
## Radboud University

# Master's Thesis

## Factors influencing the likelihood of retirement: Evidence from South Africa.

**Abstract:** Owing to advances made in medicine and science, recent trends have shown an unprecedented increase in the average life expectancy of humans. The increased life longevity, compounded by inadequate financial planning and an increased strain on pension systems, has increased old-age poverty. This paper plans to extend prior research by investigating whether factors that influence pre-retirement behaviour mirror reality when it comes to retirement decisions. Using baseline survey responses from 1,926 individuals between the ages of 60 and 105 obtained from the HAALSI study conducted in Agincourt (South Africa), a probit model was estimated to investigate whether various demographic and self-reported measures influenced the likelihood of being retired. The results indicated that, in addition to demographic covariates, certain ratings of self-reported health, memory, and perceived life longevity had a significant influence on the likelihood of being retired. Whereas self-reported measures of belongingness and HIV status were not influential predictors. Therefore, the findings indicate that in addition to re-structuring pension systems in response to changes in life longevity, behavioural policies should be implemented to ensure that more elderly can comfortably retire in the developing world.

**Specialization:** Economics, Behaviour and Policy

**Author:** Nicholas Veenis - s1030608

**Supervisor:** Dr. Sven Nolte

15/08/2021

# Table of Contents

**Key words:** Retirement, Life Expectancy/Longevity, Psychological Attitudes, HAALSI, Demographic Characteristics, South Africa.

# 1. Introduction

Over the past decade, advances made in the fields of medicine and science have resulted in an unprecedented increase in the life expectancy of humans. Since progress challenges age-old perceptions of human longevity by pushing the pre-conceived boundaries of reality, people are finding that it is becoming increasingly difficult to plan for retirement in the face of an uncertain future horizon. In line with the increasing average life expectancy, trends have shown that more old-aged people are living in poverty. In turn, research regarding retirement behaviour has garnered increased attention among policymakers and academics over the years. However, such research is somewhat complicated because the composite individual decision of when to retire differs among individuals based on various demographic and personal characteristics. Therefore, research into the retirement behaviour of individuals (especially in the developing regions of the world) deserves renewed attention to ensure that actions can be taken to negate the pitfalls of medical and scientific progress.

Since retirement, as defined by the Oxford Lexico dictionary is "the action or fact of leaving one's job and ceasing to work" (Lexico, 2021), the decision to retire is molded by an individual's ability and/or propensity to save for retirement whilst they are employed and relies on the amount of financial support received during retirement (whether private or government). Therefore, inadequate pre-retirement behaviour combined with an increasing tendency to retire earlier has become a major economic problem in developed countries as people are living for longer and need to be financially supported by those who are working (Hershey and Mowen, 2000). The increasing strain on pension system reliance means that the "onus is inevitably shifting on individual savings to supplement the necessary minimum provided by public pension schemes" (Fernández-López et al., 2010) in many developed countries. Moreover, retirement decisions are even more complex in developing countries that lack the same kinds of broad-based public pension schemes and research has shown that many elderly people have been "left to rely on their own current and accumulated earnings and support from children as means of old-age support" (Cameron and Cobb-Clark, 2008). Therefore, research into retirement behaviour will provide useful insights for both developed and developing regions of the world.

Retirement in the developed world is vastly different from retirement in the developing world. According to the increasing life expectancy, both the amount of time spent retired and the

number of retired people has increased in the developed world (Topa, Lunceford and Boyatzis, 2017). A possible reason for the increase is that "retired elderly people work because they want to, not because they need to" (Ford, Park and Sen, 2009) in developed countries. However, the same is not true for developing countries where individuals are prone to run out of money during retirement and are therefore "compelled to re-join the workforce even once they have reached retirement age" (Zeka, 2016). Interestingly, even though "South Africa's elderly have access to an old-age pension system that is among the most generous in the developing world" (Lam, Leibbrandt and Ranchhod, 2006), they are still more likely to work much longer time horizons than those in the developed world.

Being able to retire is by no means a universal guarantee. Government laws and regulations play a major role in guaranteeing retirement. Retiring in the Netherlands, where there is a Dutch state pension scheme, is comparatively easier than being able to retire in South Africa, where the pension system is "composed of a non-contributory, means-tested public benefit program" (Pension system in South Africa, 2021). The composition of the pension system in South Africa largely shifts the responsibility to save for retirement to the individuals. The high levels of unemployment and inability to save for retirement is a direct reflection of why SA "has one of the lowest saving rates in the world, and less than 4% of the citizens are adequately prepared for retirement" (Gore, 2018). Therefore, the inability to adequately prepare for retirement due to the increased life expectancy influences both developed and developing regions of the world. Research into the causes could be used to combat extremely low saving rates and ensure that more people are able to retire.

The World Health Organization has estimated that "by 2050, the world's population aged 60 years and older is expected to total 2 billion, up from 900 million in 2015" (WHO, 2018). Furthermore, the WHO estimates that roughly 80% of the 2 billion people will live in low- and middle-income countries. Furthermore, due to the demographic transition that South Africa is currently experiencing, changing the behaviour of the working-age population would ensure that future generations are able to retire. South Africa is currently in stage three of the demographic transition and "taking advantage of its growing working-age population could help accelerate South Africa's growth to 5.4 percent a year and double per capita incomes by 2030" (The World Bank, 2015). According to data published by the World Bank, South Africa's age dependency

ratio has decreased from 84.64% in 1966 to 52.4% in 2019. This indicates that the working-age population has less economic dependencies and is, therefore, able to allocate more of their income towards preparing for retirement. Individual's attitudes shape their likelihood of being able to retire.

Although global life expectations have increased considerably over the past decade, there is still a massive difference between the developed and developing regions of the world. For example, the life expectancy at birth in the Netherlands, in 2018, was 81.812 years whereas, it was only 63.857 years in South Africa (The World Bank, 2019). However, from 2005 to 2018, South Africa exhibited a far larger increase in life expectancy at birth of 10.41 years than the 2.466-year increase experienced in the Netherlands over the same period. The steep increase in the life expectancy of South Africans is a result of a scale-up of antiretroviral therapy (ART) access (Bor et al., 2013). The WHO estimates that the downward trend present in the estimated number of people newly infected with HIV, and the number of deaths due to AIDS will continue to decrease as the ART coverage increases. Health and quality of life are significant factors to consider when analyzing retirement decisions for a country like South Africa where HIV is an epidemic.

Approximately half of the OECD countries have been fast to respond to the increase in human life longevity by introducing "elements in their mandatory retirement-income provision that provide an automatic link between pensions and a change in life expectancy" (OECD, 2011). However, developing countries that do not have advanced pension systems have not been able to react. Since "families that receive a pension [in South Africa] are 11% less likely to become poor" (Stewart and Yermo, 2009) research that can be used to positively re-structure pension systems for the developing regions of the world should be prioritized to reduce old-age poverty as well to ensure that the labour force remains productive.

This paper investigates whether various demographic and self-reported factors have a significant influence on the likelihood of being retired for old-aged individuals living in a rural area of South Africa. Although, due to a shortage of literature on retirement behaviour in the developing world and since decisions are molded by attitudes, this paper is primarily concerned with the influence that self-reported health, HIV status, memory (as a proxy measurement of cognitive ability), belongingness, and life longevity have on the likelihood of being retired in South Africa. Therefore, I postulate that since individuals have different expectations, perceptions, and

experiences throughout their lives, the probability of being retired, during old age, is influenced by psychological attitudes.

The likelihood of being retired was predicted using a probit model which dichotomously classified individuals, between the ages of 60 and 105, who were able to work (i.e., not on sick leave or disabled) as being either retired or not retired (currently working or seeking employment). The model was built using baseline survey responses obtained from the 'Health and Aging in Africa: A longitudinal Study of an INDEPTH Community' (HAALSI) study of elderly individuals living in Agincourt, South Africa. Findings revealed that adding self-reported ratings of health, cognition, belongingness, and perceived life longevity to a model containing demographic characteristics, increased the ability to accurately predict who was retired. Therefore, self-reported attitudes had a significant influence on the likelihood of being retired in a rural area of South Africa. Moreover, in addition to an inverse U-shape relationship between age and the likelihood of retirement, statistically significant results revealed that unmarried, partially educated, and comparatively wealthier individuals who have equally bad ratings of their health and cognitive ability to remember, as well as a superior rating of perceived life longevity were on average the most likely to be retired. Also, the results showed that gender, HIV status, and belongingness did not have a significant influence on the likelihood of retirement.

Section 2 of this paper provides a critical review of the relevant literature streams related to retirement behaviour. In section 3, various main and sub-hypotheses were elaborated. Section 4 describes the data and variables used to specify the model. Section 5 presents the descriptive and empirical findings. Section 6 provides a conclusion wherein strengths, limitations, and motivations for future research are discussed. Section 7 consists of a list of references. Finally, section 8 consists of an Appendix containing various statistical tests.

# 2. Literature Review

There is a plethora of theoretical and empirical literature related to retirement behaviour. Primary research revealed an apparent shortage of literature that investigates the probability of being retired in old age. Further, inspection revealed that most of the research examines the pre-retirement behaviour of individuals (leading up to retirement age), centered in the developed regions of the world, and to a lesser extent on the developing regions. Various sociodemographic, psychological, and health-related factors emerged as consistent factors used to examine retirement behaviour among prior literature. Researching the influence that such factors have on the probability of being retired will contribute to understanding the difference between theory and reality. The axiom that a synthesis of such important factors is the best predictor for the likelihood of being retired is justified by the results obtained in prior literature, as highlighted in the following section. The various streams of research regarding pre-retirement behaviour and its influences on the timing of retirement amalgamate under a central theme of retirement behaviour.

Failing to prepare for retirement has an influence on an individual's likelihood of being retired and as a result, they will be more likely to be employed or seeking employment in their old age. Pang, Warshawsky and Weitzer, (2008) support this stance by linking various factors used to assess pre-retirement planning with the likelihood of being retired. Zeka (2017) found evidence of a significant relationship "between retirement age decisions and financial preparations on the one hand, and the attitudes of individuals towards retirement on the other", which further provides support to critically analyze both streams of research together. In line with prior literature, an explicit expectation would be that individuals who are more prone to saving for retirement, will have a higher likelihood of being retired in their old age than individuals who have a negative attitude towards preparing for retirement. The current literature is linked to previous literature as follows:

## 2.1 Demographic Characteristics

The factors, influences, and implications of the rapidly increasing average life expectancy have become the main topic of discussion among academics and researchers over the past two decades. Research has increased considerably as retirement has become less of a rarity and more people are able to retire than at the beginning of the 20th century (Kalemli-Ozcan and Weil, 2010). Uccello (1998) hypothesized that the normal retirement age in the United States of America would

"increase gradually over the next 25 years, from age 65 to 67" to reduce the strain placed on the social retirement fund. Further research by Davey in 2008, justified Uccello's (1998) stance by conducting a study in New Zealand to investigate the factors that influence retirement decisions. The results indicated that government "policies will [likely] seek to influence decisions about retirement and prolong workforce participation" (Davey, 2008). Instead of trying to prolong workforce participation by extending the retirement age and having a less productive labour force, research should focus on analyzing factors that influence retirement decisions.

A central theme in the literature is the assumption that individuals will behave according to the life-cycle hypothesis, which suggests that they will tend to be savers during their employment years and become dis-savers during retirement years. In addition to income, Richardson and Kilty (1989) concluded that age was an important predictor of retirement planning. Lakshmi-Malroutu and Xiao (1995) furthered the research and found that "the closer people [are] to retirement, the more likely they [are] to invest or save". Hershey and Mowen (2000) mentioned that "the fact that most people wait too long before establishing a personal savings program" has become a major economic problem for developed countries (Singleton and Keddy, 1991).

Furthermore, Fernández-López et al. (2010) found that "the probability of saving for retirement follows an inverted U-shaped pattern in age", but when considering developing countries, Ting and Kollamparambil (2015) argued against the existence of such an inverse U-shape in South Africa. Although age is used extensively in the literature to investigate the life-cycle hypothesis when it comes to preparing for retirement, Pang, Warshawsky and Weitzer (2008) demonstrate the importance of analyzing the nonlinear effect that age has on predicting the retirement timing of individuals in old age. Therefore, including age to test for the existence of a relationship between predicting the likelihood of being retired, and when the tradeoff between employment and retirement occurs is vindicated by prior research.

Gender, marital status, and education are demographic characteristics that are used extensively in the research to derive meaningful insights into retirement decisions. One of the key findings of the research by Singleton and Keddy (1991) was that age, income, gender, and education all had an influence on the expected retirement age of a sample of 219 Canadian University employees. Shifting the focus to the developing world, research by Lidi, Bedemo and Belina (2017) also found strong empirical evidence that age, income, and education positively influenced a household's

decision to save in a Rural area of Ethiopia. Their findings were not restricted to Ethiopia as Zwane, Greyling and Maleka (2016) also concluded that the same factors (in addition to employment status) had an influence on South African households' saving behaviour. More specifically, Willows (2019) found that "single, younger and male respondents [were more] likely to have sufficient retirement savings". The same is not true for factors that influence retirement decisions in the USA since Uccello (1998) found that marital status "does not appear to be a significant factor [related to retirement] for either men or women". Accordingly, an expectation of this research (following prior literature) is that age, gender, and education will be significant variables and that marital status will be an insignificant variable when it comes to predicting the likelihood of retirement.

Education was used consistently as a predictor for retirement behaviour in past literature. A noteworthy paper by McAllister et al. (2020) investigated the probability of employment for individuals older than 60 living in Denmark, Sweden, Canada, and England. They found that "the odds of employment beyond age 60 is lower for groups with low education, particularly women, and those with physical-mental health comorbidities". They further advocated that policies put in place to extend working lives should be revised to ensure equality for all. A prior expectation of this paper is that individuals with less education will be more likely to be retired than those with more education and that gender will have a significant influence on the likelihood of being retired.

## 2.2 Socioeconomic Indicator of Wealth

Over the past few decades, behavioural economics has developed and grown as a school of economic thought and as a result, there has been an increase in research regarding the role that behaviour plays in an individual's tendency to save a portion of their income. Bozio et al. (2013) used consumption and wealth data from the UK to find that there is "a positive relationship between savings rates (and wealth accumulation) and levels of permanent income". Their findings were consistent with prior research conducted in the USA by Dynan, Skinner and Zeldes (2004). Further research revealed that "saving outside of pensions [in the UK] has become more common since 2012, having [briefly] fallen between 2008 and 2012" (Crawford et al., 2020). In addition to being associated with an increase in the saving rate, Bloemen (2009) showed that an increase in wealth resulted in an increase of the early retirement rate in the Netherlands. Although the average

level of wealth in developed countries is not comparable to developing countries, the shift in mentality and responsibility of using wealth to save for retirement warrants universal research.

