In addition to these factors, the pension system was limited primarily to government employees and party organizations or large urban collectives. Therefore, the system's coverage began to shrink in its coverage as other forms of employment began to grow, such as self-employment and private enterprises activities. The changing structure of China's economy and population led researchers and officials to examine the existing system and propose major changes.

In addition to the crumbling fiscal sustainability of the old decentralized pension systems, the demographics of the country were rapidly changing in the 1980s. China's one-child policy was introduced in 1979 in an effort to meet a population target of 1.2 billion by 2000. In addition to this target population size, the government expected zero population growth by 2000 and its targeted growth rate for the 1980s was between 0.5 percent and 1 percent. Because of the combination of considerably lower fertility due to the one-child policy and reduced mortality, the population structure in the country shifted towards older age groups and the end result was a rapidly aging population. The growth rate of the population in most age groups remained stable in the period from 1950 to 1980. This pattern produced an expansion of the age pyramid and resulted in the relative stability of the age-sex distribution of the population. However, from 1964 to 1982 the oldest age groups did experience a considerable proportional increase while the percentage distribution of the two youngest age groups declined substantially.

With the growing demographic changes, the need for the Chinese government to tackle the demographic challenges and old-age poverty in rural areas was growing. The government introduced a rural pension program in 1986 by piloting the rollout among rural residents. Financed by individual voluntary contributions and matching funds from local governments, the program covered state enterprise employees and individuals previously covered by the Basic Old-Age Insurance Scheme, a program mainly designed for urban employees (Liu and Sun 2016). Under the new system in the 1980s, the pension scheme introduced coverage quotas in urban and rural systems. Following a decade of pension reforms throughout the country, the Third Plenary Session of the 14th Communist Party Central Committee in 1994 additionally set targets for expanding the existing old-age social insurance system. The framework adopted by the party called for a multi-pillared system combining a social basic pillar with supplemental enterprise-sponsored pensions and individual savings for old age. By 1998, the pension system covered two-thirds of rural counties or 2,123 counties within 31 provinces. However, a combination of poor governance and additional financial challenges, complicated by the Asian financial crisis in 1997 halted the expansion of the rural pension program and it was substantially scaled back by 1999 and in the early years of the 2000s. Pension coverage under the system declined from 80.25 million participants in 1998 (approximately 11 percent of the total rural population) to mid-50 million in 2007.

In the early 2000s, the rural pension system faced continuing challenges related to financial sustainability. Program participation was extremely skewed toward wealthier regions. Poor provinces failed to match contributions and because of low-interest rates tied to individual accounts, benefits were not paid in full. Starting from 2003 to 2008, China witnessed a resurgence in rural pension scheme participation. More than 300 districts in 25 provinces established experimental new pension schemes by 2008 (Dorfman et al. 2013).

The government, under the new Hu-Wen administration that assumed leadership in 2003, ostensibly aimed a reform-oriented approach for the country's social protection system. The ambitious transformation proposed by the new administration was based on a 2008 pilot project, initiated in the city of Baoji in the northwestern province of *Shaanx*. The local administration experimented with a new pension scheme under which the local rural residents were provided coverage under the new social pension fund. Funding for the plan was provided via voluntary contributions from individuals and heavy subsidies financed by local tax revenues. This two-pillar system covered the rural residents of Baoji. The pilot project, referred to as the "Baoji model", served as a template for the Hu-Wen administration to scale up the model nationwide.

In 2009, China launched the NRPS. Participation in the NRPS was available to all rural residents over the age of 16 years so long as they are not enrolled in an urban pension scheme. The roll-out of the program occurred based on administrative areas called *Hukou,* a system of household registration used in mainland China. Participation in the new program was voluntary and individuals who were 15 years or older could contribute towards benefits they could receive once they reach the age of 60. The rural program also provided grandfathering conditions for residents who had already reached the age of 60 at the start of the program. They were eligible to receive a basic monthly benefit of 55 RMB so long they had children and their children made monthly contributions toward the program.[8] Participants between the ages of 45 and 60 years, with fewer than 15 years of contributions, were encouraged to increase their monthly payments so they could cover the absence of prior contribution before age 45.

NRPS expanded rapidly and from its inception, it aimed to achieve full geographic coverage by 2020 (Dorfman et al. 2013; Cai, Giles, O'Keefe and Wang 2012). The program covered 23 percent of districts at the end of 2010, and over 60 percent of districts by early 2012. In Figure 1, we depict program coverage expansion between 2009 and 2013. Over 50 percent of rural residents contributed to the NRPS by the end of 2011. Total participation grew from 87 million to 326 million people from 2009 to the end of 2011 (Quan 2012).

[Figure 1 about here]

[8] The central government fully subsidizes the basic pension in Central and Western provinces, and splits the cost with local governments in Eastern provinces (Cai et al 2012).

Three major factors account for the NRPS expansion between 2009 to 2001. First, the government's strong commitment to rural pension reform and the country's high economic growth rate played a considerable role. Between 2009 and 2012, China's economy grew an annual rate of 9 to 10 percent, which provided robust fiscal capacity for the roll-out of the massive social protection program. Second, because of the growing income inequality and demographic pressures, demand for the program and its basic monthly benefit was large. Third, expanding the pension program into rural areas was a key agenda element of the Hu-Wen administration.

Financed by the central government, individual contributions and matching funds by local governments determined the total amount available for payouts. Local governments must match 30 RMB annually per individual contribution. Based on data collected from early program implementation, nearly 50 percent of participants opted for the minimum contribution of 100 RMB (Dorfman et al. 2013). The pension program provided a fixed monetary pension payment. Individual contributions are voluntary and they range from 100 to 500 RMB, on an annual basis. Monetary benefits follow the "139 Rule", which is a number based on the average life expectancy (in months) at age 60. The rule follows a basic formula for the calculation of the monthly payment: it takes the accumulated balance in the individual account and divides it by 139.

B. Data

Our empirical analysis draws on data obtained from the CHARLS, which sampled 17,708 individuals across 28 of China's 30 provinces excluding Tibet. Due to the survey's rich data on cognition, we draw on this source for our main analysis in which we compare the cognitive outcomes of individuals with access to NRPS program benefits. The survey also directly captures a person's participation in the NRPS program. Measuring program participation important is essential in our empirical approach, which we outline in the next section. The second part of our analysis draws on data from the China Health and Nutrition Survey. The main use of this survey source in our analysis is its coverage of data years prior to the start of the NRPS. Therefore, the individual survey data can enable a thorough analysis of the necessary assumptions embedded in our estimation approach. We also use the survey to conduct several additional robustness checks, which we describe in detail in Section V.

The China Health and Retirement Longitudinal Studies (CHARLS). The CHARLS is nationally representative survey that collects information on households that comprise at least one person who is 45 years or older. The timing of the CHARLS is ideal for our analysis, since it was conducted approximately a year after the NRPS implementation. Figure 2 shows the geographic coverage map for the survey. The CHARLS provides data on demographic characteristics, family structure, cognition, health, pension and retirement, work, household wealth, income, and consumption.

That sample totaled 17,708 individuals living in 10,287 households in 450 villages/urban communities in 150 cities/districts across 28 of China's 30 provinces excluding Tibet. The survey sampling occurred in three stages. First, all city-level units were stratified into eight regions, by rural and urban districts, and by city/district gross domestic product per capita. After this initial step, 150 cities were randomly chosen using probabilities proportional to size (PPS). Within the 150 cities, three primary sampling units (PSUs) were randomly selected using the same PPS method. Once the 450 PSUs were selected, age-eligible households were interviewed.

The 2011 baseline wave interviewed 10,257 households with 18,245 respondents age 45 and over.[9] The follow-up 2013 wave covered 10,979 households (or 19,666 respondents). The interviewers followed up with 88.6 percent of the original respondents and 89.6 percent of original households.[10] The 2013 CHARLS added 2,053 new households with 3,507 individuals. The 2016 harmonized CHARLS dataset merged several modules from the individual annual waves.

An important feature of the CHARLS is that it directly collects information on individual participation in various government programs, including the NRPS. Respondents are asked, "Do you participate in the New Rural Social Pension Insurance program?" Using this question, we identify which individuals participate in the NRPS. For our main analysis sample, we drop

---

[9] Initially, 19,081 households were sampled where 12,740 had age-eligible members, of which 10,257 responded.
[10] 16,159 of the original 18,245 respondents and 9,185 of the original 10,257 households.

observations with an urban *Hukou* status because individuals who are attached to an urban *Hukou* are ineligible to participate in the NRPS but participate in urban pension schemes.[11] We can directly observe NRPS participants and non-participants. Our analysis sample consists of 15,990 individuals from 429 communities in 121 cities across 28 provinces.

Proxy Measures of Cognition. A second attractive feature of the CHARLS is that it directly tests cognition based on several proxy measures extensively based on comprehensive research on aging and cognition, and measures used in the HRS (Ofstedal 2005).[12] One cognition measure tests episodic memory, captured via several recall tasks. The second cognition measure tests one's mental intactness.

The first cognition measure, episodic memory, is assessed through verbal learning and recall tasks. CHARLS uses the HRS version of the CERAD immediate and delayed word recall to measure episodic memory (Ofstedal et al. 2005). *Episodic memory* is a necessary component of reasoning in many dimensions. The two tasks that capture verbal learning and recall are immediate and delayed recall. After approximately four minutes after other questions, the respondent is asked again to recall the nouns, without reading the words a second time. Word recall tests are collected to assess individuals' short-term and long-term cognitive impairment. For the immediate recall test, surveyors randomly assign respondents with a list containing 10 common words. The respondent is given two minutes to recall as many words as he/she can remember. The immediate recall score ranges from 0 to 10 and provides the number of words recalled correctly. Following this recall, the respondent continues to answer unrelated questions for several minutes until prompted to recall the original word list. This provides the delayed recall score, ranging from 0 to 10.

[11] The Urban Social Pension Scheme was established in 2011 and rapidly expanded through cities with robust fiscal capacity. The program is voluntary and is offered to urban residents aged 16 and over who are not employed in the formal sector. The program features a two-tier system, which consists of a pay-as-you-go social pooling component and individually funded accounts.

