Measurement Models for Understanding the Social Challenges of Caring for the Elderly in Brazil and England

A thesis submitted to Lancaster University in accordance with the requirements for the degree of Doctor of Philosophy

by

Rebecca Louise PATTINSON

MATHEMATICS AND STATISTICS DEPARTMENT

Supervisor:

Dr. Andrew TITMAN

Co-supervisors:

Prof. Gillian LANCASTER

Dr. Dirley Moreira dos SANTOS

Dr. José Rodrigo de MORAES

8th June, 2018

Declaration

I declare all work in this thesis is my own original work, except where referenced. None of the work is the result of collaborative work with others and no part of it has been submitted for the award of a higher degree elsewhere.

Abstract

An ageing population is a challenge faced by many developed and developing countries. By understanding the needs in health and well-being of the elderly governments can ensure policies and services are as efficient as possible, in order to provide the best care amidst growing demand. The thesis analyses categorical data from the English Longitudinal Study of Ageing (ELSA) and Brazil's National Household Sample Survey (PNAD) to form measurement models for aspects of health, economic well-being and subjective well-being in the elderly, in England and Brazil.

Applying structural equation modelling (SEM), a measurement model for health and economic well-being in both countries was constructed using categorical variables from the survey data, as well as for subjective well-being in England. For health the variables represented self-reported morbidity and functional status, while for economic well-being variables represented housing quality and durables owned within the household. Psychometric measures of satisfaction with life, quality of life, loneliness and depression constituted subjective well-being. Based on empirical evidence from exploratory factor analysis (EFA), the latent structure of each aspect was defined. Multilevel structural equation modelling was applied to the PNAD to capture the hierarchical structure, whereby individuals were clustered by household then the households were clustered by sector and the sectors clustered by municipality. Meanwhile, the longitudinal dynamic of the ELSA allowed for a multivariate latent growth modelling, using multiple indicators to model the trajectories of subjective well-being in multiple aspects.

Health and economic well-being were positively associated in both countries. Metabolic health was a factor that was identified in both England and Brazil, while musculoskeletal health was identified only in Brazil. In England, economic well-being had two separate factors relating to the inside and outside of the household. Health and household quality significantly influenced subjective well-being, but not neighbourhood quality, with better health and economic well-being associated with greater subjective well-being. In the multilevel SEM a different structure of latent constructs was identified for the health of the sectors of elderly individuals, and the variables had different priorities.

In both countries, health was significantly better for men, but worse for those of non-white race and with older age. Regional disparities were also present in health and economic well-being. In Brazil, economic well-being was worse with older age, in rural areas, male gender and non-white race. In England, economic well-being was better for older age with no difference between the genders and races.

Subjective well-being was persistently better for men and those that were married, while being persistently worse for those of non-white race or divorced/separated marital status. Those that were widowed recovered from initially lower subjective well-being. Marital status was highly influential to the subjective well-being of the elderly with significant impact from becoming divorced or widowed and benefits from getting remarried.

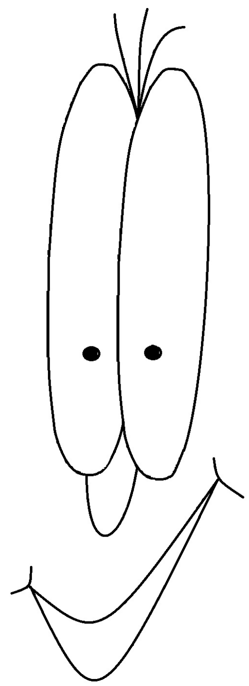
Acknowledgements

I would first and foremost like to thank my supervisor Andrew Titman for all his irreplaceable help and support throughout the project. I'd also like to thank my co-supervisor Gill Lancaster for all her help and advice, and my Brazilian co-supervisors Dirley dos Santos and Rodrigo de Moraes for all their expertise, and for hosting me and Abbie Jones during our visit to Rio de Janeiro, Brazil.

I am very grateful to the Economic and Social Research Council for funding this work, with the additional Advanced Quantitative Methods award.

On a more personal note, thanks are owed to my many colleagues in B18 over the years, who have provided a great collaborative spirit in office that has always been a joy to be part of. I'd like to thank Laura, Ben, Anna and Jess for all their support and for humouring the many adventures of Bob (featured at the bottom of this page). Special thanks must be given to Sarah and Anna for joining me on many lunches and stomps in the woods that saw me through the days when faced with adversity.

Acknowledgement of course also goes to my parents, Alan and Denise, and brother Dean, with our many adventures along the way, demonstrating great results comes from a lot of hard work and persistence. Even if it does mean living in a caravan, through two winters, on a building site, while working on this thesis.



Contents

Abbreviations

ADL - Activities of Daily Living

CES-D - Center for Epidemiological Studies Depression scale

CFA - Confirmatory Factor Analysis

CFI - Comparative Fit Index

CGA - Comprehensive Geriatric Assessment

COPD - Chronic Obstructive Pulmonary Disease

EFA - Exploratory Factor Analysis

FS - Functional Status

HSE - Health Survey for England

IADL - Instrumental Activities of Daily Living

ICC - Intraclass Correlation Coefficient

ICH - Infectious and Chronic Health

IRT - Item Response Theory

MCH - Metabolic and Cardiovascular Health

MIMIC - Multiple Indicator and Multiple Causes model

MSH - Musculoskeletal Health

NVQ - National Vocational Qualifications

PNAD - Pesquisa Nacional por Amostra de Domicílios (National Household Sample Survey)

RMSEA - Root Mean Square Error of Approximation

SEM - Structural Equation Model

SRH - Self-Rated Health

SRMR - Standardised Root Mean Square Residual

SWLS - Satisfaction With Life Scale

TLI - Tucker-Lewis Index

WLSMV - Weighted Least Squares with Mean and Variance adjustment

List of Tables

List of Figures

Chapter 1

Introduction

An ageing population is a challenge faced by many developed and developing countries. By understanding the problems in health and well-being of the elderly, governments can ensure policies and services are as efficient as possible, in order to provide the best care amidst growing demand.

The first focus of this thesis is the construction of measurement models for the health, economic well-being and subjective well-being of the elderly, to provide insight into the state of the elderly in England and Brazil. The state of the elderly for each country was analysed separately, with the aim to capture common features between the countries as well as the different challenges that each country faces. For example, in Brazil tuberculosis is an important issue, as one of the top five countries for incidence of the disease, but it is rare in the UK (World Health Organization [WHO], 2016). Nationally representative survey data was used from the English Longitudinal Study of Ageing (ELSA) and Brazil's National Household Sample Survey (PNAD; *Pesquisa Nacional por Amostra de Domicílios*, in Portuguese).

The research questions to be addressed are:

1. What is the scale of the association between the health and economic well-being of the elderly? Is it similar in England and Brazil?

2. What affects do health and economic well-being have on the subjective well-being of the elderly?

3. How do health, economic well-being and subjective well-being differ between elderly individuals of different demographics?

4. How do the demographics of an area affect the health and economic well-being of elderly individuals within it?

5. What are the changes in the subjective well-being of elderly individuals over time? How is this affected by demographics of the individual and what is the effect of a change in marital status?

Economic well-being has previously been considered in terms of an elderly persons financial wealth, income or satisfaction with their financial situation (Mullis, 1992; Douthitt et al., 1992; Deaton, 2007; Hansen et al., 2008). In this thesis, economic well-being is considered as the outcomes of financial wealth and literacy that ensure the living of standards of the elderly individual in terms of the household environment, for example the basic amenities and the state of repair.

Structural equation modelling (SEM) is a statistical method used to measure unobserved constructs through observed variables, assessing the measurement quality of the observed variables and allowing for causal relationships between the constructs. Unlike more traditional statistical methods, such as regression and ANOVA, the measurement error of the observed variables is taken into account in SEM that is otherwise ignored; leading to biased parameter estimates and misleading conclusions (Wang and Wang, 2013).

Here health, economic well-being and subjective well-being are unobserved constructs for which there is not direct observable measure. In the surveys, there are many groups of items relating to the health, economic well-being or subjective well-being of the individual. For example, self-reported diagnosis of comorbidities and difficulties with performing activities of daily living. SEM is then able to provide a simpler interpretation the interrelationships between this large number of variables by reducing the dimensionality (Bartholomew et al., 2011).

In addition, SEM can be extended to multilevel with structural equation modelling (multilevel SEM) (Muthén, 1994) and to longitudinal analysis with latent growth modelling (Bollen and Curran, 2006). The PNAD has a complex sampling plan in which individuals are clustered by household, locality and municipality. Multilevel SEM is required if clusters exist, otherwise model fit indices can be inflated leading to false conclusion about the goodness-of-fit (Muthén, 1994). Another benefit of multilevel SEM is the ability to determine a different set of latent variables for the individuals and the clusters, which in this case are the household, sector or municipality. Thus, the effect of area demographics on the cluster-level latent variables can be determined. The ELSA, however, did not have clustering but did have repeated measures, where the same variables were observed at multiple waves of the survey. These can, therefore, be used to determine trajectories of the unobserved constructs of elderly individuals over time in latent growth modelling.

The focus of structural equation modelling applications in ageing research has previously been either health or well-being, but not their interrelationship, while economic well-being has only previously been considered by Yan et al. (2014) in a non-financial concept. This thesis will determine the relationships between all three domains, and the impact of demographic and socioeconomic covariates.

The application of multilevel SEMs has been limited, particularly in ageing research, despite the need to correct for clustering in the data (Muthén, 1994). There is also the advantage of overcoming limitations of multilevel regression in mediation analysis (Preacher et al., 2010), such as biased indirect effects.

Health has been the primary focus of the application of latent growth modelling in the elderly, with the majority modelling the repeated measurement as a single observed score (Muniz-Terrera et al., 2009; Rzewuska et al., 2015; Chen et al., 2017). In this thesis we will explore subjective well-being to identify the trajectories of various aspects of subjective well-being applying multivariate latent growth modelling, with multiple indicators. The effect of baseline characteristics on the trajectory and the impact of time-varying marital status are also newly explored.

After this introductory chapter, this thesis will begin with discussion of the current literature in Chapter 2, reviewing the current understanding of health and well-being in the elderly and the problems faced in each area. Chapter 3 then introduces the two surveys, the English Longitudinal Study of Ageing (ELSA) and Brazil's Household Sample Survey (PNAD), that are used throughout the thesis; specifying the variables of interest for the analyses. The complex sampling plan of the PNAD is also detailed. The statistical methodology is presented in Chapter 4, with the general structural equation modelling applied to both datasets given in the first section. Followed by sections for the separate methods used for Brazil then England, which includes the multilevel cumulative logistic regression and multilevel structural equation methods applied to the PNAD and the item response analysis and latent growth modelling methods applied to the ELSA. The results from the analysis of the PNAD are then given in Chapter 5. The cross-sectional results of the analysis using the ELSA are then presented in Chapter 6 followed by the latent growth modelling in Chapter 7. Finally, the results are concluded with a discussion and comparison of the analyses for England and Brazil in Chapter 8, with a final comment on potential further work.

Chapter 2

Literature Review

In both the UK and Brazil the proportion of the population that is elderly has been increasing both in size and as a proportion relative to younger generations. This is shown by the old age dependency ratio, which relates the number of persons aged 65 or more years old (i.e. dependants) to the number of persons aged 20-64 years old for whom the burden of the provision of care lies. This ratio has been increasing since 1965, therefore the working population has greater demand for the provision of services, such as medical and financial services. According to estimates from (UN, 2013) the old age dependency ratio in Brazil increased from 8.9 in 1990 to 11.7 in 2010 (per 100 population aged 20-64) and in the UK from 26.9 to 27.8 in the same year.

This burden is further exacerbated by declining birth rates and child dependency ratios over the same period. Although the total dependency ratio fell from 71.0 in 1990 to 67.7 in 2010 in Brazil and from 98.9 to 69.1 in the UK, it highlights declining trend from the child population. This projects an increased burden on the working population to come as the number entering the working population from childhood is outweighed by the number entering old age.

Brazil and the UK both face similar issues due to increasingly elderly populations. In Brazil this responsibility is expected to be taken up by the older individual and their family (dos Santos and do Rosário de Fátima e Silva, 2013), on the other hand, in the UK the responsibility is expected of the Government in the form of public health and social services. The advantages of these contrasting approaches may therefore be highlighted with a single-minded aim.

Firstly, it is important be able to identify the need for care and ensure efficient practices can meet their full potential to improve and sustain the health and well-being of the elderly. Aimed at extending healthy life expectancy and promoting quality of life, the (WHO, 2014) defines active ageing as:

> "the process of optimizing opportunities for health, participation and security in order to enhance quality of life as people age".

where active refers to physical, social and economic activity. From this definition the focus of this thesis will be the obtaining of measurements for the relevant areas in order to assess active ageing in Brazil and England.

To outline the issues faced by the elderly this chapter will give a general overview, of the current understanding of the issues surrounding health and well-being, in a review of the existing literature. This is followed by a review of the policies of the Brazilian and UK governments that relate to the provision of care for the elderly. Then the current measures used to analyse well-being in the elderly, in terms of psychometric measures, are presented. Finally, previous applications of structural equation modelling in ageing research is reviewed.

2.1 Understanding of current issues affecting the elderly

There are a large number of morbidities that impact upon the lives of the elderly population, with greater prevalence and incidence than in the younger adult population. The severity of the comorbidities is also greater, in terms of symptoms and complications, due to comorbidity and polypharmacy. In this section we will discuss some of the key areas of health, including musculoskeletal disorders of the skeleton and muscular systems, metabolic disorders such as diabetes, cardiovascular disorders of the heart and circulatory system, respiratory diseases such as asthma, neurodegenerative disorders affecting cognitive function, and cancer, all of which have impact on the lives of the elderly. Comorbidities are a strong influence on the subjective well-being of the elderly, thus the resulting depression and loneliness are also discussed.

2.1.1 Musculoskeletal health

Musculoskeletal disorders, such as osteoporosis and osteoarthritis, are common in the elderly (Gheno et al., 2012) with chronic musculoskeletal disorders the most prevalent group of diseases in Brazilians aged 60 or more years old. Some have estimated around 80% of those aged 65 or over have a musculoskeletal disorder (Smith et al., 2016), and, in some regions of Brazil, 86%

suffer from musculoskeletal pain (Instituto Brasileiro de Geografia e Estatística (IBGE), 2010; Miranda et al., 2012). The most common disorder is osteoarthritis affecting around 70% of people aged 65 or more years old (Kee and Epps, 2001), often resulting in reduced function and loss of independence. Women have greater risk of osteoarthritis (Salter, 2002; Zhang, 2002; Haq et al., 2003; van 't Land et al., 2010; Duncan et al., 2011) with 90% having radiographic osteoarthritis, aged 65 or more years old, compared to 80% in men of the same age (Gheno et al., 2012). Other risk factors include excessive mechanical stress, metabolic disorder and genetics (Salter, 2002).

Osteoporosis is another common musculoskeletal disorder with a prevalence of 21.2% in women and 6.3% in men between the ages of 50-84 years old (Kanis et al., 2000; Woolf and Pfleger, 2003). The risk of osteoporosis also increases with age with a prevalence of 50% in women and 20% in men at the age of 85 years old (Gheno et al., 2012). The impact of osteoporosis mainly arises from low bone density which leads to increased risk of fractures, including spontaneous fracture without any associated trauma. Worldwide, in 2000, there were an estimated 9 million osteoporotic fractures in people aged 50 and over, of which 34% occurred in Europe (Johnell and Kanis, 2006). Fractures can also arise as a result of a fall that may occur due to impaired mobility or vision, medication or environmental factors. Fractures with the greatest impact occur in the hip, pelvis, vertebral or hand, resulting in the inability to walk and perform either basic or instrumental activities of daily living, alongside the increased risk of further fractures due to immobility exacerbating the osteoporosis (Holroyd et al., 2008). Mortality has also been associated to fractures of the hip, pelvis and vertebral (Chang et al., 2004; Holroyd et al., 2008). In addition, complications in the healing of fractures are more common in the elderly with either a longer healing period called delayed union (Nikolaou et al., 2009) or non-union with a lack of radiographic improvement (Egol et al., 1997; Lu et al., 2005; Gruber et al., 2006; Foulke et al., 2016). This leads to longer periods of immobility and can require surgery including amputation.

The elderly more commonly have musculoskeletal infections such as acute hematogenous osteomyelitis and tuberculosis than younger adults. Although tuberculosis is rare in the UK, Brazil is one of five countries (Brazil, the Russian Federation, India, China and South Africa) that account for 50% of the cases in the world. In 2015, there were over 5,000 notified cases of tuberculosis in those aged 65 or more years old (World Health Organization [WHO], 2016). In 50% of cases the infection spreads from the lungs to the spine, through the blood stream, (Moore and Rafii, 2001; Gheno et al., 2012). Signs of tuberculosis can easily be missed with subtle manifestation or symptoms that mimic other infections (Gheno et al., 2012).

Pain is a common aspect of musculoskeletal disorders, in particular lower back pain, that can be mechanical or idiopathic in nature (Jones et al., 2014). Low back pain can impact on elderly individuals by lowering their perceived independence and ability to participate in activities of daily living (Smith et al., 2016), and result in physical disability, disturbed sleep, psychosocial disruption and increased healthcare usage (Weiner et al., 2006).

2.1.2 Metabolic health

Musculoskeletal disorders have a strong link with metabolic disorders. Diabetes has been found to be associated with increased risk of osteoarthritis, risk of healing complications for fractures and the susceptibility to infection (Gheno et al., 2012; Smith et al., 2016). Likewise, disturbance of purine metabolism is associated with the development of microcrystal disorders, such as gout and calcium pyrophosphate dihydrate deposition which are more common in the elderly (Gheno et al., 2012).

Diabetes is associated with the increased risk of a wide range of medical conditions. Prevalence is estimated to be around 20-38% of the elderly, aged 65 or more years old, depending on the diagnostic criteria used (Kirkman et al., 2012; Tuomilehto et al., 1986), and is higher for those of Native American, Hispanic, black or Micronesian ethnicity (Sánchez Martínez et al., 2014). Compared to other age groups with diabetes, the elderly have a higher risk of major lower-extremity amputation, heart attack, visual impairment and end-stage renal disease (Kirkman et al., 2012). Diabetes is also associated with increased risk of geriatric syndromes including incontinence, falls, vision and hearing impairment, functional disability, cognitive impairment and dementia, frailty and functional impairment along with other medical conditions such as cardiovascular disease, cerebrovascular disease (stroke), peripheral vascular disease (blocked arteries near the heart), microvascular complications (such as retinopathy, nephropathy, neuropathy) and depression (Kirkman et al., 2012; Corriere et al., 2013).

Another prevalent condition in the elderly, similar to diabetes, is metabolic syndrome (He et al., 2006; Maggi et al., 2006). No universal definition of metabolic syndrome exists, making estimations of epidemiology difficult, however, it is commonly accepted that metabolic syndrome is a cluster of heart disease risk factors related to insulin resistance (DeFronzo and Ferrannini, 1991; Lindblad, 2001; Beilby, 2004; He et al., 2006; Lechleitner, 2008; Huang, 2009; Dontsov and Vasilyeva, 2013). Risk factors that are consistently considered in the elderly include abdominal obesity, hypertension, elevated low-density lipoprotein (LDL) cholesterol and low levels of high-density lipoprotein (HDC) cholesterol and insulin resistance (Lindblad, 2001; Alexander et al., 2003; He et al., 2006; Dontsov and Vasilyeva, 2013). Despite the different criteria, prevalence in the elderly was estimated to be 42% in those 70 or more years old (Alexander et al., 2003) and 36.8% in those 60 or more years old (He et al., 2006), with greater prevalence in women and in diabetics. The elderly with metabolic syndrome are at significantly greater risk of diabetes, stroke, chronic heart disease, and peripheral arterial disease, and also mortality in elderly men (He et al., 2006; Maggi et al., 2006).

2.1.3 Cardiovascular health

Cardiovascular disease is the leading cause of death in the world (World Health Organization [WHO], 2013) including coronary heart disease, cerebrovascular disease (stroke) and peripheral arterial disease. The elderly aged 65 or more years old, experience greater incidence, functional limitation and mortality from cardiovascular disease with increasing age (Ahto et al., 1998; Mehta et al., 2001; Hollander et al., 2003; Gabriel et al., 2009), and account for more than 85% of coronary heart disease deaths (Gabriel et al., 2009). Myocardial infarction (heart attack) has a higher incidence rate, 7-9%, and mortality rate in men aged more than 64 years old than women with incidence rate between 3-5% (Ahto et al., 1998; Gabriel et al., 2009). Following an event of myocardial infarction, the elderly also suffer more cardiovascular complications such as hypertension, shock, atrial fibrillation, heart failure and stroke (Mehta et al., 2001). Chronic heart disease, including myocardial infarction, affects approximately 38% of men and 42% of women aged 64 or more years old (Ahto et al., 1998). Cerebrovascular disease has increasing incidence rate for older ages with, for men, 1.7 cases per 1000 person-years aged 55-59 years old and 69.8 for persons aged 95 or more years old. Similarly in women, aged 55-59 years old the incidence rate was 1.2 per 1000 person-years and 33.1 for those aged 95 or more years old. Mortality is the result of 32.5% cases of stroke, for ages 55 or more years old, of which 12.4% from cerebral infarction (blocked blood vessel depriving part of the brain of oxygen) and 33.3% from cerebral haemorrhage (blood onto the brain) (Hollander et al., 2003). Stroke can result in multiple limitations to an elderly person's functional ability, and commonly results in depression and cognitive deficits (Koenig and Studenski, 1988; Coletta and Murphy, 1994). Communication is affected by slurred or slow speech (dysarthria) and receptive and expressive language deficits (aphasia), and even difficulty swallowing (dysphagia) or loss of the gag reflex. Physical function can be reduced by weakness of the muscles (paresis), disorder of purposeful movement (apraxia), loss of balance and trunk control and loss of visual-spatial or visual-motor abilities (Coletta and Murphy, 1994).

2.1.4 Respiratory health

Common respiratory diseases in the elderly include asthma, chronic obstructive pulmonary disease (COPD) and pneumonia. There are four types of pneumonia; community acquired, hospital acquired, nursing home acquired and ventilator acquired. Community acquired pneumonia has an estimated incidence rate of 25-44 cases per 1000 persons aged 65 or more years old (Janssens and Krause, 2004; Stupka et al., 2009), however the incidence increases with age such that the estimated incidence rate for those aged 65-69 years old is 18 cases per 100 persons compared to 52 cases per persons aged 85 or more years old (Jackson et al., 2004; Stupka et al., 2009; Cacciatore et al., 2017). Similarly, the incidence rate of hospitalisation from community acquired pneumonia is 8.5 per 1000 persons aged 65-69 years old and 48.5 per 1000 persons aged 90 or more years old, while mortality within those hospitalised is 7.8% aged 65-69 years old and 15.4% aged 90 or more years old (Kaplan et al., 2002). However, signs and symptoms of pneumonia in the elderly may be subtle, with common symptoms such as a cough, fever and dyspnea (shortness of breath) being uncommon in the elderly (Cacciatore et al., 2017), while common signs in the elderly include falls, sudden decline in functional status, lower appetite, urinary incontinence, delirium or confusion, and the worsening of pre-existing conditions (Faverio et al., 2014; Simonetti et al., 2014; Cacciatore et al., 2017). Many age-related diseases are associated with increased risk of pneumonia including cardiovascular diseases, cerebrovascular diseases, malignancy, Parkinson's disease and dementia, HIV infection and chronic renal or liver disease, as well as increased risk to those with dysphagia (difficulty swallowing) or are immunosuppressed (Stupka et al., 2009; Torres et al., 2013; Dang et al., 2015). Besides, pneumonia can often be the terminal event of comorbidities such as diabetes, COPD, heart failure, malignancies and dementia (Cacciatore et al., 2017). The biggest risk factor of COPD is smoking, which has been linked to a third of community acquired pneumonia cases in the elderly (Kaplan et al., 2002; Stupka et al., 2009).

COPD affects around 14% of the elderly, aged 65 or more years old (Halbert et al., 2006), resulting in loss of general health and functional status and increase in frailty and disability (Crisafulli et al., 2014). Again, the main risk factor is smoke inhalation commonly from tobacco smoking, but additionally in Brazil from exposure to wood or coal fuelled stoves (Halbert et al., 2006; Akinbami and Liu, 2011; Chang et al., 2012; McKay et al., 2012; Bentayeb et al., 2013; Taffet et al., 2014; Battaglia et al., 2015; Barbosa et al., 2017). COPD is also commonly associated with comorbidities such as hypertension, diabetes, coronary artery disease, congestive heart failure and arrhythmias (Battaglia et al., 2015). However, symptoms of COPD have been shown to be reduced by exercise, nutritional support and education (Crisafulli et al., 2014).

Asthma in the elderly is often misdiagnosed as COPD resulting in mistreatment and undertreatment (Bellia et al., 2003; Stupka and DeShazo, 2009; Scichilone et al., 2014; Yáñez et al., 2014). This is because the elderly with asthma may have presentations similar to those with COPD, such as irreversible, or partial reversible, airway obstruction and chronic bronchial inflammation (Reed, 2010; Scichilone et al., 2014), in addition to a lack of recognition of atypical asthma symptoms such as chest tightness with bronchospasm and night-time bronchoconstriction by the elderly (Scichilone et al., 2014). Due to immunosenescence (age-related changes to the immune system) the elderly with asthma also lack allergen sensitivity (Scichilone et al., 2014). The estimated prevalence of diagnosed asthma in the elderly is between 4% and 13% (Kitch et al., 2000; Reed, 2010; Blackwell et al., 2014; Scichilone et al., 2014; Yáñez et al., 2014; Abrantes, 2015; Zein and Erzurum, 2015), while a further 27% are predicted to be undiagnosed (Blackwell et al., 2014; Zein and Erzurum, 2015). However, late on-set asthma (after age 65 years old) and longstanding asthma are greater in severity, more progressive and less reversible (Reed, 2010; Scichilone et al., 2014), resulting in more hospitalisations and mortality than in younger asthmatics (Yáñez et al., 2014). Risk factors of late onset asthma are obesity and pollution, occupational and tobacco exposure (Zein and Erzurum, 2015).

2.1.5 Neurodegenerative health

Dementia, Alzheimer's disease and Parkinson's disease are neurodegenerative disorders that occur as a result of ageing (Farooqui and Farooqui, 2009) and become increasingly prevalent with advancing age (de Rijk et al., 1995; Kukull et al., 2002; Sabbagh et al., 2007; Parkinson's UK, 2009; Hindle, 2010; Prince et al., 2014). Dementia is the progressive decline in cognitive function that reduces an elderly person's ability to perform activities of daily life due to brain damage or disease (Haaksma et al., 2017) that affects approximately 7% of people aged 65 or more years old (Nitrini et al., 2009; Prince et al., 2014). Dementia consists of various types including Alzheimer's disease which accounts for 62% of those with dementia while 2% have Parkinson's dementia (Prince et al., 2014). Parkinson's dementia is a common manifestation of Parkinson's disease (Sabbagh et al., 2007). Alzheimer's disease is a type of dementia that affects around 0.3% of those aged 65-69 years old and 5.6% of those aged 90 or more years old (Kukull et al., 2002). There are multiple stages of progression in cognitive decline in learning and recall memory, fluency of language, apraxia (ability to do particular, purposeful, movements) and executive function, as well as stages of functional decline (Sabbagh et al., 2007). Deterioration was also associated with comorbidity (Haaksma et al., 2017). Parkinson's disease is also a progressive neurodegenerative disorder. Prevalence has been estimated at around 0.6% in those aged 60-79 years old and up to 2% in those aged 80 or more years old (Parkinson's UK, 2009). Those affected have symptoms such as a tremor at rest, rigidity, bradykinesia (slow movement), impaired postural control, gait difficulty, and have increased risk of falls (Hung et al., 2010); while those with Parkinson's dementia additionally have treatment related confusion, psychosis, severe motor symptoms and depression (Sabbagh et al., 2007). Neuropathy (numbness and tingly of the extremities) is, also, more prevalent in those with Parkinson's disease (Rajabally and Martey, 2011). Dementia was associated with increased risk of mortality, with an estimated hazard ratio of 5.14 in those aged 65 or more years old, compared to those without dementia (Nitrini et al., 2005).

2.1.6 Cancer

Those aged 65 or more years old account for 65.2% of cancer incidences in England (Office for National Statistics, 2015), however these individuals are often excluded from participation in clinical trials of treatments. Elderly patients are often excluded from trials of treatment efficacy due to age, comorbid conditions or cognitive function, leading to the registration of drugs that have only marginally been tested on the greatest population that will receive it (Marosi and Köller, 2016). The effects of these treatments on the elderly remains relatively unknown with suboptimal doses often given due to reduced organ function, that alters the pharmacokinetics such as drug absorption, distribution, metabolism and excretion (Given and Given, 2008). Examples given by Marosi and Köller (2016) include the decreased production of digestive enzymes for absorption of treatments taken orally, increased toxicity due to prolonged metabolism and reduced renal function affecting excretion of the drug.

A comprehensive geriatric assessment (CGA) assesses a patient in multiple domains, such as cognitive performance, mobility, emotional health and comorbidity to aid the decisions about the treatment plan; primarily the role of palliative care. For some elderly patients palliative care is preferable to treatment. This is discussed by Marosi and Köller (2016). Elderly cancer patients may have reduced physical function and functional reserve needed to manage toxic side effects of chemotherapy, or access to the psychological, emotion and social support required for daily administrations such as with radiation. Impaired cognitive function can leave some elderly patients unable to follow complex treatment plans or recognise and communicate toxicity and other complications. However, some older patients have been shown to respond well to treatment if they are in good health or administered reduced dosages.

2.1.7 Depression

Frequently comorbid with the physical conditions is depression, of which around 25% of the elderly with comorbidity had major depression (Cole and Dendukuri, 2003; Blazer, 2003). Generally, in the elderly clinically significant depression affects around 8-16% and 1-4% have major depression (NIH Consensus Development Conference, 1992; Blazer, 2003; Casey, 2012). Depression is not associated with increasing age (Blazer, 2003), however is associated with age-related events such as bereavement, disability, cognitive impairment (Cole and Dendukuri, 2003; Blazer, 2003; Blazer and Hybels, 2005; Vink et al., 2008) and comorbid conditions including cardiovascular health, stroke, neurodegenerative disorders, respiratory diseases and chronic pain (Almeida, 2014). Other risk factors include social stress, early life adversities and harmful lifestyle (Almeida, 2014).

The core symptoms of depression are a depressed mood or lack of interest, but in addition depressive symptoms include feelings of worthlessness or inappropriate guilt, lacking in concentration, fatigue, agitation, inability to sleep or sleeping too much, significant change in appetite or weight and thoughts of death or suicide (Blazer, 2003). Some of these symptoms, however, crossover with symptoms of many comorbidities in the elderly and depression in the elderly is under-recognised and under-treated (Vink et al., 2008). Alternatively, depression can be diagnosed using the Center for Epidemiologic Studies Depression Scale (CES-D, Radloff (1977)), where a summative score of 11-15 indicates sub-threshold depression and a score of 16 or over indicates depression (Blazer, 2003).

2.1.8 Loneliness

Loneliness is a problem faced by many elderly people with 7-10% in the UK reporting they often feel lonely, and 30% displaying some signs of loneliness (Andersson, 1998; Luanaigh and Lawlor, 2008; Crewdson, 2016). Loneliness can be experienced as either emotional or social isolation. Emotional loneliness derives from a lack of an attachment figure that shares a deeper connection, such as a spouse or close friend, whereas social loneliness stems from a lack of belonging and acceptance within a community (Andersson, 1998; Crewdson, 2016). Men, for the most part, reported less loneliness than women (Andersson, 1998; Luanaigh and Lawlor, 2008; Nummela et al., 2011), however Dahlberg and McKee (2014) found being male to be a risk factor of social loneliness. Other risk factors of loneliness include bereavement or widowhood, non-married marital status, low income or income discomfort, low activity levels, and low self-esteem

(Andersson, 1998; Luanaigh and Lawlor, 2008; Iecovich and Biderman, 2012; Crewdson, 2016), although the direction of association between loneliness and income discomfort, low activity levels or low self-esteem are yet undetermined. Increasing age is frequently associated with greater levels of loneliness (Holmén et al., 1992; Dugan and Kivett, 1994; Fees et al., 1999; Andersson, 1998; Drennan et al., 2008; Iecovich and Biderman, 2012) with increasing need for intervention to provide social support, group-based support and therapies (Andersson, 1998; Landeiro et al., 2017).