In line with the evolution of behavioural economics, the wealth effect has become an integral economic theory used to form the assumption that people tend to spend more as the value of their assets increases. Salotti's (2010) empirical analysis of 18 developed countries, from 1980 to 2005, found that household savings were adversely affected by an increase in wealth (contrary to results obtained by Bozio et al. (2013) and Dynan et al. (2004). However, when population dependency ratios and government savings were included in the model, wealth did not negatively influence household savings in the USA (Salotti, 2010). Additional research by Aron and Muellbauer (2000) validated the existence of a wealth effect in South Africa. Simleit, Keeton and Botha (2011) confirmed their findings as their results suggested that a decline in household savings is partly influenced by the wealth effect. The paper by Pang, Warshawsky and Weitzer (2008) contributed to understanding the link between preparing for retirement and the likelihood of retirement. Their paper found that an increase in wealth resulted in an increase in the probability of retiring for old workers in the USA. Even though an increase in wealth is linked to a decrease in household savings leading up to retirement, the same increase later in life has a significant impact on the decision to retire. A prior expectation of this paper is that an individual's household wealth level will have a significant impact on their likelihood of being retired in old age.

## 2.3 Self-Reported Health and HIV Status

### 2.3.1 Self-Reported HIV Status

Hitherto, there has been no literature published that examines the influence that HIV has on perceived life expectancy and the likelihood of being retired in South Africa. As mentioned above, a large majority of the existing literature examines the developed regions of the world. Shifting focus away from Europe, New Zealand, and the USA towards Africa will further research by including factors (such as HIV/AIDS) that affect both the quality and expectancy of life for individuals living in developing regions and have little to no significance for individuals living in developed regions. Therefore, including a measure for HIV status (albeit self-reported) is of paramount importance as the access to antiretroviral therapy (ART) increases. Bor et al. (2013), measured the influence that access to ART had on the adult life expectancy of South Africans from 2000 to 2011. The authors were able to conclude that "the survival benefits of ART far outweigh

the costs of providing treatment" (Bor et al., 2013), and that the life expectancy increased by roughly 11 years from 2003 to 2011. Prior research forms the expectation that HIV status will have an interaction effect with other measures (such as subjective life expectancy) and that being HIV negative will have a positive influence on the probability of being retired.

Furthermore, since being HIV positive negatively impacts health status, which in turn adversely affects the ability to work, individuals who know that they are HIV positive will be less likely to be employed or seeking employment in their old age. Independent studies by Rohr et al. (2017) and Chasimpha et al. (2020) assessed the accuracy of self-reported HIV statuses in South Africa and Malawi respectively. On the one hand, Rohr et al. (2017) reported that, of the respondents who were tested for HIV, only 50.9% knew their HIV status and were able to accurately report it. On the other hand, Chasimpha et al. (2020) reported that 86.4% of HIV-positive individuals correctly reported their HIV status. Whereas Lorem et al. (2020) contributed to the research by investigating whether there was a link between mortality and self-reported health in Norway. Their findings indicated that self-reported health was "affected by disease, mental health, and other risk factors" and that it was able to predict mortality with a time-dependent effect. Therefore, due to its prevalence in South Africa, including self-reported HIV status (in addition to health rating) is of paramount importance for research into the retirement behaviour of the elderly to understand the long-run impacts of the HIV epidemic.

### 2.3.2 General Health Rating

A copious amount of literature seeks to examine the influence that health has on retirement behaviour. After analyzing interview responses conducted in New Zealand, Davey (2008) was able to confirm that respondent's health had an impact on their temporal decision to retire. Moreover, her research found a correlation between poor health and a withdrawal from work. A Swedish cohort study by Sousa-Ribeiro et al. (2021) corroborated her findings by concluding that individuals who maintained a better health prospectus were more likely to be "pulled toward working until 65 and beyond". Further findings by Disney, Emmerson and Wakefield (2003) provided support that adverse health shocks were able to predict retirement behaviour. Overall, there was a shortage of literature regarding the influence that health rating had on retirement behaviour in the developing regions of the world where healthcare is less advanced.

Although the literature unanimously supports the finding that poor health influences retirement behaviour, as mentioned by Meng, Sundstrup and Andersen (2020), other lines of research explored whether retirement had a reverse impact on the health of individuals. Stolzenberg (2011) found evidence that, on average, retirement decreased health ratings for a sample of US supreme court justices. His findings were substantiated by an "elevated mortality hazard and a diminished [number of] years left alive" for the occupational group. However, reverse causality is an important challenge to consider when assessing the effect that retirement has on health since "not only [does] retirement affect health, but health also affects retirement" (Kapelyuk, 2017). Prior literature confirms that a lower health rating could be the result of, or the reason for, retirement. Since this paper is interested in the likelihood of being retired in old age, I expect that individuals who have an inferior subjective health rating will be more likely to be retired than those who have a superior rating.

Self-rated health has become a widely accepted and commonly used indicator of health in prior literature. By investigating whether there is a relationship between self-reported health and recorded biomarkers for three different European samples, Kananen et al. (2021) was able to conclude that self-reported health "has a solid biological basis and it is a valid but non-specific indicator of the biological condition of the human organism". In addition to being a valid biological indicator, Payne et al. (2017a) also found evidence that self-reported measures are invariant to gender and are not as jointly fixed to socioeconomic conditions as measured physical performance in South Africa. Separate accompanying research that also used the HAALSI cohort study found that "frailty was associated with worse subjective wellbeing, and worse self-reported health" (Payne et al., 2017b). Prior literature justifies the use of self-reported health for studies regarding retirement behaviour.

## 2.4 Psychological Factors

### 2.4.1 Cognition - Memory

Deciding on what cognitive psychological factors to include in research regarding retirement behaviour has led to discourse among academics. A leading paper by Thaler and Benartzi (2007) argued that inertia, loss aversion, and myopia were the main behavioural barriers when it came to preparing for retirement. A recent study by Tomar et al. (2021) used financial literacy as a cognitive characteristic to reveal "a positive association of future time perspective,

retirement goal clarity, and social group support with retirement planning behaviour". The results of the paper serve as a link between psychological characteristics and retirement behaviour. However, a weakness of their paper was that they only considered the retirement behaviour of women. Retirement goal clarity emerged as an important variable in most of the literature, but retirement was far less guaranteed in the developing regions than the developed regions. Understandably, setting retirement goals is exponentially harder in developing regions due to poor financial support structures and other exogenous variables such as crime and unemployment. Ultimately, the link between cognitive characteristics and the likelihood of retirement should be developed further to address the gap in the literature.

The deteriorating impact that ageing has on the cognitive ability of individuals is not widely used as a predictor for retirement behaviour in prior literature. However, from the limited research that did include cognitive functioning, there were notable results for both the professionally active and inactive elderly. A recent study by Sarabia-Cobo et al. (2020), found that elderly individuals who voluntarily opted to remain employed outperformed those who were retired in cognitive ability tests of attention, memory, and problem-solving. In addition to proving that there was a correlation between the amount of time spent retired and a lower cognitive performance rating, their analysis also revealed that mentally demanding jobs resulted in having a lower memory performance and did not affect attention and problem-solving. Moreover, a meta-analysis of early retirement by Topa, Depolo and Alcover (2017) highlighted research proving that "long-term exposure to low job complexity… has detrimental effects on cognitive functioning… and would have an impact on retirement decision making" (Oltmanns et al., 2017). Therefore, I expect that individuals who have a lower cognitive ability rating will be more likely to be retired than those who have a higher rating.

### 2.4.2 Village Estrangement (Belongingness)

An important element to consider regarding retirement decisions is the overlap between work and life. The authors De Preter et al. (2013), provided an event history analysis study, using data from the European Community Household Panel, which suggested that an interaction between work and life has an influence on the retirement timing of individuals. Furthermore, "when a sense of belonging and connection is recognized and nurtured, seniors wish to contribute to each other and the community as a whole" (Jakubec et al., 2019). Further research by Steffens et al. (2016)

examined whether belonging to a social group had an impact on the quality of life and health of retired individuals. Their results indicated that the "effect of social group membership on mortality was comparable to that of physical exercise". Much of the available literature deals with a sense of belonging to a company and not a community (or village), but Wise (2014) fills the gap by linking an increase in cognitive aging to a corresponding size reduction of a social network. My prior expectation is that individuals who have a feeling of estrangement in the village (do not feel that they belong), will be more likely to be retired than those who feel as though they are still members of their community in old age. Since belonging is shown to have a link to health and cognitive ability in prior literature.

### 2.4.3 Subjective Life Longevity

Considering ways in which to bring the retirement age more in line with increases in human life expectancy has laid a foundation for further research into individuals' perceptions of their actual retirement age. The research conducted by van Solinge and Henkens (2010) explored whether retirement planning varied based on a sample of employed Dutch individuals (between the ages of 50 and 60) expectations of how long they would live. They found that "subjective life expectancy is a factor that is taken into account in retirement decision making, at least as far as retirement intentions are concerned" (van Solinge and Henkens, 2010). Although those who believed they would live for longer expressed a desire in retiring later, their desires did not mirror reality as they did not end up retiring later. Furthermore, Yusof and Sabri (2017) mentioned that individuals who had a positive attitude towards planning for the future and a positive perspective of their life longevity would "have adequate financial support during retirement" by referring to prior literature. Therefore, previous literature advocates that planning for retirement, and financial adequacy during retirement are both not only influenced by an individual's attitude towards retirement, but also by their future time perspective.

Consequently, there is increasing support to include subjective life longevity in studies concerning retirement behaviour. Pang, Warshawsky and Weitzer (2008) study of the factors that influence retirement behaviour included self-reported longevity as an independent variable. Their results indicated that individual self-reported subjective probability of living to 75, or older, was significant (on a 99% confidence level) in every iteration of the models that predicted retirement. Correspondingly, this research will include subjective life expectancy to predict the probability of

retirement. However, since the study by Kim and Kim (2017) found that "subjective life expectancy is associated with self-rated health", the interaction of the two was assessed and evaluated using sub-hypotheses and statistical tests.

# 3. Hypotheses

Combining the various expectations formed from prior research resulted in the development of several main and sub hypotheses to answer the following central research question:

*To what extent do self-reported attitudes towards cognition, life longevity, belongingness, health, and HIV status influence the likelihood of retirement for old-aged individuals living in a rural area of South Africa?*

Essentially, this paper seeks to investigate which of the various demographic and self-reported variables are significant retirement predictors for old-aged individuals living in South Africa.

## 3.1 Formation of Main Hypotheses

The combination of various literature streams on retirement behaviour was used as a justification for the development of a predictive model that includes both demographic characteristics and self-reported attitudes as independent variables. I postulate that individuals, between the ages of 60 and 105, who report having superior rating of health, memory, and certainty of life longevity will be more likely to be either working or looking for work (as opposed to being retired) than those who have inferior ratings. In addition, individuals who have a feeling of estrangement in the village will be more likely to be retired than those who have a feeling of belonging. The postulate branches to form the following main hypotheses:

***Hypothesis 1****:* Individuals who have a superior general health rating will be less likely to be retired than those who have an inferior general health rating.

Prior literature confirms that, on the one hand, lower ratings of health are associated with an increase in labour force withdrawal, and on the other hand, retirement is linked to a further deterioration of health. Ultimately, the two streams are linked by labour force productivity. Individuals who perceive a lower health rating will be less motivated (or able) to work and more likely to retire. Then once retired, they would be less eager to look for work unless their perceived

health improves. Therefore, I expect those who have an inferior subjective health rating to have a higher likelihood of being retired than those who have a superior health rating.

***Hypothesis 2:*** Individuals who have a superior rating of their memory will be less likely to be retired than those who have an inferior memory rating.

To the same extent, prior literature provides evidence of a link between cognitive abilities and retirement decisions. The link shows that individuals who are retired scored lower on cognitive ability tests than individuals who were employed and that certain occupations were associated with cognitive deterioration. The literature failed to examine whether self-reported cognitive abilities (psychological constructs) influence retirement decisions. Since cognitive tests for memory are measurable, I anticipate that self-reported memory ratings will show similar results to actual cognitive ability tests as used in past literature. Moreover, I assume that those who have a lower memory rating, are more likely to be retired than those with a higher memory rating.

***Hypothesis 3:*** Individuals who are more certain that they will live to be 85, or older, will be less likely to be retired than individuals who are less certain.

In line with recent literature, I expect that individuals who maintain a myopic attitude and display a lack of concern, towards their future will be more likely to be retired earlier than those who are actively looking forward to and preparing for, their future. Consistent with the findings of van Solinge and Henkens (2010), a desire to retire at a later stage in life is often not actualized. Furthermore, if an individual who can retire values the present at the expense of the future, they would be less willing to work hard and save for an uncertain future. An assumption in prior literature is that such individuals would rather favour living a leisurely life in the present. Ultimately, I anticipate that deviations from a very probably expectation of living to be 85 or older, will result in a decrease in the probability of being retired.

***Hypothesis 4:*** Village estrangement will have a significant influence on the likelihood of being retired.

In consideration of the literature review, I formed the prior expectation that old-aged individuals who felt that they belonged to the village would be more likely to be employed or seeking employment as opposed to being retired. The assumption is based on findings that confirm the positive association between increased group support and belongingness. Furthermore,

alternative streams of prior literature verified that an increase in workforce participation was spurred on by a greater sense of belonging. Concisely, I assume that a feeling of estrangement (as opposed to belonging) will have a significant influence on the likelihood of being retired.

## 3.2 Formation of sub-hypotheses

Following the literature review, various sub-hypotheses were formed from the main hypotheses. Since "older adults have been neglected in HIV testing initiatives" (Rohr et al., 2017), the degree of HIV prevalence in the elderly cannot be definitively reported. Therefore, hypotheses 1 and 3 were extended in sub-hypotheses 1.1 and 3.1 to analyze the impact that self-reported HIV status has on the likelihood of retirement while considering self-reported health and perceived life longevity (respectively).

By the same token, prior literature illustrates that self-reported health is linked to both cognition and feeling of belonging (by either forward or reverse causality). Therefore, hypotheses 2 and 4 were extended in sub-hypotheses 2.1 and 4.1 to analyze the impact that self-reported health has on the likelihood of retirement while considering self-reported memory (cognition) and feeling of belonging (respectively). A succinct summary of the hypotheses is provided in table 1.

**Table 1: Summary of Hypotheses**

| | |
|---|---|
| **Main Hypotheses:** | |
| H1 | Individuals who have a **superior** general health rating will be **less** likely to be retired than those who have an **inferior** general health rating. |
| H2 | Individuals who have a **superior** rating of their memory will be **less** likely to be retired than those who have an **inferior** memory rating. |
| H3 | Individuals who are **more** certain that they will live to be 85, or older, will be **more** likely to be retired than individuals who are **less** certain. |
| H4 | Village estrangement will have a **significant** influence on the likelihood of being retired. |
| **Sub-Hypotheses:** | |
| H1.1 | For all individuals who have the same general health rating, those who self-report having a **positive** HIV status will be **more likely** to be retired (on average) than those who self-report having a **negative** HIV status for every category of health. |
| H2.1 | For all individuals who have the same memory rating, those who have a **lower** general health rating will be **more likely** to be retired (on average) than those who have a higher general health rating. |
| H3.1 | For all individuals who have the same confidence of living to be 85 or older, those who self-report having a **positive** HIV status will be **more likely** to be retired (on average) than those who self-report having a **negative** HIV status for every perceived life expectancy category. |

| | |
|---|---|
| H4.1 | For all individuals who have the same feeling of belonging in the village, those who have a **lower** general health rating will be **more likely** to be retired (on average) than those who have a **higher** general health rating. |

# 4. Research Method

## 4.1 Research Design

Empirical evidence from a survey conducted in South Africa was used to assess whether various demographic and self-reported factors are determinants (which significantly affect the likelihood) of retirement for old age individuals (between the ages of 60 and 105). Although multiple factors were analyzed, the main objective of this paper was to answer the research question. Given that the research question consisted of self-reported attitudes that potentially influence an individual's behaviour and are susceptible to change over time (Cherry, 2021), using panel data would have been optimal. However, due to inconsistencies in certain questions asked during different waves of the survey study, only the baseline responses were used to create the empirical model and draw inferences about the cross-sectional data. The following section outlines the research method to increase the reliability and validity of the results.