[12] The HRS cognition measures, and the ones used in the CHARLS, accounted for the several important considerations. First, the measures represent the major dimensions of cognitive functioning and can differentiate across a range of cognitive abilities. Second, the measures can identify respondents who exhibit some form of cognitive impairment. This second consideration guided the choice for inclusion of a traditional mental status measure that can differentiate individuals at the low functioning end of cognitive abilities. A third consideration included screening for early signs of dementia, or in the case of onset, for its subsequent progression.

The second measure, mental intactness, is measured in CHARLS using a series of questions from the Telephone Interview of Cognition Status (Brandt et al. 1988). This includes recognition of date: month, day, year, season (CHARLS allows use of the lunar calendar in addition to the Gregorian calendar), day of the week, how the respondent rates their own memory (excellent, very good, good, fair, poor), and serial subtraction of 7s from 100 (up to five times). The respondent is also asked to redraw a picture of overlapping pentagons. Based on these measures, we compute the sum of two scores—the immediate and delayed recall—for a total word recall score, ranging from 0 to 20. Low scores on this total word sum are indicative of low memory capacity and short storage duration.

Although we analyze all cognition measures in the survey, we focus on the episodic memory domain for two reasons. First, several studies highlight that this domain is the first to exhibit setbacks as aging sets in (Souchay et al. 2000; Tulving and Craik 2000; Prull et al. 2000). The second reason relates to its provision of high variation across individuals as opposed to other cognitive measures. For example, the word learning and recall tasks do not exhibit floor or ceiling effects (excess of maximum or minimum values) and the individual distribution of the score does not exhibit extreme observation bunching around minimum and maximum values. The CHARLS collects additional cognitive measures elicited by the survey respondent. In addition to several cognitive tests, respondents are asked to rate their memory based on a 5-point scale.[13] Based on this scale, we create a binary indicator variable that equals one if the perceived memory status is "at least good."

We proxy performance in cognitive memory—the memory domain is responsible for the brain's information capacity and storage duration—by collating information from several variables that each capture various facets of the memory domain. We combine data from the following factors: perceived memory status (subjective status), knowing the current month (orientation), serial 7 score (working memory), immediate recall score (memory capacity) and delayed recall score (memory duration). Using principal component analysis (PCA), we reduce these multiple measures into one composite index. This index provides a normalization of cognitive memory status, where negative (or low) values are associated with poor memory

[13] The 5-point scale is as follows: (1) Excellent (2) Very Good (3) Good (4) Fair (5) Poor.

functioning.[14,15] This index is an overall cognition proxy in the analyses that we present in the subsequent section.[16]

Based on the measures in the CHARLS, we report summary statistics of the analysis sample in Table 1. Among the eligible sample of participants and non-participants, 70 percent and 69 percent were employed in the baseline respectively. About three-quarters of the sample work in agriculture: 72 percent among participants and non-participants alike. The rural sample reported low levels of educational attainment—approximately 46 percent to 48 percent report having completed at least a secondary level of education. In terms of health, approximately one-quarter of participants and non-participants report being in “poor/fair” health status, 27 percent, and 26 percent respectively.

[Table 1 about here]

Regarding the variables of interest in the study, survey participants report a low average on various cognitive measures in the baseline period. Program participants and non-participants perceive their memory status as subpar—15 percent and 18 percent, respectively, reported their memory as being “at least good.” Participants scored slightly higher on the word recall tests. The average score on immediate word recall task for participants was 3.93 out of 10 (non-participants average 3.77 out of 10). Similarly, the delayed recall score was higher for participants than it was for non-participants, 2.91 and 2.89 respectively. Approximately 84 percent of participants and non-participants correctly named the current month. The cognitive memory index, based on the PCA, exhibits a higher average for participants than it does for non-participants, 0.06 and 0.00 respectively.

---

[14] We usa a PCA method to transform the set of proxy variabels for cognition into an aggregate index. We do so by first standardizing each cognition proxy. We then compute the covariance matrix for all cognition measures. Third, we compute the eigenvectors; combined, they contain the same information as the original variables. By design, the first component, based on the largest eigenvalue, contains the most information, whereas the last component contains the least. We reduce the set of original cognition proxies into the one index by retaining the component with the largest variance (eigenvalue).The overall PCA index, Cognitive Memory Index has a mean of 0 and a standard deviation of 1.42.

[15] We use the STAT 15 package, *pca,* to create the index variable.

[16] Online Appendix B, Table B1 reports the index component loadings based on the survey’s cognitive measures.

China Health and Nutrition Survey. As the CHARLS does not collect cognition prior to the start of the NRPS, we rely on data from the China Health and Nutrition Survey (CHNS), a survey conducted by the University of North Carolina at Chapel Hill. CHNS is a panel survey covering 1989 to 2011, a period overlapping with the start of NRPS. Additionally, the survey collects cognitive measures, very similar to those collected by the CHARLS. Therefore, the survey can allow for analysis of data for our main outcomes prior to the introduction of the pension benefits.[17]

The CHNS covers approximately 19,000 individuals in 15 provinces spanning 216 PSUs.[18] The survey's first wave started in 1989 and the survey's aim was to collect data on economic and social determinants of individual health and nutritional status. To this end, the individual survey modules focused on the following topics: food choice, nutritional intake, health behaviors, physical activities, work activities, time usage, and nutritional status. The survey sample was selected based on a multistage random selection process. First, community districts were stratified by income level, followed by a weighting scheme that selected four communities from each province (CHNS Research Team 2010).

The main component of the CHNS is the individual module. From the 2004 survey onward, all questions in the survey related to individual activities, lifestyle, health status, demographic status, body shape, and mass media exposure. The individual questionnaire consisted of two components: one component interviewed adults aged 18 and older and the second collected information on children under 18 years. Adults age 55 and older were asked to provide daily living activities and were given various cognition tests.

The CHNS sampling areas overlap with the ones sampled by the CHARLS. Figure 2 depicts the geographic coverage of the CHNS. In addition to the overlap of geographic coverage between the two surveys, the CHNS also collected information on proxy measures of memory and cognition, very similar to those collected by the CHARLS.

[17] We use this survey for analysis on pre-trends because there are no data in CHARLS for the outcomes we analyze prior to the baseline period, which is 2011.

[18] The survey covered the following provinces (reported in Figure 2): Beijing, Chongqing, Guangxi, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shaanxi, Shandong, Shanghai, Yunnan, and Zhejiang.

[Figure 2 about here]

The CHNS adopted similar cognitive screening items because its cognition proxies were also based on the US HRS survey. The same cognitive screening test was used in the three waves of the CHNS among adults aged at least 55 years. The cognition tests tested immediate and delayed recall of a two-word list, counting backward from 20, serial 7 subtraction, and memory orientation. The scores for immediate and delayed recall ranged from 0 to 10. Counting backward and serial 7s were used to assess attention and calculation, with scores ranging from 0 to 7. Orientation was assessed by asking the participant the current date (1 point each for a correct response on the year, month and date), and the name of the tool usually used to cut paper (1 point). Higher scores on all items suggest better cognitive performance.

III. Empirical Strategy

Our main identification strategy relies on within-country variation in the NRPS implementation due to program's staggered rollout across the country. Using data from the CHARLS, in Figure 1, we show how the coverage rates change at the community level between 2011 and 2013. In 2010, approximately a quarter (or 23 percent) of communities implemented the NRPS program; by 2012, this number grew to approximately 60 percent (Dorfman et al. 2013). This staggered program implementation, between 2011 and 2013, provides a source of identifying variation to examine how the program affected cognition outcomes between individuals who live in communities that adopted the NRPS and individuals whose communities did not offer the NRPS benefits.

A. Estimating Equations and Triple Difference Estimation

Given these variations, we able to use difference-in-difference-in-difference (DDD) estimations in our empirical analysis. Therefore, our identification strategy relies on the timing differences among communities for pension policy adoption. By the end of 2010, approximately

23 percent of all communities were covered. By 2012, the coverage rate reached 60 percent (Dorfman et al. 2013; Cai et al. 2012), as Figure 1 shows.[19] It is important to underscore that our DDD analysis is conducted at the community (*shequ*) level, an administrative level within a county that encompasses several neighborhoods. For each community, the CHARLS administered a community questionnaire, which collected data on its natural environment, employment, financial status, and social protection program coverage. Based on information from the CHARLS, we construct a variable, $OfferNRPS_{ct}$, which indicates the participation status (whether a community c implemented the NRPS program at time t). Due to linkage between the various administrative layers within the CHARLS, we can link each community identifier within the survey with a person's place of living and his/her response to whether they participate in various government programs, including the NRPS. This process allows us to define the variable $OfferNRPS_{ct}$ based on responses from the individual-level data.[20] We next examine the impact of the NRPS provision on cognition in the following regression:

$$(1) \qquad Y_{ict} = \beta_0 + \beta_1 (OfferNRPS_{ct} \times Above60_{ict}) + \beta_2 Above60_{ict} + \beta_3 \boldsymbol{X}_{ict} + \phi_c + \mu_t + \phi_c \times \mu_t + \varepsilon_{ict},$$

where Y_{ict} is the cognition outcome and $Above60_{ict}$ is equal to 1 if the respondent is age 60 and over.

The coefficient of interest is β_1 in (1). It captures the intent-to-treat estimate of the average effect of the NRPS program on the average outcomes of eligible individuals age 60 and over who live in a treated community, regardless of whether the individual decides to participate in the program (i.e., the ITT effect). $\boldsymbol{X}_{ict}$, is a vector of individual-level controls, education, gender, age, age squared, household size, and marital status. ϕ_c and μ_t are community-level and time fixed effects, respectively. Community-level fixed effects allow us to control for time-

---

[19] Demographic information in the CHARLS is only available at the community (*shequ*) level.

[20] If no individuals indicate having NRPS at time t in community c, then $OfferNRPS_{ct}$ equals 0. If at least one person reports participating in the NRPS, $OfferNRPS_{ct}$ is set to 1. We address potential concerns regarding measurement error and associated bias in the estimated coefficients based on this approach with additional robustness checks that we present in Section V.

invariant characteristics that affect the likelihood that a community implements the NRPS. The year fixed effects account for characteristics common across time in communities.[21,22] In addition, we use community-time fixed effects, $\phi_c \times \mu_t$, to control for community differences during the implementation of the NRPS. The DDD design is the most appropriate choice, as it controls for potential region-specific effects and is based on a similar policy rollout in other empirical studies, such as Katz (1996), Gruber (1994), and Rossin (2011).