In addition, loneliness can impact physical, mental and cognitive health as a mediating factor (Luanaigh and Lawlor, 2008; Crewdson, 2016). Inflammation has been found to be a physical response to loneliness and stress, causing decline in cognitive function (Boss et al., 2015) and reducing the effectiveness of the immune system. Besides, loneliness is a risk factor of several morbidities including cardiovascular diseases such as hypertension, myocardial infarction and stroke, and neurodegenerative diseases such as dementia and Alzheimer's disease (Holmén et al., 2000; Crewdson, 2016; Landeiro et al., 2017). Physical health is also impacted via salubrious behaviours which loneliness affects by removing social motivation for physical activity, increasing the risk of alcoholism, drug dependency and problems with sleep (Crewdson, 2016). In addition, loneliness has strong associations with anxiety (Fees et al., 1999) and depression (Alpass and Neville, 2003; Luanaigh and Lawlor, 2008; Aylaz et al., 2012; Crewdson, 2016; Landeiro et al., 2017), with interrelationships between the three aspects of mental health. Hence, loneliness remains an important factor in the care of elderly that can be the result of, and a contributing factor to, many health problems.

2.2 Policies for the Elderly

The UK and Brazilian governments have differing approaches to the legislation of provisions for the care of the elderly. In Brazil, there are specific laws for the elderly that have each contributed to the wider frame of legislation; for example, the National Policy of the Elderly (1994) and (1996) and the National Health Policy for the Elderly (1999). By comparison, legislations in the UK are often replaced by new ones and the focus is all adults in need of care and support, which includes the elderly. In this next section we will outline the historical progression of legislation in Brazil, and the current legislation in the UK that applies to the elderly.

2.2.1 Brazil

Policy on the elderly is advancing in Brazil with much focus on providing the elderly with rights and citizenship, as well as providing health care that helps them to maintain and improve their autonomy and independence. Responsibility for the elderly lies with the family, community and the State, but as discussed by dos Santos and do Rosário de Fátima e Silva (2013) and Fernandes and Soares (2012), the State fails to meet its responsibilities in the practice.

The first landmark in Brazil's policy for the elderly was in 1974 when a life-long monthly income was guaranteed by Law No. 6,179 (1974), through the National Institute of Social Welfare (INPS, Instituto Nacional de Previdência Social in Portuguese). This was aimed at the elderly, aged 70 or more years old, that were either unable to work, providing them with an income that was proportional to the minimum wage, at the time of payment, but no more than the minimum wage. In 1975 rural workers were also guaranteed a pension, and a retirement age of 65 years or older, in Law No. 6,620 (1975), thus providing social security to the elderly living in rural areas.

In 1977 the National Social Welfare and Assistance System was established, in Law No. 6,439 (1977), to unify social security assistance, ensuring financial protection for the elderly in the funding of pensions.

The Single Health System was then created in 1988 to ensure a universal health service.

The Federal Constitution in 1988 stated it was the duty of the family, society and the state to assist the elderly, ensuring their participation in the community, defending their dignity and well-being, and guaranteeing them the right to life. Preference was given to care provided in their own homes. The constitution also guaranteed free urban public transport to those aged 65 or older.

The social assistance, organised by Law No. 8,742 (1993), provided the elderly with the equal rights and citizenship, and to treatment without discrimination. A minimum monthly pension was set for the elderly, which ensured financial security for those aged 70 years or more that were unable to gain maintenance for themselves or from their family.

The first policy specifically for the elderly was the National Policy of the Elderly (1994) which changed the definition of elderly to those aged 60 or older, and created the National Council for the Elderly (o Conselho Nacional do Idoso, in Portuguese). The aim was to create conditions that would promote the autonomy, integration and effective participation in society of the elderly. This law extended the notion of the Federal Constitution (1988) stating that it is the responsibility of the family, the society and the state to ensure the elderly's rights of citizenship, participation in the community, and to defend their dignity, well-being and right to life.

Law No. 8,842 (1994) provided guidelines and actions to promote the social assistance and health of the elderly, and the improvement in knowledge and training in areas of geriatrics and gerontology, which was then extended in the National Policy for the Elderly (1996). This decree gave specific methods in how care should be given to those in need of it, such as; living centres, where the elderly stay during the day where recreational activities took place; day care centres, where the elderly stay during the day for medical assistance and care; house-home care, where the elderly have residence in participative systems; sheltered workshop work, where the elderly can earn a small income doing productive activities; and home care, where the elderly are visited at home by a healthcare professional or a member of the community.

The National Policy for the Elderly (1996) stated that the Ministry of Health, through the Department of Health Care, are responsible for guaranteeing the provision of preventative and curative services within various levels of the SUS, ensuring access to hospital care, the provision of drugs, orthotics and prosthetics that are required for recovery and rehabilitation. Along with providing health programs that encourage the elderly to remain in the community with self-care and informal care.

The National Health Policy of the Elderly (1999) set out the essential guidelines for the healthcare of the elderly, which consisted of; the promotion of healthy ageing; the maintenance of functional capacity; the assistance to meet the health needs of the elderly; rehabilitation of compromised functional capacity; the training of specialised human resources; support for the development of informal care; and support for studies and research. The aims and guidelines for each of these areas are now outlined.

The promotion of healthy ageing set about guiding the elderly in the improvement of their functional ability, and the prevention of compromising their functional ability. Improvements included the adoption of healthy life-style habits such as healthier diet and physical exercise and the elimination of habits that are harmful to health such as smoking and alcoholism, which were promoted via leaflets, radio, TV and community health agents. The prevention of falls, which could compromise functional ability, consisted of the removal of environment risks with the installation of access ramps and railings.

Maintenance of functional capacity was conducted via two stages to prevent the loss of functional ability. The first stage was to prevent health problems by such measures as vaccinations for tetanus, pneumococcal pneumonia and influenza, which are recommended by the World Health Organisation (WHO). The second stage involves the early detection of potential health problems, these encompass actions to detect non-communicable diseases, such as diabetes and hypertension, early, along with; early screening for hearing, visual and balance impairment; use of the protocols in situations common in the elderly, such as falls, mood changes and cognitive loss; dental and oral loss prevention; prevention of nutritional deficiencies; assessment of functional skills to prevent the loss of autonomy or independence; and creation of social opportunities, such as clubs, to prevent social isolation.

The assistance to the health needs of the elderly provides guidelines based on outpatient, hospital and home care. The guidelines for outpatient care focussed on the use of the geriatric consultation and therapeutic orientation. The geriatric consultation was aimed at the collection of data and recording of information, in order to identify problems through routine screening and the diagnosis of diseases, in particular diseases that are uncommonly diagnosed, such as thyroid diseases, Parkinson's disease, dementia and hypertension.

Therapeutic orientation involved the tackling of the problems identified in the geriatric consultation by informing patients on how to prevent health problems and corrective measures to encourage rehabilitation through making lifestyle changes.

Guidance for the assistance that is based in hospital includes the participation of doctors and nurses, as well as other professions such as physiotherapists, social workers, psychologists, occupational therapists, speech therapists, dentists and nutritionists. The guidance also highlights the needs of the totally dependent elderly and those with serious health problems who have no possibility of recovery.

Home-care is described to be a cheaper alternative to prolonged hospitalisation for those who are expected to have a lengthy stay in hospital, with even those requiring intravenous medication or hydration not being justifying the stay in hospital. However, an intermediate care in the form of a geriatric day-hospital is suggested for those in need of therapy.

Rehabilitation of compromised functional capacity involved the recovery from the loss of functional capacity and the prevention of further loss, through the practice of multi-professional work in medicine, nursing, physiotherapy, occupational therapy, nutrition, speech therapy, psychology and social service. The rehabilitation is aimed at reversing and preventing functional limitations as well as to reduce the amount of dependency; therefore, appropriate treatment should be given to the elderly.

The training of specialised human resources is set out by guidelines for the structure of responsibilities for the government departments involved, for example the Ministry of Education is responsible for ensuring that training is provided in higher education institutions, and the Collaborating Centres of Geriatrics and Gerontology should be established by the Ministry of Health. It is these centres that are responsible for the support of study and research.

The studies and research focus on four key areas; the profiling of the elderly in different regions health and social factors; evaluation of functional capacity and disease prevention; models of care including the implementation, monitoring and evaluation of interventions; and hospitalisations and alternatives to hospital care to improve efficiency and reduce costs.

Support for the development of informal care provides the guidelines of supporting the caregiver at home, usually female, through the means of providing health care for the carers and a partnership between the caregivers and health professionals.

The Elderly Statute was established in 2003 to regulate the rights of the elderly, who were those aged 60 and over. This statute specifies that it is the responsibility of the family, community, society and public authorities to ensure the elderly are given the right to life, health, food, education, culture, sports, leisure, work, citizenship, freedom, dignity, respect and family and community life. It defends the elderly by defining negligence, discrimination, violence, cruelty or oppression, and every violation of the rights of the elderly to be punishable by law. Overall, this statute guarantees the elderly the right to life, freedom, dignity and respect.

The Public Policy of Social Assistance (2004) was aimed at those in need of social protection such as the elderly. The policy set out the objectives to provide social protection through services and projects, help with inclusion and equity by expanding the access to goods and services and assurance that family and community life are central in social assistance.

In 2006 the National Elderly Health Policy was then reviewed and updated by Ordinance No. 2,528 (2006). The guidelines from the policy were revised to include; the promotion of active and healthy aging; comprehensive, integrated health of the elderly; encouragement of intersectoral actions aimed at comprehensiveness of care; provision of resources capable of ensuring quality of health care for the elderly; encouraging participation and strengthening of social control; training and continuing education of health professionals in the SUS health care of the elderly; information about the National Health Policy for the Elderly health professionals, managers and users of SUS; promote national and international cooperation of experience in health care of the elderly; and support the development of studies and research.

Ordinance No. 2,528 (2006) expanded the focus from the functional capacity to include the autonomy of the elderly. There were also additions that considered the rights and freedoms of the elderly, as well as combating discrimination and violence against the elderly. The existing preventative health care was recognised to be inefficient on an individual basis, therefore a collective functional assessment was introduced, considering the proportion of elderly that are unable to perform activities of daily living and those that are classed as frail.

2.2.2 England

Policy from the UK Government has a less direct approach to the elderly than the Brazilian Government. In the UK, the elderly are encompassed within policies surrounding health and social care services for all adults who are in need of care and support, which also includes adults such as the disabled. Recent legislation affecting the elderly are the Care Act (2014) and the Pensions Act (2014).

The majority of the current elderly population were previously subject to the Default Retirement Age at the age of 65 years old. This allowed for compulsory retirement to be pushed onto the elderly by employers. However, given the increase in life expectancy, this was fully replaced by the State Pension age in 2011, which is the age at which a person has access to the state pension. One of the key points in the Pensions Act (2014) was the increase in the state pension age to 67 years old (in 2026-28), with reviews set to ensure that on average persons will reach the pensionable age with the expectation of living a specified proportion of their life in retirement. Thus, enabling the elderly to maintain employment in society and ensuring financial support.

The Care Act (2014) set out the duties of the local authorities in relation to the care and support of adults, as well as support for carers. The main focus of the legislation was the promotion of individual well-being, which includes the promotion of: personal dignity, physical, mental and emotional health, protection from abuse and neglect, control of one's own day-to-day life, participation in work, education and training, social and economic well-being, family and personal relationships, suitable accommodation, and contribution to society. This ensures local authorities are responsible for the provision of care for the elderly, which can be achieved by either arranging for a person to provide the service required, provide the service themselves or by making direct payments to the individual.

Another key aspect of the Care Act (2014) was the integration of health and health-related (including housing) provisions. This was to promote well-being, prevention or delay of the development of the need for care and support, and the quality of care and support for adults.

A recent project by the Government was the Future of an Ageing Population (Government Office for Science, 2016), which collected evidence reviews of the current elderly population and future projections in order to develop informed policies for an ageing population. These policies and actions are aimed at maintaining well-being throughout life, improving quality of life for older people and enabling full participation in society. Key findings in the report by the Government Office for Science (2016) include the need to address issues of varying employment rates of older workers, negative attitudes to the needs of older workers in the workplace, barriers to life-long and late life training and education particularly in technological skills, support families with more living generations increasing pressure to care for dependants and provide suitable housing within neighbourhoods that meet the demand for transport, socialisation and access to services.

2.3 Review of Elderly Specific Measures

To be able to accurately formulate policies that will obtain their purpose it is important to have an understanding of the interrelationships of the many complex health and well-being problems discussed above. Many of these problems are exacerbated in the elderly by age-related complications, and the impact of such problems can vary greatly with individual factors such as resilience and reserve. For this reason, it is important to consider subjective aspects such as perception of health, loneliness, depression and life satisfaction. To ascertain estimates for these subjective aspects psychometric measures are required. Cognitive decline and ill health can make psychometric measures, that are usually well performing in younger age groups, unreliable in the elderly. Below we discuss the development and provision of popular measures for self-rated health, loneliness, depression and life satisfaction that have been validated for use with elderly samples.

2.3.1 Self-rated Health

Self-rated health (SRH) has been vastly studied as a sensitive measure of overall health in the elderly (aged 60 or more years old). In a review of SRH in elderly adults, Ocampo (2010) discussed the multiple factors that influenced SRH which included demographic variables, socioeconomic status, geographical location, and health related variables.

In support of SRH as a measurement of health of the elderly LaRue et al. (1979) and Giltay et al.

(2011) studied the association between SRH and the rating of health from a physician, where a significant correlation between the two ratings was discovered. LaRue et al. (1979) also studied the relationship of SRH with five-year survival, determining a significant association for elderly adults aged 77-84 years old.

Mortality has been one of the key areas of research concerning SRH, a number of studies have found an ability to predict mortality from SRH (Shen et al., 2014; Elder, 2013; Menec et al., 1999). Idler and Benyamini (1997) conducted a review of 27 studies and DeSalvo et al. (2006) conducted a systematic review of 22 studies into mortality and SRH, both concluding SRH to be an independent predictor of mortality in the elderly; even when controlling for other known indicators of mortality such as morbidity, cognitive function and disability.

More recently the effect of demographic and socioeconomic factors on the prediction of mortality has been a key focus of the literature; for example, Helmer et al. (1999) discovered a difference in the prediction of mortality between elderly men and women; for elderly men poor SRH was a predictor of five-year mortality, however for elderly women the relationship was explained by physical health and disability.

As an alternative Vogelsang (2014) expanded the single measurement of SRH to the change in SRH. The results showed a greater risk of mortality for those with a change in SRH, positive or negative, than those with consistent SRH. Additionally, Menec et al. (1999) extended the predictive abilities of SRH, in the elderly, to morbidity and self-perceived control. SRH was predictive for the loss in perceived control; however, morbidity was not predicted by SRH.

The identification of factors that significantly impact the SRH in the elderly in general has been popular in the literature (Craigs et al., 2014; Dujardin et al., 2014; Chiavegatto Filho et al., 2012; Borim et al., 2012; Arnadottir et al., 2011; Brenes-Camacho, 2011; Nummela et al., 2011; Alves and Rodrigues, 2005); as well as those that specifically influence the prediction of mortality (Helmer et al., 1999; Franks et al., 2003; Lima-Costa et al., 2012; Nishi et al., 2012; Nummela et al., 2012; Zheng and Thomas, 2013). These included Chiavegatto Filho et al. (2012) and Brenes-Camacho (2011) that both assessed the effect of income on the SRH of the elderly; their results showed higher income as positively associated with SRH and greater inequality of income to be associated with poor health ratings. Nummela et al. (2011) found loneliness to be a significant contributor to poor SRH of the elderly, whereas Dujardin et al. (2014) discovered only a weak association between self-rated and environmental factors, in the elderly. There are many different combinations of socioeconomic and health related factors that have been considered from which education, income, living alone, physical activity and functional capacity were consistently found to be important (Borim et al., 2012; Alves and Rodrigues, 2005; Arnadottir et al., 2011). In addition, Borim et al. (2012) observed that for those without religion, who had a computer at home, drank 1-4 times per month compared with none or 2-3 times per week, participated in physical activity, those not obese or those who consumed fruit or vegetables at least 4 times a week, the prevalence of excellent/very good SRH was higher. Whereas, Alves and Rodrigues (2005) showed marital status and the number of chronic diseases were important to the prevalence of poor health. Specific aspects of function were analysed by Arnadottir et al. (2011) of which depressive symptoms, activities that involved lower extremities and higher levels of exertion, household based physical activities, were highlighted as the important factors. There were various different factors shown to be important to the SRH of the elderly, however, only characteristics of the individual have been considered; we aim to expand into characteristics of the household and wider areas.

In addition, much of the previous research into the SRH of the elderly dichotomised the Likert scale variable, often to represent good or bad health, which was highlighted in a review by DeSalvo et al. (2006). For example; Alves and Rodrigues (2005) had good health as excellent, very good and good, and poor health as regular and bad responses; Dujardin et al. (2014) classed very bad, bad and fair defined as poor health with good and very good defined as good health; Camargos et al. (2009) considered regular, bad and very bad defined as poor health with very good and good defined as good health; and Nummela et al. (2011) had average, rather poor and poor defined as poor health and good and rather good defined as good health. Alternatively, Chiavegatto Filho et al. (2012) analysed the prevalence of poor responses out of a scale of excellent, very good, good, fair and poor; similarly, Borim et al. (2012) analysed the prevalence of excellent and very good responses from a scale of excellent, very good, good, bad and very bad. In most of these scales regular was an ambiguous category in the scale which could be interpreted as either bad or good health, although in the examples above regular was classed as poor health. Conversely, Arnadottir et al. (2011) and Brenes-Camacho (2011) used SRH as an ordinal scale, removing the question of whether to define regular as good or poor health.

2.3.2 Depression

Generally, depression is measured in the elderly using either the Geriatric Depression Scale (Yesavage et al., 1982) or the Center for Epidemiology Studies Depression Scale (CES-D, Radloff (1977)). Depression scales developed for younger populations can often be misleading in the elderly with many somatic symptoms, such as sleep disturbance, cognitive impairment, aches and pains, being common ailments of ageing and age-related disease. The original Geriatric Depression Scale consisted of 30 items (Yesavage et al., 1982) and a shorter version of 15 items (Sheikh and Yesavage, 1986), with yes or no responses, covering topics such as self-image, cognitive complaints and motivation. The CES-D is a self-reported scale of depression for the general population, that originally consisted of 20 items with a 4-point scale of responses from ‘rarely or none of the time’ to ‘most or all the time’. The items contained four factors; depressed affect, positive affect, somatic and retarded activity, and interpersonal. The larger number of items and response scale proved problematic when used in the elderly (Turvey et al., 1999; Karim et al., 2014). A shortened eight version was proposed by Turvey et al. (1999) using items from depressed affect, positive affect and somatic factors of the original scale. The eight-item version has been found to be more better for use in the elderly than the original scale Karim et al. (2014). In addition to reducing the number of items, the four-point scale was reduced to simpler yes or no responses.

2.3.3 Loneliness

The most commonly used measure of loneliness is the UCLA Loneliness Scale, developed as a unidimensional scale with 20 items. Each item required a response the four-point scale; often, sometimes, rarely and never. The original scale was proposed in Russell et al. (1978), where all the items were positively worded, however this was changed in the revised UCLA Loneliness Scale (Russell et al., 1980) so that the items were equally split between satisfaction with social relationships and dissatisfaction. A shorter four-item version was also suggested that was equally balanced between the positively and negatively worded items. The revised UCLA Loneliness Scale contained items with double negatives that were confusing to the elderly, aged 65 or older, (Russell, 1996). The third version (Russell, 1996) altered the items to all begin with the phrase “How often do you feel...” and validated in samples of students, nurses, teachers and the elderly.

The equal divide between positive and negative items is emphasised by (Russell, 1996) in the UCLA and in other loneliness scales, such as the Social and Emotional Loneliness Scale (DiTommaso and Spinner, 1993). (Russell, 1996) discussed the presence of two method factors found in exploratory and confirmatory factor analysis, as well as a single overall loneliness factor.

DiTommaso and Spinner (1993) proposed a multidimensional measure of loneliness, the Social and Emotional Loneliness Scale, as alternative to the unidimensional approach of the UCLA Loneliness Scale that is often criticised (Russell, 1996; DiTommaso and Spinner, 1993). DiTommaso and Spinner (1993) suggests the existence of two types of loneliness, social and emotional, where emotional loneliness can be further broken down into family and romantic dimensions. The original Social and Emotional Loneliness Scale consisted of 37 items; 11 items of the family subscale, 12 items of the romantic subscale, and 14 items of the social subscale. Each item was responded to on a 7-point Likert type scale from 'strongly disagree' to 'strongly agree' (DiTommaso and Spinner, 1993). A shortened version was then proposed in order to be of similar length as the UCLA Loneliness Scale and allowed for greater efficiency for clinical and research settings (DiTommaso et al., 2004). A subset of 15 items from the original scale are used in the shortened version, with equal balance of 5 items per subscale; family, romantic and social. Neither versions (DiTommaso and Spinner, 1993; DiTommaso et al., 2004) were specifically validated for use on a sample of the elderly.

2.3.4 Life Satisfaction

Life satisfaction is frequently measured using the Satisfaction With Life Scale (Diener et al., 1985), in particular as an aspect of subjective well-being in combination with the Positive and Negative Effects Scale (Watson et al., 1988; Tian, 2016; Wang, 2014; Carmel et al., 2017). The Satisfaction With Life Scale is a popular five item measure that was developed and validated by Diener et al. (1985); including specific validation for an elderly sample. Responses to each item are on a seven-point scale from 'strongly disagree' to 'strongly agree'. An extension of the Temporal Satisfaction With Life Scale was developed by Pavot et al. (1998) as a temporal measure of satisfaction with life, which the scale contained the wording of the SWLS in the three tenses past, present and future. In a review, conducted by the authors of SWLS, of the research into life satisfaction and the use of SWLS found strong evidence of test-retest reliability in the measure and strong temporal stability (Pavot and Diener, 2008).

2.3.5 Quality of Life

The CASP-19 is a multidimensional measure for quality of life specifically in the elderly. CASP stands for the four dimensions control, autonomy, self-realisation and pleasure of the scale that was developed by Higgs et al. (2003) and Hyde et al. (2003), which additionally has been translated and validated in Brazilian-Portuguese (Lima et al., 2014). The 19 items are evenly distributed among the dimensions, with only the first four items belonging to the control dimension. Responses are given on a four-point Likert type scale to a series of statements, with the preceding question "How often do you feel like this?".

Concerns about the proposed factor structure for the CASP-19, however, appear frequently within the literature. Sexton et al. (2013) conducted a psychometric analysis but re-modelled the data, excluding items 6 and 12, considering the eigenvalues and factor loadings from an exploratory factor analysis. Whereas, Sim et al. (2011) concluded a close fit to their data in a confirmatory factor analysis. Wiggins et al. (2008) evaluated their original factor structure of the CASP-19 and concluded that from confirmatory factor analysis "the relative weakness of the CFI and RMSEA criteria suggest that as far as model fit is concerned we could do better!"; in reference to the fit of to data from the ELSA Wave 1 and British Household Panel Survey Wave 11. In addition, (Wiggins et al., 2008) proposed a reduced form CASP-12 measure that removed poor performing items, however model fit in the subsequent confirmatory factor analysis was still poor.

2.4 Previous Methods in Analysis of the Elderly

Commonly used methods for analysing interrelationships between subjective measures are structural equation models (SEMs). This section will now review some of the previous research that have applied SEMs in ageing research, with specific discussion of the few that apply multilevel and longitudinal SEMs.

2.4.1 Structural Equation Modelling

Subjective well-being and self-rated health (also known as subjective health) are the popular focuses for the use of structural equation models in the research of the elderly. Subjective well-being is a multifaceted concept that is primarily a person's assessment on their experience of life, where satisfaction with life is a dominant aspect (Tran et al., 1991; Tian, 2016; Carmel et al., 2017). Other aspects include morale and depression (Carmel et al., 2017), happiness and goal attainment (Tran et al., 1991), and alongside life satisfaction the positive and negative effects suggested by Diener (2000) that were used by Tian (2016); Wang (2014). The purpose of SEM with subjective well-being has been to model the direct, indirect and mediating effects of an array of factors including, but not limited to: health status, stress and psychological resources (Tran et al., 1991), size of the social network (Wang, 2014), intergeneration social support (Tian, 2016), and personal resources and coping mechanisms (Carmel et al., 2017)and in studies that focused only on life satisfaction: psychological well-being (Meléndez et al., 2009), social factors, functional disability, risk behaviours and chronic disease experience (Hirve et al., 2014), elderly services and living environments (Yan et al., 2014), and chronic illness, depression and physical functioning (de Guzman et al., 2015a). Related to subjective well-being, Delhom et al. (2017) analysed the relationships between emotional intelligence, coping and depressed mood. The purpose of SEM analysis with self-rated health as the focus was similar, Hirve et al. (2014) also analysed the effect of social factors, functional disability, risk behaviours and chronic disease experience on self-rated health, while Golini and Egidi (2016) similarly analysed the effect of chronic morbidity, functional abilities and emotional health. The mediating effect of economic factors on the impact of widowhood on self-rated health was analysed in a SEM by Pandey and Jha (2012).

Other applications of SEM in elderly research had a medical focus. Masumori et al. (2005) investigated the distribution of lower urinary tract symptoms in elderly Japanese men, Lee et al. (2015) predicted health-related quality of life in patients with Parkinson's disease, Johansson et al. (2014) explored the mediating effect of inflammation on impaired cardiac function and a symptom cluster, and most relevantly, Bo et al. (2009) evaluated the existence of metabolic syndrome as a single factor influencing the metabolic components considered in the National Cholesterol Education Program Adult Treatment Panel III; and failed to find significant evidence of such.

In the studies mentioned above, representation of a general elderly population is limited with relatively small samples, ranging from 125 to 1,216 in size and focussed in small areas within China, Italy, India, Spain, Israel and the Philippines. Exception was Pandey and Jha (2012) with a large nationally representative sample of elderly individuals ($n = 29,542$) from India's National Household Sample Survey.

Summative scores and indexes were common features in SEM of the elderly, summarising psychometric measures with multiple items into a single observed variable (Tran et al., 1991; Meléndez et al., 2009; Wang, 2014; de Guzman et al., 2015b,a; Kimura et al., 2015; Liu et al., 2016; Tian, 2016; Carmel et al., 2017; Delhom et al., 2017). This approach attempts to maintain the multivariate normal assumption of the distribution of the observed variables, but with the compromise assumption that every item within the measure is influenced equally by the latent construct, and each item measures the construct with equal accuracy. Retaining the items as separate variables maintains the ability to identify the specific aspects of a construct that are influential or lacking in performance.

2.4.2 Longitudinal analysis

Latent growth modelling has only recently been applied to research in the elderly, analysing trajectories of health outcomes such as cognitive decline, disability, and mental and chronic health (Muniz-Terrera et al., 2009; Yang et al., 2016; Chen et al., 2017; Pongiglione et al., 2017). Although it is also a key part in growth mixture modelling that extends the latent growth model by categorising different trajectories with an overall categorical latent variable. This method has been used to identify patterns in health outcomes such as in functional ageing (Han et al., 2013) and distinguish between patterns of persistent and transient anxiety or depression (Rzewuska et al., 2015). Chen et al. (2017) found that pre-existing chronic and mental health problems significantly contributed to the development of further chronic and mental health problems respectively. Although this was only found in a joint multivariate latent growth model with parallel processes for chronic health and mental health, while the univariate growth models had poor goodness of fit. Muniz-Terrera et al. (2009) and Yang et al. (2016) on the other had looked at the impact of socioeconomic status on the trajectory of cognitive decline. Yang et al. (2016) showed higher socioeconomic status to only be associated with better cognitive function at baseline and not the decline, with and without controlling for health behaviours such as smoking and alcohol consumption. Conversely, Muniz-Terrera et al. (2009) found education and social class to be associated with the decline and rate of change in cognitive function as well as baseline cognitive function in an age-based quadratic trajectory. Thus, the application of latent growth modelling has allowed for the identification of trends in the health of the elderly, and determinants of greater decline.

2.4.3 Multilevel Structural Equation Modelling

Applications of multilevel structural equation models are scarce. The only application in the elderly found in the literature search was Wilson-Genderson and Pruchno (2015). The aim of this study was to examine effects of individual and neighbourhood characteristics that are associated with functional limitations that might explain the differences in functional status between men and women. Neighbourhood factors considered were availability of physicians, crime, supermarkets, fast-food establishments and store fronts (grocers, convenience stores, bar and pubs), of which fast-food establishments were significant to greater functional limitation in both genders and store fronts was only significant in women. There is, however, need for more research into the environmental influences on the health and well-being of elderly individuals.

Summary of the Chapter

In brief, the literature review has given insight into the vast network of interconnected issues faced in the care of the elderly. Health problems have been shown to be more prevalent and severe in the elderly, with complications unique to the elderly population. These health problems each come with increased risk of functional decline that can impact upon an elderly person's well-being and ability to maintain independence. Understanding these issues is key to the development of efficient policy that can sustain an ageing population.

Chapter 3

Summary of Data

This chapter will introduce Brazil's National Household Sample Survey (PNAD, *Pesquisa Nacional por Amostra de Domicílios* in Portuguese) and the English Longitudinal Study of Ageing (ELSA). The surveys are each introduced with a brief overview, which is followed by an introduction to the variables of interest to the analysis contained within this thesis. In the overview of the PNAD, the complex sampling plan and hierarchical structure will be discussed. The last section will introduce the repeated variables of the ELSA for the longitudinal analysis.

3.1 Pesquisa Nacional por Amostra de Domicílios (National Household Sample Survey)

The National Household Sample Survey of Brazil, (known as *Pesquisa Nacional por Amostra de Domicílios* in Portuguese, PNAD) is an annual cross-sectional survey that has been conducted in Brazil since 1967, by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística (IBGE), 2008). The survey is conducted every year except in the years ending in -0, when the national demographic census takes place.

The PNAD aims to investigate the general characteristics of the population in terms of, for example, demographics, education, income and housing. The main sections of the survey were repeated every year with supplementary sections that are included at different frequencies. The supplementary sections contain detailed characteristics in specific areas such as health, fertility, smoking and food security. The health supplement, that was of particular interest in this thesis, was repeated every five years; in the years ending in -3 or -8. However, in 2013 the National Health Survey (Pesquisa Nacional de Saúde, PNS) was introduced to replace this health supplement.

The survey conducted in the 2008 PNAD was of particular interest due to the health supplement. This provided additional information, for example activities of daily living, the importance of which to ageing research is highlighted in the previous chapter (Chapter 2). Also, for the first time, both urban and rural areas of all major regions were sampled, providing nationally representative data for the population of Brazil.

3.1.1 Complex Sampling Plan

Brazil is the union of 27 states, including the Federal District, that are within five major regions, as shown in Figure 3.1. Each state is divided into municipalities and metropolitan areas; which are the administrative subdivisions. The number of municipalities within a state ranged from 15 in Roraima to 853 in Minas Gerais. Each of the municipalities was then divided into census tracts, that are referred to as sectors in this thesis due to only approximate identification being given in the PNAD; for confidentiality purposes.

In 2008 the sampling plan of the PNAD was complex, containing stratification and multiple stages of cluster sampling with probabilities proportional to size. Systematic sampling was used at each stage. Municipalities and sectors were ordered by urban-rural classification, except in the North region in which sectors and households were stratified by urban-rural classification before sampling.

At the first stage of the cluster sampling the municipalities were stratified by their status as a self-representative municipality. Self-representative municipalities either had a large population or were a metropolitan region. Therefore, each self-representative municipality defined a stratum and was sampled with certain probability. Non self-representative municipalities were stratified according to location and formed smaller regional groups of municipalities. Within each stratum of non self-representative municipalities one or two municipalities were sampled with probability proportional to size; with the number of residents at the last census measuring size.

Sectors were selected at the next stage in the sampling plan from within the sampled municipalities. This was the primary sampling unit for the self-representative municipalities. Sectors were sampled with probability proportional to size, measured by the number of households. A register of households is kept by the local authority of the sector.