## 4.2 Data

The data used in this paper was obtained from the HAALSI study, which is a longitudinal study of a community in South Africa that is part of the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH). Historically, the dataset has been used predominantly for medical publications and has, to my knowledge, never been used to investigate the retirement behaviour of individuals in the community. However, the questions asked during the survey allow the dataset to be research extended beyond the study's main focal areas, such as retirement behaviour in developing regions.

### 4.2.1 Population

The HAALSI community-based cohort study produced a dataset containing survey responses from 5,059 men and women aged 40 years or older that live in the rural town of Agincourt, located in the Mpumalanga province of South Africa. The baseline survey interviews were conducted in the local language Tsonga (one of South Africa's eleven official languages) and were held from November 2014 to November 2015 with an 86% response rate. Even though data

from the second wave was available, only the baseline responses were used since the question regarding subjective life longevity was not included in the second wave of the survey.

Since there is no mandatory retirement age in South Africa, and since "it is not possible to claim the public pension before the normal [age of] eligibility" (OECD, 2017), constraints were placed on the dataset according to the qualifying age for a public pension, which is set at 60 for both men and women. In addition to removing respondents below the age of 60, those that have never worked or cannot work were also excluded to adhere to the definition of retirement. The constraints allowed for a model to be fit that calculates the predicted likelihood of retirement for individuals older than 60 who are legally allowed to receive a public pension.

### 4.2.2 Data Analysis

The data was analyzed in two stages. First, descriptive summary statistics were used to make inferences regarding the characteristic and attitude variables of the sample. Then, a multiple nonlinear probit regression model was used to estimate the probability of being retired based on such factors. Two nested models were fitted, one only containing demographic and household characteristic covariates, and the other extended to also contain self-reported explanatory variables. The nested models were compared to assess whether including self-reported variables increased the predictive ability for the likelihood of being retired. In addition, average marginal effect estimates of the probit model that included demographic and household characteristic, and self-reported independent variables were used to statistically test the hypotheses and provide meaningful conclusions. Lastly, numerous model tests were conducted (Appendix C) to critically evaluate the validity of the findings.

## 4.3 Definition and Measurement of Variables

For the most part, data acquired from the HAALSI study's survey questions followed a coherent structure with easily understandable measurements. In addition to the summary statistics provided in the baseline codebook, the dataset included clear variable labels that further contributed to understanding the data. Upon deeper inspection of the dataset, I found that there were missing and often unrealistic responses to the question asking how old the participant was. Since the goal of this paper was premised on the age of individuals, corrections were made according to their age at the time of the sample selection in 2014 by using data from a recent census and accounting for the year in which the respondent participated in the survey. In addition to

reordering certain binary response variables to ease understanding of what the model considers a success or failure, a subjective life expectancy variable was created by recoding continuous data responses to form a five-point Likert scale which will be discussed below.

### 4.3.1 Outcome Variable

The dependent variable (Retired) is a binary discrete variable. The variable was constructed by combining the responses from a subset of the population (aged 60 or older) to the following three questions:

1. *"What would be the best description of your current work status? (Unable to work (disabled), unemployed, homemaker, working full time (35 or more hours per week), working part-time (less than 35 hours per week), retired, sick or other leave, other)"*
2. *"Are you currently looking for work? (Yes or no)"*
3. *"Did you ever work? (Yes or no)"*

The last question was added to adhere to the definition of retirement since some respondents reported that they were retired, despite having never worked. Therefore, the composite measure of retirement classifies individuals aged 60 or older that are able to work (i.e., not on sick leave or disabled) by the following:

$$Y = \begin{cases} 0\ (Not\ retired) - is\ currently\ working\ or\ looking\ for\ work \\ 1\ (Retired) - is\ not\ currently\ working\ or\ looking\ for\ work\ and\ has\ worked\ in\ the\ past \end{cases}$$

Accordingly, for individuals who are entitled to receive old-age pension (older than 60), the dependent variable is a direct measure of retirement that differentiates those who are retired from those who are either employed or seeking employment. As a result, the dependent variable is also a proxy measurement of retirement confidence as an assumption would be that those who are happily retired will be less likely to be employed or seeking employment in their old age.

### 4.3.2 Primary Explanatory Variables

In line with the research question of this paper, the primary explanatory variables were self-reported measurements of general health, memory (used as a construct measurement of self-perceived cognitive ability), subjective life expectancy, HIV status, and belongingness. Self-reported HIV status and belongingness were coded as binary response variables (either yes or no) and the self-reported ratings of general health, memory, and subjective life expectancy were coded

as five-point categorical Likert scale variables. Contrary to the method implemented by Fernández-López et al. (2010), using categorical dummy variables, in this paper, was favoured over creating subsamples since all respondents were from the same area and comparisons did not have to be made between areas.

Self-reported ratings of general health, memory (cognition), and subjective life expectancy were used as psychological construct measurements for the attitudes of individuals and were not as simple to measure as demographic variables (Price, Jhangiani and Chiang, 2015). Therefore, five-point Likert scales were used to "allow the individual to express how much they agreed or disagreed with a particular statement" (Mcleod, 2019). Self-reported ratings of general health and memory were already developed as five-point categorical Likert scales in the dataset and did not need to be manipulated.

However, the subjective life expectancy variable was created by transforming continuous data answers to the question "What is the percent chance that you will live to be 85… or more?" into a categorical scale out of five (from not probable to very probable). Corrections were made for those that did not respond, either due to misunderstanding the question or because they were older than 85 and were thereby extremely certain that they would live to be 85 (or older). I assume that such individuals did not answer as they felt that a response was not needed. Hence the variable was corrected by way of imputation if the respondents were older than 85 and did not answer the question. The missing responses serve as both a limitation of the survey and as a foundation for future improvement of the clarity of the question.

### 4.3.3 Covariates

As shown in prior literature, the decision to retire depends on multiple factors. Thus, age, age squared (to test for a nonlinear relationship between age and retirement), sex, marital status, education, and wealth asset index were included as covariates in the model to examine their influence on the likelihood of retirement. Table 2 provides a detailed description of all the independent variables used in the model.

**Table 2: Descriptions of the Independent Variables**

| Variable | Description | Exp. | Nature of the variable |
|---|---|---|---|
| Age | Age in complete years | + | Years |
| Age Squared | Age squared | - | |
| Sex | Gender | Ref | 0 = Male |
| | | + | 1 = Female |
| Marriage status | "What is your current marital status?" | Ref | 0 = Currently married |
| | | + | 1 = Not married |
| Education Category | Categorization of answers to: "What is the highest level of education you have attained?" | Ref | 1 = No formal education |
| | | + | 2 = Some primary |
| | | + | 3 = Some secondary |
| | | + | 4 = Secondary or more |
| Wealth asset index | Constructed from answers to asset ownership questions. Quintiles illustrate household cumulative living standards. | Ref | 1 = Quintile 1 (Lowest) |
| | | + | 2 = Quintile 2 |
| | | + | 3 = Quintile 3 |
| | | + | 4 = Quintile 4 |
| | | + | 5 = Quintile 5 (Highest) |
| HIV status | "Have you ever tested positive for HIV?" | Ref | 0 = No |
| | | + | 1 = Yes |
| Current general health rating | "In general, how would you rate your health today?" | Ref | 1 = Very good |
| | | + | 2 = Good |
| | | + | 3 = Moderate |
| | | + | 4 = Bad |
| | | + | 5 = Very bad |
| Cognition (Memory rating) | "How would you rate your memory at the present time? Would you say it is excellent, very good, good, fair or poor?" | Ref | 1 = Excellent |
| | | + | 2 = Very good |
| | | + | 3 = Good |
| | | + | 4 = Fair |
| | | + | 5 = Poor |
| Subjective life expectancy | Categorization of answers to: "What is the percent chance that you will live to be 85… or more?" | Ref | 0 = Very probable |
| | | - | 1 = Somewhat probable |
| | | - | 2 = Neutral |
| | | - | 3 = Somewhat improbable |
| | | - | 4 = Not probable |
| Feeling of belonging | "Do you feel as if you are really part of (a member of) this village?" | + | 0 = No |
| | | Ref | 1 = Yes |

*Notes:* Expectation (Exp.) shows a prior expectation, following a review of prior literature, of the influence (positive or negative) that a variable will have on the probability of being retired. For the categorical variables, the expectation is based on departures from the reference category.

## 4.4 Probit Model Specification

Owing to the fundamental purpose of analyzing individuals' retirement decisions, a binary choice model was used to test the hypotheses. Since the dependent variable represents a Bernoulli random variable where:

$$Y = \begin{cases} 1\ (retired)\ with\ probability\ p, \\ 0\ (not\ retired) with\ probability\ 1-p. \end{cases}$$

Considering that the linear probability model fails to account for the discreteness of the binary dependent variable by not constraining "the predicted probabilities to be between zero and one" (Cameron and Trivedi, 2005), a nonlinear multiple probit model was chosen. Conditional likelihood models are consistently used in relevant literature around retirement behaviour. A probit model, as named by Chester Bliss in 1934, was used to establish a link between the binary dependent variable, which could only be either zero or one, and a set of independent variables. The probit model was chosen over the logit model since it was able to solve a limitation of the logit model by accounting for random taste variation (Train, 2012). Furthermore, statistical comparisons of the two models revealed support for specifying the probit model, as shown by table 9 in Appendix C.

Following Cameron and Trivedi (2005) the probit model specifies the conditional probability given by:

$$P \equiv \Pr[y = 1|x] = F(x'\beta)$$

Where $F(\cdot)$ is the standard normal cumulative distribution function, x is a regressor vector, and $\beta$ is a Kx1 parameter vector. More specifically, the probability is calculated according to the following equation:

$$p = \phi(x'\beta) = P(Z \leq x'\beta) = \int_{-\infty}^{x'\beta} \phi(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x'\beta} \exp\left\{-\frac{v^2}{2}\right\} dv$$

### 4.4.1 Model Estimation

Since the probit model deals with the standard normal distribution, interpreting beta coefficients of the model are not as straightforward as in an OLS model. The beta coefficients illustrate changes in the z-score along the standard normal distribution as opposed to direct changes of the dependent variable in response to the independent variables. Therefore, marginal effects will

be estimated to interpret the likelihood of retirement. Again, as mentioned by Cameron and Trivedi (2005), the marginal effects of the probit model were calculated using the following equation:

$$\frac{\partial \phi(x_i{}'\beta)}{\partial x_{ij}} = \phi(x_i{}'\beta)\beta_j$$

Therefore, marginal effect estimates show how the likelihood of being retired (Y = 1) is influenced by discrete changes (for example, from category 1 to category 2) of the independent variables. Figures 4, 5 and 6 in Appendix A provide visualizations for both the average marginal effects and the predictive margins of the primary explanatory variables to increase understanding of the results. Furthermore, model tests were provided in Appendix C and statistical tests for the overall significance of the main variables were provided in Appendix B.

# 5. Findings

## 5.1 Descriptive Results

Table 3 provides descriptive statistics of the sample before and after incorporating the dependent variable. Comparing the two nested samples showed that roughly 38% of the respondents (1,926 out of 5,059) were between the ages of 60 and 105 and could be classified as being either retired or not retired (according to the definition of retirement). As expected, the average age of the participants increased from 62 to 71 years. Furthermore, the average frequencies of individuals distributed by marital status, indicator of wealth, and feeling of belonging remained somewhat consistent after the dependent variable was added. Also, self-reported ratings of health, cognition, and life longevity did not vary significantly. However, compared to the full sample, more elderly: were males (possibly due to the low labour force participation rates of women), had less education (possibly due to the trickle-down effects of apartheid), and reported being HIV negative.

The final sample used in this paper (constrained sample) showed that 1,926 individuals could be classified according to the definition of retirement. The sample consisted of mostly married (54.39%) men (63.55%) that were divided between five various wealth asset index quintiles. A large majority of respondents (55.42%) reported having attained no formal education, felt that they belonged to the community (93.96%) and that they were not HIV positive (91.6%). Furthermore, on a five-point Likert category scale ranging from lowest to highest (1 being lowest

and 5 being the highest), the majority self-reported as being in category 4 (45.19%) for general health rating, 3 (47.75%) for memory rating, and 5 (57.38%) for life longevity. The average values provide a foundation from which individual deviations in attitudes can be compared.