Specification (1) will produce an unbiased estimate of β_1, our coefficient of interest, if the variation in program implementation across communities is unrelated to other community-related shocks. The identification *common trends* assumption, which underlies the DDD design in our estimation, posits that important factors that influence the study outcomes are either time-invariant group attributes or time-varying factors that are group invariant. In summary, the identification assumption implies that communities that happened to adopt the NRPS program would otherwise have changed in a manner similar, on average, to the communites that did not adopt the NRPS. To check whether the triple difference is an appropriate strategy to examine the effect of the NRPS program, we test the common trends assumption for the pre-policy survey data based on the empirical approach in Autor (2003). We examine the trends of various cognition measures between treated and non-treated areas *prior* to the launch of the NRPS program in 2009. Since all survey data from the CHARLS is collected post-NRPS program, we analyze data on the pre-trends of our study outcomes for the three CHNS waves that collected cognition measures. The CHNS dataset is also an ideal alternative survey source because both areas are sampled by the CHARLS but also because it directly measures cognition outcomes similar to the ones collected by the CHARLS.

The primary data challenge for this particular analysis (using the CHNS) is that the community identifiers or geographic-level variables do not match in a one-to-one fashion between the two surveys. Therefore, only for this empirical exercise using the CHNS, we redefine "treated" and "control" units at the province level (as opposed to the community *shequ* level) to rely on the geographic variables available in the CHNS. Once we reconstruct the analysis on the province

[21] We cluster the standard errors by community and age groups based on Bertrand, Duflo, and Mullainathan (2003).
[22] In Online Appendix B, we report additional robustness checks in which we cluster the standard errors by community and age. Our results are robust to community and age-specific clusters.

level, we proceed with testing the common trends assumption, at the province level, using the CHNS data for 2000 to 2009. Furthermore, and only for this empirical test, we underscore that the treatment status definition for a province in the CHNS data for 2004 to 2009 is based on two important features. First, it is based on the available baseline data from the CHARLS. Second, the definition is based on the percentage, within a province, of communities that report NRPS implementation. Therefore, our definion of a "treated" province relies on the percentage of communities, within a province, that indicate (based on survey data from the CHARLS) that participate in the NRPS program. In other words, our treatment definition of a province is defined continuously as the "treatment intensity" of a given province. Based on that continuous variable, we then code the province's treatment status with a binary variable. The province's treatment status is set to 1, if more than a given threshold of communities reported participating in the NRPS, and zero otherwise. Based on this reconstructed definition of a "treatment status" for a province, and only for the purpose of this formal test of the common trends' assumption, we then proceed to using data from the CHNS and we analyze data from the CHNS prior to the NRPS's introduction to test for any potential common trends between treated and control provinces. To define a "treated" province, we choose a threshold based on the percentage of communities (within a province) that indicate that they participate in the NRPS to define a province's "treatment" status.[23,24] Our analysis for this test uses data on cognition outcomes from the CHNS that mirror the cognition proxies collected by the CHARLS. Using the CHNS data, prior to 2011, on cognition measures from the 2004, 2006, and 2009 waves, we can estimate the following specification:

$$(2) \qquad Y_{ict} = \beta_0 + \beta_{-3} D_{ct} + \beta_{-1} D_{ct} + \phi_c + \mu_t + \varepsilon_{ict} \,,$$

[23] We use a binary definition of treatment status for each province. We define a province as "treated" (=1) if more than 67 percent of all communities within this province implemented the NRPS, based on information collected from the baseline data. In this addition to using this threshold, we conduct additional sensitivity analyses based on alternative threshold choices. In additional sensitivity analyses, we vary the threshold choice to a lower (50 percent coverage rate) or higher value (70 percent coverage rate). Based on these alternative threshold choices, we redefine the treatment status for each province and we re-estimate our specifications. Based on these additional threshold choices, the results do not provide support for violation of the common trends assumption.

[24] The CHNS does not sample from the same communities/villages as the CHARLS, so we rely on our definition of treated and control provinces based on the CHARLS to test data in the CHNS.

where Y_{ict} is the cognition proxy, and ϕ_c and μ_t are community-level and time fixed effects, respectively. We include the interactions D_{ct} of the time-period dummy variables and the treatment indicators for the first pre-treatment period and last pre-treatment period.[25] The results reported in online Appendix Table A1 provide clear evidence that β_{-3} and β_{-1} are insignificant. Therefore, we fail to reject the hypothesis that trends in the outcomes between the treatment and control areas are the same. Therefore, this exercise provides no empirical evidence that the common trends assumption is violated.

B. Two-Stage Least Squares Estimation

The introduction of the program may have witnessed endogenous selection related to program targeting. To address this possibility, we augment the DDD analysis by instrumenting the program availability at the community level. Specifically, we re-estimate specification (1). We use $OfferNRPS_{ct}$ to instrument for a person's participation in the NRPS, an approach similar to that used in Nunn and Qian (2014). We code $OfferNRPS_{ct}$ as a binary variable and the variable is set to 0 if no individuals participate in the NRPS. It is set to 1 if the community witnesses at least 1 participant. We re-estimate the following specification:

$$(3) \qquad Y_{ict}=\beta_0+\beta_1(\widehat{NRPS}_{ict}\times Above60_{ict})+\beta_2 Above60_{ict} + \beta_3 \boldsymbol{X}_{ict}+\phi_c+\mu_t+\phi_c\times\mu_t+\varepsilon_{ict}.$$

$\widehat{NRPS}_{ict}$ represents individual enrollment in NRPS, and we instrument it with $OfferNRPS_{ct}$. $\boldsymbol{X}_{ict}$ is a vector of individual-level controls, and ϕ_c, μ_t, and $\phi_c \times \mu_t$ are community-level, time, and community-time fixed effects, respectively.

IV. Main Results

A. Impacts on Cognition Measures

[25] In this specification, the second pre-treatment period is omitted.

We start by examining the program impacts on cognitive outcomes based on specifications (1) and (3). Table 2 reports the results. Column 1 through 4 provide the results for the various cognition proxies for the immediate recall measure, delayed recall, total recall and memory index, respectively. These results are based on specification (1); therefore, they are the intent-to-treat estimates on the effect of program availability in a community on the various cognition measures. The results in all columns provide striking evidence of negative cognitive impacts among individuals aged 60 and above who live in NRPS program areas. In Table 2, we report the results based on the 2SLS approach.

The results in both tables report a striking pattern of negative effects on all cognition measures. On the immediate recall test, individuals in treated areas score, on average, 0.14 points less (see Table 2) than do individuals in non-treated areas. For the delayed recall test, individuals in treatment areas score approximately 0.23 to 0.35 points less than individuals living in areas that do not offer NRPS. Program offer also had a considerably negative effect on the cognitive index. The index combines the cognition measure on mental intactness, described in Section II. On average, the provision of the NRPS benefits leads to a 0.10-point reduction in the composite score (equivalent to about 7 percent of 1 standard deviation) for the intent-to-treat specifications; the effect size is doubled in magnitude for the 2SLS estimations reported in Panel B (in Table 2). When comparing the effect size estimates across all columns, the most striking negative effect is on the delayed recall cognition measure. The associated effect size for this particular outcome is approximately double the effect size for the other two cognition measures. Neurological studies show that this proxy particular measure of cognition among the elderly is an effective predictor that can distinguish between normal aging and earlier onset of dementia in adulthood (Welsh et al. 1991, Laakso et al. 2000).

[Table 2 about here]

In addition to the impacts of program provision regardless of program participation, Table 2 reports the treatment-on-treated estimates based on specification (3). The effect size estimates reported in Panel B are approximately double the size of the effect size estimates based on the ITT specification.[26] The results reported in Panel echo the pattern reported in Panel A—the effect of program participation on measure cognition is statistically negative on all on CHARLS cognition measures: the delayed word recall task, total word recall score, and composite memory index. Furthermore, the reported coefficients for the delayed word recall, total word recall and memory index are all statistically significant.

The analysis so far focuses on data from the CHARLS, based either on whether the community implemented the NRPS or the individual reported participating in the program. The CHARLS also collects data on actual retirement but it is important to underscore that this particular variable is sparsely populated. Despite the lack of data for this variable and its likely data quality issues, we report results based on specifications (1) and (3) using data on the self-reported retirement status. Table 3 reports the results.

[Table 3 about here]

The pattern of the reported coefficients associated with the cognition measures echoes the negative effect pattern highlighted by the results reported in Table 2. The reported results show that program participation leads to faster cognitive decline via the influence the program exerts on individuals' retirement decisions. Although the results do not pass standard tests of statistical significance, it is important to stress that very few individuals respond to the retirement status question in this particular dataset. Program participants who retire exhibit decline (i.e., negative coefficients reported in Columns 1 through 3) in their performance for the cognition tests, from the immediate recall task, the delay word recall to the memory intactness measures captured in the total index. In fact, the effect size associated with the composite memory measure exhibits a large negative coefficient on the retirement status variable.

B. Cognitive Decline: Heterogeneity by Gender

Next, we report estimates on the heterogeneity of the NRPS impact. We repeat specifications (1) and (3) for the sample of females and males. Specifically, we focus on a comparison of the treatment effects (for either of the specification) by gender. Table 4 reports the

[26] Table 2 reports the F-statistic associated with the first-stage estimation in the 2SLS specification. The F-statistic is considerably above the usual rule of thumb value of 10.

results for the heterogeneous treatment analysis. Panel A reports the intent-to-treat DDD estimates, whereas Panel B provides estimates based on the 2SLS estimation approach. Both panels report separately by gender.

[Table 4 about here]

The results reported in Table 4 echo the negative pattern of the effect of the NRPS program on cognitive measures. Although the sample sizes are smaller, Columns 1 through 4 demonstrate that the NRPS program was associated with a steeper cognitive decline among individuals who live in areas that implemented the NRPS program (reported in Panel A) or among individual participants in the NRPS (reported in Panel B). When we turn attention to the magnitude of the effect size estimates, the program effects on the female sample are considerably larger. Comparing effect size estimates across columns 1 through 4, the difference between estimates based on the female sample is approximately double or more compared to the effect size for the male sample (Columns 1 and 4). The implication of these results is that the effect of the NRPS program is more considerably negative among females in rural China.