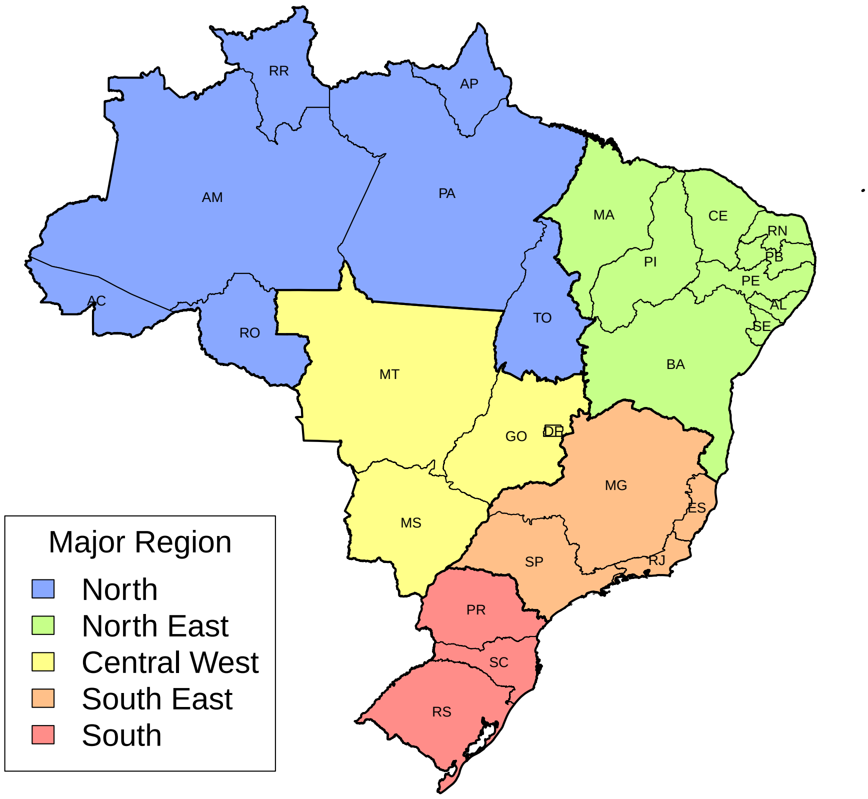


Figure 3.1: Map of the major regions in Brazil.

List of State Codes:

AC	Acre	PB	Paraíba
AL	Alagoas	PE	Pernambuco
AM	Amazonas	PI	Piauí
AP	Amapá	PR	Paraná
BA	Bahia	RJ	Rio de Janeiro
CE	Ceará	RN	Rio Grande do Norte
DF	Federal District	RO	Rondônia
ES	Espírito Santo	RR	Roraima
GO	Goiás	RS	Rio Grande do Sul
MA	Maranhão	SC	Santa Catarina
MG	Minas Gerais	SE	Sergipe
MS	Mato Grosso do Sul	SP	São Paulo
MT	Mato Grosso	TO	Tocantins
PA	Pará		

In the last stage of cluster sampling, households were sampled from within each selected sector, with equal probability. For sectors within self-representative municipalities this was fixed at 16 households per sectors, and within non self-representative municipalities the number of households differed according to the size of the sector.

At the final stage all individuals within the selected households completed the survey.

Additionally, the PNAD contained a supplementary sample of households containing the newly constructed residencies that were not on the sector's register of households at the time when the sample was drawn. New construction households were sampled separately from within each selected sector. This could be considered equivalent to stratification of the households within each sector according to the new construction status.

The complex sampling plan imposed a hierarchical structure within the data. The structure comprised of individuals residing within households, households within sectors, sectors within municipalities; and then non self-representative municipalities within strata.

Sampling weights are typically provided by surveys with complex sampling, alongside identifiers for the units in clustered sampling. Unfortunately, not all the required identifiers were given in the 2008 PNAD. For confidentiality purposes, only approximate identifiers for the sectors were available. A general weight for the households was available, however despite direct discussions and continued communications with IBGE the derivation of these weights remained unclear. Thus, it was uncertain whether these weights were specifically the sampling weights for the correction of the unequal sampling probabilities.

3.1.2 The Sample and Relevant Variables

In this next section, variables of interest are introduced that were used in the analysis presented in later chapters. The areas of particular interest were the demographics, socioeconomic characteristics and health characteristics of both the individual and the household.

The 2008 PNAD sample contained 424,321 individuals, of which 9.73% were elderly; aged 60 or more years old (following the definition of old age discussed in Chapter 2).

The survey was sometimes completed by proxy, with another person responding on the behalf of an individual. For example, a reference person was selected to represent each household, who on occasion completed the survey on the behalf of all other members of the household.

It was decided to select only elderly individuals that responded to the health section for themselves. Several of these variables referred to the individual's subjective perception of their health and the reporting of the diagnosis of morbidities, which may be inaccurate or biased; depending on the intimacy of information shared by the elderly individual with the proxy responder.

This resulted in a sample of 29,400 elderly individuals, residing in 24,766 households of 7,135 sectors and 817 municipalities. The representation of the national elderly population and distribution of these individuals in terms of the major regions was 9.64% in the Central West, 17.14% in the South, 33.29% in the South East, 30.72% in the North East and 9.20% in the North; and 81.76% lived in urban sectors.

The selected elderly individuals consisted of 31.62% aged 60-64 years old, 26.15% aged 65-69 years old, 18.78% aged 70-74 years old, 12.95% aged 75-79 years old, 6.74% aged 80-84 years old 2.72% aged 85-89 years old and 1.03% aged 90 or more years old. Further, 60.26% were female and 51.60% were of white race. Race was aggregated to a binary variable that indicated white or non-white race. The non-white category was a conglomeration of mixed race (39.49%), black race (7.74%), Asian (0.78%) and indigenous (0.39%).

The distribution of the age groups was approximately representative of the national elderly population. The population estimates given by United Nations: Department of Economic and Social Affairs. Population Division (2013) show the distribution of the national elderly population in 2008, in terms of the same age groups, was 31.37%, 23.71%, 18.19%, 12.28%, 7.99% 3.97% and 2.50%, respectively. There were, however, small differences that indicate the younger age categories were over represented in the sample, and the older age categories under represented. This may be the nature of self-reported data, as self-reporting becomes increasingly difficult with increased age and the consequential reduced health.

Health variables

The health supplement covered many areas, including some of particular interest in ageing research. These areas were the perception of general health, self-reported morbidity, activities of daily living, physical activity, smoking status and access to health services.

Self-rated health (SRH) was a single item worded as "as a general rule, considers their own state of health as..." (translated from the Portuguese *"De um modo geral, considera seu próprio estado de saúde como"*) with the given responses very bad, bad, regular, good and very good as responses. The majority of responses were either regular or good, as shown in Figure 3.2, with regular as the most frequent response, 42.43%, and only 2.64% having responded with very bad.

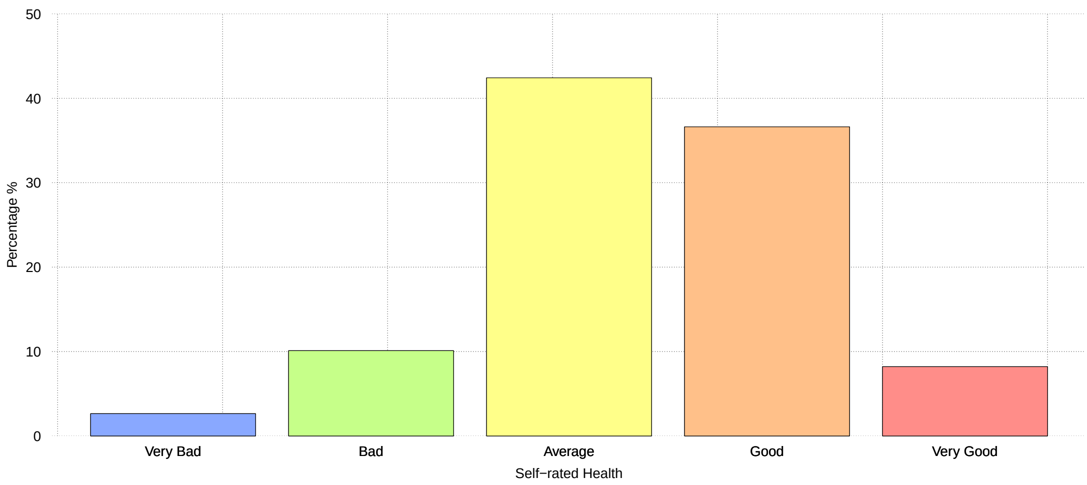


Figure 3.2: Distribution of self-rated health category responses in the 2008 PNAD.

Self-reported morbidity was represented by twelve variables. In the items, individuals were asked if "a doctor or health professional told them that they have [...]" (*"Algum médico ou profissional de saúde disse que tem* [...]*"*, in Portuguese) a comorbidity or one of a type of comorbidities. These comprised of; disease of the column or back, arthritis or rheumatism, cancer, diabetes, asthma or bronchitis, hypertension, heart disease, chronic renal failure, depression, tuberculosis, tendonitis or tenosynovitis, and cirrhosis.

Across Brazil the most prevalent comorbidities in the sample were hypertension, disease of the column or back and arthritis or rheumatism, as shown in Table 3.1, while cirrhosis and tuberculosis both had prevalence less than 0.5%. For the majority of comorbidities, the prevalence was noticeably lower in the North and North East regions, except for arthritis or rheumatism, tuberculosis and cirrhosis. The prevalence of arthritis was contrarily higher in the northern regions.

A summary variable to indicate multi-morbidity was created that categorised individuals according to the number of comorbidities they reported to be diagnosed with. Depression was considered as a separate variable to give indication of mental health. In Figure 3.3, it can be seen that the vast majority of the sample was diagnosed with at least one comorbidities, and 286 (1.15%) individuals reporting six or more comorbidities. To obtain a variable with sufficient frequencies in each category, the responses were conglomerated into the categories;

Comorbidity	Prevalence (%)					
	Total	South East	South	Central West	North	North East
Disease of the back or column	38.00	37.36	40.44	41.92	35.97	36.71
Arthritis or rheumatism	26.20	24.28	27.00	25.05	34.09	25.84
Cancer	2.16	2.51	3.43	1.94	1.40	1.36
Diabetes	15.39	17.06	16.03	16.02	13.90	13.47
Asthma or bronchitis	5.96	6.77	8.37	6.35	4.92	3.92
Hypertension	54.13	57.27	54.52	54.66	48.50	52.02
Heart disease	17.13	18.38	22.38	20.71	13.90	12.68
Chronic renal failure	3.42	3.59	4.17	4.73	3.29	2.45
Depression	9.12	10.34	14.60	9.03	4.21	6.25
Tuberculosis	0.30	0.35	0.26	0.21	0.22	0.33
Tendonitis or tenosynovitis	5.46	6.77	8.67	4.59	2.22	3.49
Cirrhosis	0.31	0.33	0.38	0.18	0.18	0.33

Table 3.1: Total and regional prevalence of self-reported morbidity in the 2008 PNAD.

no comorbidities, 1 comorbidity, 2 or 3 comorbidities, 4 or 5 comorbidities and 6 or more comorbidities.

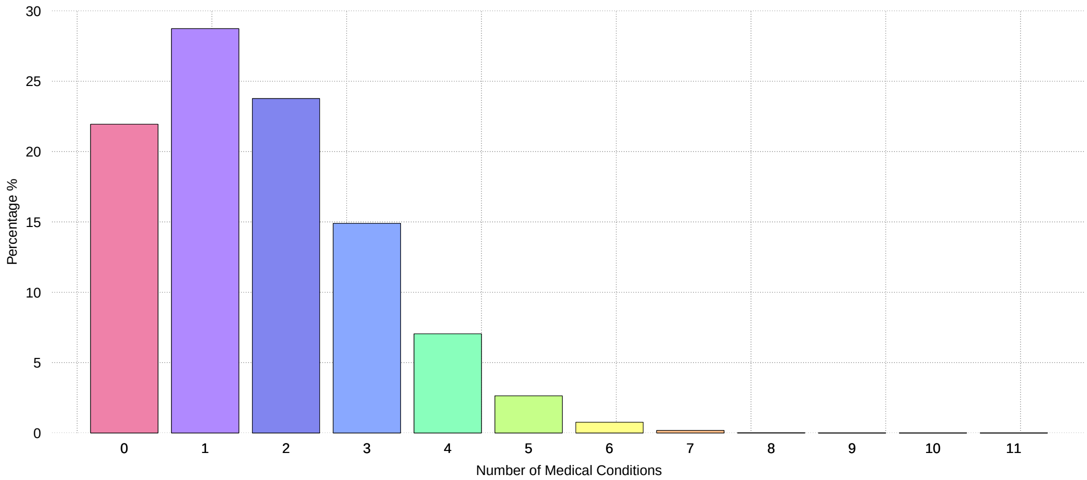


Figure 3.3: Distribution of multi-morbidity in the 2008 PNAD.

The impact of general health and morbidity on the individual's functional status was available in the form of eight items regarding difficulty performing activities of daily living. Each question asked "Normally, due to health problems, has difficulty with [...]" followed by a set of activities (*"Normalmente, por problema de saúde, tem dificuldade para* [...]*"* in Portuguese). These activities were (1) eating, bathing or using the bathroom, (2) running, lifting heavy objects, playing sports or doing heavy jobs, (3) pushing tables or performing home repairs, (4) climbing a hill or stairs, (5) stooping, kneeling or bending, (6) walking more than one kilometre, (7) walking about 100 metres and (8) making purchases of food, clothing and medicines, without help. To each item the responses were; can not, has great difficulty, has little difficulty, has no difficulty, to indicate the amount of the difficulty the individual had performing the activities listed in the item. However, not all individuals responded to all the items, with progression through the items determined by responses to items one and six.

An illustration of the decision process is given in Figure 3.4. All individuals respond to the first item (eating, bathing or using the bathroom), then if the individual can not perform these tasks or had great difficulty performing them they were not eligible for the later items (2-8). It was assumed that these individuals had lower functional status than the individuals who were eligible. Further, individuals who had no difficulty performing item 6 (walking more than one kilometre), were not applicable for the last two items (7 and 8). It was assumed that these individuals had better functional status than the individuals that were eligible.

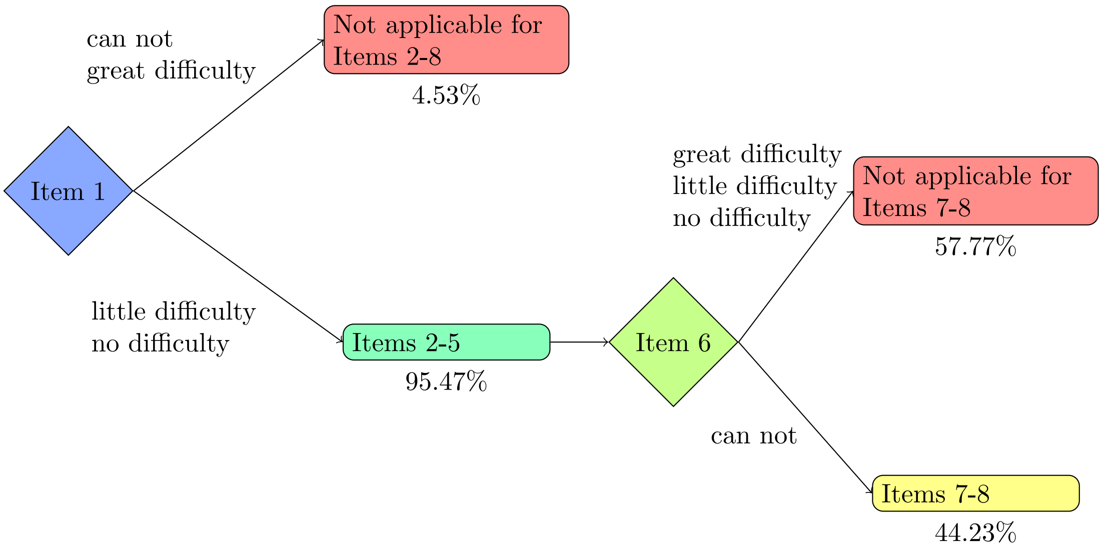


Figure 3.4: Decision diagram for responses to the activities of daily living items, with values from the 2008 PNAD.

Firstly, functional status was considered through a single aggregated variable rather than as separate independent observed variables. The aggregated variable gave indication of the general level of limitation in performing activities of daily living, categorising individuals according to their responses to all eight items; using the same methods as de Moraes et al. (2011). There were four categories; great limitation, some limitation, little limitation and no limitation. Firstly, individuals were categorised as great limitation if their response was either great difficulty or can not to item 1, or a response of can not was given to any of the later items (2-8). The individuals in the following categories all had little or no difficulty with the first item. Individuals were then allocated to the some limitation category if they had at least great difficulty for any of items 2-8. The remaining individuals then only responded to items 2-8. The little limitation was allocated to those with a little difficulty with any of items 2-6 and no difficulty was allocated to those that had no difficulty with any of the items. The resulting variable then consisted of 27.33% without limitation, 34.31% with little limitation 22.93% with some limitation and 15.43% with great limitation.

The eligibility criteria creating missingness that affected items 2-8, which was not completely at random. In total, 59.68% older individuals had missing values due to the eligibility criteria. Firstly, missingness on items seven and eight was present for the majority (55.15%) of individuals, thus these items could not be considered in the SEM; where the variables were considered as separate dependent variables. For the 4.53% of individuals that had missingness conditional on item one it was known that the level of difficulty performing the basic activities, eating, bathing and toileting, was greater than the difficulty incurred by the eligible individuals. It was, therefore, assumed that difficulty performing the activities listed in items 2-8 was also greater. A fifth category was added to items 2-8 and allocated as the response for these individuals, to ensure those with the lowest functional status were represented, and not excluded from the analysis. This category was placed at the end of the ordinal scale that indicated lower functional status. However, measurement error was not included with this approach and had the potential to introduce bias.

Lifestyle habits of physical activity and smoking status were considered as covariates. Physical activity was a single item that asked, "In the last three months, practised some type of physical exercise or sport" (*"Nos três últimos meses, praticou algum tipo de exercício físico ou esporte"*, in Portuguese) with 72.41% not participating and 20.01% participating. The further 7.57% were not eligible for this particular item, conditional on the activities of daily living aforementioned. Eligibility was conditional on items 1 and 8, with the condition on item 1 the same as before. Those that had great difficulty or can not as their response to item 1 were also not eligible. In addition, those with some level of difficulty with the activities in item 8 were not eligible.

There were two items related to the smoking status of the individual that were aggregated to create a single categorical variable indicating an individual's smoking status as either a smoker, ex-smoker or non-smoker. The first item asked "currently, smokes some product of tobacco [...]" (*"Atualmente, fuma algum produto do tabaco[...]"*, in Portuguese), and the second asked "in the past, has smoked some product of tobacco [...]" (*"No passado, fumou algum produto do tabaco[...]"*, in Portuguese). The responses were either daily, less than daily or don't smoke. An additional response of don't know was available but was not used by the selected elderly individuals. All individuals that currently smoke either daily or less than daily were classed as smokers (15.03%), individuals that currently don't smoke but in the past smoked daily or less than daily were classed as ex-smokers (34.51%), and individuals that said they don't smoke currently, nor smoked in the past, were classed as non-smokers (50.46%). There were 54 individuals with missing data on this variable.

Other health variables included as covariates were; possession of a health plan (yes, 28.42%; no, 71.58%), obtained medical consultation within the last 12 months (yes, 81.28%; no, 18.72%), participation in the family health program discussed in Chapter 2 (yes, 50.34%; no, 49.66%).

Economic variables

The household section of the PNAD contained several variables concerning the economic well-being including variables on the quality of the housing, variables on the durables owned within the household, and other items concerning the socioeconomic status of the individual or household. The quality of the housing variables covered, in terms of the household, included the construction materials of the walls and roof, the method of lighting and the number of occupants per room, and in terms of the infrastructure in the sector, included the source for the water supply, management system for sewage and the disposal of household waste.

The housing quality variables were each reduced to binary variables that indicated whether the element was adequate or inadequate in accordance to definitions in the standards of living survey (*Pesquisa sobre Padrões de Vidas* in Portuguese, Instituto Brasileiro de Geografia e Estatística (IBGE) (1999)). A summary of these variables is given in Table 3.2, showing the adequate or inadequate classification of the categories from the original variable and the percentage of individuals with each classification. Note, the management system for sewage was dependent on the location of the toilet, as well as the system used.

Table 3.2: Summary of the variable aggregation for the housing quality variables, in the 2008 PNAD.

	Adequate		Inadequate	
	Percent	Response	Percent	Response
Wall construction material	98.23%	Masonry; or Fitted wood.	1.77%	Uncoated rammed earth; Seized wood; or Straw; or Other material.
Roof construction material	99.07%	Tile; Concrete slab; Fitted wood; or Zinc.	0.93%	Seized wood; Straw; or Other material.
Water supply	91.72%	General Network with internal plumbing; or Well or spring with internal plumbing.	8.28%	Another origin with internal plumbing; or General Network without internal plumbing; or Well or spring without internal plumbing; or Another origin without internal plumbing.

Table 3.2 *Continued on next page.*

Continued from previous page.

	Adequate		Inadequate	
	Percent	Response	Percent	Response
Sewage management	70.46%	Toilet located within the household; and Network pick-up or sewer systems; or Septic tank connected to network pick-up or sewer systems; or Septic tank not connected to network pick-up or sewer systems.	29.54%	Toilet not located within the household; or Rudimentary cesspools; or Ditch; or Straight to the river, lake or sea; or Another way.
Rubbish disposal	84.40%	Collected directly; or Collected indirectly.	15.60%	Burned or buried on the property; or Played in wasteland or public space; or Played on a river, lake or sea; or Another destination.
Illumination	98.22%	Electrical (network, generator, solar).	1.78%	Oil, kerosene or bottled gas; or Another way.
Occupation rate #residents ⁄ #rooms*	99.98%	≤ 2	0.02%	> 2

*including bathrooms

Table 3.2

An overall indicator for the adequacy of the household quality was aggregated from these seven binary variables for each household. The housing quality was classed as adequate if all seven areas were adequate, which was 65.83% of the elderly individuals, whereas 34.17% lived in a household where at least one area was classed as inadequate and so the household quality was classed as inadequate.

The durables owned within the household that were of interest included; stove, fridge, television (black and white or colour), radio, landline telephone, water dispenser (an appliance that provides safe drinking water from tap water), freezer, washing machine and computer (with or without the internet). Percentages of ownership are shown in Table 3.3.

	Owned (%)	Not owned (%)
Stove	99.32	0.68
Fridge	92.58	7.42
Television	93.47	6.53
Radio	88.51	11.49
Telephone	51.96	48.04
Water dispenser	55.38	44.62
Freezer	19.91	80.09
Washing machine	39.37	60.63
Computer	20.67	79.33

Table 3.3: Percentages for the ownership of the considered durables in the 2008 PNAD.

The durables were further aggregated into two binary variables; one indicating the number of essential goods (stove, fridge, water dispenser, freezer and washing machine) owned, and the other indicating the number of communication goods (television, radio, telephone and computer) owned. The number of essential goods owned was distributed with 0.38% of individuals living in household with none of the goods, 28.31% with one or two goods, 63.17% with three or four goods and finally 8.14% with all the essential goods considered. Similarly for the communication goods, 1.48% owned none of the goods, 28.34% owned one or two goods, 52.98% owned three or four goods and 17.20% owned all the considered goods.

Other socioeconomic variables available included the education and employment status of the individual, and the tenure of the household, urban-rural classification and household income per capita of the household. The urban-rural classification was the same for all households within the same sector.

Corresponding to the individual was the number of years of education and the status of employment. Education was a derived variable for the number of years of education the individual had received, which we further aggregated to correspond to no education, primary, secondary and further education. The resulting variable had the categories; less than one year (32.96%), 1-7 years (43.91%), 8-14 years (17.39%) and 15 or more years (5.73%). Current employment showed 30.26% were employed and 69.74% were not employed during the reference week (21st-27th

September 2008); where employment refers to general economic activity.

In relation to the household were the tenure of the household, the household income per capita and the urban-rural classification. Household tenure was the condition of occupation of the household which had the categories: owned (paid or still paying, 87.47%), rented (7.14%), granted (by the employer or another person, 4.94%) and other (0.45%); where other is thought to primarily refer to squatting. There were 48 individuals with missingness. Household income per capita was measured in relation to the minimum wage (R$441 per month, Trade Economics (2017)) such that 44.68% lived in a household with less than one minimum wage per capita, 28.40% with between one and two minimum wages, 9.37% with between two and three minimum wages, 6.79% with between three and five minimum wages and finally 7.06% had more than five minimum wages; although 3.69% did not declare the income.

The urban-rural classification was distributed with 81.76% of elderly individuals living within urban sectors and 18.24% lived in rural sectors.

It was possible to also derive the sample estimated old age dependency ratio from the number of individuals aged 60 or more years old divided by the number of individuals aged 18 to 59 years old. The ratio gives an indication on the ability of the sector to provide for the needs of the elderly in terms of finances, goods and services. A plot of the old age dependency ratios is given in Figure 3.5. The percentage on the vertical axis is given in terms of the number of individuals living within a sector with the respective ratio. It can be seen that the largest percentage of elderly individuals lived in sectors with around 20 elderly individuals per 100 individuals aged 18-59 years old. That is, the working age population provided for 120% of the number of working age individuals. In summary, the minimum was 4.24, the mean was 20.25 and the maximum was 55.75 elderly persons per 100 individuals aged 18-59 years old.

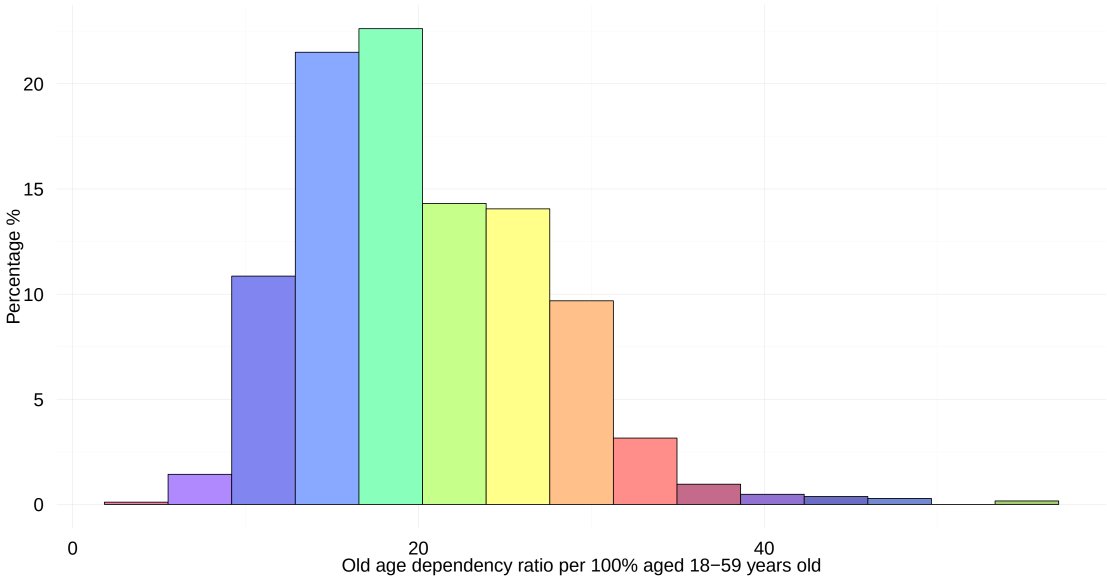


Figure 3.5: Histogram of old age dependency ratio, per 100 persons aged 18-59 years old, in the 2008 PNAD.

3.2 English Longitudinal Study of Ageing

The English Longitudinal Study of Ageing (ELSA) was designed to provide information on the health, socialisation, well-being and economic circumstances of the older adult population (aged 50 or more year old), in England (Marmot et al., 2016). The study began collecting data in 2002 repeating data collection every two years. Both objective and subjective data were collected through computer assisted interviews, self-completion questionnaires and nurses' visits.

The original cohort and the supplementary cohorts were sampled from the respondents to the Health Survey for England (HSE). The original cohort (Cohort One) was recruited from respondents to the HSE in 1998, 1999 and 2001. Supplementary samples were then drawn from HSE respondents in later years to boost the representation of the younger ages. In Wave 3 (2006-2007) the ages 50-52 years old were boosted with samples from 2001 to 2004 respondents of the HSE (Cohort Three), in Wave 4 (2008-2009) boosted ages 50-74 years old from the 2006 HSE and in Wave 6 (2012-2013) ages 50-55 years old were boosted from the 2009-2011 HSE (Cohort Four).

For the interviews, the respondent was visited in their own home by an interviewer who conducted the interview using a Computer-Assisted Personal Interviewing system. Both core members and their partners were interviewed, with partners only to provide additional information surrounding the core member. The interview including the sections on individual demographics, health, social participation and work and pension were conducted concurrently, whereas sections for cognitive function, expectations, psychosocial health, effort and reward and the final questions were conducted in private.

To supplement the data for each core member a self-completion questionnaire and a nurses' visit were given in Waves 2 and 4. These were mainly administered to only the core members but were also administered to partners if requested. The self-completion questionnaire contained questions on social participation, family and friends, and fruit and vegetable consumption, along with questions designed to measure well-being. The nurses' visit included measures of blood pressure, grip strength, a blood sample, standing and sitting height, weight, waist and hip measurement, lung function, balance, leg raise, chair rise and a saliva log.

A proxy survey was conducted if a person was unable to respond to the survey for medical reasons (cognitive impairment, physical or mental ill health) or being away for reasons such as temporary care or in hospital. This required for there to be someone who knew the respondent well enough to respond to the survey on their behalf. Obviously, some sections such as cognitive function could not be conducted in a proxy interview. The proxy informant had to be an adult of 16 years or older, and preferably a close relative such as a son or daughter.

3.2.1 Relevant Content

Data from the ELSA was analysed both cross-sectionally and longitudinally. Firstly, the variables are introduced from the more general cross-sectional subset of data and then the specific variables used in longitudinal analysis are extended. One aim of the cross-sectional analysis was to compare the results with that from the PNAD, therefore variables were selected to be as similar as possible between the two surveys.

In the cross-sectional analysis data from Wave 4 (2008-2009) of the survey was used to coincide with the 2008 PNAD data. Individuals that were core members aged 60 or more years old and gave a full or partial interview in person were selected for analysis. This provided a sample of 6,742 elderly individuals, made up of 5,503 from the original cohort and 1,239 from the Wave 3 supplementary sample; while only 43 individuals gave a partial interview.

The demographic variables considered were age, race and gender. The sample was predominantly white (97.85%) with only 145 non-white individuals and 7 with missing values. In terms of gender the sample had slightly more females with 55.24% and 44.76% were male, although a smaller majority than in the PNAD selection. Age was aggregated into groups of five years; 29.29% were 60-64 years old, 21.61% were 65-69 years old, 21.17% were 70-74 years old, 12.98% were 75-79 years old, 8.54% were 80-84 years old, and finally the oldest age group was aggregated into the remaining 6.41% aged 85 or more years old.

A single geographic variable, the Government Office Region, was available. These regions are illustrated in Figure 3.6. The selected elderly individuals were distributed across the nation with 6.33% in the North East, 11.98% in the North West 10.61% in Yorkshire and the Humber, 10.20% in the East Midlands, 11.00% in the West Midlands, 12.38% in the East, 8.64% in London, 16.96% in the South East and 11.89% in the South West, as well as 15 individuals that had missing values.

The following sections will outline and introduce the variables of interest. This covered the three areas of health, economic well-being and subjective well-being. Similarly to the PNAD, the health section covered topics of self-rated health, self-reported morbidity and functional status and the economic sections covered problems in the accommodation and the durables owned within the household. Additionally, the ELSA contained subjective data on individual's quality of life, loneliness, satisfaction with life and depression.

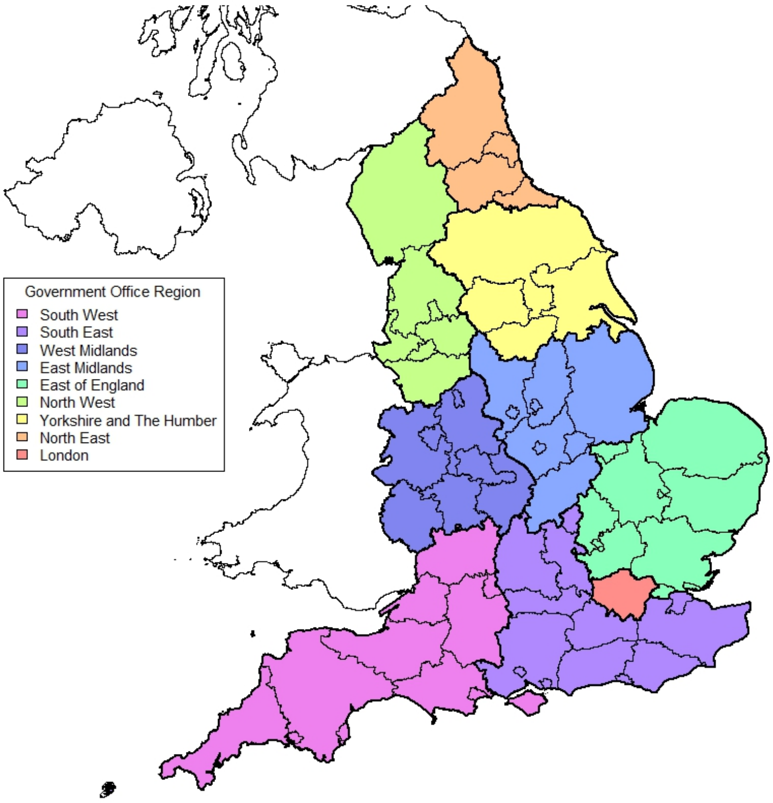


Figure 3.6: Map of the Government Office Regions in England.