**Table 3: Descriptive Summary Statistics of Nested Probit Models**

| | **Base Sample** *(Age: 40-105)* | | | **Constrained Sample** *(Age: 60-105)* | | |
|---|---|---|---|---|---|---|
| | N | Freq. % | Mean (Std. Dev) | N | Freq. % | Mean (Std. Dev) |
| **Dependent Variable:** | | | | | | |
| Retired | | | | 1,926 | | 0.910 |
| Yes (1) | | | | 1,753 | 91.02 | |
| No (0) | | | | 173 | 8.98 | |
| **Independent Variables:** | | | | | | |
| Age | 5,059 | | 62.471 (13.024) | 1,926 | | 71.377 (08.535) |
| Age Squared | 5,059 | | 4072.237 (1693.27) | 1,926 | | 5167.469 (1274.48) |
| Sex | 5,059 | | 0.536 | 1,926 | | 0.364 |
| Male (0) | | 46.35 | | | 63.55 | |
| Female (1) | | 53.65 | | | 36.45 | |
| Marriage status | 5,055 | | 0.491 | 1,925 | | 0.456 |
| Currently married (0) | | 50.94 | | | 54.39 | |
| Not married (1) | | 49.06 | | | 45.61 | |
| Education category | 5,042 | | 1.833 | 1,920 | | 1.557 |
| No formal education (1) | | 45.74 | | | 55.42 | |
| Some primary (2) | | 34.03 | | | 35.78 | |
| Some secondary (3) | | 11.38 | | | 6.51 | |
| Secondary or more (4) | | 8.85 | | | 2.29 | |
| Wealth asset index (Individuals categorized into 5 quintile (1-5)) | 5,059 | | 2.989 | 1,926 | | 2.998 |
| HIV status | 5,025 | | 0.124 | 1,917 | | 0.084 |
| Positive (1) | | 12.40 | | | 8.40 | |
| Negative (0) | | 87.60 | | | 91.60 | |
| Current general health rating | 5,056 | | 2.328 | 1,925 | | 2.490 |
| Very good (1) | | 20.19 | | | 15.95 | |
| Good (2) | | 47.88 | | | 45.19 | |
| Moderate (3) | | 12.86 | | | 15.12 | |
| Bad (4) | | 17.05 | | | 21.35 | |
| Very bad (5) | | 2.02 | | | 2.39 | |
| Cognition (Memory rating) | 4,939 | | 2.760 | 1,887 | | 2.897 |
| Excellent (1) | | 9.44 | | | 6.47 | |
| Very good (2) | | 26.40 | | | 23.37 | |
| Good (3) | | 45.41 | | | 47.75 | |

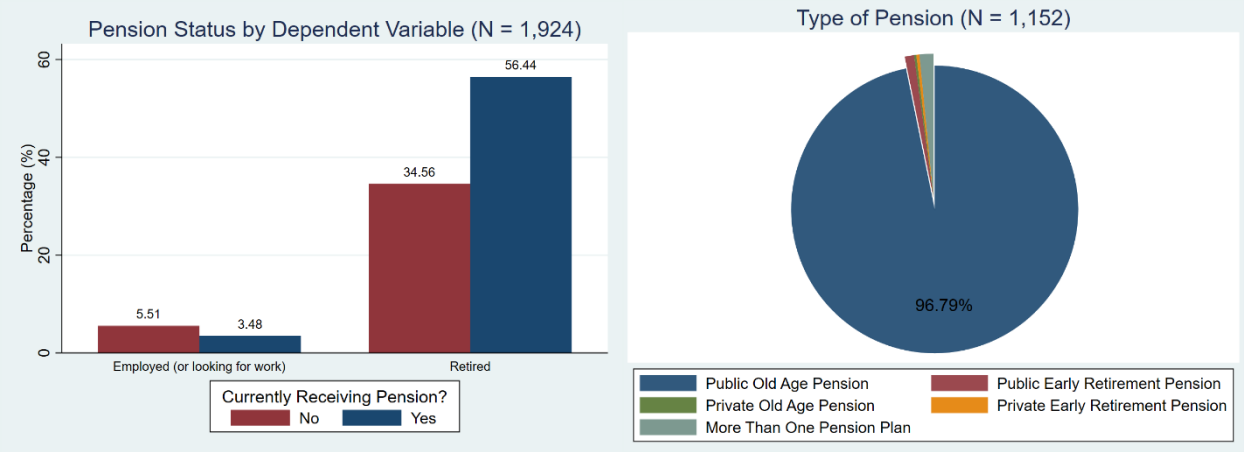
| | | | | | |
|---|---|---|---|---|---|
| Fair (4) | | 16.26 | | 18.87 | |
| Poor (5) | | 2.49 | | 3.55 | |
| Subjective life expectancy | 4,042 | | 0.777 | 1,497 | 0.796 |
| Very probable (0) | | 57.50 | | 57.38 | |
| Somewhat probable (1) | | 19.74 | | 18.50 | |
| Neutral (2) | | 13.38 | | 14.30 | |
| Somewhat improbable (3) | | 6.28 | | 6.75 | |
| Not probable (4) | | 3.09 | | 3.07 | |
| Feeling of belonging | 4,938 | | 0.927 | 1,886 | 0.940 |
| Yes (1) | | 92.67 | | 93.96 | |
| No (0) | | 7.33 | | 6.04 | |

*Notes*: The codes of the variables were included in parentheses to make sense of the mean values.

The bar graph in figure 1 illustrates the percentages of respondents either currently receiving or not receiving a retirement pension based on their employment status. Moreover, of those that are currently receiving a retirement pension, the pie chart in the same figure shows what type of pension they are receiving. The figure shows that only 59.92% of the sample (either retired or not retired) were receiving a pension and 40.07% were not receiving a pension (2 individuals (0.01%) did not respond). The fact that 34.56% of retired individuals were not receiving a pension is cause for concern and shows a lack of preparation for retirement. A possible reason that the respondents were not receiving pension during retirement is that the value of their income and/or assets were too high to qualify for the public old-age pension. This is substantiated by the heavy reliance (96.79%) on the public old-age pension fund by those that do receive a pension. The minuscule portion of individuals that have a private retirement pension plan further shows a lack of preparation for retirement due to education or behavioural norms.

The difference between the number of retirees receiving a pension and those not receiving a pension (as seen in Figure 1) indicates that there is an increased reliance on informal savings or on family members to finance retirement. Such reliance could be a viable way to guarantee retirement confidence; however, the poor saving culture of South Africans (Gore, 2018) suggests that people are not only forced to work for longer (Sousa-Ribeiro et al., 2021) but also that retirees may be forced to re-enter the workforce Zeka (2016) as they live for longer and eventually run out of money.

**Figure 1: Division of Respondents by Status and Source of Retirement Pension**




## 5.2 Multiple Nonlinear Regression Results

Table 4 presents the results for the multiple nonlinear regression analysis using two nested probit models. The two models differed in that probit model 1 did not include self-reported health, cognition, life longevity, and estrangement as independent variables. Therefore, probit model 1 served as a base "characteristic" model where HIV status was included due to its high demographic prevalence. Model 1 had 1,911 observations compared to 1,461 observations in model 2. The reduction in observations was due to missing data responses to the subjective life longevity question. Since both McFadden's $R^2$ and the AIC criterion suggested that model 2 was a better fit, including self-reported ratings of health, cognition, life longevity, and belongingness were seen to have an influence on the likelihood of being retired. The statistical tests were elaborated further in Appendix C. Overall, including the self-reported rating variables did not drastically change the significance of the characteristic variables compared to probit model 1. Other than a change in the significance of having some primary education (compared to no formal education) and being in the highest wealth asset index (compared to the lowest).

Since the beta coefficients in table 4 represent changes in the z-scores, interpretation of the results became troublesome. Therefore, the average marginal effects provided in table 5 were used to investigate the hypotheses.

**Table 4: Probit Regression Estimates of the Likelihood of Retirement**

| | Probit Model (1) | | Probit Model (2) | |
|---|---|---|---|---|
| Variables | Beta Coef. | Std. Err. | Beta Coef. | Std. Err. |
| **Demographic Characteristics** | | | | |
| Age | 0.378*** | 0.089 | 0.371*** | 0.097 |
| Age Squared | -0.002*** | 0.001 | -0.002*** | 0.001 |
| Female (vs Male) | 0.054 | 0.106 | 0.031 | 0.123 |
| Not Married (vs Currently Married) | 0.221** | 0.104 | 0.269** | 0.119 |
| Education Category (vs No Formal Education) | | | | |
| Some primary (1-7 years) | 0.181* | 0.100 | 0.255** | 0.115 |
| Some secondary (8-11 years) | -0.228 | 0.162 | -0.159 | 0.180 |
| Secondary or more (12+ years) | 0.117 | 0.295 | 0.125 | 0.309 |
| **Household Characteristics** | | | | |
| Wealth Asset Index (vs Quintile 1: Lowest) | | | | |
| Wealth quintile 2 | -0.014 | 0.135 | 0.025 | 0.155 |
| Wealth quintile 3 | 0.027 | 0.136 | 0.082 | 0.156 |
| Wealth quintile 4 | 0.371** | 0.148 | 0.400** | 0.167 |
| Wealth quintile 5 (Highest) | 0.220 | 0.151 | 0.339* | 0.175 |
| **Self-Reported Health and HIV Status** | | | | |
| HIV Positive (vs HIV Negative) | -0.051 | 0.144 | -0.038 | 0.172 |
| Current General Health Rating (vs Very Good) | | | | |
| Good | | | 0.202 | 0.142 |
| Moderate | | | 0.237 | 0.186 |
| Bad | | | 0.410** | 0.192 |
| Very Bad | | | - | - |
| **Self-Reported Psychological Attitudes** | | | | |
| Current Memory Rating (vs Excellent) | | | | |
| Very good | | | 0.416** | 0.179 |
| Good | | | 0.533*** | 0.188 |
| Fair | | | 0.587*** | 0.226 |
| Poor | | | 0.395 | 0.348 |
| Certainty of Living to 85 (vs Very Probable) | | | | |
| Somewhat probable | | | -0.273** | 0.135 |
| Neutral | | | -0.155 | 0.150 |
| Somewhat improbable | | | -0.422** | 0.201 |
| Not probable | | | -0.516* | 0.280 |
| Village Estrangement (vs Belongingness) | | | 0.052 | 0.213 |
| Constant | -14.588*** | 3.210 | -14.758*** | 3.521 |
| Observations | 1,911 | | 1,461 | |
| McFadden's R2 | 0.136 | | 0.155 | |
| AIC*n | 1035.238 | | 861.087 | |

*Notes*: There are differences in the sampling sizes between the nested models due to missing values in some of the subjective variables. Since finding a false negative (someone who is not retired but is reported as being retired) would be more misleading than a false positive, the Akaike Information Criterion (AIC) was used to test the goodness of fit (alongside McFadden's R2). Due to the differences in sampling sizes, the AIC*n values were calculated and the smaller value for model 2 indicates a better fit. Model 2 will serve as the main model henceforth. Further extensive model tests for goodness-of-fit, specification, predictive power, and multicollinearity are provided in appendix C. Statistical significance is shown as follows:
*** p<0.01, ** p<0.05, * p<0.1

Congruent with prior literature, table 5 shows that age is not only significantly associated with planning for retirement (as shown by Richardson and Kilty (1989) and Lakshmi-Malroutu and Xiao (1995)) but it is also significantly associated with the likelihood of being retired (as shown by Pang, Warshawsky and Weitzer (2008) and Singleton and Keddy (1991)). Furthermore, the significance of the positive age and a negative age squared coefficients suggests that an inverse U-shape relationship exists between age and retirement which is comparable to the U-shape association between age and saving for retirement as shown by Fernández-López et al. (2010). Additionally, using the test proposed by Lind and Mehlum (2009), I found strong evidence to reject the existence of a monotone or U-shape relationship between age and retirement in favour of an inverse U-shape relationship at a 0.05 level of significance. The test revealed that 86 years was an extreme point. Therefore, leading up to the age of 86, the probability of being retired increases, however, for those over the age of 86, the probability of being retired decreases. To the same extent, the probability of being either employed or looking for work increased when individuals were over the age of 85. The need to re-enter the workforce is a sign that old age poverty is a persistent concern as individuals are not comfortable with their quality of life or their financial adequacy during retirement as the average life expectancy increases. Figure 2 provides a visualization of the inverse U-shape based on age and the various fitted/predicted values. The graph shows that the probability of being retired starts to decrease as age increases past roughly 84 years.

**Figure 2: Probability of Being Retired by Age**

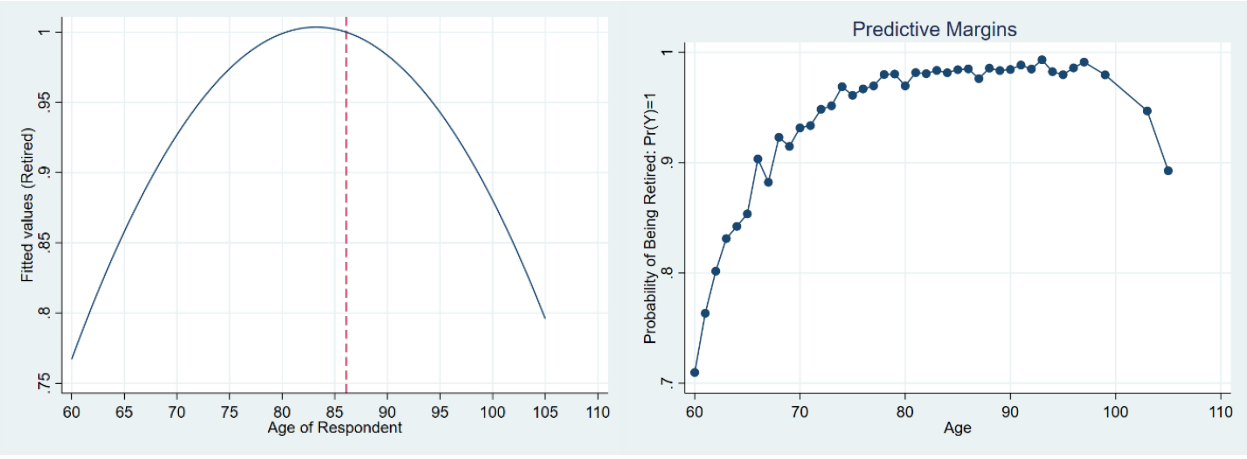



Table 5 indicates that being not married (compared to being married), having some primary education (compared to no formal education), and belonging to the 4$^{th}$ wealth asset quintile

(compared to the lowest quintile 1) were statistically significant predictors of retirement with a 95% level on confidence. On the contrary, being female (compared to male) and HIV positive (compared to negative) were not statistically significant predictors of retirement.

**More specifically, when considering demographic and household characteristic covariates as well as self-reported HIV status, the results of table 5 indicate that:**

***Marital Status:*** Individuals who are not married were 3.8% more likely to be retired (on average) than those who were married. Although the empirical results are consistent with research into retirement saving (Willows, 2019), it differs from the findings of (Ucello, 1998) as marital status was a significant predictor of retirement.

***Education:*** Individuals who have attained some primary education (1-7 years) were not only the most likely to be retired (out of all education categories) but were also 3.6% more likely to be retired (on average) than individuals who had no formal education. The empirical findings differ from those obtained by McAllister et al. (2020).

***Wealth:*** Individuals who belong to the second-highest quintile (4/5) of the wealth asset index had the highest likelihood of being retired and were 5.7% more likely to be retired than those in the lowest quintile. Therefore, an increase in wealth is not directly associated with a higher likelihood of being retired. The findings differ from the results of Pang, Warshawsky and Weitzer (2008) and Bloemen (2009). This suggests that further research into the link between wealth and retirement needs to be done in developing countries.

***HIV status:*** The likelihood of being retired was not significantly influenced by the self-reported HIV status. Further research into the influence that HIV status has on the likelihood of retirement needs to be done to make comparisons.

Tests for the overall statistical significance of the categorical independent variables were reported in table 6 of appendix B. The main explanatory variables for self-reported cognition and subjective life longevity were significant with an alpha value of 0.1. However, the self-reported health rating was not significant since the ‘very bad’ category was omitted from the model as it was a perfect predictor of retirement. Therefore, there were only 3 degrees of freedom for the five-

point category scale. Fortunately, the academic world is placing less reliance on statistical significance when it comes to evaluating results (Wasserstein, Schirm and Lazar, 2019).

**Moreover, considering the main hypotheses, table 5 and Appendix A show that:**

***Hypothesis 1:*** The results reveal support for hypothesis 1 since individuals who have inferior health ratings are more likely to be retired than those who have superior ones. The relationship between discrete changes in the categories is proportional to increases in the likelihood of retirement. Furthermore, compared to those who have a very good health rating, those who have a bad rating were statistically significant at a significance with a 95% level of confidence. Those who have a bad health rating were 6% more likely to be retired than those who have a very good health rating. The results are consistent with those obtained by Davey (2008), Sousa-Ribeiro et al. (2021), and Disney, Emmerson and Wakefield (2003).