This particular finding may have alarming implications. Recent empirical studies, based on the CHARLS on other studies, show that Chinese women perform considerably worse on cognition test than do Chinese men (Lei et al. 2012). This cognition gap is particularly exacerbated among older survey cohorts. As women in China have longer life expectancy at birth or conditional on reaching the age of 60 (Liu et al. 2009), if older women witness a faster cognitive decline due to program participation, the implications of this effect are unambiguous—earlier retirement could be a source for expansion of morbidity in older age among women. Put differently, if these gender differences due to the NRPS are indeed real, it implies that encouraging early retirement will lead to greater life expectancy but one with greater disability among women in rural China (Wang 1993).

V. Robustness Checks

In this section, we perform a series of exercises aimed to re-examine the validity of our empirical approach. Specifically, we re-estimate our main empirical estimations with an individual sample based on individuals who are not eligible to benefit from the NRPS. Therefore, estimating the main specifications on this sample of non-eligible individuals should yield non-significant results. We also specifically address the possibility of measurement error related to either individuals misreporting their participation in the NRPS program or a community incorrectly indicating a program implementation within its boundaries. The main results survive these extension exercises, suggesting that our main results are not driven by secular trends, alternative contemporaneous policies in the same areas where NRPS was implemented, or unobserved shocks to the study outcomes.

A. Falsification Exercises

We now turn to several falsification exercises to bolster the validity of our estimated results. Specifically, we construct a falsification exercise based on an alternative sample of individuals, which comprises a subset of individuals who are neither eligible for NRPS participation nor its program benefits. Therefore, in theory when we rerun specifications (1) and (3), the coefficients of interest discussed earlier in Section III for this alternative study sample should not be significant.

As we underscored earlier in Section III, the NRPS program is only available to individuals who live in rural administrative districts as long as they are not enrolled in an urban pension scheme. In the main analysis and results presented in Section IV, we excluded urban pensioners and elderly individuals without children who live in rural administrative districts (rural *Hukou*) because these individuals are ineligible for the NRPS. However, for this particular falsification exercise, we reconstruct the analysis sample and employ the opposite approach. Only for this exercise, we include respondents who indicate they are pensioners who participate in an urban pension program or are elderly individuals (aged 60 and above) who did not contribute to the NRPS prior to reaching age 60 and report no current children who live in rural administrative districts. In other words, in the falsification exercise, individuals included in the sample are: (1) those who live in rural areas but obtain benefits from an urban pension system; or

(2) those who are elderly and without children who happen to live in rural administrative districts. The main objective of this falsification exercise is to examine whether the effect of the NRPS on the cognition outcomes for this "placebo" sample would differ between individuals who live in areas that offered NRPS coverage and individuals who live in areas that did not. If indeed specifications (1) and (3) yield no spurious results, this specific falsification exercise should produce non-significant coefficient estimates for the coefficients associated with the NRPS effect on cognition outcomes. We re-estimate specifications (1) and (2), as described in the main empirical approach, but based on the placebo sample. In Appendix A Table A2, we report the results based on this falsification exercise.

Table A2 reports non-significant estimates for the effect of the NRPS program on the various cognition measures: immediate recall score, total recall score, and cognitive memory index. In other words, these results imply that urban pensioners who live in communities that offer the NRPS—relative to urban pensioners who live in communities that do not offer the NRPS—do not exhibit statistically significant differences in cognitive performance. The results based on this additional robustness check further bolster the validity of our main results presented in Tables 2 and 4; they are unlikely to be based on a spurious specification choice.[27]

B. Alternative Measures of NRPS Participation

In this subsection, we further explore the possibility that our main analysis relies on mismeasured individual participation in the NRPS or administrative communities' incorrect reports of NRPS program implementation within their boundaries. Either of these possibilities will yield measurement error in our program impacts. The presence of measurement error in self-

[27] We also conduct an additional falsification test in which we re-estimate specifications (1) and (3) on a set of placebo outcomes. The selection of these placebo outcomes was based on no conceptual mechanism linking pension and program impacts. This additional falsification exercise was another attempt to examine the credibility of our main results. In Appendix Table A3, we report the results based on a set of four "placebo" outcomes. The four placebo outcomes are: one's nationality being Han, the number of female household members, the number of daughters in the household, and the mother's educational level. Appendix A Table A3 reports the results. In Panel A, we report the ITT results, where Panel B reports the TOT results. In both panels, the results provide no empirical support of program effects on the set of placebo outcomes.

reporting could yield biased impact estimates. Therefore, we perform additional consistency checks based on alternative approaches intended to measure NRPS participation.

Propensity Score Method Definition. First, it is possible that survey responses based on the CHARLS are incorrect, resulting in possible mismeasurement of how NRPS participation is recorded in our analysis. We address this possibility with an alternative measurement of the individual NRPS participation status. To do so, we use data on individual characteristics and we reconstruct the likelihood that an individual participates in the NRPS program. We define the NRPS participation status based on a propensity score matching approach. Specifically, we predict the NRPS participation status based on a combination of the respondent's characteristics, such as education, gender, parental education, and nationality. We use data on these variables based on the baseline collected in the CHARLS. Using these characteristics, we then predict the propensity of NRPS participation, $\widehat{NRPS}_{ic}$, based on the propensity score matching method. The predicted likelihood, based on this estimation technique, is $\mathrm{PrNRPS}_{\mathrm{ic}}$. We construct an alternative measurement of the NRPS participation status variable by defining $PrNRPS_{ic}$ = 1 if $\widehat{NRPS}_{ic}$ is greater than one standard deviation above the average value of $\widehat{\mathrm{NRPS}}_{\mathrm{ic}}$.

Next, we proceed with the DDD analysis as in the main portion of our results section based on specifications (1) and (3). However, in these new additional analyses, we rely on the definition of program participation based on this reconstructed variable ($PrNRPS_{ic}$) regarding NRPS participation status. This participation variable is based on estimation from the propensity score approach rather than the estimation approach reported in Section IV, which draws from the self-reported variable in the CHARLS. We report the results based on the alternative approach in Online Appendix Table B2. The reported results demonstrate that the pattern of our findings is indeed robust to this alternative definition of NRPS program participation.

Community NRPS Participation Definition. Next, we consider the possibility that there is a measurement error, due to misreporting in the individual surveys, for the NRPS program participation variable. If this is indeed the case, the measurement error in the individual-level variable in the CHARLS can subsequently creep into the community variable, measuring whether a given community implements the NRPS program or not, an indicator used in specification (3).[28]

Therefore, we verify the robustness of our approach with an alternative definition of the variable $OfferNRPS_{ct}$. Specifically, the objective of this additional exercise is to correct for possible contamination of what communities are coded as treated (or indicating implementation of the NRPS program). The source of this measurement could be due to false reporting at the individual level. To tackle this potential issue, we re-estimate specifications (1) and (2) but rely on a higher threshold that defines when the variable $OfferNRPS_{ct}$ (the variable that indicates community participation in the NRPS) switches from zero to one. Instead of relying on a threshold of at least one individual reporting NRPS participation to set $OfferNRPS_{ct}$= 1, we now use an alternative (and higher) threshold of at least four participants in community *c* to set $OfferNRPS_{ct}$= 1. Furthermore, in yet another more stringent definition, we rely on a definition for when the community indicator switches from zero (non-participating) to 1 (participating) based on an even higher threshold, of least seven individuals within the community, for each community in the CHARLS, reporting participating in the NRPS.

We report the results from these additional analyses in Online Appendix Table B3. The results demonstrate that our original estimates are robust to alternative and much more conservative definitions of the threshold, which determines when the variable $OfferNRPS_{ct}$ switches its binary values.

Using Online Administrative Data. Using additional data from online sources, our final robustness check aims to bolster the direction of the estimated program effects based on the main

[28] We conduct an additional extension exercise to address another potential source of measurement error in the variable that measures whether a community implements the NRPS. In the main analysis, we define the implementation of the NRPS program at the community level based on survey data at the individual level. In this empirical approach, if at least one individual in the community reports participating in the NRPS, then we define the community as one having implemented the NRPS. However, it is possible that communities with a very small number of NRPS participants are systematically related to a set of other factors that we can not observe in our survey data and account in our analysis. This scenario could produce a measurement error affecting our measurement of the instrument. Therefore, in an additional extension exercise, we re-estimate by removing communities that report very few NRPS participants within their boundaries. We then proceed by re-estimating the main specifications reported in Section III. Online Appendix B Table B4 reports the results for this extension exercise. The pattern of the results remains consistent with the main results.

analysis. For this empirical exercise, we comb data available via Chinese newspapers (online or in paper format) and we specifically search for information regarding public announcements regarding geographic areas announcing their participation in the NPRS. The two levels for which such data on public announcements is available is only at the city and community levels. We specifically focus on the public announcements for NRPS implementation in Heilongjiang Province. Specifically, based on public announcements, we are able to identify whether a city (or communities within a city) participates in the NRPS in a given year for the period 2009 to 2013, which is the analysis period in our main estimation approach. Based on data from the public announcements, we are able to identify the exact timing of when specific cities (and communities within these cities) switched from non-participation to participation in the NRPS.

However, we face a challenge related to the definition of the *community* unit between the CHARLS and actual administrative units from the public announcements regarding NRPS implementation. We are unable to map the actual communities (within cities) to the communities (or the variable *community ID*) in the CHARLS survey, which is analysis unit in our main analyses, because the definition of a community in the CHARLS differs from the definition of an administrative *community unit* (available in online records). Although we are unable to re-estimate the main analysis (which we perform done at the community ID level) and replace the survey data from the CHARLS with administrative data for community-level participation, we are able to rerun our previous specifications at the city level (a higher level than the community level) for two reasons: (1) we are able to observe the number of communities within each city that definitively implemented the NRPS program based on the online public announcements, and (2) we are able to observe the total number of all communities within each city; the total number of these communities is a fixed number.