Health variables

Health variables used in the analysis included two collections of self-reported morbidities (cardiovascular comorbidities and chronic comorbidities) and a collection of functional status variables with both basic and instrumental activities of daily living.

The cardiovascular conditions were shown to the respondent as a single list and asked, "Has a doctor ever told that you have any of the [heart] conditions on this card?". The conditions listed on the card were high blood pressure or hypertension, angina, a heart attack (including myocardial infarction or coronary thrombosis), congestive heart failure, a heart murmur, an abnormal heart rhythm, diabetes or high blood sugar, a stroke (cerebral vascular disease), and high cholesterol. The prevalence of each of these conditions is shown in Table 3.4. Due to the longitudinal nature of the survey, some of the responses were brought forward from previous Waves, unless disputed at the current Wave. After taking this into account, all conditions had 6 individuals with missing values except hypertension which had 4, diabetes had 22 and high cholesterol had 5 individuals with missing values.

Comorbidity	Prevalence (%)
Hypertension	47.49
Angina	10.96
Heart attack	6.32
Congestive heart failure	0.79
Heart murmur	5.66
Abnormal heart rhythm	10.82
Diabetes	10.77
Stroke	5.55
High cholesterol	38.34

Table 3.4: Prevalence of self-reported cardiovascular morbidities in the ELSA (Wave 4).

Similarly, for the chronic conditions the list contained: chronic lung disease such as; chronic bronchitis or emphysema; asthma; arthritis, osteoarthritis or rheumatism; osteoporosis or thin/brittle bones; cancer or malignant tumour, including minor skin cancers; Parkinson's disease; emotional, nervous or psychiatric problems; Alzheimer's disease; dementia, senility or serious memory impairment; and malignant blood disorder, including leukaemia or lymphoma. The prevalences are given in Table 3.5. The most frequent chronic comorbidity was arthritis (43.1%), followed by asthma (13.5%), while the prevalence of the other chronic conditions was less than 10% each. The prevalence of the neurodegenerative disorders (Parkinson's disease, Alzheimer's disease and dementia) was each very low, particularly Alzheimer's disease, therefore, as only 4 individuals were diagnosed with more than one of the three conditions, the variables were aggregated into a single binary variable indicating the presence of a neurodegenerative disorder 2.99%. The majority of the variables had 6 individuals with missing values, except asthma and arthritis with 5 individuals and dementia with 4 individuals with missing values.

Comorbidity	Prevalence (%)
Chronic lung disease	7.99
Asthma	13.54
Arthritis	43.08
Osteoporosis	9.75
Cancer	9.77
Parkinson's disease	0.95
Psychiatric problems	9.90
Alzheimer's disease	0.13
Dementia	0.88

Table 3.5: Prevalence of self-reported chronic morbidities in the ELSA (Wave 4).

The activities of daily living (ADLs) were also asked "Please tell me if any difficulty with these because of a physical, mental, emotional or memory problem. Again exclude any difficulties you expect to last less than three months. Because of a health or memory problem, have difficulty doing any of the activities on this card?" The activities on the card were;

1. Dressing, including putting on shoes and socks
2. Walking across a room
3. Bathing or showering
4. Eating, such as cutting up food
5. Getting in and out of bed
6. Using the toilet, including getting up or down
7. Using map to figure out how to get around strange place
9. Recognising when in physical danger
10. Preparing a hot meal
11. Shopping for groceries
12. Making telephone calls
13. Communication (speech, hearing or eyesight)
14. Taking medications
15. Doing work around the house or garden
16. Managing money, such as bills and expenses

Activities 1-6 were the basic ADLs whilst activities 7-15 were the instrumental ADLs. Respondents were not told this distinction. In Figure 3.7 the frequency of each activity is shown, which highlights dressing, bathing and doing house or garden work to be the three most frequent tasks the elderly respondents had difficulty performing; while recognising danger was the least frequent.

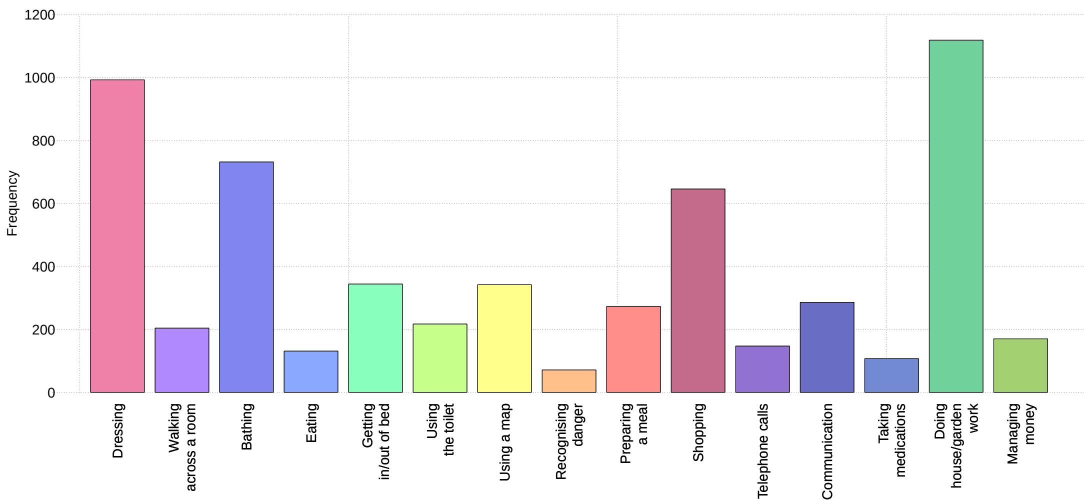


Figure 3.7: Frequency of difficulty with performing activities of daily living in the ELSA (Wave 4).

Economic variables

Economic variables in the ELSA included two collections of variables; one collection of problems with the accommodation and the other of durables owned.

Respondents were shown a list of accommodation problems and asked to identify which were present in their own household. Each problem was represented by a binary observed variable. The problems with accommodation in the list, and the percentage with the problem, were: shortage of space (7.08%), noise from neighbours (4.61%), other street noise (7.83%), too dark/not enough light (2.53%), pollution/grime (2.44%), rising damp (2.65%), water leaks from roof/gutters/windows (2.41%), bad condensation problems (1.99%), electrical or plumbing problems (1.21%), rot and decay (0.79%), insects, mice or rats (2.92%), and too cold in winter (4.20%).

Similarly, respondents were shown a list of durables that they may own within the household and asked which they had. The durables, with the percentage that owned the durable, were; television (87.02%), video recorder (72.21%), CD player (75.86%), deep freeze or fridge freezer (84.34%), washing machine (80.50%), tumble dryer/washer-dryer (45.07%), dishwasher (28.94%), microwave oven (79.26%), computer (48.26%), online-digital/satellite/cable TV (35.92%), telephone (landline) (84.76%), and DVD player (69.01%).

Variables of wealth were also available in the ELSA which will be used later for comparison purposes. Wealth was indicated by two variables; net financial wealth and total wealth. Net financial wealth was calculated as the difference between gross financial wealth and financial debt. Financial wealth was the total value of cash investments such as savings, cash ISA and shares *etc.*, while financial debt included credit card debt and private loans. A full break down is given in Figure 3.8 which shows the make up of the total wealth and net financial wealth. From the diagram it can be seen that total wealth was calculated from both housing and non-housing wealth. Total financial wealth contributed to the net non-housing wealth alongside the value of businesses and property, while net housing wealth was the difference between the value of the primary house and the value of the mortgage.

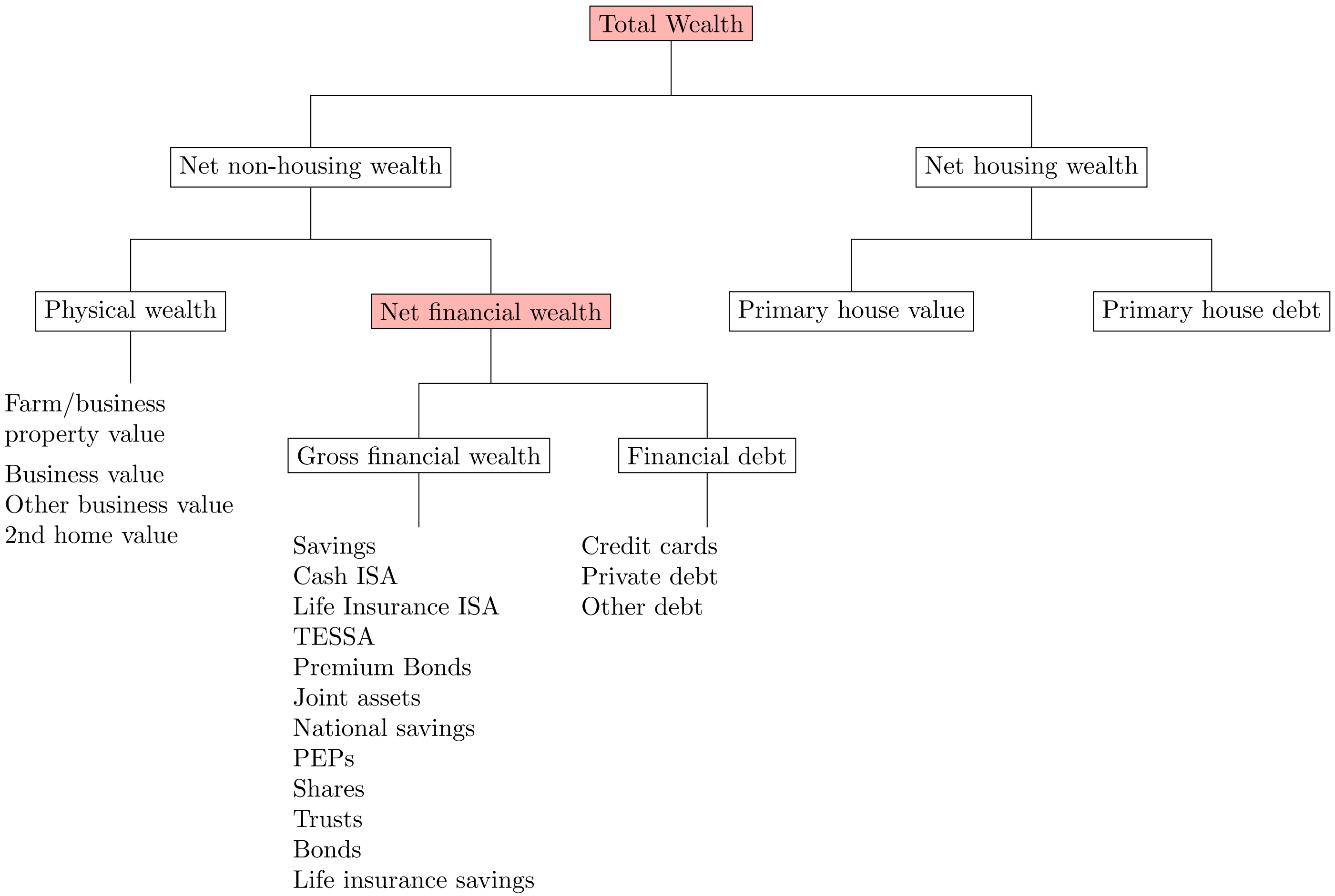


Figure 3.8: Dissection of the wealth variables in the ELSA (Wave 4).

Well-being variables

In addition to health and economic variables the ELSA contained subjective data on well-being comprising of the CASP-19 measure of quality of life in older adults (Higgs et al., 2003), the Satisfaction With Life Scale (SWLS) (Diener et al., 1985), eight items of the Center for Epidemiological Studies Depression scale (CES-D) (Radloff, 1977) and five items on loneliness of which four were from the UCLA loneliness scale (Russell, 1996). Apart from the latter, which was part of the interview, these were included in the self-completion questionnaire.

The CASP-19 was a measure of quality of life in later life developed by Higgs et al. (2003) and Hyde et al. (2003). The measure contained 19 items split into the four domains; control, autonomy, self-realisation and pleasure. The question asked "How often, do you feel like this?" followed by the 19 statements in Table 3.6. For each item the possible responses were never, not often, sometimes or often.

Domain	Item Number	Content
Control	1[a]	My age prevents me from doing the things I would like to.
	2[a]	I feel that what happens to me is out of my control.
	3	I feel free to plan for the future.
	4[a]	I feel left out of things.
Autonomy	5	I can do the things that I want to do.
	6[a]	Family responsibilities prevent me from doing what I want to do.
	7	I feel that I can please myself what I do.
	8[a]	My health stops me from doing things I want to do.
	9[a]	Shortage of money stops me from doing the things I want to do.
Pleasure	10	I look forward to each day.
	11	I feel that my life has meaning.
	12	I enjoy the things that I do.
	13	I enjoy being in the company of others.
	14	On balance, I look back on my life with a sense of happiness.
Self-Realisation	15	I feel full of energy these days.
	16	I choose to do things that I have never done before.
	17	I feel satisfied with the way my life has turned out.
	18	I feel that life is full of opportunities.
	19	I feel that the future looks good for me.

a - negative wording and reverse scoring.

Table 3.6: Item descriptions of the CASP-19, measure of quality of life in older adults, from Higgs et al. (2003).

Several of the items, indicated in Table 3.6, were negatively worded thus were reverse coded. That is, a response of never indicated a better quality of life than often; the opposite of the positively worded items. Therefore, these items were coded to reflect this.

The frequency of responses to each category for the 19 items are shown in Figure 3.9. Most of the items appear to be centred around the middle two categories (sometimes and not often), however it is clear to see that some were heavily skewed as the majority of individuals gave often as their response (Items 5, 7, 10-14 and 17), while less than 10% used the lower categories; never and not often combined.

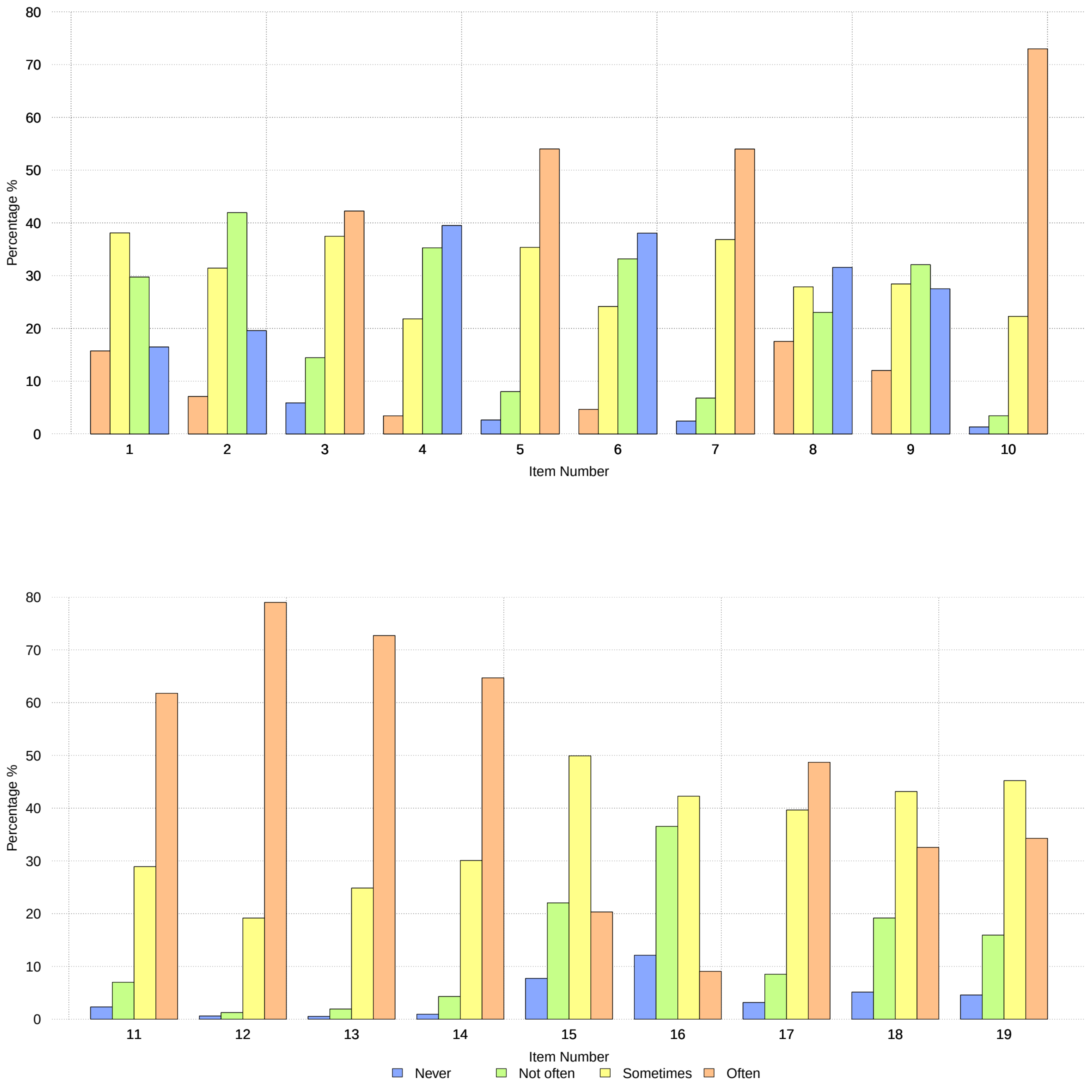


Figure 3.9: Frequencies of the category responses to the CASP-19 items in the ELSA (Wave 4).

The Satisfaction With Life Scale (SWLS) was developed by Diener et al. (1985) to determine an individual's "overall judgement of their life", without the focus on domains such as health or energy that may be valued differently in terms of importance by different individuals. The five items, described in Table 3.7, were preceded by "Please say how much you agree or disagree with the following statements". Items 1, 2 and 3 focused on satisfaction with current life, whereas items 4 and 5 focused on satisfaction with life compared with the individual's ideals. Responses were given on the 7-point Likert scale; strongly agree, agree, slightly agree, neither agree nor disagree, slightly disagree, disagree and strongly disagree, which are shown in Figure 3.10.

Item Number	Content
1	In most ways my life is close to ideal.
2	The conditions of my life are excellent.
3	I am satisfied with my life.
4	So far I have gotten the important things I want in life.
5	If I could live my life over I would change almost nothing.

Table 3.7: Description of the items in the Satisfaction with Life Scale in the ELSA (Wave 4).

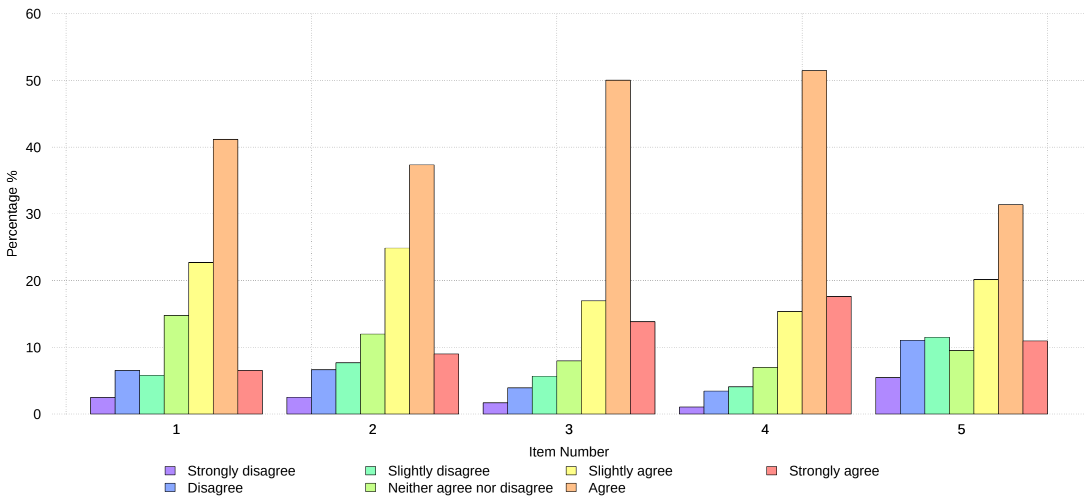


Figure 3.10: Frequency of the category responses to the Satisfaction With Life Scale items in the ELSA (Wave 4).

Loneliness was covered by five items, four from the revised UCLA Loneliness Scale and an additional fifth item that had the same format. The revised UCLA Loneliness Scale set out by Russell (1996) consisting of 20 items with responses on the 4-point Likert scale; never, rarely, sometimes and always. The revised scale was "developed to create a revised version of the ULCA Loneliness Scale that included positively worded or non-lonely items" (Russell, 1996) due to the possibility of systematic biases in negatively worded items or items that mentioned loneliness in the original scale. However, contrary to this the additional fifth item asks "How often do you feel lonely?" and the categories of response were the 3-point Likert scale; hardly ever or never, some of the time and often. A summary of the items is given in Table 3.8 and the distribution of responses in Figure 3.11.

Item Number	Content
1	How often do you feel you lack companionship?
2	How often do you feel left out?
3	How often do you feel isolated from others?
4^a	How often do you feel in tune with the people around you?
5	How often do you feel lonely?

a - negative wording and reverse scoring.

Table 3.8: Description of the loneliness items in the ELSA (Wave 4).

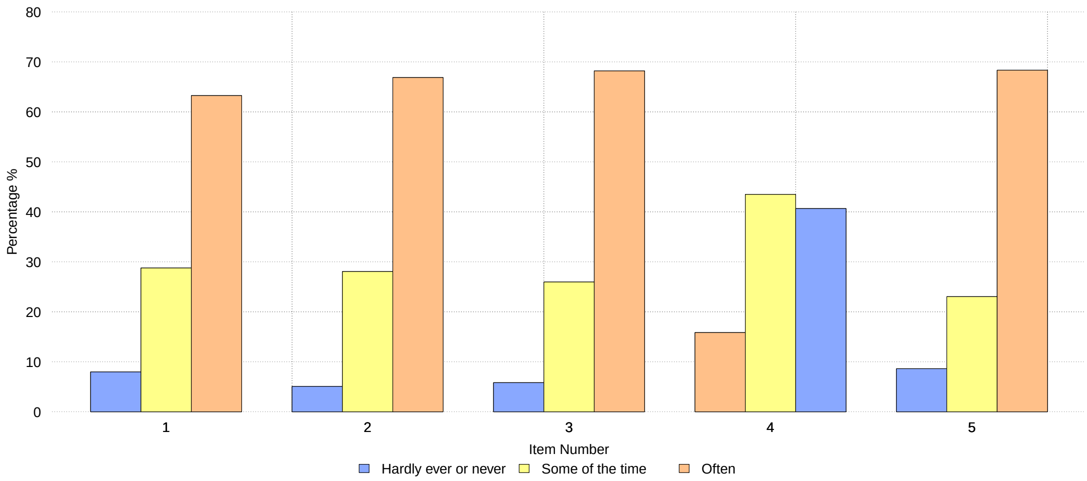


Figure 3.11: Frequency of the category responses to the loneliness items in the ELSA (Wave 4).

Lastly, a shortened version of the Center for Epidemiologic Studies-Depression scale (CES-D) was included in the interview. This was originally developed by Radloff (1977) and later the shortened eight item version was validated among older adults by Karim et al. (2014). These items are described in Table 3.9 and were binary variables with yes/no responses; shown in Figure 3.12. The majority of responses to each item was for the category that did not indicate depression, especially in the positively worded items where over 90% responded with yes. From the original scale, items 1, 5 and 7 were from the depressed affect factor, item 4 and 6 were both of the factor positive affect and items 2, 3 and 8 of the somatic factor. There were no items from the Interpersonal factor.

Item Number	Content
1	Whether felt depressed much of the time during past week.
2	Whether felt everything they did during past week was an effort.
3	Whether felt their sleep was restless during past week.
4[a]	Whether was happy much of the time during past week.
5	Whether felt lonely much of the time during past week.
6[a]	Whether enjoyed life much of the time during past week.
7	Whether felt sad much of the time during past week.
8	Whether could not get going much of the time during past week.

a - negative wording and reverse scoring.

Table 3.9: Description of the CES-Despression Scale items in the ELSA (Wave 4).

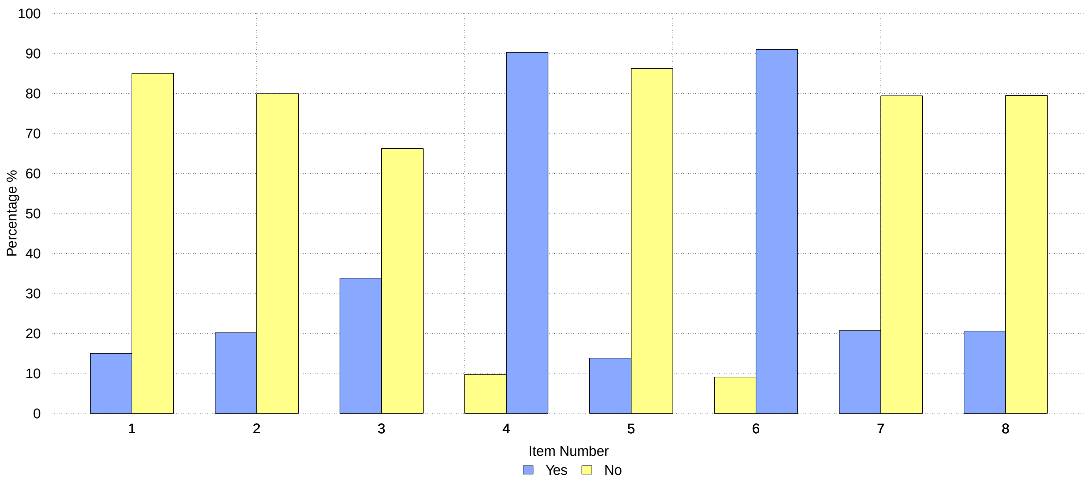


Figure 3.12: Frequency of category responses to the CES-Depression Scale items in the ELSA (Wave 4).

3.2.2 Longitudinal data

Longitudinal analyses focus predominantly on the well-being variables in the ELSA. These variables are collected in every even numbered Wave (2, 4 and 6). For reasons explained later in Chapter 7, the CASP-19 was not included in the longitudinal analysis.

To ensure a consistent selection in all analysis, only individuals that responded to at least one item in each of the three remaining areas of well-being (life satisfaction, loneliness and depression) at all three time points were considered.

The individuals selected were core members of the first cohort and aged at least 60 years old at Wave 2. All individuals gave a full or partial interview at all three waves, also including those that later gave an interview from an institution. Similarly to the cross-sectional data, the selected individuals consisted of 55.39% females and 44.61% males, 98.39% white race and 1.61% non-white race (with 1 missing value) and the ages at Wave 2 were distributed as: 25.11% aged 60-64 years old, 23.90% aged 65-69 years old, 19.99% aged 70-74 years old, 15.62% aged 75-79 years old, 10.10% aged 80-84 years old and 5.29% aged 85 or more years old. The final sample size was $n = 5,604$.

Drop-out is a prominent issue in longitudinal analysis of the elderly, particularly from deteriorating health or death. The profiles of drop-out and the percentages are given in Table 3.10, indicating the majority of individuals were present at all three waves with a third dropping out after Wave 2; a small percentage did return at Wave 6.

Wave 2	Wave 4	Wave 6	Percent
✓	✓	✓	54.12%
✓	✓		17.13%
✓			25.87%
✓		✓	2.87%

Table 3.10: Percentages of patterns of drop-out in the ELSA (Waves 2,4 & 6).

To be expected, drop-out was greater for those in the older age groups. Table 3.11 shows a minority of the younger age groups dropped out, while the drop out was over 50% in the age groups over 75 or more years old; while over 50% aged 85 or more years old dropped out after the first time point.

In the longitudinal analysis, changes in marital status were considered a time-varying variable, with the variable indicating the transitions between Waves 0 and 2, 2 and 4, and 4 and 6. Wave 0 marital status was taken from the responses to the Health Survey for England from which the

Pattern of Presence	Age (in years)					
	60-64	65-69	70-74	75-79	80-84	≥85
✓ ✓ ✓	66.31%	63.26%	55.54%	46.40%	30.92%	16.89%
✓ ✓	10.59%	12.99%	16.43%	24.11%	28.27%	27.36%
✓	19.83%	20.16%	25.18%	27.18%	28.87%	54.39%
✓ ✓	3.27%	3.58%	2.86%	2.29%	1.94%	1.35%

Table 3.11: Patterns of drop out by the 5-year age groups, from the ELSA (Waves 2, 4 & 6).

respondent was sampled. Single was specified as those that had never been married, therefore the 9 observations that changed marital status to single were excluded. The possible marital statuses (with percentages from Wave 0) were; single (4.68%), married or in a civil partnership (70.04%), divorced or separated (9.65%), or widowed (15.63%). Only those that were present at all three waves were included. The transitions of interest were from married to divorced, from married to widowed, and from divorced or widowed to married; the frequencies of these transitions are given in Table 3.12. Otherwise too few observed transitions were made.

Transition	Waves 0 to 2	Wave 2 to 4	Waves 4 to 6
Married to divorced	16	13	9
Married to widowed	159	130	140
Divorced or widowed to married	26	24	16

Table 3.12: Frequencies of marital status transitions in the ELSA (Waves 2, 4 & 6).

Summary of the Chapter

In this chapter we have introduced two datasets, the National Household Sample Survey for Brazil and the English Longitudinal Study of Ageing. Both datasets contained comparable variables concerning health in relation to self-reported morbidity and functional status, and economic circumstances in relation to housing quality and durables owned within the household. Individuals were selected from both datasets that were aged 60 or more years and provided their own responses to the subjective variables. However, each dataset presented separate issues with the complex hierarchical structure of the PNAD and the repeated measures of the ELSA.

Chapter 4

Methodology

This chapter provides detail of the methods used in the analysis of the data. The first section outlines the structural equation methods applied to both datasets, starting with the layout of the exploratory and confirmatory factor analysis, then more detailed descriptions of areas such as path diagrams, assessing model fit and model estimation methods are given. The second section sets out the methods applied only to Brazil's National Household Sample Survey (PNAD, Pesquisa Nacional de Amostra por Domicílios) which includes multilevel cumulative logistic regression and multilevel structural equation modelling. Finally, the methods applied only to the English Longitudinal Study of Ageing (ELSA) data are outlined. These methods include item response analysis and latent growth modelling.

All statistical analyses are completed using R version 3.2.3 and Mplus Version 6 (Muthén and Muthén, 1998). All structural equation models detailed later in Sections 4.1, 4.2.3 and 4.3.2 are conducted in Mplus, with data preparations performed using the Mplus Automation package in R (Hallquist and Wiley, 2016). The multilevel cumulative logistic regression models, described in Section 4.2.1, are analysed using the ordinal package in R (Christensen, 2015). In Mplus, delta parametrisation is used in all analyses.

4.1 General Methods - Structural Equation Modelling

In this section we look at the structural equation modelling methods applied to data from both the PNAD and ELSA. Firstly, the methods of exploratory and confirmatory factor analysis used to empirically define the measurement model for each domain are detailed. Then the structural equation modelling methods are described for the analysis of the interrelationships between the domains and the covariate effects, within each country.

Structural equation modelling enables us to measure and identify relationships between unobservable constructs, from a collection of observed variables. An advantage over standard regression is the ability to model a large number of observed variables by the reduction of the dimensionality, with the avoidance of multi-collinearity that would otherwise be introduced (Kline, 2010).

4.1.1 Tetrachoric and polychoric correlation coefficients

The observed variables detailed in Chapter 3 are all categorical, the first stage of SEM with binary and ordinal variables is the estimation of the tetrachoric or polychoric correlation coefficients, for each pair of observed variables (Lee et al., 1992). It is assumed that every observed categorical variable, Y, is the discretized manifestation of a latent response variable with standard normal distribution Y^*. The tetrachoric or polychoric correlation coefficient between two observed categorical variables can then be estimated using rectangle probabilities of the standard bivariate normal distribution. From this point onwards observed categorical variables will simply be referred to as observed variables, and the variable assumed to be categorical unless specified otherwise.