***Hypothesis 2:*** The results did not find conclusive support for hypothesis 2 since the probability of being retired does not proportionally increase over each category of reported memory. This is shown by the inverse U-shape in the top right graph of figure 6 in Appendix A. Although, having a poor memory rating (compared to excellent) was not statistically significant, the predictions did not show convincing support for the relationship. Those who have a fair rating of memory were 10.3% more likely to be retired than those who have an excellent rating. The findings show that individuals who report having the two extreme cases of memory rating (excellent and poor) were the most likely to be employed or seeking employment. The findings are somewhat supported by Oltmanns et al. (2017) since individuals who have low cognitive ratings would be more likely to have been employed in jobs with low complexity and therefore would not be able to retire. Furthermore, the results show support for the findings of Sarabia-Cobo et al. (2020) up to the second last cognitive measure (fair).

***Hypothesis 3:*** The results from table 5 fail to provide conclusive support for hypothesis 3 since the marginal effects of subjective life expectancy on retirement decreased less for those that had a neutral view (compared to a very probable view) towards living to 85 than for any other category. Figure 4 in Appendix A shows the deviation in the

probability of being retired over the categories. Individuals who had a very probably (superior) certainty of living to be 85 or older were the most likely to be retired on average. For the most part, the findings agree with the paper by van Solinge and Henkens (2010), however, the insignificant discrete departures of the categorical variables mean that we cannot convincingly accept the hypothesis. Furthermore, as expected, the subjective life expectancy was invariant with an increase in age, as shown by the convergence to complete retirement for each category as age increases in figure 5 of Appendix A. Therefore, the question was not as applicable for the elderly as it was for the average of the sample.

***Hypothesis 4:*** Compared to a feeling of belonging, estrangement was not a significant predictor of retirement and therefore we fail to accept hypothesis 4. This goes against the ideas of prior literature by (Jakubec et al., 2019).

**Table 5: Average Marginal Effect Estimates of the Likelihood of Retirement: Probit Model**

| Variables | Probit Model (2) Delta-method M.E. (dy/dx) | Std. Err. |
|---|---|---|
| **Demographic Characteristics** | | |
| Age | 0.055*** | 0.014 |
| Age Squared | -0.000*** | 0.000 |
| Female (vs Male) | 0.005 | 0.018 |
| Not Married (vs Currently Married) | 0.038** | 0.017 |
| Education Category (vs No Formal Education) | | |
| Some primary (1-7 years) | 0.036** | 0.016 |
| Some secondary (8-11 years) | -0.028 | 0.033 |
| Secondary or more (12+ years) | 0.019 | 0.044 |
| **Household Characteristics** | | |
| Wealth Asset Index (vs Quintile 1: Lowest) | | |
| Wealth quintile 2 | 0.004 | 0.027 |
| Wealth quintile 3 | 0.014 | 0.026 |
| Wealth quintile 4 | 0.057** | 0.024 |
| Wealth quintile 5 (Highest) | 0.050* | 0.026 |
| **Self-Reported Health and HIV Status** | | |
| HIV Positive (vs HIV Negative) | -0.006 | 0.026 |
| Current General Health Rating (vs Very Good) | | |
| Good | 0.033 | 0.024 |
| Moderate | 0.038 | 0.030 |
| Bad | 0.060** | 0.028 |
| Very Bad | - | - |
| **Self-Reported Psychological Attitudes** | | |
| Current Memory Rating (vs Excellent) | | |
| Very good | 0.079** | 0.038 |
| Good | 0.096** | 0.040 |
| Fair | 0.103** | 0.043 |
| Poor | 0.076 | 0.062 |
| Certainty of Living to 85 (vs Very Probable) | | |
| Somewhat probable | -0.041* | 0.021 |
| Neutral | -0.022 | 0.022 |
| Somewhat improbable | -0.068* | 0.038 |
| Not probable | -0.088 | 0.058 |
| Village Estrangement (vs Belongingness) | 0.007 | 0.030 |
| Observations | 1,461 | |

*Notes*: The average marginal effects were estimated using a method of maximum likelihood. The categorical variables show the change in probability of being retired according to discrete deviations from the base levels. Visualizations of the main average marginal effects for each main variable are provided in figure 6 of Appendix A.
*** p<0.01, ** p<0.05, * p<0.1

Overall statistical significance tests were done on the interaction effects between variables in table 7 of Appendix B. Although HIV status was an insignificant predictor, the probabilities of being retired still differed according to self-reported status, and therefore the findings of the sub-hypotheses (using figure 3) were as follows:

***Hypothesis 1.1:*** The results fail to find support for hypothesis 1.1. Those who self-reported having a negative HIV status were more likely to be retired than those who self-reported having a positive HIV status for every category of health. The proportional differences in the probabilities of being retired based on HIV status are because there was not a significant interaction between self-reported HIV status and health (as shown in table 7 of Appendix A). Those who self-reported having a bad health rating and a negative HIV status were the most likely to be retired as seen in figure 3. This result is against the findings of Lorem et al. (2020).

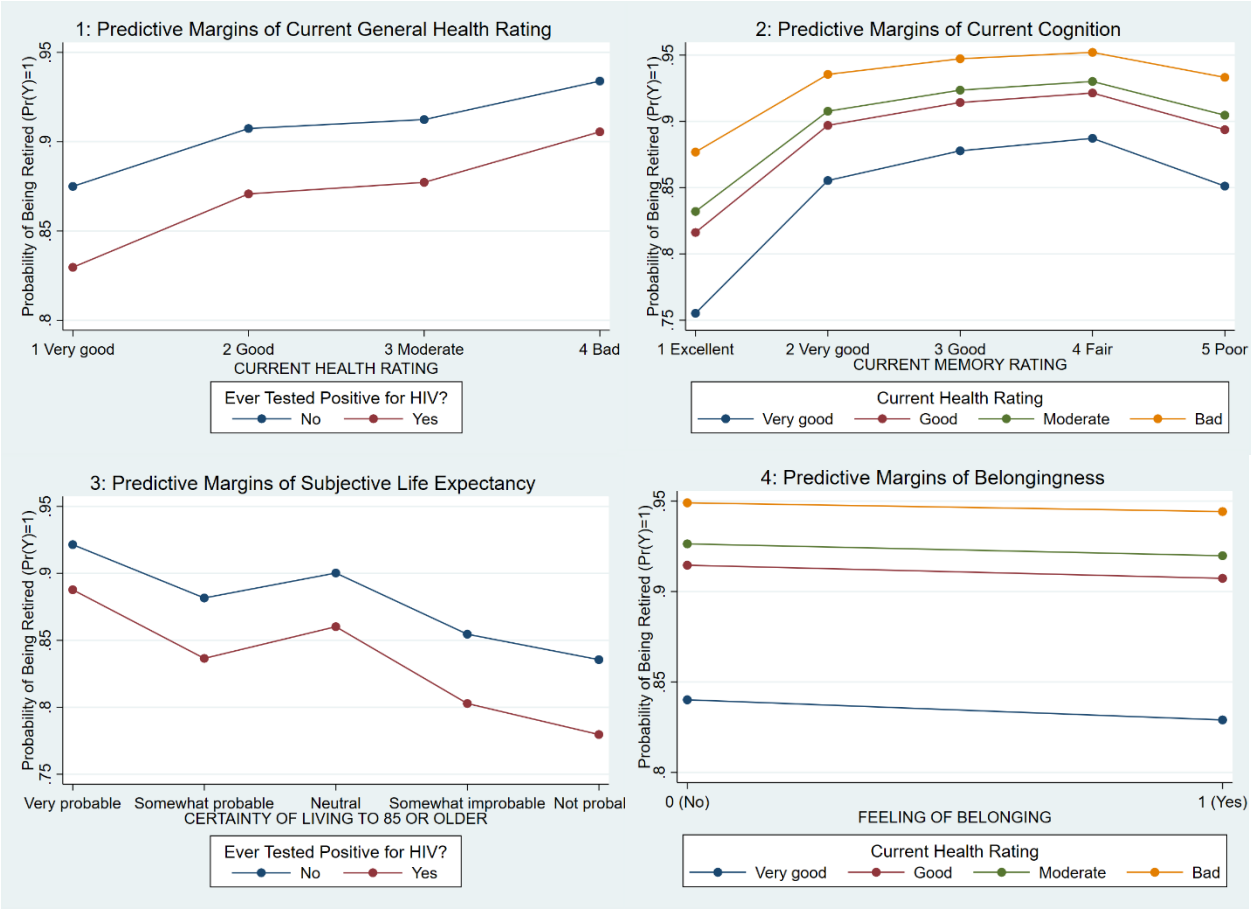
***Hypothesis 2.1:*** The results provide support for hypothesis 2.1. Those who have a bad general health rating were far more likely to be retired than those who had a very good general health rating for every level of current memory rating. The findings are consistent with prior literature that links cognitive deterioration to poor self-reported health.

***Hypothesis 3.1:*** The results fail to support hypothesis 3.1. Those who reported having a negative HIV status were marginally more likely to be retired than those who self-reported being HIV positive. The vertical comparisons are consistent for every category of life certainty. However, when the interaction term between HIV positive and subjective life expectancy was added to the model, the main variable of subjective life expectancy became more significant. This lays a foundation for further research into whether reality differs from self-reported HIV status as suggested by Rohr et al. (2017).

***Hypothesis 4.1:*** The results provide support for hypothesis 4.1. Those with a ‘bad’ general health rating were more than 10% more likely to be retired than those who had a ‘very good’ health rating. The difference in extreme echelons provided support for the hypothesis. However, there was not a significant difference between a feeling of

belonging and a feeling of estrangement (shown by the minor probability differences in figure 3).

**Figure 3: Predictive Margins of Main Explanatory Variables by HIV Status and Health**



# 6. Conclusion

This paper investigated whether various demographic and self-reported factors had a significant influence on the likelihood of being retired for old-aged individuals living in a rural area of South Africa. Although multiple factors were considered, the main purpose of this paper was to assess whether self-reported attitudes towards cognition, life longevity, belongingness, health, and HIV status had an influence on the likelihood of retirement. A sub-sample of 1,926 individuals between the ages of 60 and 105 who were able to work, was created using baseline survey responses from the 'Health and Aging in Africa: A longitudinal Study of an INDEPTH Community' (HAALSI) study conducted in Agincourt, South Africa and collected from November 2014 to November 2015. A probit model which dichotomously classified individuals of the sample as being either retired or not retired (currently working or seeking employment) revealed that self-reported attitudes influenced the likelihood of being retired to a great extent.

This paper found that there is an inverse U-shape relationship between age and likelihood of retirement which suggests that more people are being forced to re-enter the workforce during retirement years. Of the statistically significant factors, empirical results revealed that on an aggregate five-point Likert scale, (where 1 is the lowest/worst and 5 is the highest/best rating) unmarried individuals who have attained some primary education (1-7 years), belong to the second-highest quintile on the wealth asset index (4), have a bad rating of health (2), a fair rating of their cognitive ability to remember (2), and a very probable subjective certainty of living to be 85 or older (5) were the most likely to be retired on average. Furthermore, gender, HIV status, and feelings of belongingness had no significant influence on the likelihood of being retired. Various main and sub-hypotheses tested for a directly proportional relationship between discrete changes from the base category of the main explanatory variables against the likelihood of being retired and only revealed support for self-reported health.

A benefit of this paper is that it was able to meaningfully contribute to the field of research. Hitherto, there has been a dearth of research regarding the factors that influence the likelihood of retirement. The field of research regarding retirement behaviour has been dominated by research into the pre-retirement behaviour of individuals. On the other hand, there has been a lackadaisical approach towards investigating whether perceptions mirror reality for retirement. The ubiquitous gap in research between preparing for retirement and retiring is emphasized when comparing

developed and developing regions. Therefore, this paper used factors that were proven to influence preparing for retirement as well as factors that were proven to influence the likelihood of retirement as a justification. This research is of paramount importance as it shows the current implications that the increasing trends in human longevity have on the elderly. Furthermore, it studies the implications in societies where there is far less societal support for the elderly. A notable strength of this paper is that it considered the average marginal effects of various factors which makes the sample comparable to more developed regions of the world. A strength of this research is that pre-retirement behaviour is largely synonymous with the likelihood of being retired in old age.

This paper was not without limitations. A major limitation was the missing responses to the question regarding subjective life longevity. Since the probit model uses maximum likelihood estimation, I favoured having more observations over having uncorrected response variables. An additional limitation was that the subjective life longevity question was omitted from the second wave of the survey (possibly due to a lack of question coherence). As a result, a cross-sectional analysis using only the baseline responses was chosen over a time series analysis. This was a major limitation since individual's attitudes generally change over time. Therefore, this study was more of a snapshot of the attitudes rather than a broader analysis.

The main recommendation of this paper is that the pension system in South Africa needs to be re-evaluated to ensure adequate financial support during retirement which in turn will mean that elderly individuals are not forced to re-enter the workforce. In addition, since there is a link between preparing for retirement and the likelihood of being retired, increasing access to education for the youth should be prioritized. Furthermore, since retirement is influenced by attitudes, behavioural policies should be implemented which positively influence the negative attitudes towards retirement. Also, increased health checkups and programs to mentally stimulate the elderly will positively influence perceptions of health and cognition in rural areas. Since this paper could not conclusively reject certain hypotheses, further research into the influence that belongingness and HIV status have on the likelihood of being retired is required. Lastly, correcting the limitations of repeated survey questions and increasing the response rate would have a positive influence on the validity of the findings of this paper.

# 7. References


Aron, J. and Muellbauer, J., 2000. Inflation and output forecasts for South Africa: monetary transmission implications. *The Centre for the Study of African Economies Working Paper Series*, WPS/2000-23.

Bartlett, J., 2014. *R squared in logistic regression – The Stats Geek*. [online] Thestatsgeek.com. Available at: <https://thestatsgeek.com/2014/02/08/r-squared-in-logistic-regression/> [Accessed 9 August 2021].

Bliss, C., 1934. The Method of Probits. *Science*, 79(2037), pp.38-39.

Bloemen, H., 2009. The Effect of Private Wealth on the Retirement Rate: An Empirical Analysis. *Economica*, 78(312), pp.637-655.

Bor, J., Herbst, A., Newell, M. and Barnighausen, T., 2013. Increases in Adult Life Expectancy in Rural South Africa: Valuing the Scale-Up of HIV Treatment. *Science*, 339(6122), pp.961-965.

Bozio, A., Emmerson, C., O'Dea, C. and Tetlow, G., 2013. Savings and wealth of the lifetime rich: evidence from the UK and US. *Institute for Fiscal Studies*, Working Paper 13(30).

Cameron, A. and Trivedi, P., 2005. *Microeconomics: Methods and Applications*. New York: Cambridge University Press.

Cameron, L. and Cobb-Clark, D., 2008. Do co-residency and financial transfers from the children reduce the need for elderly parents to works in developing countries?. *Journal of Population Economics*, 21(4), pp.1007-1033.

Chasimpha, S., Mclean, E., Dube, A., McCormack, V., dos-Santos-Silva, I. and Glynn, J., 2020. Assessing the validity of and factors that influence accurate self-reporting of HIV status after testing: a population-based study. *AIDS*, 34(6), pp.931-941.

Cherry, K., 2021. *How Can Our Attitudes Change and Influence Behaviors?*. [online] Verywell Mind. Available at: <https://www.verywellmind.com/attitudes-how-they-form-change-shape-behavior-2795897> [Accessed 10 August 2021].