Therefore, we redefine our main treatment. Specifically, and only for this additional empirical exercise, we change the definition of the treatment variable from the binary variable used in our main analyses (at the community level) to a continuous variable that measures treatment intensity (at the city level)[29]. Based on this reconstructed definition of the treatment

[29] The Heilongjiang province is ideal for this empirical exercise for a number of reasons. First, online announcements regarding city-level implementations of the NRPS are readily available regarding NRPS implementation between 2011 and 2013 at the city level. As outlined in Section II, this is the period for which data is available from the two

variable, we re-estimate this additional robustness check at the city level as opposed to at the community level, an estimation approach performed in the main analysis.[30]

Using data for Heilongjiang Province, in this final robustness check we re-estimate the empirical specifications outlined in Section III but we rely on a continuous definition of the main treatment variable. We report the results for this final robustness check at the city level in Online Appendix Table B5. This additional analysis (at the city level) relies on a very limited sample but despite this statistical power limitation, the pattern of the results echo the pattern reported in the main analysis. Both the effect size and the direction of the program effects are consistent with the direction of the main estimates based on survey data from the CHARLS. The results based on these additional analyses provide additional evidence regarding the direction of the program impacts reported in our main analyses.

VI. Discussion and Conclusions

In this paper, we investigate the effect of China's NRPS on cognition among the elderly in rural China. We specifically examine the effects on two categories of cognitive functioning among the elderly: episodic memory and intact mental status. We do this by using new longitudinal data available from the CHARLS for older individuals. We find large and significant negative impacts of the pension program on cognition outcomes. Albeit surprising, the estimated program impacts are similar to the negative findings in other settings in the context

---

CHARLS waves. Second, the province is one of the largest provinces in China.This factor can considerably facilitate the re-estimation exercise because our main empirical approach relies on identifying variation based on both time and space. Third, most of the city-level implementation of the NRPS in this province around 2013. Other cities or areas within provinces had either already adopted the the NRPS program prior to 2013, or information on city-level NRPS implementation by year was not readily available online or via posted public records.

[30] In this additional estimation exercise, we can estimate the original specifications at the city level. Specifically, we can compute the treatment intensity (for city participation) as follows: $city\_participation_{t}$=(# communities in a city offering $NRPS_{ct}$)/(# total communities in a city)$_{ct}$. The main advantage of this additional robustness check is that we are able to observe the number of communities that implement the NRPS program based on public announcements (the numerator). The denominator of the fraction presented above is the total number of communities and that number is a constant. A major disadvantage of this approach is that are able to re-estimate the specifications from Section III only at the city level (and only for the Heilongjiang Province, for which we are able to obtain data on city or community announcements regarding NRPS implementation). This implies that in this additional analysis the number of number of observations is low, which limits the statistical power for statistical inference.

of high-income countries, such as the US, England and the European Union (Rohwedder and Willis 2010; Mazzonna and Peracchi 2012). Individuals in the areas that implement the NRPS score considerably lower than individuals who live in areas that do not offer the NRPS program.

Two findings deserve special emphasis. First, we find considerably larger program impacts on the cognition measure that tests delayed word recall. Previous neurological research documents the importance of this measure, particularly in detecting the difference between the process of normal aging and individuals more likely to witness an earlier onset of dementia in adulthood. We also find stark program impact differences by gender. The effect of the NRPS program is considerably more negative among females in rural China. This result has important implications for the welfare of women in rural China. Women's longevity is considerably higher than men's, a pattern that is also true for individuals in our rural Chinese sample. Our findings suggest that early retirement is likely to accelerate cognitive declines in adulthood, which is likely to result in lower healthy life expectancy among women in rural China.

Furthermore, our findings support the *mental retirement* hypothesis that decreased mental activity results in atrophy of cognitive skills and suggest that retirement plays a significant role in explaining cognitive decline at an older age. However, further studies would be necessary to specify the effect of professional activities on cognition and, in particular, on other cognitive domains. Specifically, two additional areas will be of particular interest regarding the nexus between retirement and cognitive decline in developing countries. First, what role does the type of job—formal versus informal or white collar versus blue collar—play in determining the speed of individual cognitive decline? Second, it is important to uncover and examine the underlying mechanisms between a person's retirement and cognitive decline. A particularly important mediating factor in developing countries is the role of informal social networks, social status, and the frequency and quality of social interactions.

Finally, our findings have policy implications that call for new policy interventions among the elderly in resource-constrained settings, such as the context of our study. Cognitive impairments among the elderly, even if not severely debilitating, bring about a loss of quality of life and can have negative welfare consequences. Policies aimed at facilitating or promoting physical activity or labor force participation, even in older ages, are likely to generate positive spillovers.

References

Adam, Stphane, Eric Bonsang, Sophie Germain, and Sergio Perelman. 2007. "Retirement and Cognitive Reserve: A Stochastic Frontier Approach Applied to Survey Data" *University of Liège, CREPP Working Paper 2007/04*.

Autor, David H. 2003. "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing" *Journal of Labor Economics*, 21(1): 1-42.

Banks, James and Zo Oldfield. 2007. "Understanding Pensions: Cognitive Function, Numerical Ability and Retirement Saving" *Fiscal studies*, 28(2): 143-170.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much should we Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249-275.

Bingley, Paul and Alessandro Martinello. 2013. "Mental Retirement and Schooling" *European Economic Review*, 63(October 2013): 292-298.

Bloom, David E. and Roddy McKinnon. 2014. "International Handbook on Ageing and Public Policy: The Design and Implementation of Pension Systems in Developing Countries: Issues and Options" *Institute for the Study of Labor, IZA Policy Paper No. 59*: 108.

Bonsang, Eric, Stphane Adam, and Sergio Perelman. 2012. "Does Retirement Affect Cognitive Functioning?" *Journal of Health Economics*, 31(3): 490-501.

Brandt, Jason, Miriam Spencer, and Marshal Folstein. 1988. "The Telephone Interview for Cognitive Status" *Neuropsychiatry, Neuropsycholy and Behavioral Neurology*, 1(2): 111-117.

Cai, Fang, John Giles, Philip O'Keefe, and Dewen Wang. 2012. *The Elderly and Old Age Support in Rural China*, 1st ed. Washington, DC: World Bank Publications.

CHARLS Research Team. 2013. "Challenges of Population Aging in China. Evidence from the National Baseline Survey of the China Health and Retirement Longitudinal Study (CHARLS)" *Beijing: School of National Development, Peking University*: 1-32.

Cheng, Lingguo, Hong Liu, Ye Zhang, and Zhong Zhao. 2016. "The Health Implications of Social Pensions: Evidence from China's New Rural Pension Scheme" *Journal of Comparative Economics*, 46(1): 53-77.

Coe, Norma B., Hans-Martin v. Gaudecker, Maarten Lindeboom, and Jürgen Maurer. 2012. "The Effect of Retirement on Cognitive Functioning" *Health Economics*, 21(8): 913-927.

Cutler, David M. and Richard Johnson. 2004. "The Birth and Growth of the Social Insurance State: Explaining Old Age and Medical Insurance Across Countries" *Public Choice*, 120(1-2): 87-121.

de Grip, Andries, Arnaud Dupuy, Jelle Jolles, and Martin v. Boxtel. 2015. "Retirement and Cognitive Development in the Netherlands: Are the Retired really Inactive?" *Economics & Human Biology*, 19(December 2015): 157-169.

Dorfman, Mark C., Robert Holzmann, Philip O'keefe, Dewen Wang, Yvonne Sin, and Richard Hinz. 2013. *China's Pension System: A Vision*, 1st ed. Washington, DC: World Bank Publications.

Eibich, Peter. 2015. "Understanding the Effect of Retirement on Health: Mechanisms and Heterogeneity" *Journal of Health Economics*, 43(September 2015): 1-12.

Feldstein, Martin and Jeffrey B. Liebman. 2002. "Social Security." In *Handbook of Public Economics*, 1st ed., ed. Auerbach, Alan J. Feldstein, and Martin, 2245-2324. Vol. 4. Amsterdam, The Netherlands: Elsevier.

Fillenbaum, G. G., D. C. Hughes, A. Heyman, L. K. George, and D. G. Blazer. 1988. "Relationship of Health and Demographic Characteristics to Mini-Mental State Examination Score among Community Residents" *Psychological Medicine*, 18(3): 719-726.

Gruber, Jonathan. 1994. "The Incidence of Mandated Maternity Benefits" *The American Economic Review*, 84(3): 622-641.

Holzman, R., D. Robalino, and N. Takayama. 2009. *Closing the Coverage Gap. the Role of Social Pensions and Other Retirement Income Transfers. World Bank*, 1st ed. Washington, DC: The World Bank.

Katz, Lawrence F. 1996. "Wage Subsidies for the Disadvantaged" *National Bureau of Economic Research, Mimeo*, No. w5679.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects" *Econometrica*, 75(1): 83-119.

Kling, Jeffrey R., Jens Ludwig, and Lawrence F. Katz. 2005. "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment" *The Quarterly Journal of Economics*, 120(1): 87-130.

Laakso, Mikko P., Merja Hallikainen, Tuomo Hänninen, Kaarina Partanen, and Hilkka Soininen. 2000. "Diagnosis of Alzheimer's Disease: MRI of the Hippocampus Vs Delayed Recall" *Neuropsychologia*, 38(5): 579-584.

Lei, Xiaoyan, Yuqing Hu, John J. McArdle, James P. Smith, and Yaohui Zhao. 2012. "Gender Differences in Cognition among Older Adults in China" *The Journal of Human Resources*, 47(4): 951-971.

Liu, Jufen, Gong Chen, Xinming Song, Iris Chi, and Xiaoying Zheng. 2009. "Trends in Disability-Free Life Expectancy among Chinese Older Adults" *Journal of Aging and Health*, 21(2): 266-285.

Liu, Tao and Li Sun. 2016. "Pension Reform in China" *Journal of Aging & Social Policy*, 28(1): 15-28.

Mazzonna, Fabrizio and Franco Peracchi. 2012. "Ageing, Cognitive Abilities and Retirement" *European Economic Review*, 56(4): 691-710.

Musalem, Alberto R. and Maribel D. Ortiz. 2011. "Governance and Social Security: Moving Forward on the ISSA Good Governance Guidelines" *International Social Security Review*, 64(4): 9-37.