The latent response variable Y^*, from which the observed variables Y manifested, can discretized based on the cumulative standard normal distribution to form the threshold model;

$$Y = c \quad \Leftrightarrow \quad \tau_{c-1} < Y^* \leq \tau_c, \tag{4.1}$$

for threshold variables $\tau_0 < \tau_1 < ... < \tau_C$, where $\tau_0 = -\infty$ and $\tau_C = \infty$; as introduced to SEM by Muthén (1984). For $c = 1, \ldots, C-1$. Note, $\Phi(\tau_c)$ is equal to the proportion of observations of Y that are greater than or equal to c; where Φ is the cumulative probability from the standard normal distribution. The thresholds are regarded as fixed for the estimation of the tetrachoric or polychoric correlation coefficient.

Let a pair of observed variables be denoted Y_r and Y_s, with corresponding categories $Y_r = 1, \ldots, C_r$ and similarly $Y_s = 1, \ldots, C_s$; where C_r, $C_s \geq 2$. If both observed variables are binary, i.e. $C_r = C_s = 2$, then the responses can be represented in a 2×2 contingency table:

where a, b, c and d are the proportion of observed observations for corresponding combination of

		y_r	
		1	2
y_s	1	a	b
	2	c	d

categories of y_r and y_s.

The tetrachoric correlation coefficient can be found by setting the following equation (Equation (4.2)) equal to the proportion of responses in cell d and solving for $\rho_{r,s}$ through methods of numerical integration (Divgi, 1979). Let $\phi\left(Y_r^*; Y_s^*; \rho_{r,s}\right)$ be the density function of the standard bivariate normal distribution, with correlation parameter $\rho_{r,s}$, such that:

$$\phi\left(Y_r^*; Y_s^*; \rho_{r,s}\right) = \frac{1}{2\pi\sqrt{1-\rho_{r,s}^2}} \exp\left\{-\frac{\left(Y_r^{*2} + Y_s^{*2} - 2\rho_{r,s}Y_r^*Y_s^*\right)}{2\left(1-\rho_{r,s}^2\right)}\right\}$$

then,

$$\mathbb{L}(Y_r^* > \tau_r, Y_s^* > \tau_s, \rho_{r,s}) = \int_{\tau_r}^{\infty}\int_{\tau_s}^{\infty} \phi\left(Y_r^*; Y_s^*; \rho_{r,s}\right) dY_r^*\, dY_s^*. \tag{4.2}$$

can be solved for $\rho_{r,s}$.

This is a special case of the polychoric correlation coefficient. For the case when at least one of Y_r or Y_s has more than two categories, the contingency table contains $C_r \times C_s$ cells. Let the vector $\boldsymbol{p}$ contain all the observed proportions $p_{11}, \ldots, p_{rs}$ from the contingency table and let the vector $\boldsymbol{A}$ contain, in the same order, the following integral for each corresponding element of $\boldsymbol{p}$:

$$\mathbb{L}(\tau_r > Y_r^* \geq \tau_{r-1}, \tau_s > Y_s^* \geq \tau_{s-1}, \rho_{r,s}) = \int_{\tau_{r-1}}^{\tau_r}\int_{\tau_{s-1}}^{\tau_s} \phi\left(Y_r^*; Y_s^*; \rho_{r,s}\right) dY_r^*\, dY_s^*.$$

A value of $\rho_{r,s}$ can then be estimated using maximum likelihood estimation that approximates $\boldsymbol{A} = \boldsymbol{p}$, however a unique solution does not always exist (Ekström, 2011).

The log-likelihood function, proportional by some arbitrary constant, for the maximum likelihood estimation of $\rho_{r,s}$ is;

$$l(\rho_{r,s}) \propto \sum_{i=1}^{C_r}\sum_{j=1}^{C_s} n_{i,j} \log \mathbb{P}(Y_r = i, Y_s = j|\rho_{r,s})$$

where $n_{i,j}$ is the number of observations such that $y_r = i$ and $y_s = j$, and

$$\mathbb{P}(Y_r = i, Y_s = j|\rho_{r,s}) = \mathbb{L}(\tau_{r,i} > Y_r^* \geq \tau_{r,i-1}, \tau_{s,j} > Y_s^* \geq \tau_{s,j-1}, \rho_{r,s})$$

is the probability that $Y_r = i$ and $Y_s = j$ given $\rho_{r,s}$ (Olsson, 1979).

The resulting matrix is then fed forward into the second stage of ordinal SEM estimation, that involves the estimation of the structural parameters (Lee et al., 1992).

4.1.2 Exploratory and Confirmatory Factor Analysis

Let H_a be the number of observed variables in an aspect, a, where each aspect represents a collection of observed variables with common construct. For example, the set of variables of self-reported morbidity. A separate exploratory factor analysis (EFA) was applied to each aspect.

The aim of EFA is to find a smaller set of factors, $\{\eta_1, \ldots, \eta_M\}$, for each set of H_a latent response variables, such that each factor has meaningful interpretation. The factors are associated with the latent response variables via the factor loadings $\lambda_{m,h}$, for $h \in \{1, \ldots, H_a\}$ and $m \in \{1, \ldots, M\}$. This indicated the level of influence that the factor has on the observed variable, via the latent response variable.

The factors, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_M)'$, are assumed to have a multivariate normal distribution with zero mean and variance-covariance matrix $\boldsymbol{\Psi}$. The measurement model for the latent response variables in the vector, $\boldsymbol{Y}^* = \left(Y_1^*, \ldots, Y_{H_a}^*\right)'$ is then expressed as:

$$\boldsymbol{Y}^* = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{4.3}$$

with the variance-covariance structure;

$$\text{Var}\left(\boldsymbol{Y}^*\right) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{4.4}$$

where $\boldsymbol{\Lambda}$ is the $H_a \times M$ matrix of factor loadings and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_{H_a})'$ the vector containing the H_a measurement error terms, that are assumed to be independent, normally distributed, with zero mean and variance-covariance matrix $\boldsymbol{\Theta} = \text{diag}\{\theta_1, \ldots, \theta_{H_a}\}$.

EFA is applied to the observed variables, within each aspect, to determine the minimum number of factors, M, required to obtain a good fit to the data; whilst maintaining meaningful interpretation of the factors. The goodness of fit is determined from the model fit indices, provided by Mplus, to be discussed in further detail later in this chapter (Section 4.1.4), which includes: the Root Mean Square Error of Approximation (RMSEA), the Standardised Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI).

The eigenvalue of each factor from the sample variance-covariance matrix $\boldsymbol{S}$ is also considered. Either, as the last factor with eigenvalue greater than one, as recommended by Kaiser (1960), or if the eigenvalues do not fall below one and the number of observed variables per factor is less than three, then the scree plot is to be consulted, taking interpretation as suggested by Cattell (1966), where the factor before a "sharp drop" in eigenvalues is considered.

Oblique rotation was used to obtain a measure of the relationship between the observed variable and the factor, while allowing overlap between the factors.

The latent structure was restricted to a simple structure, such that only one factor loadings per observed variable was non-zero. This was to prevent double counting and ensure a clear conceptual interpretation for each of the factors as a distinct dimension of the domain. It was intended that the factors identified in the EFA would be associated via their relationship to the single concept of the domain, which would form the basis of a measurement model.

A factor loading was considered large enough to be interpreted as a measure of the factor if the overlapping variance was greater than 10%. This corresponds to a factor loading greater than 0.316 (0.32 to two decimal places, as suggested by Tabachnick (2001) on page 702).

Negative factor loadings, with magnitude greater than 0.316, were not considered in the definition of the factors. The existence of large negative factor loadings can indicate over-factoring, or a negative association between two observed variables.

This is repeated for each aspect $a \in \{1, \ldots, A\}$ until a simple factor structure is defined for every aspect.

A measurement model is formed for each domain (health, economic well-being or subjective well-being) by combining these latent structures, defined in the EFA, for all the aspects of that domain. The confirmatory factor model (CFA) is as defined in Equation (4.3) with a single non-zero factor loading estimated for each latent response variable. For the purpose of model identification, the factor loading of one item per factor is fixed to one; which sets the scale of the factor.

The next aim is to model a single factor for each domain, representing the overall construct of either health, economic well-being or subjective well-being. If the domain consists of three or more factors with strong and positive correlations, then the single second-order factor is included in the measurement model; less than three would lack model identification. The factor correlations are obtained through the standardised CFA.

The parameters only require standardisation in relation to the factor variances; as the variances of the latent response variables are already assumed to be one. Standardised factor loadings, $\boldsymbol{\Lambda}^*$, indicate the amount of change in the latent response variable given unit change in the factor. In this case, the standardised factor loadings are estimated by the equation;

$$\lambda^*_{m,h} = \lambda_{m,h}\sqrt{\psi_m} \tag{4.5}$$

for item h and factor m, with ψ_m the variance of factor m. The square of the standardised factor loadings, is also called the R-squared, that indicates the percentage of variance of the item h, that is explained by the factor m.

The standardised parameters $\boldsymbol{\Psi}^*$ of $\boldsymbol{\Psi}$ give the estimated factor correlations, such that for factors r and s :

$$\psi^*_{r,s} = \frac{\psi_{r,s}}{\sqrt{\psi_{r,r}}\sqrt{\psi_{s,s}}}. \tag{4.6}$$

4.1.3 Structural Equations

In the measurement models for the overall domain (health, economic well-being or subjective well-being) and structural equation models (SEMs) the directional relationships between the factors $\boldsymbol{\eta}$ are expressed by the structural equation system:

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{4.7}$$

where $\boldsymbol{B}$ is the matrix of regression slope parameters between the factors, with diagonal elements equal to zero, and $\boldsymbol{\zeta}$ the H_a-dimensional vector of residuals. The β parameters are the factor loadings of the second-order factor, that represents the domain, onto the first-order factors which are all the factors from the aspects within that domain; which previously defined in the EFA. Let d denote the second-order factor of the domain and m the first-order factor that loads onto item h. The variance of the first-order factors is then:

$$\begin{aligned}\psi_m &= \mathrm{Var}(\beta_m \eta_d + \zeta_m)\\ &= \beta_m^2 \psi_d + \phi_m.\end{aligned}$$

Thus, the standardised parameters of the factors loadings $\boldsymbol{\Lambda}$ and variance-covariance matrix $\boldsymbol{\Psi}$

are expressible as:

$$\lambda^*_{m,h} = \lambda_{m,h}\sqrt{\beta^2_m\psi_d + \phi_m}$$
$$\psi^*_{r,s} = \frac{\psi_{r,s}}{\sqrt{\beta^2_r\psi_d + \phi_r}\sqrt{\beta^2_s\psi_d + \phi_s}}$$

and the standardised parameters of the regression slopes, $\boldsymbol{B}^*$ of $\boldsymbol{B}$, are given by:

$$\beta^*_m = \frac{\beta_m\sqrt{\psi_d}}{\sqrt{\beta^2_m\psi_d + \phi_m}}.$$

The standardised regression weights can be interpreted similarly to the standardised factor loadings as they indicate the amount of change in the first-order factor given a unit increase in the second-order factor.

Despite the large sample size, frequencies for some of the categorical responses were small. Therefore, as mentioned by Titman et al. (2016) the standard errors of the measurement models may require bootstrapping in order to obtain valid standard errors for the parameter estimates, in particular because the standard errors produced by Mplus are based on asymptotic theory. Non-parametric bootstrapping based on resampling subjects with replacement was applied using 500 draws.

The estimated measurement models are displayed in a path diagram for which a general example is given in Figure 4.1. Circles represent the latent variables and rectangles represent the observed variables. Estimated thresholds, $\tau_{m,h,c}$, for item h are given below the corresponding observed variable $Y_{m,h}$. The estimated factors loadings $\boldsymbol{\Lambda}$, and respective standardised factor loadings $\boldsymbol{\Lambda}^*$, are given on the arrows between the factors, $\{\eta_1, \ldots, \eta_M\}$, and the respective latent response variables, $Y^*_{m,h}$. The estimated regression slopes $\boldsymbol{B}$, and standardised regression slopes $\boldsymbol{B}^*$, are given on the arrows between the domain factor η_d and the factors $\{\eta_1, \ldots, \eta_M\}$. The residual variance $\theta_{m,h}$, of the latent response variable $Y^*_{m,h}$, is indicated by the small arrow; similarly for the residual variances $\{\phi_1, \ldots, \phi_M\}$ of the factors $\{\eta_1, \ldots, \eta_M\}$. The factor variance ψ_d of the factor η_d is located next to the factor.

In the diagram, subscripts denoting the factor m have been dropped from the corresponding latent response variables $Y^*_{m,h}$, observed variables $Y_{m,h}$, residual variances $\theta_{m,h}$ and thresholds $\tau_{m,h,c}$ in the diagram; the subscript of $\tau_{m,h,c}$ denoting item h is also dropped.

Multiple indicator multiple cause models are fitted for each SEM to determine the effect of covariates on the domains. The interrelationships of the domains are defined as above with the additive presence of covariates, $\boldsymbol{X}$, with respective regression coefficients $\boldsymbol{\Gamma}$. Hence, the structural model is expressed as:

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{X} + \boldsymbol{\zeta}. \tag{4.8}$$

where $\boldsymbol{\Gamma}$ is the vector of regression coefficients for the covariates $\boldsymbol{X} = (x_1, \ldots, x_Q)'$.

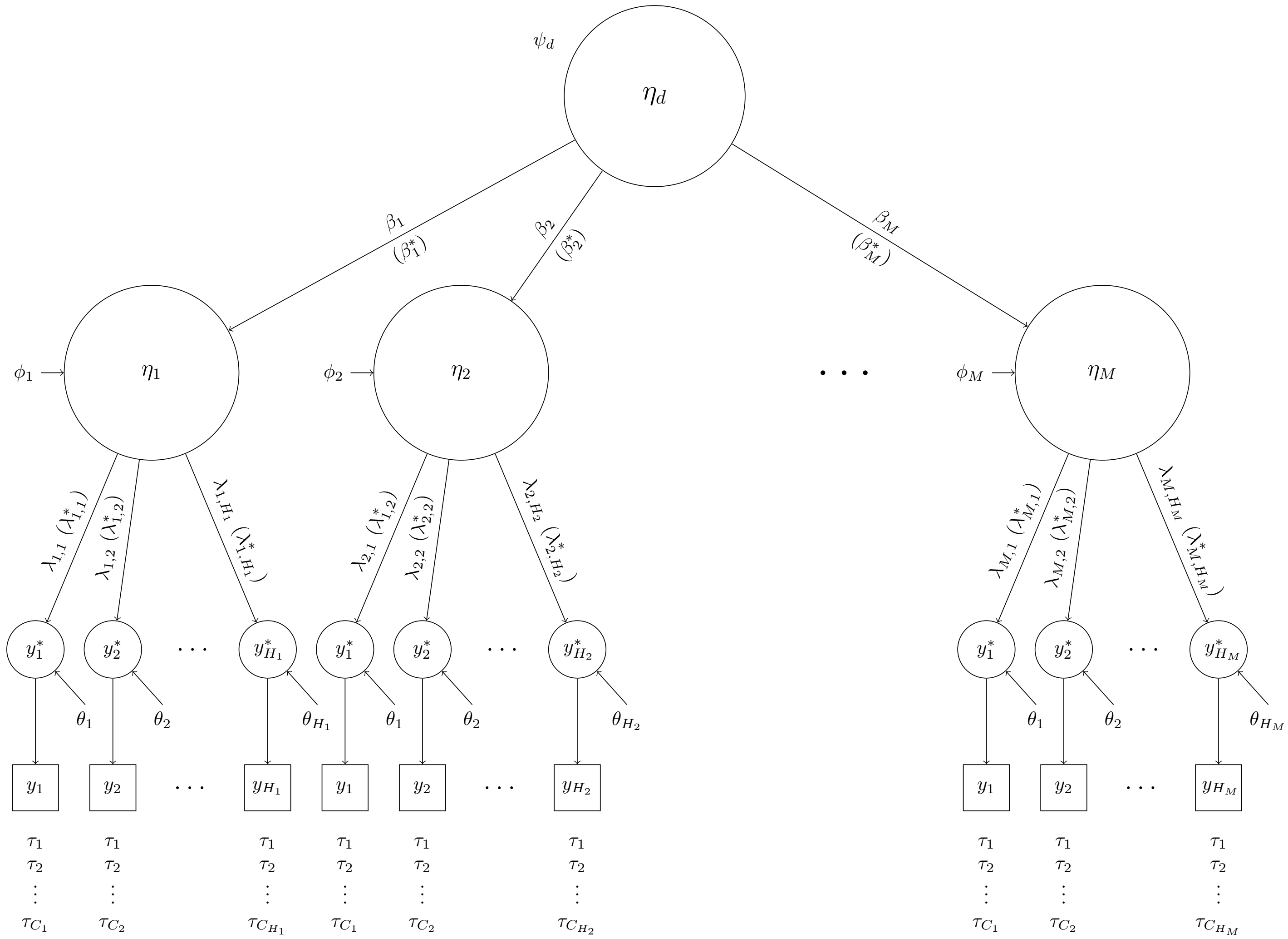


Figure 4.1: Example path diagram for the measurement model of a single domain, d.

4.1.4 Model Fit

To test the fit of the model to the data the null hypothesis that the fitted model fully captures the true covariance structure is determined on the basis of the sample variance-covariance $\boldsymbol{S}$ and the model estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$. The formal test of model fit in structural equation modelling is the chi-square statistic, however the χ^2 is often highly sensitive to minor deviations from the true model, resulting in the rejection of the null hypothesis due to minor deviations that do not have practical significance (Wang and Wang, 2013). Alternative model fit indices are therefore used, which will now be discussed further.

Absolute measures of model fit include the Root Mean Square Error of Approximation (RMSEA) and the Standardised Root Mean Square Residual (SRMR). The RMSEA, defined below, is based on the non-centrality parameter of the chi-square adjusted for the sample size, favouring parsimonious models, although this favouring is reduced by increasingly large sample sizes (Kline, 2010):

$$\text{RMSEA} = \sqrt{\frac{1}{N}\frac{\max\{\chi^2 - df, 0\}}{df}}.$$

The SRMR, on the other hand, is defined as;

$$\text{SRMR} = \sqrt{\frac{\sum_j \sum_{k<j} (s_{jk} - \hat{\sigma}_{jk})^2}{p(p+1)/2}},$$

when all observed variables are categorical (with unit variance of the latent response), which is based on the difference between the sample covariance matrix $\boldsymbol{S}$ and the model estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$, adjusting for the number of observed variables, p, in the model. It is, therefore, more sensitive to misspecification of the latent structure (Hu and Bentler, 1999).

Relative fit indices compare the discrepancy of the model from that of a baseline model. The baseline model in the case of the Comparative Fit Index (CFI, Bentler (1990)) and the Tucker-Lewis Index (TLI, Tucker and Lewis (1973)) assumes independence between all of the latent variables; *i.e.* that the covariance matrix is diagonal. The CFI is defined as:

$$\text{CFI} = 1 - \frac{\chi^2_{\mathcal{M}} - df_{\mathcal{M}}}{\chi^2_{\mathcal{N}} - df_{\mathcal{N}}}$$

for model $\mathcal{M}$ and null model $\mathcal{N}$, while the TLI is defined as:

$$\text{TLI} = \frac{\frac{\chi^2_{\mathcal{N}}}{df_{\mathcal{N}}} - \frac{\chi^2_{\mathcal{M}}}{df_{\mathcal{M}}}}{\frac{\chi^2_{\mathcal{N}}}{df_{\mathcal{N}}} - 1}.$$

which is also known as the Non-Normed Fit Index. Both indices are roughly on the normed 0-1 scale with values nearer 1 indicating better fit, although it is possible for the TLI to have value greater than one.

Cut-off values considered for the indication of good fit are; RMSEA< 0.05, SRMR< 0.06, CFI$>$ 0.90 and TLI> 0.95, as recommended by Hu and Bentler (1999), Hooper et al. (2008) or Kline (2010). However, each index is sensitive to different conditions and, as discussed by Hu and Bentler (1999), the cut-off values for combinations of fit indices varied. The conditions of each model are therefore taken into consideration before the goodness of fit is accepted or rejected.

4.1.5 Model Estimation

All exploratory and confirmatory factor analysis, structural equation modelling (including multilevel), item response analysis and latent growth modelling are conducted using Mplus Version 6 (Muthén and Muthén, 1998). Mplus has strong functionality for handling a large number of categorical observations, and large complex models.

As introduced by Muthén (1984), the model estimation was a multi-step process that applied the Weighted Least Squares with Mean and Variance adjustment (WLSMV) estimator. This was later extended to the multilevel case by Asparouhov and Muthén (2007). Firstly, the threshold structures of the latent response variables are estimated. Then these estimates are fed forward into the estimation of the first and second moments of the latent response variables, $\boldsymbol{Y}^*$;

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{Y}^*|\boldsymbol{X}\right) =& \Delta\left\{\boldsymbol{\Lambda}\left(\boldsymbol{I}_m-\boldsymbol{B}\right)^{-1}\boldsymbol{\Gamma}\boldsymbol{X}\right\},\\ \mathrm{Var}\left(\boldsymbol{Y}^*|\boldsymbol{X}\right) =& \Delta\left\{\boldsymbol{\Lambda}\left(\boldsymbol{I}_m-\boldsymbol{B}\right)^{-1}\boldsymbol{\Psi}\left(\boldsymbol{I}_m-\boldsymbol{B}\right)'^{-1}\boldsymbol{\Lambda}'+\boldsymbol{\Theta}\right\}\Delta.\end{aligned}$$

These provide the respective tetrachoric/polychoric correlations between the latent response variables, and the regression slopes, $\boldsymbol{\Gamma}$, when covariates are present (i.e. $\boldsymbol{X} \neq \boldsymbol{0}$). The Δ parameters are the scale factor parameters used by Mplus for alternative parametrisations, in this case, when all observed variables are categorical, $\Delta = 1$, however, using the theta parametrisation $\Delta = 1/\sqrt{\lambda^2\Psi + 1}$ (Muthén and Asparouhov, 2002).

The third step of the estimation process minimises the weighted least squares estimator,

$$(\boldsymbol{S}-\boldsymbol{\Sigma})'\boldsymbol{W}^{-1}(\boldsymbol{S}-\boldsymbol{\Sigma}),$$

where $\boldsymbol{S}$ is the sample variance-covariance matrix, $\boldsymbol{\Sigma}$ is the model estimated variance-covariance matrix and, for WLSMV, W is the diagonal matrix with the estimated variances from the sample covariance matrix, $\boldsymbol{S}$, for robust estimation (Muthén et al., 1997; Muthén, 1998). This is in contrast with using the inverse of the estimated covariance matrix from the polychoric correlations, which may be singular, in full weighted least squares. Whilst this is theoretically more efficient, it performs worse in practice.

Ideally, full information maximum likelihood would be used to estimate the model, however this would require a high amount of computation time and resources due to the high dimensional integration and large number of parameters.

Mplus provides robust inferences, and estimates of standard errors and chi-square tests of model fit (Muthén et al., 1997; Muthén and Muthén, 1998). Where possible, standard errors are bootstrapped with the resampling of cases as mentioned above.

The missing data was assumed to be missing at random given the covariates (MARX), which is a more restrictive assumption than missing at random. MARX assumes that the missingness occurred at random conditional on the covariates but not on the dependent variables. For missing values with the WLSMV estimator, Mplus used the methods that were presented by Asparouhov and Muthén (2010), where full information maximum likelihood is used at the first and second stages of the WLSMV estimation, to estimate the univariate and bivariate models respectively.

4.2 Methods for Brazil Analysis

Multilevel modelling of the PNAD is of key interest due to the complex sampling plan and hierarchical structure of the PNAD, presented in the previous chapter (Section 3.1.1). In order to correct for the underestimation of standard errors in the presence of a clustering, multilevel modelling was applied in the analysis of the 2008 PNAD. This allowed for inferences at the cluster level, which was either the households, sectors, or municipalities in Brazil. Multilevel modelling is first explored with the application of multilevel cumulative logistic regression to the ordinal self-rated health (SRH) variable. In the presence of a complex sampling plan, the typical approach is to a weighted analysis using the sampling weights to correct for the unequal sampling probabilities. This was only possible in a non-multilevel analysis of the PNAD due to limited availability of the sampling variables, as discussed in Section 3.1.1. However, for comparison of parameter estimates a weighted single level cumulative logistic regression was also applied.

The structural equation model (SEM) discussed in the previous section, was also extended a multilevel analysis in order to account for the clustering, which could inflate model fit indices (Muthén, 1994). An additional benefit of multilevel SEM is the ability to define a separate factor structure for the clusters and for the older individuals.

4.2.1 Multilevel Cumulative Logistic Regression

Multilevel cumulative logistic regression is applied for analysis of self-rated health (SRH) of the elderly in Brazil, determining the amount of variation that is explained by clustering and how characteristics of the clustering impact upon SRH. The variance partitioning coefficients are first estimated to identify the amount of variation explained by each possible clustering to be considered, and determine the clustering that requires inclusion in the multilevel model to adjust for the hierarchical structure in the PNAD. Without correction for the hierarchical structure, the standard errors are underestimated leading to the exaggeration in the significance of the variable. The effect of important factors is then ascertained in univariate and a multivariate multilevel cumulative logistic regression models are fitted with the variables of interest, analysing. A cumulative model is most appropriate given the ordinal variable for self-rated health variable in the PNAD.

The variance partitioning coefficient (VPC) for each potential higher level is calculated from the two level null model. Let $Y_{i,j}$ be the ordinal dependent variable, SRH, for individual i in cluster j, that had categories $c \in \{1, \dots, C\}$. In this section, the levels are denoted by the superscripts (1) and (2) for the first level (individuals) and the second level (cluster). The two level null model is defined as:

$$\log\left(\frac{\mathbb{P}(Y_{i,j} \leq c)}{1 - \mathbb{P}(Y_{i,j} \leq c)}\right) = \alpha_c + \beta^{(2)} + \epsilon_j^{(2)}$$

where α_c is the intercept corresponding to the category c, and from the cluster level $\beta^{(2)}$ is the intercept term, representing the mean value for the cluster, and $\epsilon_j^{(2)}$ is the random term, that is assumed to be normally distributed with zero mean and variance $\theta^{(2)}$. Note the unconventional notation for multilevel regression that has been adopted in order to maintain similar notation to the SEM convention used elsewhere in this thesis.

The amount of variance explained by the cluster is:

$$\text{VPC} = \frac{\theta^{(2)}}{\theta^{(2)} + \frac{\pi^2}{3}},$$

where $\frac{\pi^2}{3} \approx 3.29$ is the variance of the logistic distribution accounting for the level 1 variance (Goldstein et al., 2002). If the higher level explained more than 5% of the variance (Hedges and Hedberg, 2007) then it was carried forward in the combinations, to a three level model.

Just as with the two level models, for each combination of higher levels, a null model of three levels is fitted in order to obtain the VPCs for each combination. The null three level model is defined as:

$$\log\left(\frac{\mathbb{P}(Y_{i,j,k} \leq c)}{1 - \mathbb{P}(Y_{i,j,k} \leq c)}\right) = \alpha_c + \beta^{(3)} + \epsilon^{(2)}_{j,k} + \epsilon^{(3)}_{k} \tag{4.9}$$

where the intercept $\beta^{(3)}$ is now the mean at the third level, denoted (3), and there exists a random intercept $\epsilon^{(2)}_{j,k}$ and $\epsilon^{(3)}_{k}$ for each cluster level, that are assumed to have multivariate normal distribution with zero means and respective variances $\theta^{(2)}$ and $\theta^{(3)}$ and covariance $\theta^{(2,3)}$.

The VPCs estimate the amount of variance explained by level 2:

$$\text{VPC}^{(2)} = \frac{\theta^{(2)}}{\theta^{(2)} + \theta^{(3)} + \frac{\pi^2}{3}},$$

and the amount of similarity of persons within the same level 2 and 3 clustering:

$$\text{VPC}^{(2,3)} = \frac{\theta^{(2)} + \theta^{(3)}}{\theta^{(2)} + \theta^{(3)} + \frac{\pi^2}{3}}.$$

Additionally, there is the amount of variance explained by the level 3 clustering

$$\text{VPC}^{(3)} = \frac{\theta^{(3)}}{\theta^{(2)} + \theta^{(3)} + \frac{\pi^2}{3}},$$

however this is similar to the VPC from the two level models.

The appropriate higher levels are identified as the combination that explains the most variance, when the $\text{VPC}^{(2)}$ and $\text{VPC}^{(2,3)}$ are greater than 5%.

Covariates are considered at all three levels in the matrices $\boldsymbol{X}^{(3)}_{k}$ for the level three cluster k, $\boldsymbol{X}^{(2)}_{j,k}$ for the level two cluster j within the level three cluster k, and $\boldsymbol{X}^{(1)}_{i,j,k}$ for the individual within level two cluster j and level three cluster k.

$$\log\left(\frac{\mathbb{P}(Y_{i,j,k} \leq c)}{1 - \mathbb{P}(Y_{i,j,k} \leq c)}\right) = \alpha_c + \beta^{(3)} + \boldsymbol{\Gamma}^{(1)}\boldsymbol{X}^{(1)}_{i,j,k} + \boldsymbol{\Gamma}^{(2)}\boldsymbol{X}^{(2)}_{j,k} + \boldsymbol{\Gamma}^{(3)}\boldsymbol{X}^{(3)}_{k} + \epsilon^{(2)}_{j,k} + \epsilon^{(3)}_{k} \tag{4.10}$$

where the random terms are as in Equation (4.9), and $\boldsymbol{\Gamma}^{(1)}$, $\boldsymbol{\Gamma}^{(2)}$ and $\boldsymbol{\Gamma}^{(3)}$ are the regression coefficients that correspond to covariates for the level indicated. The covariate effects are interpreted in terms of the odds ratio which are calculated by taking the exponential of the regression coefficients.

The model is estimated using maximum likelihood estimation. The log-likelihood function, $l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta^{(2)}, \theta^{(3)}; \boldsymbol{y})$, is given by;

$$\sum_k \sum_j \log \int \int \prod_i \mathbb{P}\left(Y_{i,j,k} = y_{i,j,k} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \epsilon^{(2)} = u, \epsilon^{(3)} = v\right) f(\theta^{(2)}) f(\theta^{(3)})\, d\theta^{(2)}\, d\theta^{(3)}$$

where $\mathbb{P}\left(Y_{i,j,k} = y_{i,j,k} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \theta^{(2)}, \theta^{(3)}\right)$ is the probability that individual i in cluster j of cluster k was observed to be in category $y_{i,j,k} \in \{1, \ldots, C\}$ conditional on values of the level 2 and 3 random effects; which are the non-zero elements of $\mathbb{P}\left(Y_{i,j,k} = c | \boldsymbol{\alpha}, \boldsymbol{\beta}, \theta^{(2)}, \theta^{(3)}\right) \mathbb{I}(Y_{i,j,k} = c)$, with $\mathbb{I}()$ the indicator function: and the random effects $\epsilon^{(2)}$ and $\epsilon^{(3)}$ have normal density function $f(\cdot)$ with mean zero and variances $\theta^{(2)}$ and $\theta^{(3)}$ respectively.

The probability $\mathbb{P}\left(Y_{i,j,k} = y_{i,j,k} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \theta^{(2)}, \theta^{(3)}\right)$ is calculated from the property:

$$\mathbb{P}(Y_{i,j,k} = c | \boldsymbol{\alpha}, \boldsymbol{\beta}, \theta^{(2)}, \theta^{(3)}) = \mathbb{P}(Y_{i,j,k} \leq c | \boldsymbol{\alpha}, \boldsymbol{\beta}, \theta^{(2)}, \theta^{(3)}) - \mathbb{P}(Y_{i,j,k} \leq c - 1 | \boldsymbol{\alpha}, \boldsymbol{\beta}, \theta^{(2)}, \theta^{(3)}).$$

The significance of each covariate was first analysed in a multilevel univariate cumulative logistic regression, using the likelihood ratio statistic to compare the estimated log-likelihood of each univariate model l_1 to that of the null model, l_0;

$$\mathrm{LR} = -2(l_0 - l_1)$$

This procedure is then repeated for the multivariate model, comparing the estimated log-likelihood of the full model, $l_{\mathcal{F}}$, to that of model with each covariate removed in turn.