Crawford, R., Cribb, J., Emmerson, C. and Simpson, P., 2020. Retirement expectations, attitudes and saving behaviour: how have these changed during a decade of pension reforms?. *The Institute for Fiscal Studies*, pp.5-7.

Davey, J., 2008. What Influences Retirement Decisions?. *Social Policy Journal of New Zealand*, (33).

De Preter, H., Van Looy, D., Mortelmans, D. and Denaeghel, K., 2013. Retirement timing in Europe: The influence of individual work and life factors. *The Social Science Journal*, 50(2), pp.145-151.

Disney, R., Emmerson, C. and Wakefield, M., 2003. Ill-health and retirement in Britain: A panel data-based analysis. *The Institute for Fiscal Studies*, WP 03/02.

Dynan, K., Skinner, J. and Zeldes, S., 2004. Do the Rich Save More?. *Journal of Political Economy*, 112(2), pp.397-444.

Fernández-López, S., Otero, L., Vivel, M. and Rodeiro, D., 2010. What Are the Driving Forces of Individuals' Retirement Savings?. *Czech Journal of Economics and Finance*, 60(3), pp.226-251.

Ford, J., Park, K. and Sen, S., 2009. All Work and No Play: Pecuniary Versus Non-Pecuniary Factors in the Labour Supply of the Elderly. *Department of Economics, University of Birmingham*, Discussion Papers 09-08.

Glen, S., 2016. *Hosmer-Lemeshow Test: Definition*. [online] Statistics How To. Available at: <https://www.statisticshowto.com/hosmer-lemeshow-test/> [Accessed 9 August 2021].

Gore, A., 2018. *Launch of Discovery Bank. The world's first behavioural bank*. [online] Linkedin.com. Available at: <https://www.linkedin.com/pulse/launch-ofdiscovery-bank-worlds-first-behavioural-adrian-gore> [Accessed 2 April 2021].

Harvard Center for Population and Development Studies, 2016, *"HAALSI Baseline Survey", doi:10.7910/DVN/F5YHML, Harvard Dataverse, V1*

Hershey, D. and Mowen, J., 2000. Psychological Determinants of Financial Preparedness for Retirement. *The Gerontologist*, 40(6), pp.687-697.

Hosmer, D., Lemeshow, S. and Sturdivant, R., 2013. *Applied logistic regression*. 3rd ed. New York: John Wiley and Sons, pp.157-160.

Jakubec, S., Olfert, M., Choi, L., Dawe, N. and Sheehan, D., 2019. Understanding Belonging and Community Connection for Seniors Living in the Suburbs. *Urban Planning*, 4(2), pp.43-52.

Kalemli-Ozcan, S. and Weil, D., 2010. Mortality change, the uncertainty effect, and retirement. *Journal of Economic Growth*, 15(1), pp.65-91.

Kananen, L., Enroth, L., Raitanen, J. et al., 2021. Self-rated health in individuals with and without disease is associated with multiple biomarkers representing multiple biological domains. *Scientific Reports*, 11(1).

Kapelyuk, S., 2017. *What are you retiring for? Health consequences in early aging country*. [online] Conference.iza.org. Available at: <https://conference.iza.org/conference_files/WoLabConf_2018/kapelyuk_s26280.pdf> [Accessed 4 August 2021].

Kim, JH. and Kim, JM., 2017. Subjective life expectancy is a risk factor for perceived health status and mortality. *Health and Quality of Life Outcomes*, 15(1).

Lakshmi Malroutu, Y. and Xiao, J., 1995. Perceived Adequacy of Retirement Income. *Journal of Financial Counseling and Planning*, 6, pp.17-23.

Lam, M., Leibbrandt, G. and Ranchhod, V., 2006. Aging in Sub-Saharan Africa: Recommendation for Furthering Research. *National Research Council (US) Committee on Population*.

Lexico Dictionary. 2021. *Retirement*. [online] Available at: <https://www.lexico.com/definition/retirement> [Accessed 13 August 2021].

Lidi, B., Bedemo, A. and Belina, M., 2017. Determinants of Saving Behavior of Farm Households in Rural Ethiopia: The Double Hurdle Approach. *Developing Country Studies*, 7(12), pp.17-26.

Lind, J. and Mehlum, H., 2009. With or Without U? The Appropriate Test for a U-Shaped Relationship*. *Oxford Bulletin of Economics and Statistics*, 72(1), pp.109-118.

Lorem, G., Cook, S., Leon, D., Emaus, N. and Schirmer, H., 2020. Self-reported health as a predictor of mortality: A cohort study of its relation to other health measurements and observation time. *Scientific Reports*, 10(1).

McAllister, A., Bodin, T., Brønnum-Hansen, H., Harber-Aschan, L. et al., 2020. Inequalities in extending working lives beyond age 60 in Canada, Denmark, Sweden and England—By gender, level of education and health. *PLOS ONE*, 15(8).

Mcleod, S., 2019. *Likert Scale Definition, Examples and Analysis*. [online] Simplypsychology.org. Available at: <https://www.simplypsychology.org/likert-scale.html> [Accessed 11 August 2021].

Meng, A., Sundstrup, E. and Andersen, L., 2020. Factors Contributing to Retirement Decisions in Denmark: Comparing Employees Who Expect to Retire before, at, and after the State Pension Age. *International Journal of Environmental Research and Public Health*, 17(9).

OECD, 2017. Country Profiles – South Africa. *Pensions at a Glance 2017: OECD and G20 Indicators, OECD Publishing, Paris.*

OECD, 2011. Linking Pensions to Life Expectancy. *Pensions at a Glance 2011: Retirement-income Systems in OECD and G20 Countries, OECD Publishing, Paris*, pp.81-102

Oltmanns, J., Godde, B., Winneke, A., Richter, G. et al., 2017. Don't Lose Your Brain at Work – The Role of Recurrent Novelty at Work in Cognitive and Brain Aging. *Frontiers in Psychology*, 8.

Pang, G., Warshawsky, M. and Weitzer, B., 2008. The Retirement Decision: Current Influences on the Timing of Retirement among Older Workers. *Wharton Pension Research Council Working Papers*, 232(WP2008-04).

Payne, C., Gómez-Olivé, F., Kahn, K. and Berkman, L., 2017a. Physical Function in an Aging Population in Rural South Africa: Findings from HAALSI and Cross-National Comparisons with HRS Sister Studies. *The Journals of Gerontology: Series B*, 72(4), pp.665-679.

Payne, C., Wade, A., Kabudula, C. et al., 2017b. Prevalence and correlates of frailty in an older rural African population: findings from the HAALSI cohort study. *BMC Geriatrics*, 17(1).

Pensionfundsonline.co.uk. 2021. *Pension system in South Africa - Pension Funds Online*. [online] Available at: <https://www.pensionfundsonline.co.uk/content/country-profiles/south-africa> [Accessed 13 August 2021].

Pregibon, D., 1980. Goodness of Link Tests for Generalized Linear Models. *Applied Statistics*, 29(1), pp.15-24.

Price, P., Jhangiani, R. and Chiang, I., 2015. *Research Methods of Psychology – 2nd Canadian Edition*. 2nd ed. [ebook] Victoria: B.C.: BCcampus. Available at: <https://opentextbc.ca/researchmethods/> [Accessed 11 August 2021].

Richardson, V. and Kilty, K., 1989. Retirement Financial Planning Among Black Professionals. *The Gerontologist*, 29(1), pp.32-37.

Rohr, J., Xavier Gómez-Olivé, F., Rosenberg, M., Manne-Goehler, J. et al., 2017. Performance of self-reported HIV status in determining true HIV status among older adults in rural South Africa: a validation study. *Journal of the International AIDS Society*, 20(1).

Salotti, S., 2010. Global imbalances and household savings: The role of wealth. *The Social Science Journal*, 47(1), pp.21-44.

Sarabia-Cobo, C., Pérez, V., Hermosilla, C. and de Lorena, P., 2020. Retirement or no Retirement? The Decision's Effects on Cognitive Functioning, Well-Being, and Quality of Life. *Behavioral Sciences*, 10(10).

Simleit, C., Keeton, G. and Botha, F., 2011. The Determinants of Household Savings in South Africa. *Studies in Economics and Econometrics*, 35(3), pp.1-19.

Singleton, J. and Keddy, B., 1991. Planning for retirement. *Activities, Adaptation & Aging*, 16(2), pp.49-55.

Sousa-Ribeiro, M., Bernhard-Oettel, C., Sverke, M. and Westerlund, H., 2021. Health- and Age-Related Workplace Factors as Predictors of Preferred, Expected, and Actual Retirement Timing: Findings from a Swedish Cohort Study. *International Journal of Environmental Research and Public Health*, 18(5), pp.27-46.

Steffens, N., Cruwys, T., Haslam, C., Jetten, J. and Haslam, S., 2016. Social group memberships in retirement are associated with reduced risk of premature death: evidence from a longitudinal cohort study. *BMJ Open*, 6(2).

Stewart, F. and Yermo, Y., 2009. Pensions in Africa. *OECD Working Papers on Insurance and Private Pensions*, 30.

Stolzenberg, R., 2011. Do Not Go Gentle into That Good Night: The Effect of Retirement on Subsequent Mortality of U.S. Supreme Court Justices, 1801–2006. *Demography*, 48(4), pp.1317-1346.

Thaler, R. and Benartzi, S., 2007. The Behavioral Economics of Retirement Savings Behavior. *The AARP Public Policy Insitute*, 2.

The World Bank, 2015. *South Africa's Changing Demographic Could Lift Growth to 5.4% by 2030*. [online] Available at: <https://www.worldbank.org/en/news/press-release/2015/08/17/south-africa-demographic-lift-growth> [Accessed 2 April 2021].

Ting, L. and Kollamparambil, U., 2015. Nature and determinants of household retirement savings behaviour in South Africa. *Development Southern Africa*, 32(6), pp.675-696.

Tomar, S., Kent Baker, H., Kumar, S. and Hoffmann, A., 2021. Psychological determinants of retirement financial planning behavior. *Journal of Business Research*, 133, pp.432-449.

Topa, G., Depolo, M. and Alcover, C., 2017. Early Retirement: A Meta-Analysis of Its Antecedent and Subsequent Correlates. *Frontiers in Psychology*, 8.

Topa, G., Lunceford, G. and Boyatzis, R., 2017. Financial Planning for Retirement: A Psychosocial Perspective. *Frontiers in Psychology*, 8.

Train, K., 2012. Discrete Choice Methods with Simulation. *Cambridge University Press*, pp.97-133.

Trevethan, R., 2017. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Frontiers in Public Health*, 5(307).

Uccello, C., 1998. Factors Influencing Retirement: Their Implications for Raising the Retirement Age. *AARP Public Policy Institute Working Paper #9810*.

van Solinge, H. and Henkens, K., 2010. Living longer, working longer? The impact of subjective life expectancy on retirement intentions and behaviour. *The European Journal of Public Health*, 20(1), pp.47-51.

Vittinghoff, E., Glidden, D., Shiboski, S. and McCulloch, C., 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. New York: Springer-Verlag.

Wasserstein, R., Schirm, A. and Lazar, N., 2019. Moving to a World Beyond "p < 0.05". *The American Statistician*, 73(sup1), pp.1-19.

WHO, 2018. *Ageing and health*. [online] World Health Organisation. Available at: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> [Accessed 23 March 2021].

Willows, G., 2019. South African individual retirement savings: An analysis of the social factors. *Risk Management and Insurance Review*, 22(3), pp.303-328.

Wise, D., 2014. *Discoveries in the Economics of Aging*. University of Chicago Press.

World Bank, 2019. *Life expectancy at birth, total (years) - South Africa*. [online] World Development Indicators. Available at: <https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=ZA> [Accessed 2 April 2021].

World Bank, 2019. *Life expectancy at birth, total (years) - Netherlands | Data*. [online] World Development Indicators. Available at: <https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=NL> [Accessed 2 April 2021].

Yusof, R. and Sabri, M., 2017. *Determinants of Retirement Savings*. [online] Research Gate. Available at: <https://www.researchgate.net/publication/321572773> [Accessed 14 August 2021].

Zeka, B., 2016. Factors Influencing the Retirement Savings of Individuals in the Eastern Cape, South Africa. *Proceedings of the 24th International Academic Conference, Barcelona*.

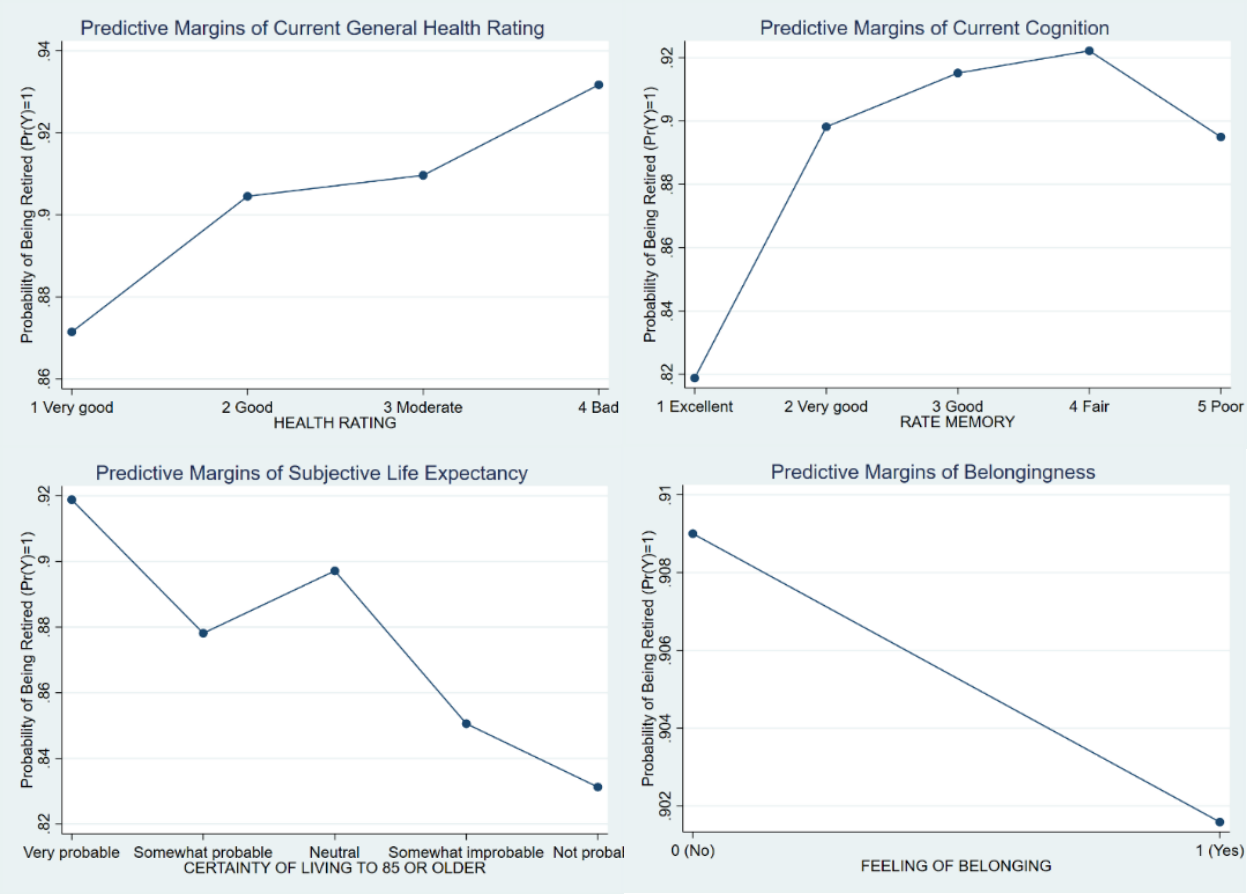
Zeka, B., 2017. Exploring the Retirement Attitudes of Pre-Retired Individuals in the Eastern Cape, South Africa. *International Journal of Management and Applied Science*, 3(9).