Nikolov, Plamen and Alan Adelman. 2019a. "Do Private Household Transfers to the Elderly Respond to Public Pension Benefits?" *The Journal of the Economics of Ageing*, 14(1): Forthcoming.

------. 2019b. "Short-Run Health Consequences of Retirement and Pension Benefits: Evidence from China" *Forum for Health Economics and Policy*, 21(2): 1-27.

Nunn, Nathan and Nancy Qian. 2014. "US Food Aid and Civil Conflict" *American Economic Review*, 104(6): 1630-1666.

Ofstedal, Mary B., Gwenith G. Fisher, and A. R. Herzog. 2005. "Documentation of Cognitive Functioning Measures in the Health and Retirement Study" *University of Michigan, HRS Documentation Report DR-006*, 10: 1-79.

Prull, Matthew W., John D. Gabrieli, and Silvia A. Bunge. 2000. "Age-Related Changes in Memory: A Cognitive Neuroscience Perspective." In *Handbook of Aging and Cognition*, 2nd ed., ed. Craik, Fergus. Salthouse, Timothy A., 91-153. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Qing, N. B. 2009. "The Useful Trial to Establish a New Rural Old-Age Insurance System: Investigation on the Pilot Project of Baoji, Shaanxi Province" *Theoretical Horizon*, 6: 40-42.

Quan, Lu. 2012. "Analysing the Coverage Gap in China" *London, HelpAge International*(Briefing no. 11): 16.

Rohwedder, Susann and Robert J. Willis. 2010. "Mental Retirement" *The Journal of Economic Perspectives*, 24(1): 119-138.

Rossin, Maya. 2011. "The Effects of Maternity Leave on Children's Birth and Infant Health Outcomes in the United States" *Journal of Health Economics*, 30(2): 221-239.

Sabates-Wheeler, Rachel and Johannes Koettl. 2010. "Social Protection for Migrants: The Challenges of Delivery in the Context of Changing Migration Flows" *International Social Security Review*, 63(3-4): 115-144.

Souchay, Cline, Michel Isingrini, and Laurence Espagnet. 2000. "Aging, Episodic Memory Feeling-of-Knowing, and Frontal Functioning." *Neuropsychology*, 14(2): 299.

The World Bank. 2017. "Population Growth" *The World Bank, Population Growth*: 1.

Tulving, Endel and Fergus I. Craik. 2000. *The Oxford Handbook of Memory,* 1st ed. Oxford, England.: Oxford: Oxford University Press.

UNC Chapel Hill, CHNS Research Team. 2015. "Map of Survey Regions" *University of North Carolina*.

------. 2010. "Survey Design" *University of North Carolina*.

United Nations. 2013. "Population Facts" *United Nations, Department of Economic and Social Affairs, Population Division*(December 2013, 8): 1-4.

------. 2007. "World Economic and Social Survey 2007, Development in an Ageing World" *United Nations, Department of Economic and Social Affairs, E/2007/50/Rev.1*(2007): 1-212.

------. 2019. "World Population Prospects 2019: Highlights" *United Nations, Department of Economic and Social Affairs*(June 2019): 1-2.

Vilela, Andrea. 2013. "Pension Coverage in China and the Expansion of the New Rural Social Pension" *HelpAge International*, ISBN 1872590799.

Wang, M. 1993. *Living Longer is Not Equal Living Healthy: Life Quality and Medical Security*, 1st ed. Beijing, China: Beijing: Chinese Economic Press.

Welsh, Kathleen, Nelson Butters, James Hughes, Richard Mohs, and Albert Heyman. 1991. "Detection of Abnormal Memory Decline in Mild Cases of Alzheimer's Disease using CERAD Neuropsychological Measures" *Archives of Neurology*, 48(3): 278-281.

Figures and Tables

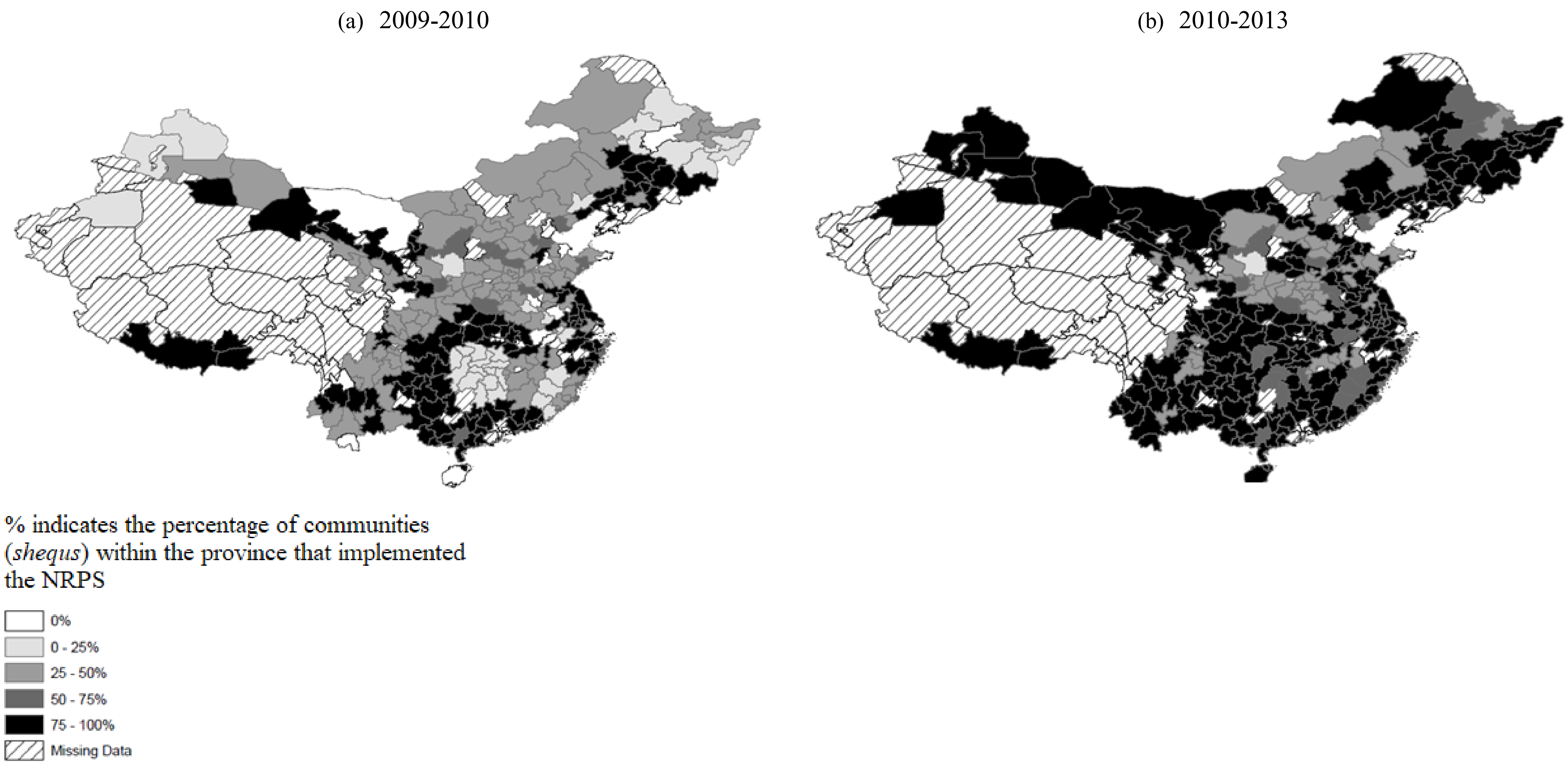

Fig 1. Geographic Implementation of NRPS. This figure shows the implementation of NRPS.