4.2.2 Weighted Cumulative Logistic Regression

Ideally, the unequal sampling probabilities present in the PNAD data would be accounted for in the multilevel cumulative logistic regression model. Due to confidentiality reasons, the sampling probabilities for each level were not available in the 2008 PNAD, for a multilevel analysis. Therefore, to provide comparison between a design-based approach with sampling weights and a model-based approach with multilevel modelling, a single level weighted cumulative logistic regression is also applied to the ordinal response of SRH.

The sampling weight ω_i for the selection of individual i, is calculated as the inverse of the probability of sampling i. All individuals within the sampled household are included, therefore the probability of sampling the individual was equal to the probability of sampling the intersection of the household j, the sector k and the municipality l within which individual i resided.

$$\begin{aligned}\omega_{i,j,k,l} &= \frac{1}{\mathbb{P}\left(\text{select } i\right)} \\ &= \frac{1}{\mathbb{P}\left(\text{select } j \cap k \cap l\right)} \\ &= \frac{1}{\mathbb{P}\left(\text{select } j|\text{select } k \cap l\right)\mathbb{P}\left(\text{select } k|l\right)\mathbb{P}\left(\text{select } l\right)}\end{aligned}$$

In the case of weighted multilevel analysis, the separate conditional probabilities $\mathbb{P}\left(\text{select } j|\text{select } k \cap l\right)$, $\mathbb{P}\left(\text{select } k|l\right)$ and $\mathbb{P}\left(\text{select } l\right)$ could be used to provide weights for the analysis at each level (Pfeffermann et al., 1998; Asparouhov, 2006). It is these separate conditional probabilities that were unavailable in the 2008 PNAD.

An overall weight for the sampled individuals was, however, available; this weight was equal for all individuals within the same household.

The model is defined similarly to above, such that;

$$\log\left(\frac{\mathbb{P}(Y_{i,j,k,l} \leq c)}{1 - \mathbb{P}(Y_{i,j,k,l} \leq c)}\right) = \alpha_c - \mathbf{\Gamma}\boldsymbol{X}_i,$$

where α_c is the threshold parameter for the category $c \in \{1, \ldots, C\}$ and $\mathbf{\Gamma}$ are the regression coefficients for the covariates $\boldsymbol{X}_i$ of individual i.

Maximum likelihood estimation of the weighted single-level cumulative logistic regression was done using the ordinal package in R, which uses a naïve case weighting method (Christensen, 2015). The log-likelihood for the weighted single level cumulative logistic regression is:

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{y}) = \sum_l \sum_k \sum_j \sum_i \log\, \omega_{i,j,k}\mathbb{P}(Y_{i,j,k,l} = y_{i,j,k,l}|\boldsymbol{\alpha}, \boldsymbol{\beta})$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the parameters of interest, and $\mathbb{P}(Y_{i,j,k,l} = y_{i,j,k,l}|\boldsymbol{\alpha}, \boldsymbol{\beta})$ gives the probability of observing that individual i is in the category, $y_{i,j,k,l} \in \{1, \ldots, C\}$, that they were observed to be in.

4.2.3 Multilevel Structural Equation Modelling

The process for defining the multilevel SEM is a multi-step process (Muthén, 1994; Grilli and Rampichini, 2007). Firstly, the within level (Level one) was defined the individual level. The process is outlined below, where steps 1-4 are applied within each aspect before the resulting structures are combined in step 5 to form the multilevel measurement models for each domain:

1. Exploration of factor structure of the total covariance structure in a non-hierarchical structural equation analysis (using the methods in Section 4.1.2);
2. Estimation of the between level variation for each variable;
3. Multilevel EFA of the within level structure in multilevel EFA with an unrestricted structure at the between level;
4. Multilevel EFA of between level factor structure in multilevel EFA with the within level factors restricted to number of factors concluded in the previous step;
5. Finally, multilevel CFA with the within and between level latent structures as identified in steps 3 and 4.

To form the multilevel structure the total covariance matrix $\boldsymbol{\Sigma}$ is expressed as the sum of the two covariances matrices $\boldsymbol{\Sigma}^{(\mathcal{W})}$ and $\boldsymbol{\Sigma}^{(\mathcal{B})}$, that represent the within level and between level covariance; denoted by the superscripts $(\mathcal{W})$ and $(\mathcal{B})$. Sample covariances matrices $\boldsymbol{S}^{(\mathcal{T})}, \boldsymbol{S}^{(\mathcal{PW})}$ and $\boldsymbol{S}^{(\mathcal{B})}$ correspond to the total, pooled-within and between level covariance matrices. The covariance matrix $\boldsymbol{S}^{(\mathcal{PW})}$ is an unbiased estimator of $\boldsymbol{\Sigma}^{(\mathcal{W})}$, while

$$\frac{1}{\varsigma}\left(\boldsymbol{S}^{(\mathcal{B})} - \boldsymbol{S}^{(\mathcal{PW})}\right)$$

is an unbiased estimator of $\boldsymbol{\Sigma}^{(\mathcal{B})}$, where ς reflects the average cluster size:

$$\varsigma = \frac{1}{N(G-1)}\left(N^2 - \sum_{g=1}^{G} N_g\right)$$

for total sample size N and clusters $g \in \{1, \dots, G\}$ of size N_g (Muthén, 1994).

To estimate the between level variation the intraclass correlation coefficient (ICC) is estimated for each variable. The ICC is

$$\frac{\sigma^{(\mathcal{B})}}{\sigma^{(\mathcal{W})} + \sigma^{(\mathcal{B})}}$$

which is estimated using the unbiased estimates of $\sigma^{(\mathcal{W})} \in \boldsymbol{\Sigma}^{(\mathcal{W})}$ and $\sigma^{(\mathcal{B})} \in \boldsymbol{\Sigma}^{(\mathcal{B})}$ from above.

The latent response variable, Y_h^*, for each item is expressed as the sum of the two latent variables corresponding to the within and between levels. Similarly to the expression for $\boldsymbol{\Sigma}$;

$$Y_h^* = Y_h^{(\mathcal{W})} + Y_h^{(\mathcal{B})},$$

where the variables $Y_h^{(\mathcal{W})}$ and $Y_h^{(\mathcal{B})}$ are independent normally distributed latent variables representing the contributions from the within and between levels respectively.

The measurement model for $\boldsymbol{Y}^*$ is, therefore, the combined within level measurement model for $\boldsymbol{Y}^{(\mathcal{W})}$ and between level measurement model for $\boldsymbol{Y}^{(\mathcal{B})}$, such that:

$$\boldsymbol{Y}^* = \left(\boldsymbol{\Lambda}^{(\mathcal{W})}\boldsymbol{\eta}^{(\mathcal{W})} + \boldsymbol{\epsilon}^{(\mathcal{W})}\right) + \left(\boldsymbol{\Lambda}^{(\mathcal{B})}\boldsymbol{\eta}^{(\mathcal{B})} + \boldsymbol{\epsilon}^{(\mathcal{B})}\right) \tag{4.11}$$

The distribution of each item is determined by the between level factors $\boldsymbol{\eta}^{(\mathcal{B})}$ through their corresponding factor loadings $\boldsymbol{\Lambda}^{(\mathcal{B})}$ and residuals $\boldsymbol{\epsilon}^{(\mathcal{B})}$, and also by the within level factors $\boldsymbol{\eta}^{(\mathcal{W})}$ through their corresponding factor loadings $\boldsymbol{\Lambda}^{(\mathcal{W})}$ and residuals $\boldsymbol{\epsilon}^{(\mathcal{W})}$. The vectors of residuals $\boldsymbol{\epsilon}^{(\mathcal{W})}$ and $\boldsymbol{\epsilon}^{(\mathcal{B})}$ are assumed to be independently normally distributed with zero mean and diagonal variance matrices $\boldsymbol{\Theta}^{(\mathcal{W})}$ and $\boldsymbol{\Theta}^{(\mathcal{B})}$ respectively. The residual variances $\boldsymbol{\Theta}^{(\mathcal{W})}$ are fixed to 1 via the scale factor Δ.

The multilevel EFA, in steps 3 and 4, follows the same procedure as the non-hierarchical EFA in Section 4.1.2. In step 3, the between level structure remains unrestricted while the number of factors at the within level is explored. Then, based on the conclusions of step 3, the number of factors at the between level is explored while the number within level factors is fixed. The latent structures are defined using the same methods as in Section 4.1.2, using the model fit indices and eigenvalues to define the number of factors, and the estimated factor loadings to ensure a simple latent structure with no negative factor loadings and no cross-loadings.

The SRMR in the multilevel analysis measures model fit at the within level and between level separately, using estimates of the two covariance matrices $\boldsymbol{\Sigma}^{(\mathcal{W})}$ and $\boldsymbol{\Sigma}^{(\mathcal{B})}$. Thus:

$$\text{SRMR} = \begin{cases} \sqrt{\dfrac{\sum_j \sum_{k<j} \left(s_{jk}^{(\mathcal{W})} - \hat{\sigma}_{jk}^{(\mathcal{W})}\right)^2}{p(p+1)/2}} & \text{Within level} \\ \sqrt{\dfrac{\sum_j \sum_{k<j} \left(s_{jk}^{(\mathcal{B})} - \hat{\sigma}_{jk}^{(\mathcal{B})}\right)^2}{p(p+1)/2}} & \text{Between level} \end{cases}$$

where $s_{jk}^{(\mathcal{W})}$ and $s_{jk}^{(\mathcal{B})}$ are the sample variance-covariances and $\hat{\sigma}_{jk}^{(\mathcal{W})}$ and $\hat{\sigma}_{jk}^{(\mathcal{B})}$ are the model estimated variance-covariances, and p is the number of observed variables in the model. The RMSEA, CFI and TLI model fit indices are based on the estimated χ^2, hence remain unchanged from the definitions in Section 4.1.4.

For Step 5 the structural equations for the multilevel measurement models of each domain is defined at both levels:

$$\boldsymbol{\eta}^{(\mathcal{W})} = \boldsymbol{B}^{(\mathcal{W})}\boldsymbol{\eta}^{(\mathcal{W})} + \boldsymbol{\zeta}^{(\mathcal{W})} \tag{4.12}$$

$$\boldsymbol{\eta}^{(\mathcal{B})} = \boldsymbol{B}^{(\mathcal{B})}\boldsymbol{\eta}^{(\mathcal{B})} + \boldsymbol{\zeta}^{(\mathcal{B})} \tag{4.13}$$

where the regression slope parameters $\boldsymbol{B}^{(\mathcal{W})}$ and $\boldsymbol{B}^{(\mathcal{B})}$ both have zeros on the diagonal, such that $\text{diag}\{\boldsymbol{B}^{(\mathcal{W})}\} = \mathbf{0}$ and $\text{diag}\{\boldsymbol{B}^{(\mathcal{B})}\} = \mathbf{0}$ respectively. It is assumed the residuals $\boldsymbol{\zeta}^{(\mathcal{W})}$ and $\boldsymbol{\zeta}^{(\mathcal{B})}$ are independent and normally distributed with zero means and diagonal variance matrices $\boldsymbol{\Phi}^{(\mathcal{W})}$ and $\boldsymbol{\Phi}^{(\mathcal{B})}$ respectively.

The covariances structure between the factors is:

$$\text{Var}(\boldsymbol{Y}^*) = \boldsymbol{\Lambda}^{(\mathcal{W})}\boldsymbol{\Psi}^{(\mathcal{W})}\boldsymbol{\Lambda}^{(\mathcal{W})\prime} + \boldsymbol{\Theta}^{(\mathcal{W})} + \boldsymbol{\Lambda}^{(\mathcal{B})}\boldsymbol{\Psi}^{(\mathcal{B})}\boldsymbol{\Lambda}^{(\mathcal{B})\prime} + \boldsymbol{\Theta}^{(\mathcal{B})}$$

where $\boldsymbol{\Psi}^{(\mathcal{W})}$ and $\boldsymbol{\Psi}^{(\mathcal{B})}$ are the factor variance-covariance matrices.

The multilevel SEM then combines the resulting multilevel measurement models for health and economic well-being, to ascertain the interrelationships between the within level factors and between level factors. Covariates are then added to the Equations (4.12) and (4.13) to form a multilevel multiple indicator and multiple causes model (MIMIC), such that covariates of the cluster are considered at the between level. The structural equations of multilevel MIMIC are:

$$\boldsymbol{\eta}^{(\mathcal{W})} = \boldsymbol{B}^{(\mathcal{W})}\boldsymbol{\eta}^{(\mathcal{W})} + \boldsymbol{\Gamma}^{(\mathcal{W})}\boldsymbol{X}^{(\mathcal{W})} + \boldsymbol{\zeta}^{(\mathcal{W})}$$

$$\boldsymbol{\eta}^{(\mathcal{B})} = \boldsymbol{B}^{(\mathcal{B})}\boldsymbol{\eta}^{(\mathcal{B})} + \boldsymbol{\Gamma}^{(\mathcal{B})}\boldsymbol{X}^{(\mathcal{B})} + \boldsymbol{\zeta}^{(\mathcal{B})}$$

where $\boldsymbol{\Gamma}^{(\mathcal{W})}$ and $\boldsymbol{\Gamma}^{(\mathcal{B})}$ are the regression coefficients for the within and between level covariates, $\boldsymbol{X}^{(\mathcal{W})}$ and $\boldsymbol{X}^{(\mathcal{B})}$, respectively.

4.3 Methods for England Analysis

The ELSA grants the opportunity to analyse the additional domain of subjective well-being in the elderly, not just in a cross-sectional ideology but in a longitudinal setting as well. An area of concern was discussed in Chapter 3, concerning the items of the CASP-19 measure of quality of life in older adults. Following concerns of the latent structure of the CASP-19 raised in Section 2.3.5, multidimensional item response theory (IRT) is applied to assess the performance of the items and the overall test.

Changes in subjective well-being are analysed using the repeated measures of subjective well-being at three time points; waves 2, 4 and 6. Univariate and multivariate latent growth modelling, are applied to the aspects of subjective well-being, looking at the change over time within individuals and how that change is affected by covariates.

4.3.1 Item Response Analysis

The aim of the IRT is to assess the performance of the 19 quality of life variables from the CASP-19, in terms of multiple dimensions. Therefore, a multidimensional two parameter item response model, with probit link function, was selected to model responses.

The multidimensional model provides an approach that is similar to the structural equation modelling, but with the ability to model the location of items and individuals on the same scale of a latent trait. For consistency with the SEM, a probit link function is chosen, although the logit link function is typically used in IRT.

Performance of the items is assessed through the discrimination and difficulty parameters α_h and $\boldsymbol{\beta}_h$. Let $D(h)$ denote the latent trait η_m, $m \in \{1, \ldots, M\}$, that corresponds to item h that has categories $c \in \{1, \ldots, C\}$. The cumulative probability of individual i with latent trait $\eta_{D(h),i}$ responding with at least category c is defined as:

$$\mathbb{P}(Y_{h,i} \geq c | \eta_{D(h),i}) = \Phi\left(\alpha_h(\eta_{D(h),i} - \beta_{c,h})\right)$$

The discrimination parameter α_h maps the steepness of the curve, which indicate the ability of item h to distinguish between two individuals at similar locations on the scale of the latent trait $D(h)$. Higher α_h estimates indicate greater discrimination ability. The difficulty parameters $\beta_{c,h}$ indicate the location of latent trait required for the individual to have 0.5 probability of responding with a category less than or equal to c for item h; such that $\beta_{1,h} < \cdots < \beta_{C-1,h}$. It is desirable for the items to have high discrimination to provide confidence to the estimated location of an individual on the scale of the latent trait from their given response.

Let $P_{c,h,i}$ be direct probability that individual i responds with category c, for $c = 2, \ldots, C-1$, where:

$$\begin{aligned} P_{c,h,i} &= \mathbb{P}(Y_{h,i} = c|\eta_{D(h),i}) \\ &= \Phi\left(\alpha_h(\eta_{D(h),i} - \beta_{c,h})\right) - \Phi\left(\alpha_h(\eta_{D(h),i} - \beta_{c+1,h})\right). \end{aligned}$$

For $c = 1$, $P_{1,h,i} = 1$ and for $c = C$, $P_{C,h,i} = \Phi\left(\alpha_h(\eta_{D(h),i} - \beta_{C,h})\right)$.

A key assumption of the item response analysis is of local independence. That is, for a given individual, the items are assumed to be independent conditional through the latent trait, η. For the unidimensional case with p items related to the single latent trait η_m, local independence means the joint probability, conditional on η, can be expressed as the product:

$$\prod_{h=1}^{p} \mathbb{P}(Y_h = y_h|\eta).$$

In the multidimensional case, this is defined by Lord and Novick (1968), and mentioned by Reckase (2009), such that all items are independent conditional on the profile of the latent traits $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_M)$. That is, the joint probability conditional on $\boldsymbol{\eta}$ can be expressed as

$$\prod_{h=1}^{p} \mathbb{P}(Y_h = y_h|\boldsymbol{\eta}).$$

The amount of information provided by item, h about a latent trait η_m, $m \in \{1, \ldots, M\}$, not necessarily $\eta_{D(h),i}$, is measured for any given value of η_m by the equation:

$$I_h(\eta_m) = \frac{\pi^2}{3}\frac{\lambda_h^2}{\theta_h}\sum_{c=1}^{C}\frac{Q_{c,h}(1-Q_{c,h}) - Q_{(c-1),h}(1-Q_{(c-1),h})}{P_{c,h}}$$

where $P_{c,h} = \mathbb{P}(Y_h = c|\eta_m)$ and $Q_{c,h} = \sum_{c=1}^{C} c \times P_{c,h}$ (Asparouhov and Muthén, 2016).

Information about each latent trait, η_m, from all the items is then given by the total information curve. The total information for any given level of η_m is the sum of all the item information curves for that respective latent trait:

$$I\left(\eta_m\right)=\sum_{h=1}^{p} I_h\left(\eta_m\right)$$

It is our desire that the items provide information across the majority of the latent trait, with focus around the mean, given that our analysis is based on the general elderly population. Therefore, a range of difficulty in the items is preferable.

4.3.2 Latent Growth Modelling

Utilisation of the longitudinal properties of the ELSA is important, as said by (Ferrer and Ghisletta, 2011) "longitudinal designs are unambiguous to ageing research with interest in both changes within individuals and comparing the difference in the changes between individuals".

The aim is to create a multivariate latent growth model that analyses the change over time of multiple aspects, and to determine the effect of covariates. Each aspect of subjective well-being is explored in univariate analysis. Firstly, a CFA model was fitted with configural invariance, such that the factor structure was the same at each time point.

A typical problem of longitudinal analysis, especially in the elderly, is drop out. The high levels of drop out in the elderly are often due to deterioration in health or to death. Due to the categorical outcomes, and thus the use of WLSMV estimator, the missing data is assumed to be missing at random only conditional on the covariates and not on the dependent variables (MARX).

The methods described below are given from a structural equation perspective and are drawn from Duncan et al. (1999); Bollen and Curran (2006) and Wang and Wang (2013).

The latent response variables of Equation (4.1) are extended to each time point t of each item h. Let $Y_h^{(t)}$ be the categorical observed variable $h \in \{1, \ldots, H\}$ with categories $c \in \{1, 2, \ldots, C\}$, at time point $t \in \{0, \ldots, T=2\}$, which is assumed discretize a latent continuous variable, $Y_h^{(t)}$ such that;

$$Y_{h,i}^{(t)} = c \quad \Leftrightarrow \quad \tau_c^{(t)} < Y_h^{*(t)} \leq \tau_{c+1,h}^{(t)}$$

where $\tau_0 = -\infty$ and $\tau_C = \infty$. The time points $t = 0, 1, 2$ correspond to the waves $w = 2, 4, 6$ respectively.

A latent growth process, is defined with multiple indicators for each aspect of subjective well-being, as seen in Bollen and Curran (2006). Similarly to Equation (4.3), the measurement model is defined for the single factor η at each time point t so that:

$$\boldsymbol{Y}^{*(t)} = \boldsymbol{\Lambda}^{(t)}\eta^{(t)} + \boldsymbol{\epsilon}^{(t)}$$

where $\boldsymbol{Y}^{(t)}$ is the vector of latent response variables at time t. The vector $\boldsymbol{\Lambda}^{(t)}$ gives the factor loadings for the factor $\eta^{(t)}$ at the respective time point and the residuals $\boldsymbol{\epsilon}^{(t)}$ are assumed to be normally distributed with zero mean and diagonal variance matrix $\boldsymbol{\Theta}^{(t)}$.

The latent variables $\{\eta_1, \ldots, \eta_T\}$ are expressed in terms of the two latent variables α and β that represent a random intercept and a random slope. The latent growth model is therefore defined as:

$$\eta^{(t)} = \alpha + \beta t + \zeta^{(t)}.$$

The random intercept α and random slope β are then expressed as:

$$\alpha = \mu_\alpha + \zeta_\alpha,$$
$$\beta = \mu_\beta + \zeta_\beta$$

for the respective factor means μ_α and μ_β and residuals ζ_α and ζ_β.

In accordance with the configural invariance, all observed variables are repeated at each of the time point, therefore there is residual dependence between the repetitions of the latent response variables $\{Y_h^{(1)}, \ldots, Y_h^{(T)}\}$. Hence, for each item h, $\text{Cov}(\epsilon_h^{(u)}, \epsilon_h^{(v)}) \neq 0$ for all $u, v \in \{1, \ldots, T\}$ where $u \neq v$. The covariance structure of the residuals is defined such that the covariance between to residuals of lag-1 are equal; $Cov(\epsilon_h^{(0)}, \epsilon_h^{(1)}) = Cov(\epsilon_h^{(1)}, \epsilon_h^{(2)})$.

Measurement invariance is an important area of latent growth modelling to ensure consistent measurement of the factor. There are three levels of invariance, weak, strong and strict, that were formally defined by Meredith (1993) and extended to ordinal variables by Millsap and Yun-Tein (2004). All levels of measurement invariance include configural invariance. At the weak level of invariance, the factor loadings are fixed across the time points. Additionally, strong measurement invariance had fixed item thresholds, or item intercepts in the case of continuous outcomes. For strict measurement invariance the residual variance of the items is also fixed.

Strong measurement invariance is imposed on the latent growth models here. Factor loadings and thresholds are constrained to be invariant between time points. The factor loadings are $\lambda_h^{(t)} = \lambda_h$ for all t. For binary variables ($C = 2$), the threshold estimates are fixed such that

$\tau_1^{(t)} = \tau_1$ for all t. In the case of ordinal variables ($C > 2$), the first two thresholds are fixed such that $\tau_1^{(t)} = \tau_1$ and $\tau_2^{(t)} = \tau_2$ for all t. Variables that are repeated are thought to have residual dependence, hence the residual covariance structure mentioned above, therefore, strict measurement invariance was not appropriate.

The latent growth model for each aspect is displayed in the form of a path diagram. A general example is given in Figure 4.2. Similarly to Figure 4.1, latent variables are represented by circles and observed variables by rectangles; with the exception of residuals where each variable is represented by just the residual variance and a small arrow directed towards the respective variable.

Estimated parameters of the latent growth model are the mean and variance of the random intercept α and the random slope β, shown to the left of the respective variable, and the covariance/correlation between the two variables. Estimates of the covariance (and correlation) between the random intercept and random slope are shown on the curved double-ended arrow connecting the two variables. The factor loadings of α are all fixed to one, as indicated on the diagram, and the factor loadings of β are equal to $t \in \{0, 1, 2\}$, representing the time scale.

The measurement models of $\eta^{(2)}, \eta^{(4)}$ and $\eta^{(6)}$ are displayed the same as in Figure 4.1. The factors loadings $\{\lambda_1, \ldots, \lambda_H\}$ are located above the respective arrow, however the factors loadings between the latent response variables and their respective observed variables are all fixed to one and are assumed in the diagram. The residual dependence between the latent response variables, of repeated items, is shown by the curved double-ended arrows connecting the residual variances; the covariance/correlation estimates will be given in a table below the diagram.

In multivariate latent growth models, the growth processes are jointly modelled, where the growth process for each aspect a is;

$$\eta_a^{(t)} = \alpha_a + \beta_a t + \zeta_a^{(t)},$$

with the random intercept α_a and random slope β_a expressed as;

$$\alpha_a = \mu_{\alpha_a} + \zeta_{\alpha_a},$$
$$\beta_a = \mu_{\beta_a} + \zeta_{\beta_a}$$

with respective factor means μ_{α_a} and μ_{β_a} and residuals ζ_{α_a} and ζ_{β_a}.

The covariance structure between the random intercepts $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_A)'$ and random slopes $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_A)'$ is defined such that the covariance between the random intercepts and random slopes of different aspects are constraint to be zero;

$$\mathrm{Cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{pmatrix} \mathrm{Cov}(\boldsymbol{\alpha}) & \mathbf{0} \\ \mathbf{0} & \mathrm{Cov}(\boldsymbol{\beta}) \end{pmatrix}.$$

Additionally, time invariant and time-varying covariates are included in the multivariate latent growth models. The time invariant covariates are modelled to affect the random intercept, α_a, and random slope, β_a, for each aspect a such that

$$\alpha_a = \mu_{\alpha_a} + \sum_{q=1}^{Q} \gamma_{\alpha_a,q} x_q + \zeta_{\alpha_a}$$
$$\beta_a = \mu_{\beta_a} + \sum_{q=1}^{Q} \gamma_{\beta_a,q} x_q + \zeta_{\beta_a}$$

for the time invariant covariates $\{x_1, \ldots, x_Q\}$ with the regression coefficients $\{\gamma_{\alpha_a,1}, \ldots, \gamma_{\alpha_a,Q}\}$ and $\{\gamma_{\beta_a,1}, \ldots, \gamma_{\beta_a,Q}\}$ on the random intercept and random slope respectively. Whereas, the time-varying covariates were modelled to affect the latent variable at each time point, t, such that

$$\eta_a^{(t)} = \alpha_a + \beta_a t + \sum_{u=1}^{U} \kappa_u^{(t)} x_u^{(t)} + \zeta_a^{(t)},$$

for the time-varying covariates $\{x_1^{(t)}, \ldots, x_U^{(t)}\}$ with regression coefficients $\{\kappa_1^{(t)}, \ldots, \kappa_U^{(t)}\}$ respectively.

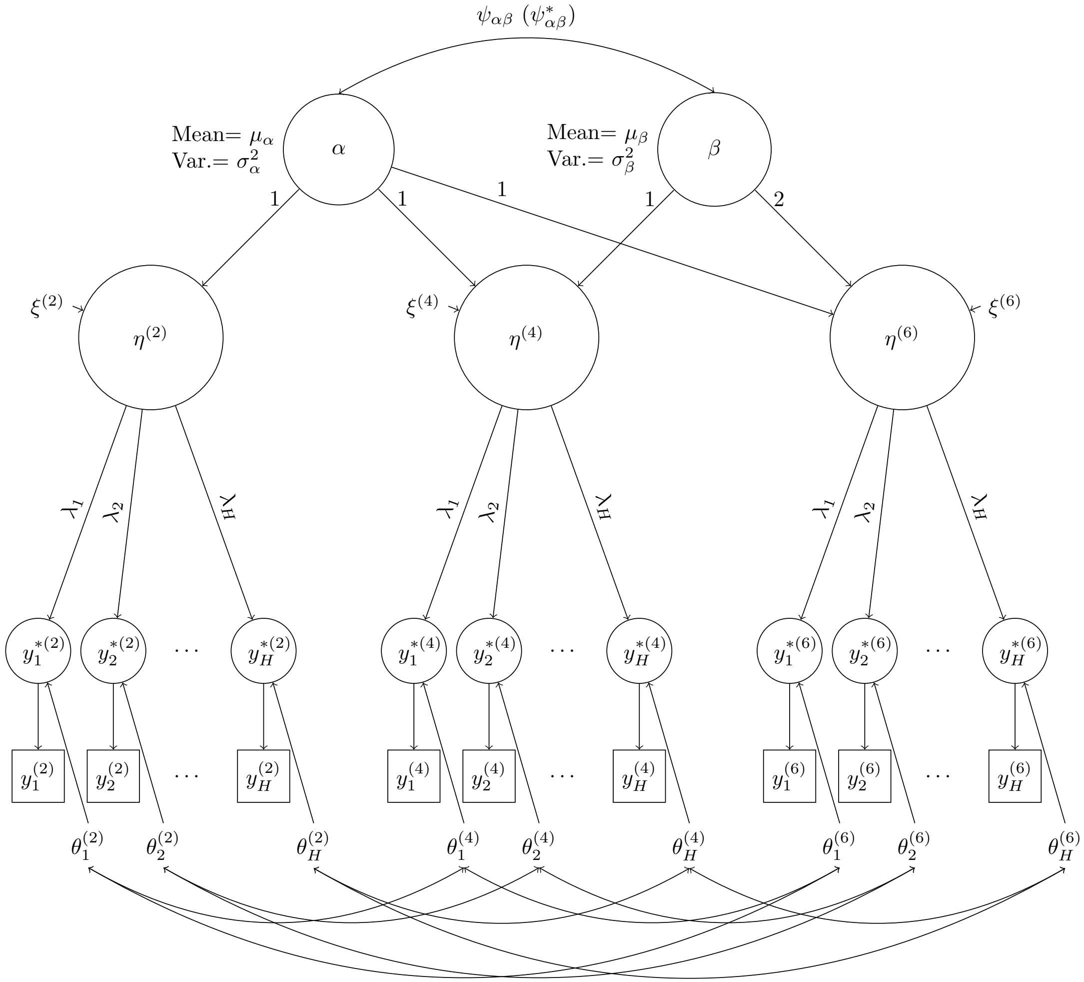


Figure 4.2: Example path diagram for the latent growth process for aspect, a.

Chapter 5

Analysis of Brazil

This chapter contains the results of analyses for Brazil, using the 2008 National Household Sample Survey (PNAD; Pesquisa Nacional por Amostra de Domicílios, in Portuguese). First of the analyses was multilevel cumulative logistic regression of self-rated health, a common method for measuring overall health in the elderly. This revealed several aggregated health and economic variables to be highly important, but consist of unrealistic assumptions; such as all comorbidities having equal impact. Secondly, analyses were aimed at the construction of a non-multilevel structural equation model (SEM) that incorporated measurement models of health and economic well-being. The latent structure of aspects of health and economic well-being were defined using exploratory factor analysis, that were then combined to create a measurement model each for the health and economic well-being domains. Covariates were then included in the non-multilevel SEM to determine differences in the health and economic well-being among the elderly population of Brazil. In the final section, the SEM was extended to a multilevel SEM that incorporated the sector level from the hierarchical structure of the 2008 PNAD, correcting for potential biases. Again, the effects of covariates at both levels were further explored.

5.1 Multilevel Cumulative Logistic Regression of Self-rated Health

Self-rated health (SRH) has been the response variable in the literature with many examples in studies of the elderly. The SRH of the elderly has previously been used to predict mortality and many determinants of SRH have been identified, as discussed in Section 2.3.1. The large number of determinants with significant impact highlights the importance of SRH as an indicator of an elderly person's perception of their own health.

Determinants of an individual's SRH from the household and wider areas, were of key interest in previous studies. Several authors have used multilevel regression for the identification of the determinants of SRH from households and regions (Höfelmann et al., 2014; Pirani and Salvini, 2012) and de Moraes et al. (2011). In a study of the elderly in Italy, Pirani and Salvini (2012) highlighted regional factors of health services to be highly significant, along with aspects of the household such as housing conditions and composition; although they were considered as individual level variables. In Brazil, Höfelmann et al. (2014) and de Moraes et al. (2011) concluded that household and regional factors were important predictors of, such as housing quality, the SRH of the elderly.

Data from the 2008 PNAD has been previously analysed by Höfelmann et al. (2014) and de Moraes et al. (2011). Höfelmann et al. (2014) dichotomised the 5-point Likert scale of the SRH variable into an indicator of 'poor' health, with responses from 'very good' to 'average' grouped together as an indicator of not 'poor' health. The limitations of which were discussed in Section 2.3.1. Höfelmann et al. (2014) then applied a multilevel Poisson model to assess the effect of factors on the prevalence of 'poor' SRH. In contrast, de Moraes et al. (2011) aggregated the variables into a 3-point Likert scale by merging the 'very bad' and 'bad' categories and the 'very good' and 'good' categories, then applied a multilevel cumulative logistic regression.

Since, dichotomising or collapsing levels potentially loses information, we aimed to model the 5-point Likert scale of SRH and assess its association with a variety of explanatory variables. Thus, similarly to de Moraes et al. (2011), a multilevel cumulative logistic regression was applied but with the inclusion of all five categories in the scale of the response variable.