Zwane, T., Greyling, L. and Maleka, M., 2016. The Determinants of Household Savings in South Africa: A Panel Data Approach. *International Business & Economics Research Journal (IBER)*, 15(4), pp.209-218.

# 8. Appendix

## 8.1 Appendix A: Visuals for Predictive Margins and AME

### Figure 4: Probability of Being Retired per Categorical Variable




Graphs showing the average predicted probabilities based on the probit regression model. The average probability of being retired was calculated by assuming that everyone within the constrained dataset was treated as if they fit into one category. More specifically (looking at the bottom left graph in figure 4), there is a 92% probability of being retired if every individual believed that it was very probable that they would live to be 85, or older. Using predictive margins to group individuals allows us to interpret the influence that discrete movements between the categories of factor variables (on average) would have on the probability of being retired.

**Figure 5: Probability of Being Retired by Subjective Life Expectancy at various Ages**

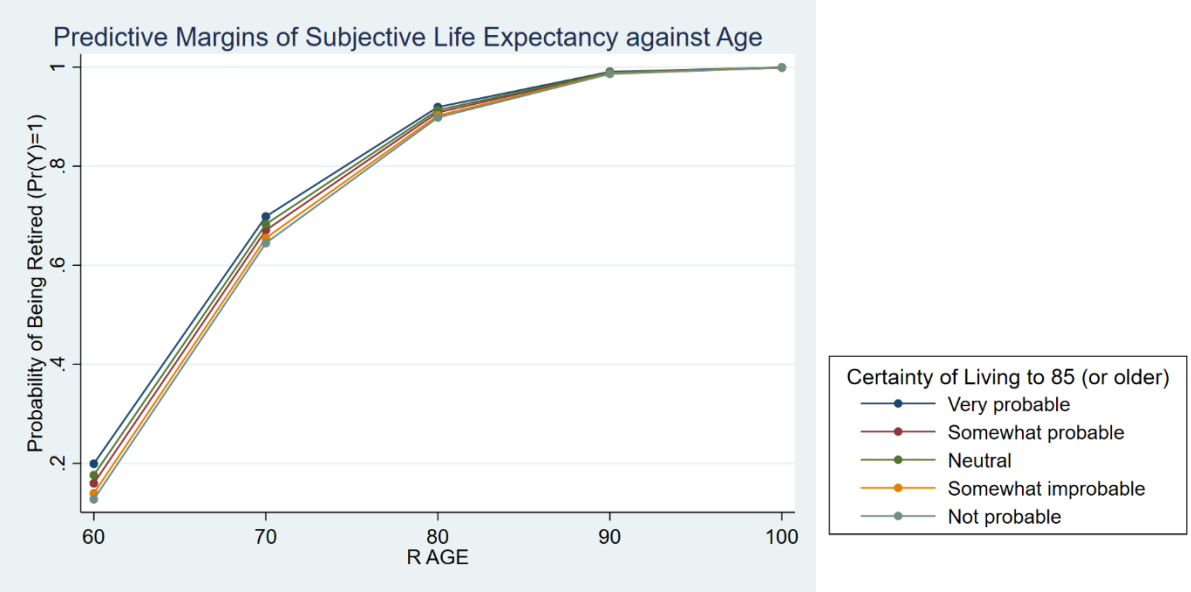
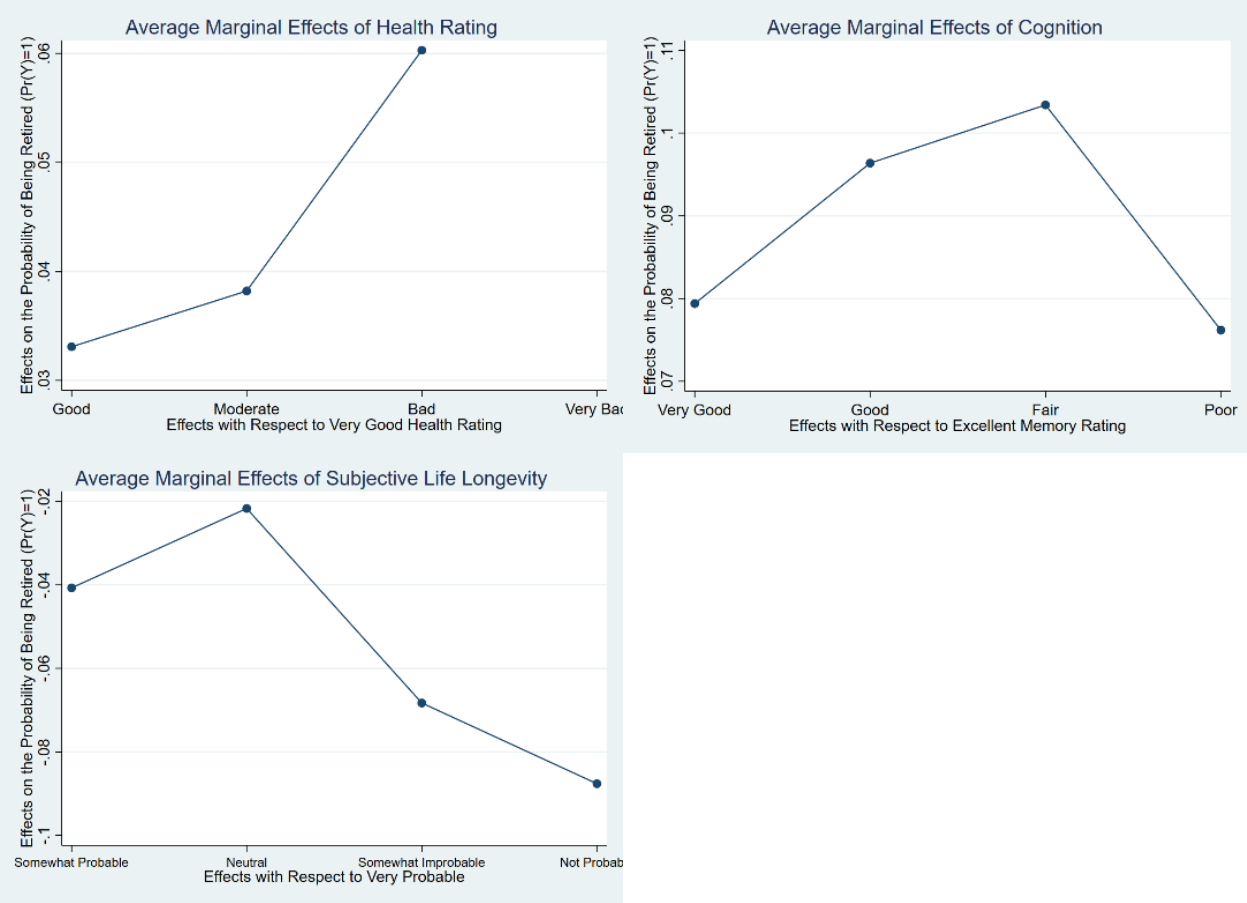



**Figure 6: Average Marginal Effects of Main Explanatory Variables**




*Notes*: Belongingness was not included because it is a binary variable (only one point on a graph). Figure 6 provides visualization for discrete changes in categorical variables with respect to the base category.

## 8.2 Appendix B: Significance Tests for the Main Factor and Interaction Variables

**Table 6: Overall Effect of the Main Factor Variables**

| | Probit Model (2) | | |
|---|---|---|---|
| Variable | Wald Chi-Squared | D.f. | Pr > Chi-Sq |
| Education Category (vs No Formal Education) | 7.57* | 3 | 0.0559 |
| Wealth Asset Index (vs Quintile 1: Lowest) | 9.20* | 4 | 0.0563 |
| Current General Health Rating (vs Very Good) | 4.68 | 3 | 0.1972 |
| Cognition – Memory rating (vs Excellent) | 8.94* | 4 | 0.0625 |
| Certainty of Living to 85 (vs Very Probable) | 8.55* | 4 | 0.0735 |

Notes: Test for the overall significance that the factor variables have in the model. Current general health rating has 3 degrees of freedom. Since general health rating 5 (Very bad) was dropped due to being a perfect predictor of retirement. *** p<0.01, ** p<0.05, * p<0.1.

Table 6 shows the overall statistical significance of the multi-category dummy variables used in the model. The Wald chi-square test statistics are measured with n-1 degrees of freedom. Furthermore, table 7 on the next page shows the overall significance of the various interaction terms and main variable effects if they were introduced into four separate nested models individually. The table of the coefficients for the interaction terms was omitted to save paper. Since the interaction terms were not significant overall, they were not included in the models. Also, interpreting results with interaction terms is much more difficult. In favour of keeping the model as simple as possible, they were not tested further.

**Table 7: Tests for the Overall Effect of Interaction Terms**

| | **Probit Model (2.1)** | | | **Probit Model (2.2)** | | | **Probit model (2.3)** | | | **Probit Model (2.4)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Chi2 | D.f. | Pr > Chi-Sq | Chi2 | D.f. | Pr > Chi-Sq | Chi2 | D.f. | Pr > Chi-Sq | Chi2 | D.f. | Pr > Chi-Sq |
| HIV positive x Health Rating | 0.87 | 3 | 0.834 | | | | | | | | | |
| Health Rating | 3.86 | 3 | 0.277 | | | | | | | | | |
| HIV positive | 0.39 | 1 | 0.530 | | | | | | | | | |
| Cognition x Health Rating | | | | 9.99 | 11 | 0.532 | | | | | | |
| Health Rating | | | | 3.36 | 3 | 0.340 | | | | | | |
| Cognition | | | | 2.87 | 4 | 0.580 | | | | | | |
| HIV positive x Life Longevity | | | | | | | 4.22 | 3 | 0.239 | | | |
| HIV positive | | | | | | | 2.66 | 1 | 0.103 | | | |
| Life Longevity | | | | | | | 11.02 | 4 | 0.026** | | | |
| Estrangement x Health Rating | | | | | | | | | | 1.05 | 2 | 0.590 |
| Health Rating | | | | | | | | | | 5.14 | 3 | 0.162 |
| Estrangement | | | | | | | | | | 0.73 | 1 | 0.394 |
| McFadden's r2 | 0.1559 | | | **0.1657** | | | 0.1601 | | | 0.1565 | | |

Notes: The factor variables are considered in separate nested models added one at a time and compared to the base levels. The separate models show all interactions and main effects of variables as compared to their base values. *** p<0.01, ** p<0.05, * p<0.1

## 8.3 Appendix C: Statistical Tests of the Model

**Table 8: Goodness-of-Fit Tests for Two Nested Probit Models**

| | **Probit Model (1)** | **Probit Model (2)** | **Difference** |
|---|---|---|---|
| N: | 1911 | 1461 | -450 |
| Log-Lik Intercept Only: | -578.173 | -468.098 | 110.075 |
| Log-Lik Full Model: | -499.619 | -395.543 | 104.076 |
| D: | 999.238(1893) | 791.087(1426) | -208.152(-467) |
| LR: | 157.108(12) | 145.110(24) | -11.998(12) |
| Prob > LR: | 0.000 | 0.000 | 0.000 |
| **McFadden's R2:** | **0.136** | **0.155** | **0.019** |
| McFadden's Adj R2: | 0.105 | 0.080 | -0.025 |
| Maximum Likelihood R2: | 0.079 | 0.095 | 0.016 |
| Cragg & Uhler's R2: | 0.174 | 0.200 | 0.026 |
| McKelvey and Zavoina's R2: | 0.239 | 0.271 | 0.032 |
| Efron's R2: | 0.092 | 0.110 | 0.018 |
| Variance of y*: | 1.314 | 1.372 | 0.058 |
| Variance of error: | 1.000 | 1.000 | 0.000 |
| Count R2: | 0.910 | 0.903 | -0.007 |
| Adj Count R2: | 0.000 | 0.007 | 0.007 |
| AIC: | 0.542 | 0.589 | 0.048 |
| **AIC*n:** | **1035.238** | **861.087** | **-174.152** |
| BIC: | -13303.100 | -9599.999 | 3703.101 |
| BIC': | -66.443 | 29.775 | 96.219 |

Notes: Warning N's do not match. Fitstat command in Stata showing all the goodness of fit statistics of the two models.