China Health and Retirement Longitudinal Study China Health and Nutrition Survey

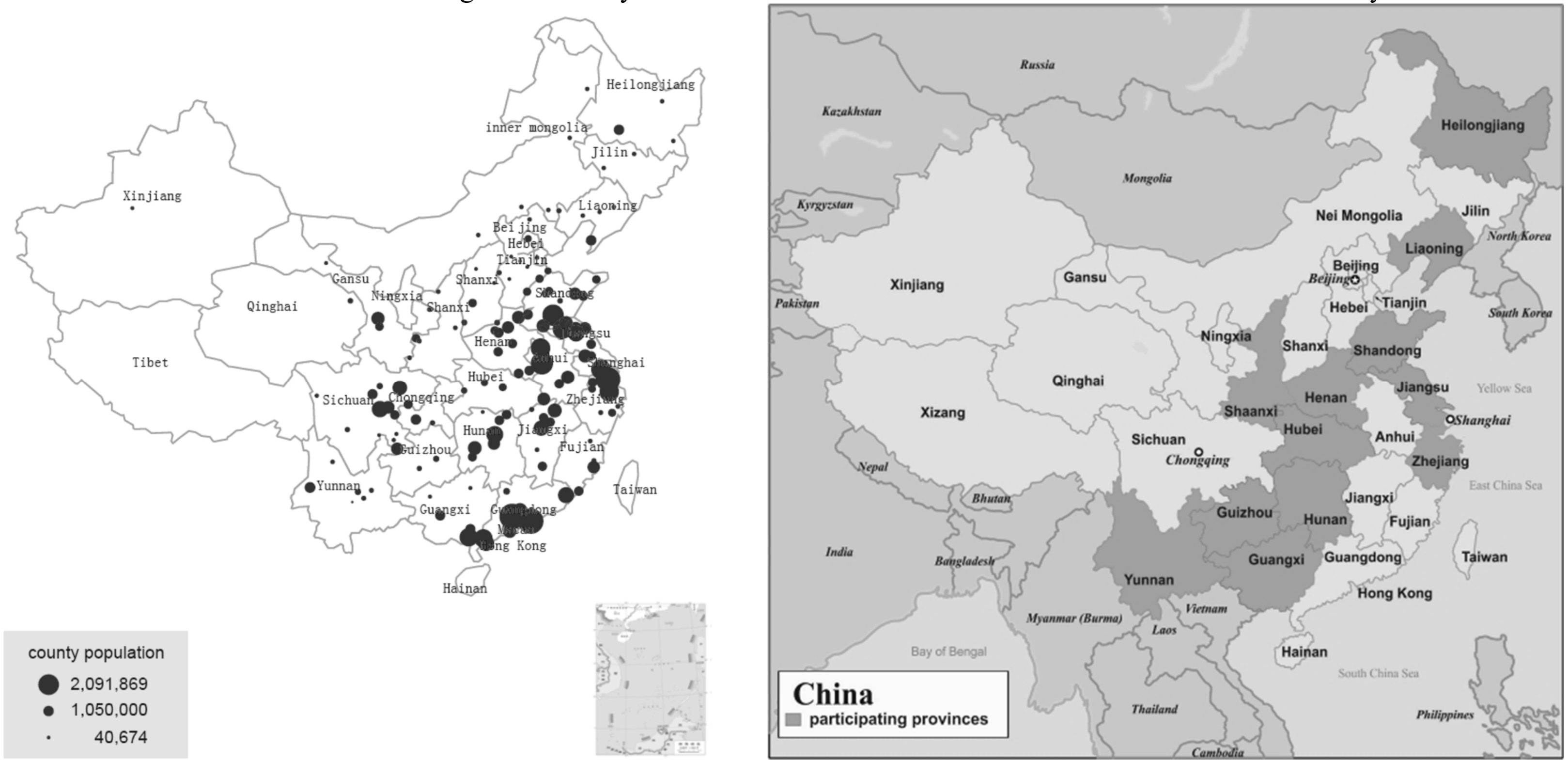

Fig 2. Coverage Maps. Source: CHARLS Research Team (2013) and CHNS Research Team (2015).

Table 1: Summary Statistics.

	Full Sample	Baseline		
		Participants	Non-Participants	p-value[a]
Demographics of Respondents				
Respondent's Age	59.31 (10.01)	58.43 (9.68)	58.44 (10.24)	0.99
# of Household Residents	3.74 (1.87)	3.68 (1.78)	3.75 (1.88)	0.04
# Living Children	2.77 (1.44)	2.81 (1.39)	2.74 (1.45)	0.07
Percent Female	0.53 (0.50)	0.54 (0.50)	0.53 (0.50)	0.38
Percent Married	0.80 (0.40)	0.81 (0.39)	0.78 (0.41)	0.00
Percent Living Near Children	0.90 (0.30)	0.91 (0.28)	0.92 (0.27)	0.40
Percent With At Least Lower Secondary Education	0.48 (0.50)	0.48 (0.50)	0.46 (0.50)	0.10
Labour Market and Health Outcomes				
Weekly Work Hours	45.45 (23.87)	47.26 (24.07)	46.89 (22.70)	0.50
Percent Currently Working	0.70 (0.46)	0.70 (0.46)	0.69 (0.46)	0.11
Percent Working in Agriculture	0.72 (0.45)	0.72 (0.45)	0.73 (0.45)	0.49
Percent Reporting Poor/Fair Health	0.25 (0.43)	0.27 (0.44)	0.26 (0.44)	0.23
Respondent's BMI	23.40 (3.84)	23.62 (3.91)	23.05 (3.81)	0.00
Percent Visited Doctor (Past Month)	0.20 (0.40)	0.20 (0.40)	0.19 (0.39)	0.08
Percent Stayed in Hospital (Past Year)	0.11 (0.31)	0.10 (0.29)	0.09 (0.28)	0.06
Percent Ever Smoked	0.41 (0.49)	0.40 (0.49)	0.40 (0.49)	0.98
Percent Smoking Now	0.25 (0.44)	0.29 (0.45)	0.30 (0.46)	0.40
Cognition				
Immediate Recall Score	3.79 (1.76)	3.93 (1.69)	3.77 (1.70)	0.00
Delayed Recall Score	2.86 (2.00)	2.91 (1.91)	2.89 (1.96)	0.61
Total Recall Score	6.67 (3.47)	6.85 (3.32)	6.68 (3.36)	0.02
Cognitive Memory Index	0.00 (1.43)	0.06 (1.38)	0.00 (1.39)	0.06
Observations	28,034	10,011	3,680	

Notes: Standard deviations are reported in parenthesis. (a) We test the null hypothesis that the difference in participant and non-participant means is equal to 0. (b) Low (or Negative) values denote lower performance on the cognition test.

Table 2: ITT and LATE Estimates on Cognitive Outcomes.

	Immediate Word Recall[a]	Delay Word Recall[a]	Total Recall[a]	Cognitive Memory Index[b]
	(1)	(2)	(3)	(4)
Panel A (ITT):				
Offered NRPS *	-0.144***	-0.230***	-0.353***	-0.103**
Above60 [c]	(0.052)	(0.052)	(0.093)	(0.040)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.230	0.215	0.247	0.313
Observations	21,202	21,202	21,202	21,202
Panel B (TOT):				
NRPS Participation * Above60 [d]	-0.208* (0.120)	-0.425*** (0.122)	-0.633*** (0.214)	-0.212** (0.087)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	241.242	241.242	241.242	241.242
R-squared	0.065	0.152	0.102	0.110
Observations	21,202	21,202	21,202	21,202

Notes: (a) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (b) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. (c) Our DDD coefficient: Policy instrument interacted with an indicator for being over 60 years old. Individual level controls: Above60 (1= Yes), Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. Columns 1-8 are estimated using Ordinary Least Squares (OLS) with Community and Year FE. Panel A is estimated using Ordinary Least Squares (OLS) with Community, Year and Community*Year FE. Panel B is estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Table 3: LATE Estimates of Retirement on All Outcomes.

	Immediate Word Recall[b]	Delay Word Recall[b]	Total Recall[b]	Cognitive Memory Index[c]
	(1)	(2)	(3)	(4)
Retired (Yes=1)[a]	-12.761	-13.290	-22.620	-13.282
	(9.746)	(10.119)	(18.320)	(8.277)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	8.47	8.56	8.62	8.88
Beta (First-Stage)	0.005	0.006	0.005	0.005
SE (First-Stage)	0.002	0.002	0.002	0.002
Observations	22,199	22,199	22,199	22,199

Notes: (a) Directly asked about retirement procedure. "Have you completed retirement procedures (including early retirement) or internal retirement (Retirement from government departments, enterprises and institutions, not including retirement in the sense of getting agricultural insurance)?" A positive answer is coded as being retired. (b) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (c) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. Individual level controls: Age, Age Squared, Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. Columns (1) through (4) are estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Table 4: Heterogeneous Treatment Effects Using Male and Female Samples.

Panel A (ITT):	**Immediate Word Recall**	**Delayed Word Recall**	**Total Recall**	**Cognitive Index**
	(1)	(2)	(3)	(4)
ITT Male Sample:				
Offered NRPS * Above60 [c]	-0.032	-0.198***	-0.230*	-0.073
	(0.071)	(0.075)	(0.131)	(0.053)
Baseline Mean	3.911	2.935	6.868	0.195
Controls	Yes	Yes	Yes	Yes
R-squared	0.235	0.236	0.256	0.290
Observations	10,121	10,121	10,121	10,121
ITT Female Sample:				
Offered NRPS * Above60 [c]	-0.140***	-0.197**	-0.336**	-0.121***
	(0.067)	(0.082)	(0.135)	(0.052)
Baseline Mean	3.689	2.798	6.514	-0.175
Controls	Yes	Yes	Yes	Yes
R-squared	0.270	0.250	0.286	0.350
Observations	11,003	11,003	11,003	11,003
Panel B (TOT):				
TOT Male Sample:				
NRPS Participation *	-0.071	-0.445***	-0.516*	-0.165
Above60 [d]	(0.159)	(0.171)	(0.296)	(0.119)
Baseline Mean	3.911	2.935	6.868	0.195
Controls	Yes	Yes	Yes	Yes
R-squared	0.235	0.234	0.255	0.289
F-Stat (First Stage)	212.541	212.541	212.541	212.541
Observations	10,121	10,121	10,121	10,121
TOT Female Sample:				
NRPS Participation *	-0.323**	-0.455**	-0.778**	-0.280***
Above60 [d]	(0.154)	(0.189)	(0.312)	(0.121)
Baseline Mean	3.689	2.798	6.514	-0.175
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	230.04	230.04	230.04	230.04
R-squared	0.269	0.249	0.284	0.349
Observations	11,003	11,003	11,003	11,003

Notes: (a) Directly asked about retirement procedure. "Have you completed retirement procedures (including early retirement) or internal retirement (Retirement from government departments, enterprises and institutions, not including retirement in the sense of getting agricultural insurance)?" A positive answer is coded as being retired. (b) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (c) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. Individual level controls: Age, Age Squared, Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. The ITT effects are estimated using Ordinary Least Squares (OLS) with Community, Year and Community*Year FE. Panel B is estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Online Appendix A

A. Robustness Checks

Table A1: Test of Common Trends Using CHNS Data.

		Immediate Word Recall[a]	Delayed Word Recall[a]	Total Recall[a]
		(1)	(2)	(3)
67% Coverage Rate Threshold	Treatment * 2004	-0.190	-0.407	-0.472
		(0.248)	(0.286)	(0.511)
	Treatment * 2009	-0.027	-0.241	-0.273
		(0.238)	(0.249)	(0.475)
	R-Squared Adj	0.127	0.142	0.147
	Year FE	Yes	Yes	Yes
	Community FE	Yes	Yes	Yes
	Observations	4,742	4,719	4,615
70% Coverage Rate Threshold	Treatment * 2004	0.094	-0.123	0.109
		(0.267)	(0.314)	(0.546)
	Treatment * 2009	0.004	-0.302	-0.303
		(0.264)	(0.265)	(0.522)
	R-Squared Adj	0.126	0.141	0.146
	Year FE	Yes	Yes	Yes
	Community FE	Yes	Yes	Yes
	Observations	4,742	4,719	4,615

Notes: Source: CHNS 2000, 2004 and 2006 Waves. Base year is 2006. (a) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. Columns 1-4 are estimated using Ordinary Least Squares (OLS) with Community and Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Table A2: Falsification Test Using Placebo Sample.