A multilevel model was adopted in order to correct for the hierarchical structure imposed by the complex sampling plan of the PNAD. The sampling plan consisted of individuals sampled from within households, that were sampled from within sectors, and sectors that were sampled from within municipalities. Municipalities were sampled from within strata dependent on their status as a self-representative municipality. Ideally sampling weights would be used for this purpose, however, these were not available as mentioned in Section 3.1.1; where a more detailed description of the sampling plan is also given. This implied the possibility for a four level model of all elderly individuals, or separate analyses for self-representative municipalities and non self-representative municipalities, with a four and five level model respectively; both of which would have been very complicated.

In order to select the required levels, the variance partitioning coefficients (VPCs) were estimated. In all models, the individuals were modelled as Level One. The VPC estimated the amount of variance in SRH that was explained by the higher level. For example, the amount an elderly person's perception of their health changed according to the household they live in.

First the VPC was calculated for each higher level, in turn, from a two level model that included no explanatory variables. For household, sector and municipality all the elderly individual in the sample were modelled. In addition, non self-representative municipalities were grouped together to form strata by their similarities in geographical location and demographics. Thus, the calculation of VPC was then separated for the self-representative municipalities and for clusters of non self-representative municipalities that were within the same stratum.

The estimated VPC are shown in Table 5.1, and were considered adequately large when greater than 5%.

Table 5.1: Variance Partitioning Coefficients (VPCs) for the two level null models of SRH, in the 2008 PNAD.

Random Intercept	Household	Sector	Municipality	SR Strata	Non-SR Strata
Variance	0.905	0.490	0.185	0.093	0.137
VPC	21.4%	13.0%	5.3%	3.3%	4.1%

SR - Self-representative

Household explained the greatest amount of variation in SRH, while strata explained less than 5% for both individuals in self-representative and non self-representative municipalities. Thus, the similarities of non self-representative municipalities used to form strata did not reflect similarities in the SRH of older individuals, so the strata were not considered as a higher level. Household, sector and municipality, however, explained more than 5% of the variation in SRH and were taken forward into the three level models.

The VPC was then estimated for each three level model. The combinations of higher levels were, household and sector, household and municipality or sector and municipality. However, households and sectors were too similar to model together as levels, with 17.7% of sectors containing only one household.

Table 5.2: Variance Partitioning Coefficients (VPCs) for the three level null models of SRH, in the 2008 PNAD.

Random Intercept	Household	Municipality	Sector	Municipality
Variance	0.692	0.217	0.314	0.150
VPC	16.5%	5.2%	8.6%	4.0%

The estimated VPC from Table 5.2 showed the combination of household and municipality, as levels, explained more variance in SRH than sector and municipality. Both levels had a larger VPC in the model with household and municipality, than the model with sector and municipality. Hence, it was decided that a three level model was most appropriate, with level one as individuals, level two as households and level three as municipalities.

Explanatory variables were then examined from the individual, household and municipality level. As mentioned in Section 4.2, an overall weight for the individuals was available in the 2008 PNAD. This weight was used to calculate weighted frequencies and means for the responses to the explanatory variables of interest. Table 5.3 contains a summary of all the variables in terms of the weighted and unweighted responses.

There was little difference between the unweighted and weighted percentages/means, except in the major region variable. The difference in the major regions indicated the sample potentially under-represented the population in the South East region and over-represented the North East region.

Table 5.3: Summary of Explanatory Variables; Categorical Variables with Percentage (%) of Responses, and Continuous Variables with Mean Responses

			% or Mean	
	Variable	Response(s)	Unweighted	Weighted
Individual level				
	Gender			
		Male	39.7%	39.0%
		Female	60.3%	61.0%
	Age			
		60-64 years old	31.6%	31.2%
		65-69 years old	26.2%	25.9%
		70-74 years old	18.8%	19.2%
		75-79 years old	13.0%	13.1%
		80-84 years old	6.7%	6.8%
		85-89 years old	2.7%	2.8%
		≥90 years old	1.0%	1.0%
	Race			
		White	51.6%	56.8%
		Non-white	48.4%	43.2%
	Education			

Table 5.3 *Continued on next page*

Continued from previous page

		% or Mean	
Variable	Response(s)	Unweighted	Weighted
	Less than 1 year or no education	33.0%	30.5%
	1-7 years	43.9%	46.3%
	8-14 years	17.4%	17.4%
	≥15 years	5.7%	5.8%
Occupation			
	Employed	30.3%	28.9%
	Not employed	69.7%	71.1%
Physical activity			
	Participation	20.0%	20.5%
	No participation	72.4%	72.1%
	Not applicable	7.6%	7.4%
Smoking status			
	Non-smoker	50.5%	52.8%
	Ex-smoker	34.5%	33.2%
	Smoker	15.0%	14.1%
Self-reported morbidity			
	No conditions	21.9%	21.5%
	1 condition	28.7%	28.8%
	2-3 conditions	38.7%	38.7%
	4-5 conditions	9.7%	9.9%
	≥6 conditions	1.0%	1.0%
Depression			
	Yes	9.1%	9.7%
	No	90.9%	90.3%
Functional status			
	Without limitation	27.3%	27.6%
	Mild limitation	34.3%	34.5%
	Moderate limitation	22.9%	22.7%
	Great limitation	15.4%	15.1%
Health plan			

Table 5.3 Continued on next page

Continued from previous page

		% or Mean	
Variable	Response(s)	Unweighted	Weighted
	Yes	28.4%	30.3%
	No	71.6%	69.7%
Medical consultation			
	Yes	81.3%	82.0%
	No	18.7%	18.0%
Household level			
Living alone			
	Yes	17.3%	17.5%
	No	82.7%	82.5%
Urban-rural classification			
	Urban	81.8%	84.3%
	Rural	18.2%	15.9%
New construction			
	Yes	0.9%	0.7%
	No	99.1%	99.3%
Household income per capita			
	Less than 1 min. wage or no income	44.7%	41.6%
	1-2 min. wages	28.4%	29.6%
	2-3 min. wages	8.4%	10.2%
	3-5 min. wages	6.8%	7.5%
	≥5 min. wages	7.1%	7.2%
	Not declared	3.7%	3.9%
Household tenure			
	Owned	87.5%	87.2%
	Rented	7.1%	7.5%
	Transfer	4.9%	4.9%
	Other	0.4%	0.4%
Communication goods			
	All	17.9%	18.0%
	3-4	54.7%	54.8%

Table 5.3 *Continued on next page*

Continued from previous page

		% or Mean	
Variable	Response(s)	Unweighted	Weighted
	1-2	26.3%	26.3%
	None	1.1%	1.0%
Essential goods			
	All	8.1%	8.8%
	3-4	63.2%	66.6%
	1-2	28.3%	24.4%
	None	0.4%	0.2%
Housing quality			
	Adequate	65.8%	68.5%
	Inadequate	34.2%	31.5%
Family health program			
	Yes	49.6%	49.0%
	No	50.4%	51.0%
Municipality level			
Self-representative			
	Yes	53.9%	50.4%
	No	46.1%	49.6%
Major region			
	North	9.2%	4.1%
	North East	30.7%	25.3%
	Central West	9.6%	6.3%
	South	17.1%	16.9%
	South East	33.3%	47.4%
Municipality housing quality			
	% of adequate housing	0.66	0.69
Old age dependency ratio			
	$\#$ aged $\geq$60 / $\#$ aged 18-59	0.20	0.21

Table 5.3

At the individual level demographic variables (age, gender, race), socioeconomic (number of years education, current employment) and health variables (physical activity, smoking, number of self-reported morbidity, depression, functional status, possession of a health plan, medical consultation) were analysed. Self-reported morbidity was sum of equally weighted medical conditions that comprised of: column or back disease, arthritis or rheumatism, cancer, diabetes, asthma or bronchitis, hypertension, heart disease, chronic renal failure, tuberculosis, tendonitis or tenosynovitis, and cirrhosis. Functional status was derived from the eight variables of activities of daily living described in detail in Section 3.1.2.

At the household level demographic variables (living alone, urban-rural classification, status as a new construction), socioeconomic (household income per capita, household tenure, number of communication goods, number of essential goods, housing quality) and health variables (self-representative municipality, major region, proportion of quality housing, old age dependency ratio) were also analysed. Living alone included lodgers, domestic employees and relatives of the domestic employees as fellow occupants. The household income per capita was with respect to the minimum wage. The communication goods counted were a land-line telephone, a mobile telephone, a television, a radio and a computer. The essential goods counted were a fridge, a freezer, a stove, a water dispenser and a washing machine. The classification of housing quality was as defined in Chapter 3. The old age dependency ratio is the number of elderly persons divided by the number of people of working age, that indicates the ability of the municipality to provide for its dependent elderly.

The reference categories for the categorical variables was the category with the highest proportion, unless the interpretation was not intuitive. This was the case for the ordinal variables: self-reported morbidity, mobility, communication goods and essential goods. To preserve the order in the interpretation of the categories, the reference category was selected to be the end of the scale. The interpretation was then on a scale for increased morbidity, limitation of mobility or number of goods not owned.

Initially, univariate models were fitted to determine the effect and statistical significance of each explanatory variable. The estimated odds ratios, 95% confidence intervals and p-values from each model are given in Table 5.4. The statistical significance of each variable was tested using a likelihood ratio test; comparing the univariate model to the null model. The likelihood ratio statistics and p-values are given in Table 5.4.

There were 28,057 elderly individuals, 23,657 households, and 817 municipalities with 257 elderly individuals living in a new construction household and 13,556 living in the non self-representative municipalities. The mean age of the respondents was 69.3 years old for both males and females, with a maximum age of 107 and 108 respectively.

All but two of the explanatory variables were statistically significant. Sex and status as a new construction were both not significant variables, in the univariate analysis, although this was marginal for the sex.

From the individual demographic variables, the odds ratios for age showed a steady decrease in SRH with increasing age, until a 'dip' at age 85-89 years old and recovery at aged 90+ years old. The odds ratios were less than 1 and decreased as the age increased. Thus, an individual was increasingly less likely to have SRH as high as those aged 60-64 years old. Race indicated that SRH was lower for non-white individuals, than white.

For education and occupation, the odds ratios showed an increased SRH for the employed and higher educated. Less than one year of education had an odds ratio of 0.67, thus individuals had lower odds of rating their health as a high as individuals with 1-7 years of education. The odds ratios were greater than 1 for those with more than 7 years of education and individuals were more likely to rate their health higher.

Lifestyle choices of physical activity and smoking were both statistically significant. Participation in physical activity improved SRH, with an odds ratio of 2.21, whereas those that were not applicable had much reduced SRH. Status as a smoker or ex-smoker also reduced an individual's SRH.

Self-reported morbidity and limitations in functional status severely reduced the odds of better SRH. The reporting of a single comorbidity had an odds ratio of 0.34, which decreased further as the number of comorbidities increased. Depression also had a similar effect with estimated odds ratio of 0.25. Increasing levels of limitation to functional status decreased the odds ratios from 0.21 with mild limitation to 0.07 with great limitation.

At the household level all but the new construction variable, were statistically significant. Living alone improved SRH with odds ratio of 1.18. Individual's living in rural sectors were less likely to report SRH as high as those living in urban sectors. The odds ratios were greater than 1 and increased as the household income per capita increased, in comparison with households of less than 1 minimum wage. Rented households had significantly improved SRH in comparison to owned households, whereas transferred and other households had no significant difference. Not owning all the communication or essential goods increasingly reduced odds ratios. Living in a household of inadequate quality had odds ratio 0.74 compared with adequate households. The household's participation in the Family Health Program significantly increased the individual's SRH with odds ratio of 1.49.

Demographics of the municipality additionally influenced an elderly individual's SRH. Living in a non self-representative municipality reduced SRH with an odds ratio of 0.76. The odds ratios were also less than 1 for all the major regions, in comparison to the South East. SRH was lowest in the North and North East regions. Municipalities with a higher proportion of adequate quality households and old age dependency ratio had significantly improved SRH.

Table 5.4: Estimates of the odds ratios, 95% confidence intervals and p-values from the univariate multilevel ordinal logistic regressions, of SRH in the 2008 PNAD, including likelihood ratio statistic (LR) and corresponding p-value.

	Variable	Odds Ratio	95% CI	p-value	LR	p-value
Individual level						
Gender					3.6	0.057
	Male	1.049	(0.999 , 1.102)	0.057		
	Female	1				
Age					177.8	< 0.001
	60-64 years old	1				
	65-69 years old	0.914	(0.857 , 0.975)	0.007		
	70-74 years old	0.774	(0.721 , 0.832)	< 0.001		
	75-79 years old	0.678	(0.625 , 0.735)	< 0.001		
	80-84 years old	0.663	(0.596 , 0.736)	< 0.001		
	85-89 years old	0.532	(0.455 , 0.623)	< 0.001		
	90+ years old	0.647	(0.504 , 0.831)	< 0.001		
Race					141.0	< 0.001
	White	1				
	Non-white	0.722	(0.684 , 0.762)	< 0.001		
Education					1581.1	< 0.001
	Less than 1 year	0.673	(0.634 , 0.714)	< 0.001		
	1-7 years	1				
	8-14 years	2.216	(2.063 , 2.381)	< 0.001		
	15+ years	4.833	(4.316 , 5.412)	< 0.001		
Occupation					473.5	< 0.001
	Employed	1.863	(1.759 , 1.974)	< 0.001		
	Not employed	1				
Physical activity					2333.9	< 0.001
	Participation	2.213	(2.071 , 2.363)	< 0.001		
	No participation	1				

Table 5.4 *Continued on next page*

Continued from previous page

Variable	Odds Ratio	95% CI	p-value	LR	p-value
Not applicable	0.147	(0.133 , 0.163)	< 0.001		
Smoking status				75.3	< 0.001
Smoker	0.839	(0.778 , 0.904)	< 0.001		
Ex-smoker	0.791	(0.749 , 0.836)	< 0.001		
Non-smoker	1				
Self-reported morbidity				5781.2	< 0.001
None	1				
1	0.344	(0.319 , 0.370)	< 0.001		
2-3	0.118	(0.109 , 0.128)	< 0.001		
4-5	0.041	(0.036 , 0.046)	< 0.001		
6 or more	0.015	(0.011 , 0.019)	< 0.001		
Depression				991.6	< 0.001
Yes	0.247	(0.226 , 0.271)	< 0.001		
No	1				
Functional status				5015.1	< 0.001
Without limitation	1				
Mild limitation	0.209	(0.195 , 0.225)	< 0.001		
Moderate limitation	0.119	(0.109 , 0.130)	< 0.001		
Great limitation	0.068	(0.061 , 0.075)	< 0.001		
Health plan				506.7	< 0.001
Yes	1.959	(1.845 , 2.080)	< 0.001		
No	1				
Medical consultation				975.3	< 0.001
Yes	1				
No	2.851	(2.660 , 3.056)	< 0.001		
Household level					
Living alone				23.8	< 0.001
Yes	1.180	(1.104 , 1.261)	< 0.001		
No	1				
Urban-rural classification				6.7	0.010
Urban	1				
Rural	0.901	(0.832 , 0.975)	0.010		

Table 5.4 *Continued on next page*

Continued from previous page

Variable	Odds Ratio	95% CI	p-value	LR	p-value
New construction				0.1	0.804
Yes	1.035	(0.791 , 1.353)	0.803		
No	1				
Household income per capita				1458.3	< 0.001
Less than 1 min. wage	1				
1-2 min. wages	1.525	(1.433 , 1.623)	< 0.001		
2-3 min. wages	2.480	(2.258 , 2.724)	< 0.001		
3-5 min. wages	3.600	(3.230 , 4.013)	< 0.001		
5+ min. wages	5.546	(4.963 , 6.198)	< 0.001		
Not declared	2.790	(2.423 , 3.213)	< 0.001		
Household tenure				11.2	0.011
Owned	1				
Rented	1.109	(1.006 , 1.223)	0.037		
Transfer	0.896	(0.795 , 1.010)	0.073		
Other	0.695	(0.462 , 1.045)	0.080		
Communication goods				657.4	< 0001
All	1				
3-4	0.468	(0.435 , 0.503)	< 0.001		
1-2	0.341	(0.312 , 0.372)	< 0.001		
None	0.336	(0.262 , 0.431)	< 0.001		
Essential goods				303.3	< 0.001
All	1				
3-4	0.548	(0.498 , 0.603)	< 0.001		
1-2	0.396	(0.355 , 0.441)	< 0.001		
None	0.398	(0.236 , 0.671)	< 0.001		
Housing quality				84.1	< 0.001
Adequate	1				
Inadequate	0.744	(0.699 , 0.792)	< 0.001		
Family Health Program				176.8	< 0.001
Yes	1				
No	1.492	(1.406 , 1.583)	< 0.001		
Municipality level					

Table 5.4 *Continued on next page*

Continued from previous page

Variable	Odds Ratio	95% CI	p-value	LR	p-value
Self-representative				33.0	< 0.001
Yes	1				
No	0.765	(0.699 , 0.836)	< 0.001		
Major region				5.0	0.025
South East	1				
South	0.916	(0.814 , 1.030)	0.143		
Central West	0.711	(0.612 , 0.826)	< 0.001		
North	0.643	(0.549 , 0.752)	< 0.001		
North East	0.578	(0.521 , 0.641)	< 0.001		
Housing quality proportion	1.517	(1.401 , 1.642)	< 0.001	103.2	< 0.001
Old age dependency ratio	2.016	(1.092 , 3.722)	0.025	5.0	0.025

Table 5.4

All the explanatory variables were then analysed jointly in a full model. Despite being insignificant in the univariate analysis gender was included as a controlling variable. The estimated odds ratios, 95% confidence intervals and p-values are given in Table 5.5. The likelihood ratio statistic and p-value are also given in Table 5, comparing the full model to the model with that variable removed.

As described in Section 4.2.2, a weighted non-multilevel model was also fitted, for comparison purposes. This method took into account the weights given in the 2008 PNAD, but were unfortunately not suitable for multilevel analysis. The aim of this analysis was to compare the parameter estimates to determine whether there was evidence of potential biases.

The results of the weighted cumulative logistic regression are given in the right-hand columns of Table 5.5. Parameter estimates were not dissimilar, with only marginal differences, between the weighted analysis and the multilevel analysis. Therefore, it is plausible that the multilevel analysis adequately adjusts for the unequal sampling probabilities by accounting for the hierarchical structure in the data.

Table 5.5: Estimates of the odds ratio, 95% confidence interval and p-value from the full multilevel ordinal logistic regression, of SRH in the 2008 PNAD, along with the likelihood ratio statistic (LR) and corresponding p-value. Followed by the odds ratio, 95% confidence interval and p-value from a weighted single level ordinal logistic regression of SRH in the 2008 PNAD.

		Multilevel Unweighted Model					Weighted Non-hierarchical Model		
	Variable	Odds Ratio	95% CI	p-value	LR	p-value	Odds Ratio	95% CI	p-value
Individual level									
Gender					315.9	< 0.001			
	Male	0.601	(0.567 , 0.636)	< 0.001			0.620	(0.588 , 0.653)	< 0.001
	Female	1					1		
Age					57.6	< 0.001			
	60-64 years old	1					1		
	65-69 years old	1.167	(1.091 , 1.248)	< 0.001			1.150	(1.081 , 1.222)	< 0.001
	70-74 years old	1.179	(1.093 , 1.271)	< 0.001			1.153	(1.077 , 1.234)	< 0.001
	75-79 years old	1.197	(1.098 , 1.306)	< 0.001			1.198	(1.108 , 1.296)	< 0.001
	80-84 years old	1.396	(1.248 , 1.563)	< 0.001			1.361	(1.230 , 1.507)	< 0.001
	85-89 years old	1.241	(1.052 , 1.465)	0.010			1.198	(1.028 , 1.384)	0.020
	90+ years old	1.832	(1.409 , 2.382)	< 0.001			1.804	(1.415 , 2.301)	< 0.001
Race					5.5	0.019			
	White	1					1		
	Non-white	0.934	(0.882 , 0.990)	0.021			0.956	(0.908 , 1.006)	0.085

Table 5.5 *Continued on next page*

Continued from previous page

	Multilevel Model					Weighted Model		
Variable	Odds Ratio	95% CI	p-value	LR	p-value	Odds Ratio	95% CI	p-value
Education				150.3	< 0.001			
Less than 1 year	0.865	(0.810 , 0.922)	< 0.001			0.839	(0.792 , 0.890)	< 0.001
1-7 years	1					1		
8-14 years	1.389	(1.284 , 1.503)	< 0.001			1.350	(1.258 , 1.450)	< 0.001
15+ years	1.846	(1.612 , 2.114)	< 0.001			1.732	(1.533 , 1.956)	< 0.001
Occupation				80.6	< 0.001			
Employed	1.331	(1.249 , 1.417)	< 0.001			1.287	(1.217 , 1.362)	< 0.001
Not employed	1					1		
Physical activity				508.7	< 0.001			
Participation	1.523	(1.422 , 1.631)	< 0.001			1.430	(1.245 , 1.520)	< 0.001
No participation	1					1		
Not applicable	0.318	(0.279 , 0.362)	< 0.001			0.333	(0.297 , 0.374)	< 0.001
Smoking status				16.5	< 0.001			
Smoker	0.852	(0.787 , 0.923)	< 0.001			0.887	(0.825 , 0.954)	0.001
Ex-smoker	0.988	(0.932 , 1.048)	0.694			1.012	(0.958 , 1.068)	0.675
Non-smoker	1					1		
Self-reported morbidity				2594.3	< 0.001			

Table 5.5

Continued on next page

Continued from previous page

	Multilevel Model					Weighted Model		
Variable	Odds Ratio	95% CI	p-value	LR	p-value	Odds Ratio	95% CI	p-value
None	1					1		
1	0.463	(0.429 , 0.500)	< 0.001			0.508	(0.474 , 0.544)	< 0.001
2-3	0.201	(0.185 , 0.218)	< 0.001			0.237	(0.221 , 0.254)	< 0.001
4-5	0.090	(0.080 , 0.102)	< 0.001			0.116	(0.105 , 0.128)	< 0.001
6 or more	0.044	(0.034 , 0.058)	< 0.001			0.058	(0.046 , 0.074)	< 0.001
Depression				297.2	< 0.001			
Yes	0.449	(0.410 , 0.493)	< 0.001			0.498	(0.459 , 0.540)	< 0.001
No	1					1		
Mobility				1167.1	< 0.001			
Without limitation	1					1		
Mild limitation	0.384	(0.357 , 0.413)	< 0.001			0.410	(0.384 , 0.437)	< 0.001
Moderate limitation	0.271	(0.249 , 0.295)	< 0.001			0.298	(0.277 , 0.320)	< 0.001
Great limitation	0.293	(0.263 , 0.327)	< 0.001			0.323	(0.293 , 0.355)	< 0.001
Health plan				26.5	< 0.001			
Yes	1.207	(1.121 , 1.291)	< 0.001			1.218	(1.145 , 1.295)	< 0.001
No	1					1		
Medical consultation				239.8	< 0.001			

Table 5.5

Continued on next page

Continued from previous page

	Multilevel Model					Weighted Model		
Variable	Odds Ratio	95% CI	p-value	LR	p-value	Odds Ratio	95% CI	p-value
Yes	1					1		
No	1.749	(1.628 , 1.879)	< 0.001			1.642	(1.540 , 1.752)	< 0.001
Household level								
Living alone				35.0	< 0.001			
Yes	1.242	(1.155 , 1.337)	< 0.001			1.229	(1.151 , 1.312)	< 0.001
No	1					1		
Urban-rural classification				0.4	0.555			
Urban	1					1		
Rural	1.026	(0.929 , 1.133)	0.615			1.034	(0.949 , 1.127)	0.445
New construction				1.3	0.251			
Yes	0.849	(0.646 , 1.116)	0.240			0.828	(0.635 , 1.081)	0.165
No	1					1		
Household income per capita				138.8	< 0.001			
Less than 1 min. wage	1					1		
1-2 min. wages	1.186	(1.111 , 1.267)	< 0.001			1.181	(1.114 , 1.252)	< 0.001
2-3 min. wages	1.474	(1.331 , 1.632)	< 0.001			1.442	(1.320 , 1.576)	< 0.001
3-5 min. wages	1.749	(1.549 , 1.974)	< 0.001			1.723	(1.550 , 1.914)	< 0.001

Table 5.5

Continued on next page

Continued from previous page

	Multilevel Model					Weighted Model		
Variable	Odds Ratio	95% CI	p-value	LR	p-value	Odds Ratio	95% CI	p-value
5+ min. wages	1.877	(1.639 , 2.150)	< 0.001			1.801	(1.596 , 2.031)	< 0.001
Not declared	1.657	(1.430 , 1.920)	< 0.001			1.576	(1.386 , 1.791)	< 0.001
Household tenure				5.9	0.118			
Owned	1					1		
Rented	1.099	(0.995 , 1.215)	0.063			1.086	(0.994 , 1.186)	0.069
Transfer	1.069	(0.945 , 1.209)	0.289			1.066	(0.957 , 1.187)	0.244
Other	0.844	(0.562 , 1.272)	0.418			0.779	(0.536 , 1.133)	0.191
Communication goods				46.0	< 0.001			
All	1					1		
3-4	0.829	(0.765 , 0.898)	< 0.001			0.860	(0.801 , 0.923)	< 0.001
1-2	0.735	(0.662 , 0.816)	< 0.001			0.773	(0.705 , 0.847)	< 0.001
None	0.735	(0.562 , 0.961)	0.025			0.729	(0.568 , 0.936)	0.013
Essential goods				12.4	0.006			
All	1					1		
3-4	0.905	(0.819 , 1.001)	0.051			0.939	(0.861 , 1.024)	0.156
1-2	0.826	(0.735 , 0.929)	0.001			0.867	(0.782 , 0.961)	0.007
None	0.874	(0.509 , 1.500)	0.625			0.800	(0.479 , 1.343)	0.196

Table 5.5

Continued on next page

Continued from previous page

	Multilevel Model					Weighted Model		
Variable	Odds Ratio	95% CI	p-value	LR	p-value	Odds Ratio	95% CI	p-value
Housing quality				0.2	0.631			
Adequate	1					1		
Not adequate	0.979	(0.883 , 1.086)	0.391			1.035	(0.940 , 1.140)	0.521
Family Health Program				0.5	0.498			
Yes	1					1		
No	0.980	(0.921 , 1.042)	0.519			0.978	(0.930 , 1.030)	0.405
Municipality level								
Self-representative				3.9	0.050			
Self-representative	1					1		
Non self-representative	0.909	(0.831 , 0.995)	0.038			0.934	(0.884 , 0.986)	0.014
Major region				62.9	< 0.001			
South East	1					1		
South	1.016	(0.906 , 1.139)	0.787			1.077	(1.006 , 1.152)	0.032
Central West	0.902	(0.773 , 1.053)	0.192			0.986	(0.888 , 1.094)	0.785
North	0.817	(0.693 , 0.964)	0.017			0.832	(0.733 , 0.946)	0.005
North East	0.679	(0.609 , 0.756)	< 0.001			0.744	(0.696 , 0.796)	< 0.001

Table 5.5

Continued on next page

Continued from previous page

	Multilevel Model					Weighted Model		
Variable	Odds Ratio	95% CI	p-value	LR	p-value	Odds Ratio	95% CI	p-value
Municipality housing quality	1.105	(0.954 , 1.280)	0.185	1.5	0.227	1.248	(1.096 , 1.422)	< 0.001
Old age dependency ratio	1.597	(0.869 , 2.936)	0.132	2.1	0.143	1.577	(1.088 , 2.285)	0.016
Variances of the Random Intercepts								
Household	0.556							
Municipality	0.124							

Table 5.5

Controlling for the effect of the other explanatory variables sex was statistically significant in the full model. The full model showed males had significantly lower SRH than females with odds ratio of 0.60.

The likelihood ratio statistics revealed some household and municipality level variables to no longer be significant to SRH; that were previously in the univariate analysis. At the household level urban-rural classification, housing quality and Family Health Program were not significant. At the municipality level self-representative, municipality housing quality and old age dependency ratio were not significant.

The estimated odds ratios reduced in effect size from in the full model for the majority of explanatory variables. For example, the odds ratios for education, in the univariate analysis, were 0.67 for less than 1 year, 2.22 for 8-14 years and 4.83 for 15+ years in comparison to the 1-7 year reference category, whereas in the full model the estimated odds ratios were 0.86, 1.39 and 1.88 respectively.

The effect of age was counter to what was seen in the univariate analysis. The odds ratios were greater than one and increased with age so, after accounting for the health variables, elderly individuals rated their health better with older age. This could be the growing adaptation to the ageing process and optimism from survival found by Ocampo (2010). For example, if an 80 year old person had the same health, and socioeconomic, characteristics as a 60 year old then they perceived a more positive outlook of their own health.

One potential of the multilevel approach was to correct for the unequal sampling probabilities from the complex sampling plan. Alternatively to the multilevel approach, the sampling weights could be used in a non-multilevel approach. From Table 5.5, the estimated coefficients from a weighted logistic regression are given, which showed little difference in the estimated parameters

In comparison to females, males were less likely to rate their health as good. This difference between genders, however, was only seen once the other factors were controlled for. There are many changes in the ageing process between males and females, for example females rate their health as poor for a larger proportion of years (Camargos et al., 2009) and have a longer life expectancy. Helmer et al. (1999) showed that the SRH of elderly women were more influenced by their physical health, which was a possible reason for why gender was only significant to the SRH once other aspects of health were taken into account.

Similarly to the results of Franks et al. (2003), those that were of non-white race showed significantly lower SRH than those of white race.

All the individual health characteristics that were considered were highly important to an elderly person's perception of their own health. The deterioration of health measured through the number of comorbidities and the presence of depression had a strong influence; those with the presence of just one comorbidity, including depression, reduced the odds of a better health rating by over 50%. Those reporting four or more comorbidities were over 90% less unlikely to rate their health any better than very bad, than those with none. The impact of health problems on an elderly person's ability to independently perform basic and instrumental activities of daily living varied. Any level of limitation saw a significantly reduced likelihood of a better health rating; however, those with the most limitation had a higher odds ratio than those with only moderate limitation. This could have been related to the amount of help the individual received, due to their health problems, as those with great limitation were unable to, or had great difficulty, in performing basic tasks such as eating, toileting and bathing; therefore, probably also received help completing the survey. Maintaining independence also proved beneficial as living alone significantly increased the odds of an elderly individual having a better health rating, although it was however possible that living alone was conditional upon the person's health status. The possession of a health plan and not seeking a medical consultation both increased the odds that an elderly person would have a better perception of their own health. Whereas, the effect of being a smoker had a significantly negative effect on how an elderly person perceived their health, although an ex-smoker rated their health similarly to that of a non-smoker.

There were the several aspects of the household that influenced the SRH of elderly residents. Household income per capita was positively associated with the SRH of the elderly, with those earning five or more times the minimum wage per capita having higher chance of having good health. The perception of health for an elderly person was also reduced by the household having all at least one of the four communication goods; landline telephone, mobile telephone, television or radio. Similarly, the households lacking at least one of the five essential goods had a significantly negative impact upon the SRH of the elderly, however, those with none of the goods had slightly higher odds than those with one or two. It was possible that the households that do not possess any of these variables were located in areas of communal living or without electricity.

The aggregation of the self-reported morbidity, functional status and housing quality variables removes the ability to determine the different effect of each individual element and makes unrealistic assumption of equal weighting. For example, the aggregated variable of self-reported morbidity gave unit weight to all of the comorbidities, therefore assuming the effect of a progressive or debilitating comorbidity, such as arthritis, that limits an elderly person's independence and ability to participate in social activities, had the same impact on perception of health as a life threatening condition such as cancer or tuberculosis.

Structural equation modelling allows for the inclusion of each of the elements as a separate variable. Inclusion of every element as a separate covariate in the regression would have the assumption that the variables were independent and would suffer from collinearity.

Hence, the following sections look at exploring the latent structures behind each aspect, self-reported morbidity, functional status, housing quality and durables owned, and the relationships between the aspects and the effect of covariates on the aspects.

5.2 Structural Equation Modelling

This section will look at measuring the latent constructs, health and economic well-being, of the elderly in Brazil, using the data available in the PNAD, and observing the relationship between the domains. A similar analysis will later be presented for England so as to compare between the nations.

As discussed in Chapter 4, the PNAD contained four aspects related to health and economic well-being. Relating to health there were the 12 variables of self-reported morbidity and the eight variables of functional status, and relating to economic well-being there were six variables of housing quality and nine variables of durables owned within the household.

The aim was to construct a measurement model for health and a measurement model for economic well-being using these variables, that later form a SEM that determines the relationships between health, economic well-being, and a selection of covariates. Firstly, exploratory factor analysis (EFA) identified the latent structure within each aspect, then those structures were fed forward into a measure of the respective domain; health or economic well-being. The SEM was then formed to analyse both the relationship between the two constructs and how each was affected by a selection of covariates such as age and gender.