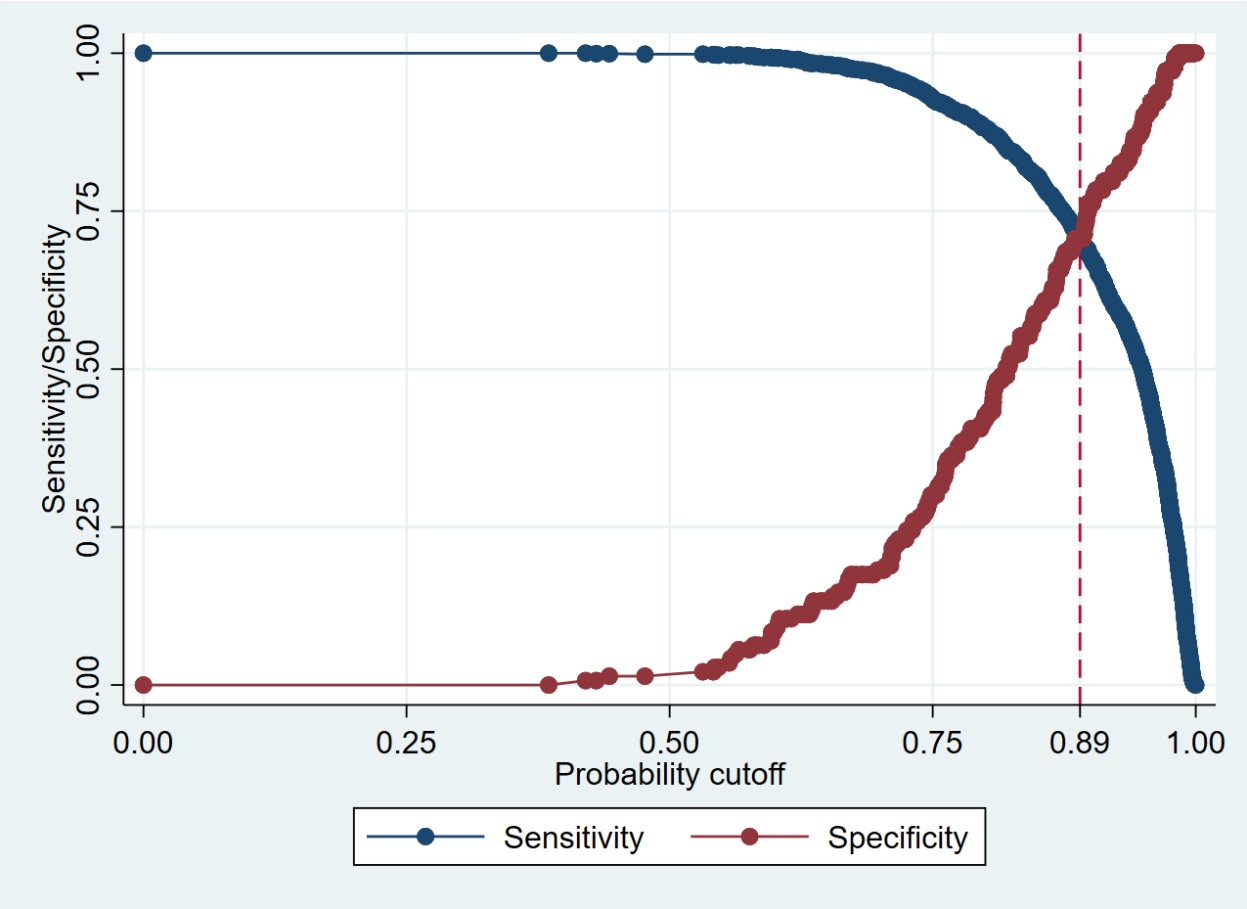
Since the dependent variable of the model is a binary variable, McFadden's $R^2$ was used (in addition to other tests) to assess the predictive ability of the model. Considering that the probit model was estimated using the method of maximum likelihood, McFadden's $R^2$ is equal to: $1 - \frac{\text{Log(Lmod)}}{\text{Log(Lnull)}}$. Where $\text{Log}(L_{mod})$ is the log-likelihood value of the fitted probit model and $\text{Log}(L_{null})$ is the log-likelihood value of a null model with only an intercept and no other predictor variables (Bartlett, 2014). Models that have a larger McFadden's $R^2$ value (closer to 1) will have a far better predictive ability than models with a smaller value (closer to 0). However, a general rule of thumb that is accepted in the literature is that a McFadden $R^2$ value between 0.2 and 0.4 implies that a model will have a very good fit. Therefore, the McFadden $R^2$ value of 0.155 associated with the probit model shows that it is somewhat of a good fit to the data.

**Table 9: Goodness-of-Fit Tests for Logit and Probit Model**

| | Logit Model | Probit Model | Difference |
|---|---|---|---|
| N: | 1461 | 1461 | 0 |
| Log-Lik Intercept Only: | -468.098 | -468.098 | 0.000 |
| Log-Lik Full Model: | -395.993 | -395.543 | -0.450 |
| D: | 791.986(1426) | 791.087(1426) | 0.900(0) |
| LR: | 144.210(24) | 145.110(24) | -0.900(0) |
| Prob > LR: | 0.000 | 0.000 | 0.000 |
| McFadden's R2: | 0.154 | 0.155 | -0.001 |
| McFadden's Adj R2: | 0.079 | 0.080 | -0.001 |
| Maximum Likelihood R2: | 0.094 | 0.095 | -0.001 |
| Cragg & Uhler's R2: | 0.199 | 0.200 | -0.001 |
| McKelvey and Zavoina's R2: | 0.300 | 0.271 | 0.029 |
| Efron's R2: | 0.113 | 0.110 | 0.004 |
| Variance of y*: | 4.700 | 1.372 | 3.329 |
| Variance of error: | 3.290 | 1.000 | 2.290 |
| Count R2: | 0.903 | 0.903 | 0.000 |
| Adj Count R2: | 0.007 | 0.007 | 0.000 |
| AIC: | 0.590 | 0.589 | 0.001 |
| AIC*n: | 861.986 | 861.087 | 0.900 |
| BIC: | -9599.099 | -9599.999 | 0.900 |
| **BIC':** | **30.675** | **29.775** | **0.900** |

Notes: The difference of 0.900 in BIC' shows weak support for the probit model over the logit model. Furthermore, when the predictions of the probit and logit model are regressed against each other, the r-squared value is 0.9975 which shows that the models are practically the same.

**Figure 7: Sensitivity/Specificity for the Probit Model**




The approximate probability cutoff value of 89% was chosen to provide the best balance between sensitivity and specificity of the model. Therefore, summary statistics of the model using a probability cutoff of 89% is as follows:

**Table 10: Percentage of Correctly Classified Variables**

| | True | | |
|---|---|---|---|
| Classified | D | ~D | Total |
| + | 932 | 42 | 947 |
| - | 386 | 101 | 487 |
| Total | 1318 | 143 | 1461 |

*Classified correctly if predicted Pr(D) >= 0.89. True D is when the respondent is retired (Y=1) and ~D is when the respondent is employed or looking for work (Y=0).*

Therefore, classifying respondents according to the "gold standard", as mentioned by Trevethan (2017), we see that 932 and 101 respondents were correctly classified as being retired and not retired (respectively). Whereas, on the other hand, 42 and 386 respondents were falsely

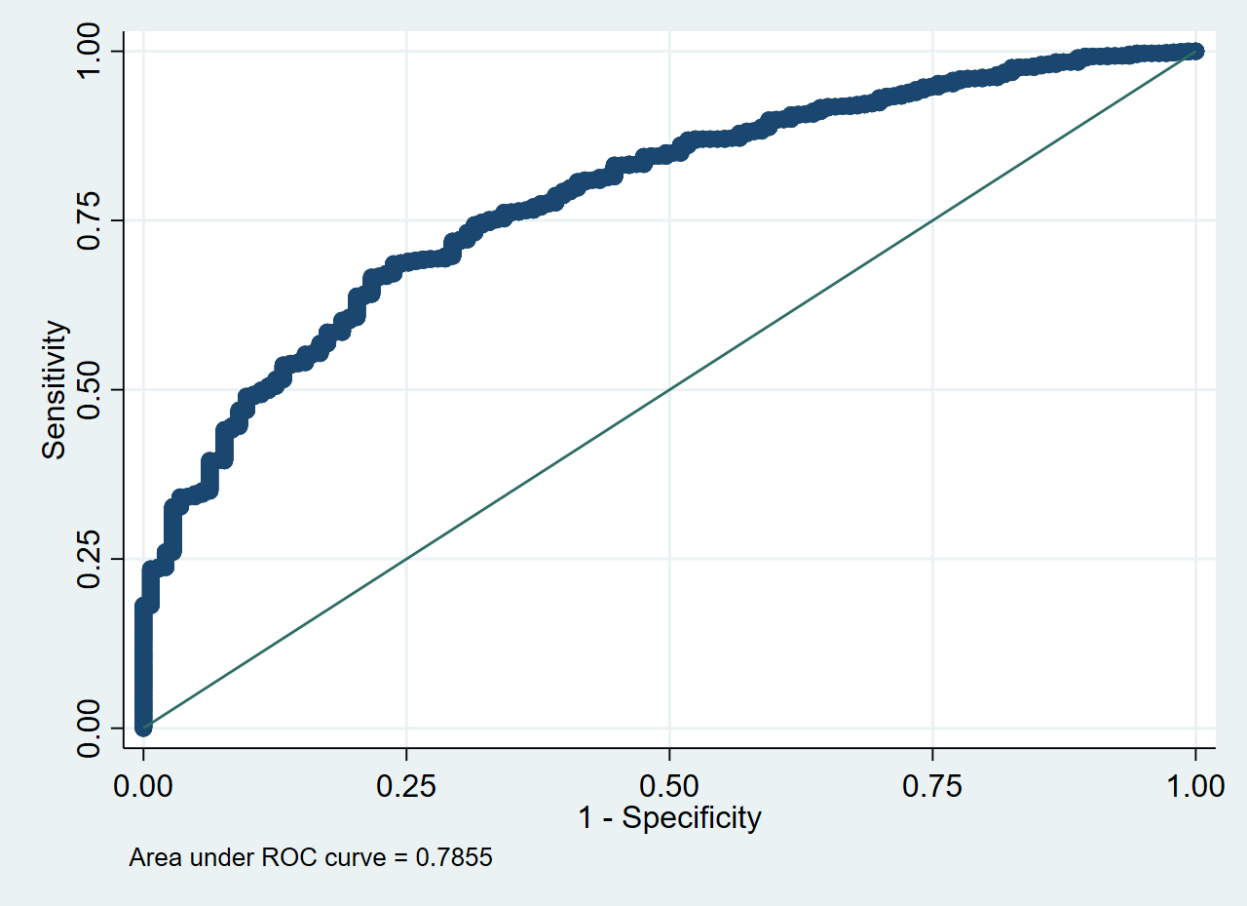
classified as being retired and not retired (respectively). The summary statistics of percentage correctly classified was as follows:

**Table 11: Summary Statistics for Classification of Data**

| | | |
|---|---|---|
| Sensitivity | Pr(+\|D) | 70.71% |
| Specificity | Pr(-\|~D) | 70.63% |
| Positive predictive value | Pr(D\|+) | 95.69% |
| Negative predictive value | Pr(~D\|-) | 20.74% |
| | | |
| False + rate for true ~D | Pr(+\|~D) | 29.37% |
| False - rate for true D | Pr(-\|D) | 29.29% |
| False + rate for classified + | Pr(~D\|+) | 4.31% |
| False - rate for classified - | Pr(D\|-) | 79.26% |
| Correctly Classified | | 70.70% |

The sensitivity value shows that 70.71% of those who are retired (Y=1) were correctly classified and the specificity value shows that 70.63% of those who are not retired (Y=0) were correctly classified. Although the model was able to correctly classify true negatives and positives, the predictive value for positive values (95.69%) was far higher than for negative values (20.74%). Since a low sensitivity is associated with many missing values and a low specificity is associated with false classifications, a combination of the two was chosen. Further tests showed that the percentage of correctly classified values was 90.28% when using the default cutoff of 5%. Changing the cutoff to the default of 5% increased the predictive power by increasing the sensitivity to 99.58% and by decreasing the specificity to 2.10%. Therefore, with a probability cutoff of 5%, the model had far fewer missing cases, but a far larger number of false classifications.

**Figure 8: Receiver Operating Characteristic (ROC) Curve**




The ROC curve is a graphical illustration of sensitivity against 1-specificity of the model. More specifically, the graph shows deviations from a model of complete random chance (shown by the 45-degree line from the origin). The shape of the model is associated with a good degree of accuracy. The area under the ROC curve (0.7855) is calculated by integrating the graph and indicates that the model has good predictive power.

**Table 12: Link-Test for Specification Error**

| | | |
|---|---|---|
| Probit Regression | Number of observations: | 1,461 |
| | LR chi2(2): | 145.25 |
| | Prob > chi2: | 0 |
| Log likelihood = -395.47344 | Pseudo R2: | 0.1551 |

| Y - (Retirement) | Coef. | Std. Err. | Z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Predicted Value (ŷ) | .8743463 | 0.3498871 | 2.5 | 0.012 | 0.1885802 | 1.560113 |
| Predicted Value Squared $(\hat{y})^2$ | .0520754 | 0.1403618 | 0.37 | **0.711** | -0.2230288 | 0.3271795 |
| Constant | .0609579 | 0.2015856 | 0.3 | 0.762 | -0.3341426 | 0.4560584 |

The table above shows the results of a link test, as described by Pregibon (1980), to investigate if there is a specification error. The null hypothesis of the test is that the model is not characterized by a misspecification error. Since the predicted value squared $(\hat{y})^2$ is not statistically significant, we fail to reject the null hypothesis and conclude that the model is correctly specified. Furthermore, the predicted value ($\hat{y}$) is statistically significant (with alpha = 0.05) which is in line with expectations of the model.

**Table 13: Pearson Chi-Squared Goodness-of-Fit Test**

| Pearson Chi-squared Goodness of Fit Test of Probit Model | |
|---|---|
| Number of observations: | 1,461 |
| Number of covariate patterns: | 1,411 |
| Pearson chi2 (1386): | 1331.17 |
| Prob > chi2: | 0.8515 |

Using the Pearson chi-squared goodness of fit test, we fail to accept the null hypothesis that there is a lack of fit (p-value = 0.85). However, the number of covariate patterns (1,411) is extremely close to the number of observations (1,461) which means that most of the covariate pattern sets will be singletons. To combat this, a Hosmer-Lemeshow test was conducted.

**Table 14: Hosmer-Lemeshow Goodness-of-Fit Test**

| Hosmer-Lemeshow Goodness of Fit Test of Probit Model | |
|---|---|
| Number of observations: | 1,461 |
| Number of groups: | 10 |
| Hosmer-Lemeshow chi2(8): | 2.83 |
| Prob > chi2: | 0.9445 |

*Notes: Table collapsed on quantiles of estimated probabilities*

As suggested by Hosmer, Lemeshow and Sturdivant (2013), regrouping the data into 10 groups (a commonly accepted number of groups) allowed the goodness of fit model to have far fewer degrees of freedom. Once again, there is not enough evidence to conclude that a lack of fit for the model exists (p-value = 0.9445). A notable limitation of having such a large p-value could indicate that the model has a weak test power (Glen, 2016).

**Table 15: Test for Multicollinearity using Variance Inflation Factors**

| **Variable** | VIF | 1/VIF |
|---|---|---|
| Age | 143.41 | 0.006973 |
| Age Squared | 90.42 | 0.01106 |
| Female | 2.05 | 0.48722 |
| Not Married | 2.51 | 0.398017 |
| **Education Category** | | |
| Some primary (1-7 years) | 1.91 | 0.522571 |
| Some secondary (8-11 years) | 1.25 | 0.801251 |
| Secondary or more (12+ years) | 1.19 | 0.842873 |
| **Wealth Asset Index** | | |
| Wealth quintile 2 | 2.11 | 0.474357 |
| Wealth quintile 3 | 2.15 | 0.464955 |
| Wealth quintile 4 | 2.29 | 0.436736 |
| Wealth quintile 5 | 2.42 | 0.413928 |
| HIV Positive | 1.13 | 0.884025 |
| **Current General Health Rating** | | |
| Good | 4.65 | 0.215046 |
| Moderate | 2.22 | 0.450146 |
| Bad | 2.8 | 0.357327 |
| Very Bad | - | - |
| **Cognition- Current Memory Rating** | | |
| Very Good | 4.95 | 0.202166 |
| Good | 9.59 | 0.104243 |
| Fair | 4.16 | 0.240537 |
| Poor | 1.71 | 0.583402 |
| **Certainty of Living to 85** | | |
| Somewhat Probable | 1.46 | 0.686114 |
| Neutral | 1.35 | 0.742665 |
| Somewhat Improbable | 1.18 | 0.851043 |
| Not Probable | 1.11 | 0.901158 |
| Village Estrangement | 1.09 | 0.915705 |
| Mean VIF | 12.05 | |

Table 15 revealed that multicollinearity was not a problem in the model since none of the independent variables (other than age and age squared) exceeded the commonly used threshold value of 10 (Vittinghoff et al., 2012). Since age squared was directly related to age, it would be expected to increase the mean VIF. When age squared was removed, multicollinearity was still not an issue.