	Immediate Word Recall[a]	Total Recall[a]	Cognitive Memory Index[b]
	(1)	(2)	(3)
Offered NRPS * Above60[c]	-0.098	-0.710	-0.183
	(0.245)	(0.494)	(0.207)
Baseline Mean	0.253	0.000	0.000
Controls	Yes	Yes	Yes
R-squared	0.611	0.625	0.620
Observations	604	594	576

Notes*:* (a) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (b) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. (c) Our DDD coefficient: Policy instrument interacted with an indicator for being over 60 years old. A significant coefficient suggests the differential treatment towards urban pensioners in treated communities relative to urban pensioner in control communities; a cause of concern for the instrument's validity. Individual level controls: Above60 (1=Yes), Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. The specifications are estimated with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Table A3: Test on Placebo Outcomes for Specifications (1) and (3).

	Han (=1 if yes)	# Dead Daughter	Mother's Educ	# of Living Sons
	(1)	(2)	(3)	(4)
Panel A (ITT):				
Offered NRPS * Above60 [a]	-0.004	-0.013	-0.015	0.010
	(0.004)	(0.029)	(0.012)	(0.024)
Baseline Mean	0.920	1.299	1.190	1.466
Controls	Yes	Yes	Yes	
R-squared	0.652	0.165	0.130	0.235
Observations	20,102	21,202	19,656	21,202
Panel B (TOT):				
NRPS Participation * Above60 [b]	-0.010	-0.032	-0.035	0.025
	(0.009)	(0.071)	(0.030)	(0.059)
Baseline Mean	0.920	1.299	1.190	1.466
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	282.617	279.6213	291.9617	279.6213
R-squared	0.652	0.165	0.130	0.235
Observations	20,102	21,202	19,656	21,202

Notes: (a) Our DDD coefficient (Policy instrument interacted with an indicator for being over 60 years old). The control group becomes individuals under the Age of 60 living in eligible communities that didn't offer NRPS between 2011 and 2013. (b) Individual participation instrumented with the policy variable. Individual level controls: Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. Panel A is estimated using Ordinary Least Squares (OLS) with Community, Year and Community*Year FE. Panel B is estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

B. Additional Robustness Checks

Table B1: PCA Weights (Component Loadings).

Cognitive Index	
Variable	**Loading**
Immediate Word Recall	0.595
Delayed Word Recall	0.588
Serial 7	0.414
Self-Reported Memory	0.137
Knows Current Month (Yes=1)	0.331

Table B2: ITT and LATE Estimates on Cognition using Propensity Score for NRPS Participation.

	Immediate Word Recall[a]	**Delay Word Recall**[a]	**Total Recall**[a]	**Cognitive Memory Index**[b]
	(1)	(2)	(3)	(4)
Panel A (ITT):				
	NRPS (=1 if Propensity >= Mean + .5 SD			
Offered NRPS *Above60[c]	-0.052	-0.182***	-0.234**	-0.054
	(0.053)	(0.060)	(0.103)	(0.042)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.238	0.220	0.254	0.317
Observations	18,487	18,487	18,487	18,487
	NRPS (=1 if Propensity >= Mean + 1 SD			
Offered NRPS * Above60[c]	-0.075	-0.198***	-0.273***	
	(0.052)	(0.059)	(0.101)	-0.079* (0.041)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.230	0.216	0.247	0.310
Observations	20,309	20,309	20,309	20,309
Panel B (TOT):				
	NRPS (=1 if Propensity >= Mean + .5 SD			
PrNRPS *Above60[d]	-0.112	-0.396***	-0.508**	-0.118
	(0.116)	(0.133)	(0.226)	(0.091)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	458.733	458.733	458.733	458.733
Observations	18,487	18,487	18,487	18,487
	NRPS (=1 if Propensity >= Mean + 1 SD			
PrNRPS *Above60[d]	-0.161	-0.427***	-0.588***	-0.170*
	(0.112)	(0.129)	(0.220)	(0.088)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	348.111	348.111	348.111	348.111
Observations	20,309	20,309	20,309	20,309

Notes: (a) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (b) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. (c) Our DDD coefficient: Policy instrument interacted with an indicator for being over 60 years old. Individual level controls: Above60 (1= Yes), Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. Columns 1-8 are estimated using Ordinary Least Squares (OLS) with Community and Year FE. Panel A is estimated using Ordinary Least Squares (OLS) with Community, Year and Community*Year FE. Panel B is estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Table B3: ITT and LATE Estimates on Cognition Omitting Mismeasured Communities.

	Immediate Word Recall[a]	Delay Word Recall[a]	Total Recall[a]	Cognitive Memory Index[b]
	(1)	(2)	(3)	(4)
Panel A (ITT):				
	Sample excluding communities with less than 4 participants			
Offered NRPS *Above60[c]	-0.093*	-0.194***	-0.287***	-0.091**
	(0.054)	(0.058)	(0.101)	(0.040)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.226	0.213	0.245	0.310
Observations	19,566	19,566	19,566	19,566
	Sample excluding communities with less than 7 participants			
Offered NRPS * Above60[c]	-0.100*	-0.198***	-0.297***	-0.089**
	(0.054)	(0.058)	(0.102)	(0.041)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.225	0.211	0.243	0.307
Observations	19,057	19,057	19,057	19,057
Panel B (TOT):				
	Sample excluding communities with less than 4 participants			
PrNRPS *Above60[d]	-0.228*	-0.472***	-0.700***	-0.221**
	(0.130)	(0.141)	(0.246)	(0.098)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	243.0807	243.0807	243.0807	243.0807
Observations	19,566	19,566	19,566	19,566
	Sample excluding communities with less than 7 participants			
PrNRPS *Above60[d]	-0.238*	-0.472***	-0.711***	-0.213**
	(0.130)	(0.139)	(0.243)	(0.098)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	260.183	260.183	260.183	260.183
Observations	19,057	19,057	19,057	19,057

Notes: (a) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (b) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. (c) Our DDD coefficient: Policy instrument interacted with an indicator for being over 60 years old. Individual level controls: Above60 (1= Yes), Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. Columns 1-8 are estimated using Ordinary Least Squares (OLS) with Community and Year FE. Panel A is estimated using Ordinary Least Squares (OLS) with Community, Year and Community*Year FE. Panel B is estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Table B4: ITT and LATE Estimates on Direct Measures of Health Varying the Definition of Instrument.

	Immediate Word Recall[a]	Delay Word Recall[a]	Total Recall[a]	Cognitive Memory Index[b]
	(1)	(2)	(3)	(4)
Panel A (ITT):				
	Offer NRPS (=1 if at least 4 in community participate)			
Offered NRPS *Above60[c]	-0.070	-0.198***	-0.268***	-0.088**
	(0.049)	(0.057)	(0.096)	(0.039)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.229	0.216	0.247	0.313
Observations	21,202	21,202	21,202	21,202
	Offer NRPS (=1 if at least 7 in community participate)			
Offered NRPS * Above60[c]	-0.061	-0.178***	-0.239**	-0.078**
	(0.049)	(0.057)	(0.096)	(0.039)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.229	0.215	0.247	0.313
Observations	21,202	21,202	21,202	21,202
Panel B (TOT):				
	Offer NRPS (=1 if at least 4 in community participate)			
PrNRPS *Above60[d]	-0.162	-0.458***	-0.620***	-0.204**
	(0.115)	(0.135)	(0.226)	(0.091)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	240.613	240.613	240.613	240.613
Observations	21,202	21,202	21,202	21,202
	Offer NRPS (=1 if at least 7 in community participate)			
PrNRPS *Above60[d]	-0.140	-0.405***	-0.544**	-0.176**
	(0.113)	(0.132)	(0.222)	(0.089)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	258.336	258.336	258.336	258.336
Observations	21,202	21,202	21,202	21,202

Notes: (a) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (b) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. (c) Our DDD coefficient: Policy instrument interacted with an indicator for being over 60 years old. Individual level controls: Above60 (1= Yes), Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. Columns 1-8 are estimated using Ordinary Least Squares (OLS) with Community and Year FE. Panel A is estimated using Ordinary Least Squares (OLS) with Community, Year and Community*Year FE. Panel B is estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level

Table B5: ITT and LATE Estimates on Cognition. City-Level Analysis (Heilongjiang).

	Immediate Word Recall[a]	**Delay Word Recall**[a]	**Total Recall**[a]	**Cognitive Memory Index**[b]
	(1)	(2)	(3)	(4)
Panel A (ITT):				
		CHARLS Data		
Offered NRPS *Above60[c]	-0.317	-0.151	-0.468	-0.213
	(0.267)	(0.263)	(0.476)	(0.189)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	0.113	0.080	0.106	0.149
Observations	178	178	178	178
		Admin Data		
Offered NRPS * Above60[c]	0.272	-0.288	-0.016	-0.057
	(0.693)	(0.762)	(1.308)	(0.507)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
R-squared	.107	.079	.099	.145
Observations	178	178	178	178
Panel B (TOT):				
		CHARLS Data		
PrNRPS *Above60[d]	-1.558	-0.741	-2.299	-1.047
	(1.374)	(1.278)	(2.369)	(0.966)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	10.075	10.075	10.075	10.075
Observations	178	178	178	178
		Admin Data		
PrNRPS *Above60[d]	-0.557	0.590	0.033	0.117
	(1.407)	(1.595)	(2.682)	(1.044)
Baseline Mean	3.792	2.862	6.678	0.000
Controls	Yes	Yes	Yes	Yes
F-Stat (First Stage)	8.526	8.526	8.526	8.526
Observations	178	178	178	178

Notes: (a) Word recall tests: Immediate Recall = [0,10], Delayed Recall = [0,10] and Total Recall = [0,20]. (b) We created the Cognitive Memory Index using principal component analysis, combing measures of short/long term memory, working memory and orientation. (c) Our DDD coefficient: Policy instrument interacted with an indicator for being over 60 years old. Individual level controls: Above60 (1= Yes), Marital Status (=1 if Married), Gender (=1 if Female), Education Levels (Base Group is illiterate with no formal education), # of Household Residents. Columns 1-8 are estimated using Ordinary Least Squares (OLS) with Community and Year FE. Panel A is estimated using Ordinary Least Squares (OLS) with Community, Year and Community*Year FE. Panel B is estimated using Two-Stage Least Squares (2SLS) with Community, Year and Community*Year FE. Clustered standard errors at the community level reported in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level