5.2.1 Measurement Model of Health

Latent structures underlying each of the two aspects of health, self-reported morbidity and functional status were determined through separate EFA. Firstly, EFA was applied to the self-reported morbidity variables: disease of the column or back, arthritis or rheumatism, cancer, diabetes, asthma or bronchitis, hypertension, heart disease, chronic renal failure, depression, tuberculosis, tendonitis or tenosynovitis, and cirrhosis. Then to the functional status variables: eat, take a bath or go to the bathroom; run, lift heavy objects, play sports or do heavy jobs; push table or perform home repairs; climb hill or stairs; stooping, kneeling or bending; walking more than a kilometre; walking about 100 metres; and make purchases of food, clothing and medicines.

	Back or column disease	Arthritis	Cancer	Diabetes	Asthma or bronchitis	Hypertension	Heart disease	Chronic renal failure	Depression	Tuberculosis	Tendonitis or tenosynovitis	Cirrhosis
Back or column disease	1.000											
Arthritis	0.490	1.000										
Cancer	0.059	0.089	1.000									
Diabetes	0.073	0.142	0.028	1.000								
Asthma or bronchitis	0.201	0.205	0.136	0.103	1.000							
Hypertension	0.170	0.207	0.051	0.353	0.094	1.000						
Heart disease	0.176	0.217	0.113	0.246	0.214	0.435	1.000					
Chronic renal failure	0.256	0.226	0.212	0.210	0.209	0.202	0.284	1.000				
Depression	0.295	0.287	0.160	0.156	0.269	0.238	0.303	0.276	1.000			
Tuberculosis	0.096	0.082	0.110	0.087	0.272	0.026	0.165	0.246	0.218	1.000		
Tendonitis or tenosynovitis	0.369	0.421	0.112	0.108	0.216	0.128	0.167	0.215	0.369	0.168	1.000	
Cirrhosis	0.114	0.075	0.231	0.172	0.213	-0.030	0.105	0.242	0.176	0.422	0.146	1.000

0.000 0.250 0.500 0.750 1.000

Table 5.6: Estimated tetrachoric correlation coefficients for the self-reported morbidity variables, in the 2008 PNAD.

Self-reported Morbidity

The self-reported morbidity variables demonstrated some visible clustering in the estimated tetrachoric correlation coefficients given in Table 5.6.

For example, back or column disease had higher correlation with arthritis and tendonitis or tenosynovitis than with cancer or tuberculosis, whereas tuberculosis had higher correlations with asthma or bronchitis and chronic renal failure.

This was confirmed by the eigenvalues and model fit indices, shown in Table 5.7. The eigenvalues were less than one after three factors, indicating three factors to sufficiently fit the data. The Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) also suggested good fit with three factors. The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) showed good fit with two or more factors.

Factor(s)	Eigenvalue	RMSEA	SRMR	CFI	TLI
1	3.213	0.038	0.085	0.834	0.797
2	1.459	0.017	0.071	0.973	0.959
3	1.315	0.008	0.028	0.995	0.991
4	0.935	0.007	0.023	0.998	0.994

Table 5.7: Eigenvalues and model fit indices from the EFA of the self-reported morbidity variables, in the 2008 PNAD.

The three factors were supported by the estimated factor loadings, shown in Table 5.8. The factor loadings indicated a clear simple structure between the variables and the three factors. Factor 1 had large factor loading with the comorbidities related to the musculature and skeleton, also known as musculoskeletal disorders (Mai et al., 2005). The musculoskeletal disorders were: disease of the back or column, arthritis or rheumatism and tendonitis or tenosynovitis. Factor 2 had large factor loading with three comorbidities, two were metabolic disorders that are known to be associated with the third comorbidity, heart disease (Pluhar, 2014; He et al., 2006). Factor 3 loaded onto the remaining infectious and chronic, but potentially life-threatening, comorbidities; cancer, asthma or bronchitis, chronic renal failure, depression, tuberculosis and cirrhosis. It was decided to name the underlying concept for these factors as ‘Musculoskeletal health’, ‘Metabolic and cardiovascular health’ and ‘Infectious and chronic health’.

Functional Status

The variables concerning functional status and ADLs had an additional complication with missingness as discussed in Section 3.1.2. The missingness of variables 2-8 was dependent on

	Factor		
	1	2	3
Disease of the back or column	**0.666**	-0.003	0.015
Arthritis or rheumatism	**0.720**	0.049	-0.023
Cancer	0.016	-0.009	**0.322**
Diabetes	-0.030	**0.403**	0.140
Asthma or bronchitis	0.170	0.009	**0.384**
Hypertension	-0.032	**0.855**	-0.008
Heart disease	0.026	**0.461**	0.298
Chronic renal failure	0.159	0.157	**0.383**
Depression	0.269	0.136	**0.393**
Tuberculosis	-0.030	-0.070	**0.590**
Tendonitis or tenosynovitis	**0.527**	-0.039	0.201
Cirrhosis	-0.022	-0.106	**0.575**

Bold highlight for factor loadings > 0.316

Table 5.8: Estimated factor loadings from the EFA with three factors, for the self-reported morbidity variables, in the 2008 PNAD.

either variable 1 or variable 6. Illustrated previously, 4.5% of elderly individuals were not eligible for variables beyond variable 1 as they had high levels of difficulty with eating, bathing or using the bathroom, and a further 55.4% were not eligible for variables 7 and 8 as they had no difficulty walking 1km (variable 6). Variables 7 and 8 were dropped from the analysis as the missingness was too high. For variables 2-6, an additional category was created that was imputed as the response from ineligible individuals. This category indicated a greater level of difficulty with the activities than those that were eligible.

	Eat, bath & toilet	Run, lift, play sport & heavy/ jobs	Push furniture & home repairs	Climb hill/stairs	Stoop, kneel & bend	Walk 1km
Eat, bath & toilet	1.000					
Run, lift, play sport & heavy jobs	0.742	1.000				
Push furniture & home repairs	0.811	0.844	1.000			
Climb hill/stairs	0.788	0.826	0.834	1.000		
Stoop, kneel & bend	0.781	0.787	0.799	0.849	1.000	
Walk 1km	0.810	0.791	0.811	0.866	0.831	1.000

0.000 0.250 0.500 0.750 1.000

Table 5.9: Estimated polychoric correlation coefficients for the activities of daily living variables, in the 2008 PNAD.

Table 5.9 shows the estimated polychoric correlation coefficients between the remaining six variables. The correlations were relatively high ranging from 0.742 to 0.866, with the highest correlation between ‘climb hill/stairs’ and ‘Walk 1km’.

The eigenvalues decreased from 5.058 for one factor to 0.268 for the second factor. This large drop to below 1.000 indicated that one factor would be sufficient, as supported by most of the model fit indices SRMR=0.016, CFI=0.997 and TLI=0.996. However, the RMSEA=0.089 indicated only satisfactory fit, at best. The estimated factor loadings were large for all the variables, suggesting the lack of fit may be from the imputed values. The imputed values removed all measurement error of the later variables by imposing the same value, which was against the assumption of conditional independence. This inflated the correlation between the variables, thus inflated the factor loadings. A possible name for the underlying concept is ‘Functional status’ (Schulz, 2006).

	Factor loading
Eat, bath & toilet	**0.867**
Run, lift, play sport & heavy jobs	**0.889**
Push furniture & home repairs	**0.909**
Climb hill/stairs	**0.934**
Stoop, kneel & bend	**0.899**
Walk 1km	**0.915**

Bold highlight for factor loadings > 0.316

Table 5.10: Estimated factor loadings from the EFA with one factor, for the functional status variables in the 2008 PNAD.

Health Measurement Model

The interrelationships between the three factors from self-reported morbidity and one factor of functional status were combined into form the measurement model of overall health. The correlation between the factors, shown in Table 5.11, were estimated in a confirmatory factor analysis, which were both moderate and positive. This illustrated the commonality between the factors that could be explained by an overall second order factor, but the correlation was not so large that the could be considered as repetitions.

Due to the imputed category values in the functional status variables, the factor loading for the second variable rather than the first was fixed to one. It was felt this variable better represented the scale of all these latent response variables.

	Musculoskeletal health	Metabolic and cardiovascular health	Infectious and chronic health	Functional status
Musculoskeletal health	1.000			
Metabolic and cardiovascular health	0.401	1.000		
Infectious and chronic health	0.693	0.634	1.000	
Functional status	0.507	0.491	0.510	1.000

0.000 0.250 0.500 0.750 1.000

Table 5.11: Model estimated factor correlation from confirmatory factor analysis of health in the 2008 PNAD.

The estimated model is given in the path diagram given in Figure 5.1, with the standard errors and R-squared estimates for the respective variables given in Table 5.12. Model was very good with RMSEA= 0.025, CFI= 0.996 and TLI= 0.996.

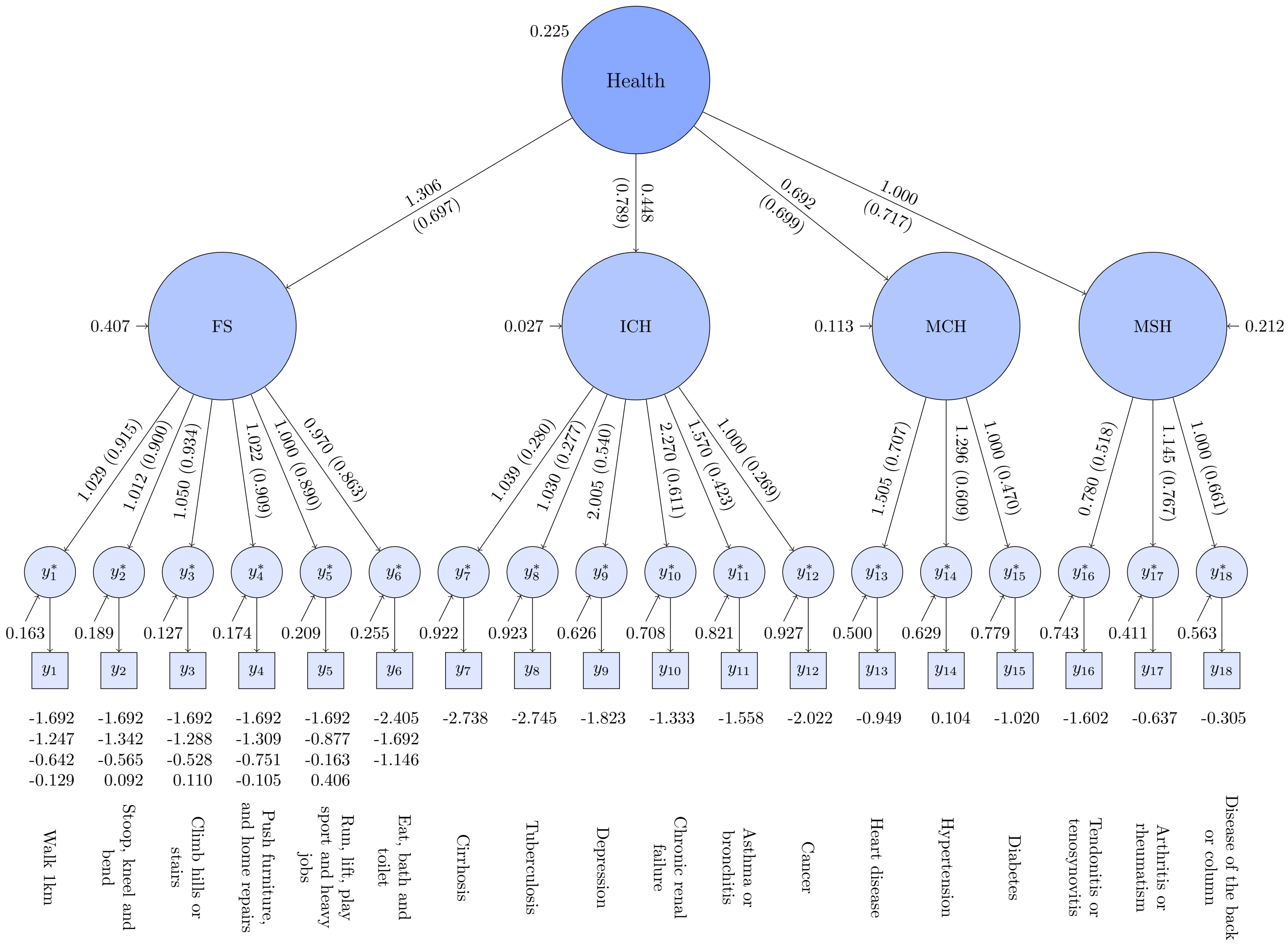


Figure 5.1: Path diagram for the estimated measurement model of health, in the 2008 PNAD.
Abbreviations: FS, Functional status; ICH, Infectious and chronic health; MCH, Metabolic and cardiovascular health; and MSH, Musculoskeletal health.

Variable	Standard Error of Factor Loadings Original	Bootstrapped	R-Squared
Musculoskeletal health			
Disease of the back or column	-	-	0.437
Arthritis or rheumatism	0.025	0.025	0.589
Tendonitis or tenosynovitis	0.027	0.027	0.257
Metabolic and cardiovascular health			
Diabetes	-	-	0.221
Hypertension	0.044	0.042	0.371
Heart disease	0.051	0.050	0.500
Infectious and chronic health			
Cancer	-	-	0.073
Asthma or bronchitis	0.156	0.164	0.179
Depression	0.216	0.236	0.374
Chronic renal failure	0.197	0.208	0.292
Tuberculosis	0.215	0.225	0.077
Cirrhosis	0.211	0.236	0.078
Functional status			
Eat, bath & toilet	0.004	0.006	0.745
Run, lift, play sport & heavy jobs	-	-	0.791
Push furniture & home repairs	0.002	0.002	0.826
Climb hill/stairs	0.002	0.003	0.873
Stoop, kneel & bend	0.002	0.003	0.811
Walk 1km	0.002	0.003	0.837
Health			
Musculoskeletal health	-	-	0.515
Metabolic and cardiovascular health	0.026	0.025	0.489
Infectious and chronic health	0.042	0.044	0.623
Functional status	0.034	0.033	0.485

Table 5.12: Estimated standard errors for factors loadings, original and bootstrapped, and R-squared estimates for the variables in the measurement model of health, for the 2008 PNAD.

Infectious and chronic health was the strongest indicator of overall health, with the largest estimated standardised factor loading. The small variance of the factor reflected the low diagnosis of more than one comorbidity in the variables of the infectious and chronic health factor (2.77%). The most influential comorbidities were the chronic renal failure and depression variables, while cancer was the least influential.

The factor loadings of the other latent responses were relatively large, with the ‘climb hills or stairs’ variable having the highest influence on the factor. Although, the factors loadings for all six variables were very large. The smallest standardised factor loading was for the basic activities variable ‘eat, bath and toilet’ with an estimate of 0.863, which was still large. Functional status, however, was the factor with least association with overall health of the elderly.

Musculoskeletal health, and metabolic and cardiovascular health factors both had strong associations with overall health, but had comparatively small associations with the respective latent response variables. The largest association was with the arthritis or rheumatism variable where the factor explained 58.9% of the variance. By comparison, only 22.1% of the variance in the diabetes variable was explained by the metabolic and cardiovascular health factor.

On average, the bootstrapped standard errors were by 13.4%. The largest percentage increases were in the functional status variables where the standard errors were very smal,l therefore a 50% increase corresponded to an increase in value by 0.001.

5.2.2 Measurement Model of Economic Well-being

The housing quality and durables owned variables were dichotomised to indicate whether the variable of housing quality was adequate or not, and whether the durable was owned by the household. The housing quality variables consisted of; the construction material of the walls, the construction material of the roof, the origin of the water supply, the sewage management system and if the toilet was located within the household, the disposal method for rubbish and method of illumination within the household. The number of residents per room was also considered, as by de Moraes et al. (2011), however there were only seven observations categorised that were not adequate, with two or more residents per room, which was insufficient. The durables considered were: stove, fridge, television, radio, landline telephone, water dispenser, freezer, washing machine and computer.

Exploration of the variables revealed adequate method of illumination (i.e. electrical) was perfectly correlated with to the possession of electrical durables, such as fridges or televisions. Therefore, the illumination variable was removed to prevent double counting.

Housing Quality

The estimated tetrachoric correlation coefficients for the housing quality variables are given in Table 5.13. These indicated higher correlation in two clusters, firstly, between variables subject to the household construction, and secondly between variables subject to services usually provided by the sector.

The resulting eigenvalues and model fit indices, shown in Table 5.14 indicated a single factor model was sufficient. Despite the eigenvalues and CFI indicating that a single factor model had good fit the RMSEA, SRMR and TLI all indicated a two factor model was required for a good

	Wall construction	Roof construction	Water supply	Sewage management	Rubbish disposal
Wall construction	1.000				
Roof construction	0.710	1.000			
Water supply	0.510	0.429	1.000		
Sewage management	0.480	0.280	0.638	1.000	
Rubbish disposal	0.567	0.533	0.870	0.763	1.000

0.000 0.250 0.500 0.750 1.000

Table 5.13: Estimated tetrachoric correlation coefficients for housing quality variables, in the 2008 PNAD.

fit. However, as there are only five variables one of the two factors loaded onto only two or less variables, as was shown in the estimated factor loadings in the two factor analysis.

Factor(s)	Eigenvalue	RMSEA	SRMR	CFI	TLI
1	3.337	0.055	0.102	0.992	0.983
2	0.923	0.034	0.026	0.999	0.993

Table 5.14: Eigenvalues and model fit indices from the EFA of the housing quality variables in the 2008 PNAD.

For a model with one factor, all the variables had strong factor loadings, except the ownership of a water dispenser. The ownership of a water dispenser had an estimated factor loading of 0.241, while ownership of a fridge had the highest loading of 0.893.

Durables owned

The estimated tetrachoric correlation coefficients of the durables owned are shown in Table 5.15. Variables such as a landline telephone, washing machine and computers had the highest correlations, whereas water dispenser had low correlation with many of the variables.

The eigenvalues and model fit indices, shown in Table 5.16, gave a mixed indication for the number of factors. The RMSEA indicated very good fit with only one factor, while the others indicated two factors were required for a good fit. Thus, the estimated factor loadings were considered in the decision to have a single factor model. The exploratory factor model highlighted a second factor with only radio and freezer as plausible indicators. This would have raised identification issues later.

	Stove	Fridge	Television	Radio	Landline telephone	Water dispenser	Freezer	Washing machine	Computer
Stove	1.000								
Fridge	0.648	1.000							
Television	0.529	0.769	1.000						
Radio	0.343	0.333	0.335	1.000					
Landline telephone	0.477	0.738	0.592	0.353	1.000				
Water dispenser	0.358	0.248	0.196	0.145	0.252	1.000			
Freezer	0.283	0.242	0.379	0.302	0.370	0.036	1.000		
Washing machine	0.480	0.710	0.554	0.391	0.702	0.137	0.475	1.000	
Computer	0.419	0.629	0.568	0.433	0.690	0.189	0.409	0.648	1.000

0.000 0.250 0.500 0.750 1.000

Table 5.15: Estimated tetrachoric correlation coefficients for the durables owned variables, in the 2008 PNAD.

Factor(s)	Eigenvalue	RMSEA	SRMR	CFI	TLI
1	4.641	0.033	0.063	0.986	0.981
2	1.081	0.024	0.043	0.995	0.990
3	0.838	0.020	0.031	0.998	0.993

Table 5.16: Eigenvalues and model fit indices from the EFA of the durables owned variables, in the 2008 PNAD.

Economic Well-being Measurement Model

The intention was to create a measurement model of a single economic well-being factor, however, with only two sections, each with a single factor, a model with an overall economic well-being construct was not identified. Therefore, the choice was to maintain the two factors previously identified or to merge the variables as a collaborative set of indicators of economic well-being. The decision was taken to maintain a single economic well-being factor, that was measured by the combination of variables from housing quality and durables owned.

The path diagram with parameters estimates are shown in Figure 5.2, alongside the standard errors and given in Table 5.17.

A good model fit was indicated by CFI = 0.940 and TLI = 0.929, however the RMSEA = 0.062 indicated only an adequate model fit.

The most influential indicators of economic well-being were adequate water supply and rubbish disposal, and the ownership of a fridge and landline telephone. Water supply and rubbish disposal were sector based services highlighting the importance of the locality to the economic well-being of the individual.

Ownership of a radio, water dispenser or freezer were poorer indicators of economic well-being, with smaller standardised factor loadings. There was a lower frequency of ownership of a freezer that could explain the lack of information. However, this could not explain the low factor loadings of the other indicators as 88.5% and 55.4% owned a radio or water dispenser respectively.

Bootstrapped estimates of the standard errors, shown in Table 5.17, were on average 11.0% larger than the asymptotically estimated standard errors. This corresponded to an increase of 0.003, which can be considered relatively small. The correction in the standard errors through bootstrapping was therefore relatively small, in both the health and economic well-being measurement models, and the estimates asymptotic standard errors were considered to give adequate representation. Given the small frequencies of some of the observed variables and the large models, bootstrapping was not possible for the models to follow.

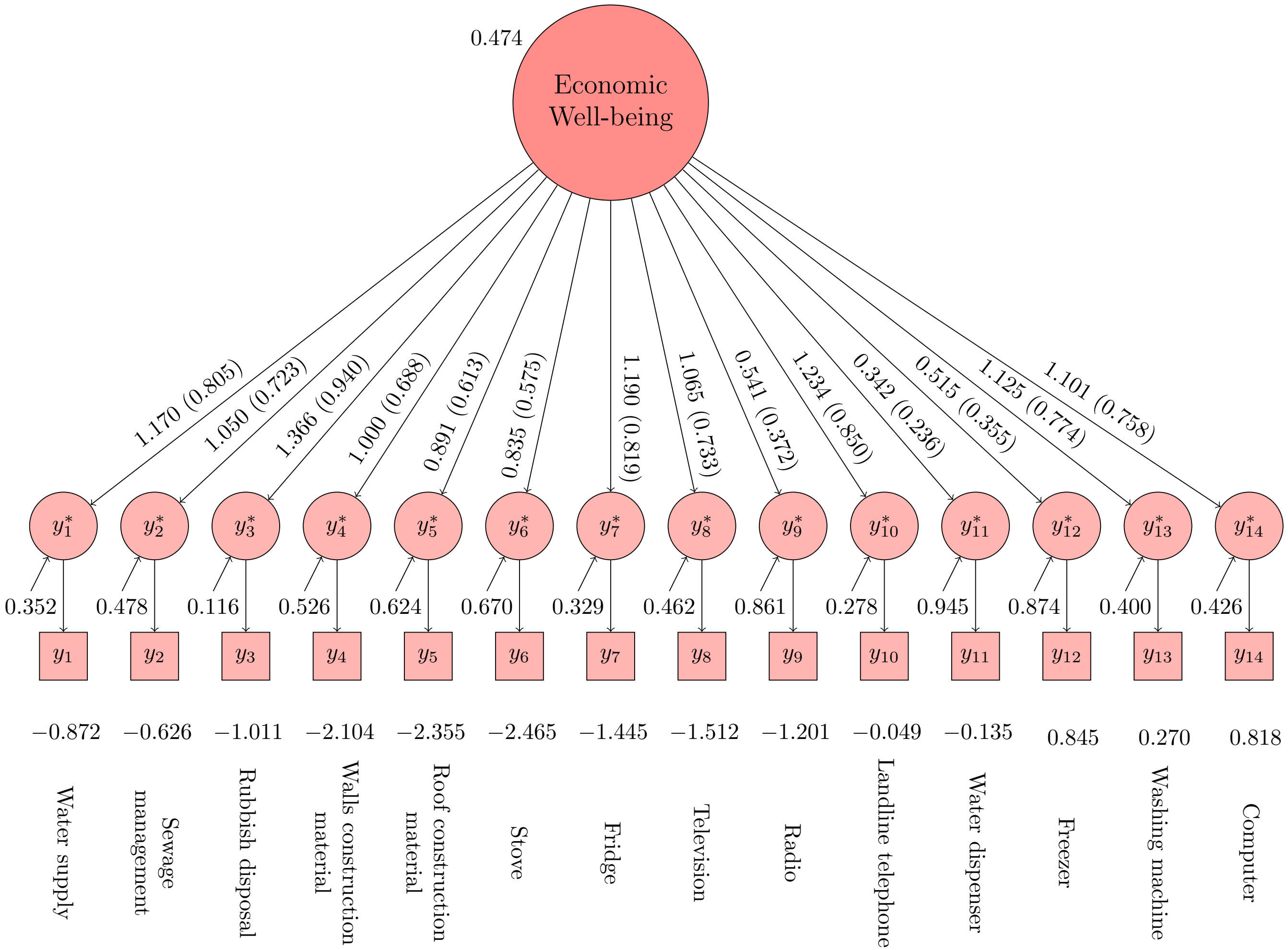


Figure 5.2: Path diagram for the estimated measurement model of economic well-being, in the 2008 PNAD.

Parameter	Standard Error of Factor Loadings Original	Bootstrapped	R-Squared
Water supply	0.026	0.029	0.648
Sewage management	0.024	0.027	0.522
Rubbish disposal	0.029	0.033	0.884
Walls construction material	-	-	0.474
Roof construction material	0.035	0.036	0.376
Stove	0.036	0.038	0.330
Fridge	0.027	0.029	0.671
Television	0.025	0.028	0.538
Radio	0.017	0.019	0.139
Landline telephone	0.027	0.030	0.722
Water dispenser	0.013	0.014	0.055
Freezer	0.016	0.019	0.126
Washing machine	0.025	0.028	0.600
Computer	0.025	0.029	0.574

Table 5.17: Standard errors of the unstandardised factor Loadings, original and bootstrapped, and R-squared estimates for the variables in the measurement model of economic well-being, for the 2008 PNAD.

5.2.3 Structural Equation Model

The interrelationship between health and economic well-being of the Brazilian elderly was modelled in the SEM that combined the measurement models of health and economic well-being. The association between health and wealth has been well established throughout history with poor health causing lower wealth via the inability to work and vice versa through the inability to afford healthcare, maintain hygienic living conditions or sustain an adequate diet. Thus, the domains were correlated and no causal direction was defined.

The tuberculosis or cirrhosis variables of self-reported morbidity were removed from the measurement model of health. Due to the low frequencies of tuberculosis and cirrhosis, some bivariate tables with these two variables contained empty cells. Testing the measurement model without the tuberculosis and cirrhosis variables showed, although marginally reduced, the model fit remained very good with RMSEA=0.030, CFI=0.997 and TLI=0.997.

The path diagram of the estimated SEM with parameter estimates is given in Figure 5.3 with corresponding standard errors and R-squared estimates in Table 5.18.

Association between health and economic well-being was positive and although statistically significant, relatively small. The positive correlation demonstrated that better health in the elderly was associated with better economic well-being, and vice versa, with poorer status in one domain also associated with poorer status in the other.

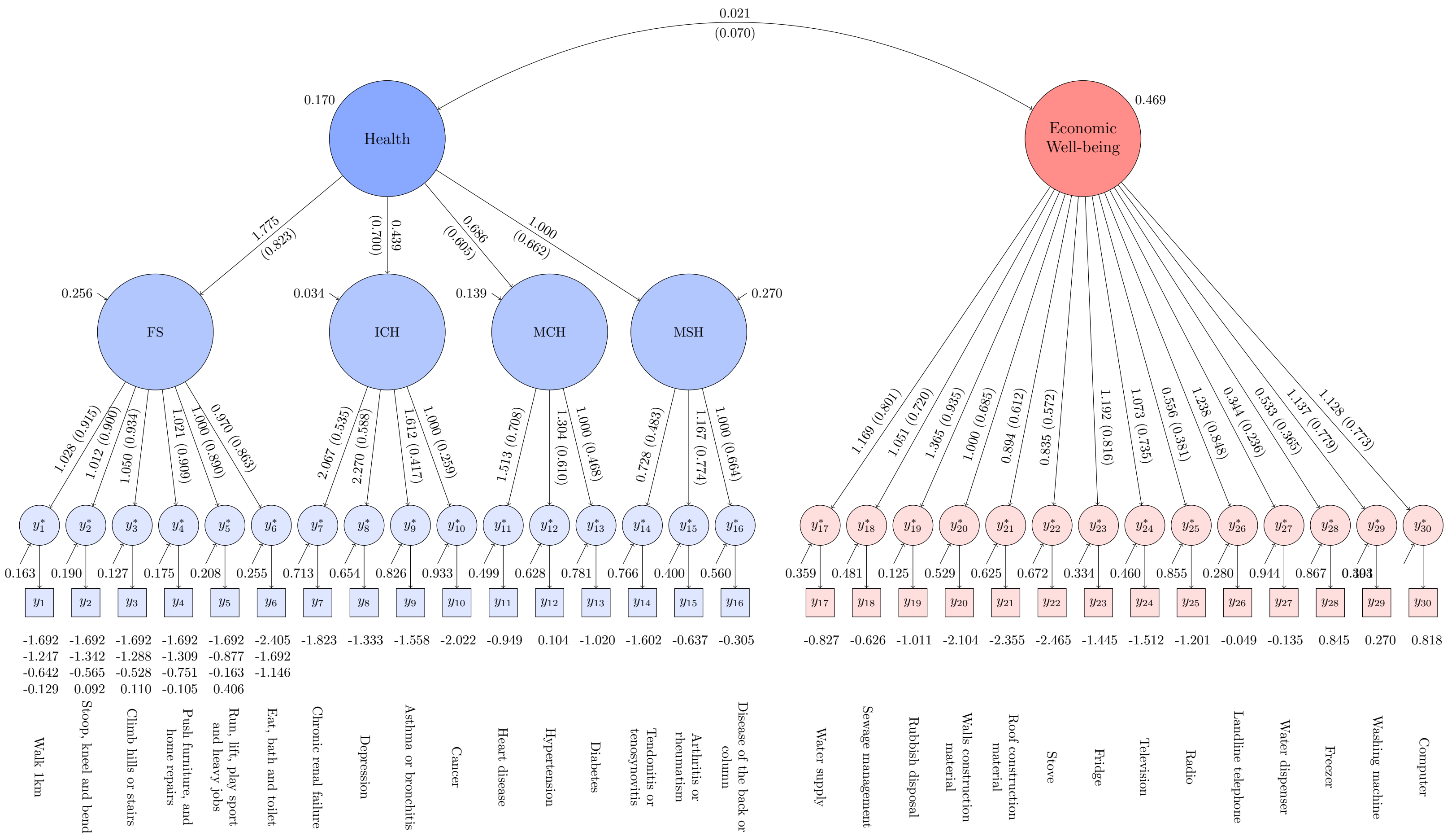


Figure 5.3: Path diagram for the estimated SEM, for the 2008 PNAD.
Abbreviations: FS, Functional status; ICH, Infectious and chronic health; MCH, Metabolic and cardiovascular health; and MSH, Musculoskeletal health.

Variable	Standard Errors	R-Squared
Musculoskeletal health		
Disease of the back or column	-	0.440
Arthritis or rheumatism	0.025	0.600
Tendonitis or tenosynovitis	0.028	0.234
Metabolic and cardiovascular health		
Diabetes	-	0.219
Hypertension	0.045	0.372
Heart disease	0.053	0.501
Infectious and chronic health		
Cancer	-	0.067
Asthma or bronchitis	0.172	0.174
Depression	0.232	0.346
Chronic renal failure	0.217	0.287
Functional status		
Eat, bath & toilet	0.004	0.745
Run, lift, play sport & heavy jobs	-	0.792
Push furniture & home repairs	0.002	0.825
Climb hill/stairs	0.002	0.873
Stoop, kneel & bend	0.002	0.810
Walk 1km	0.002	0.837
Health		
Musculoskeletal health	-	0.387
Metabolic and cardiovascular health	0.026	0.366
Infectious and chronic health	0.044	0.490
Functional status	0.051	0.677
Economic well-being		
Water supply	0.026	0.641
Sewage management	0.024	0.519
Rubbish disposal	0.030	0.875
Walls construction material	-	0.469
Roof construction material	0.035	0.375
Stove	0.036	0.328
Fridge	0.027	0.666
Television	0.025	0.540
Radio	0.018	0.133
Landline telephone	0.027	0.720
Water dispenser	0.013	0.056
Freezer	0.017	0.133
Washing machine	0.026	0.606
Computer	0.026	0.597

Table 5.18: Estimated standard errors of the unstandardised factor loadings, R-squared estimates and scale factors for the parameters of the SEM for the 2008 PNAD.

The next SEM estimated the effect of covariates on health and economic well-being in the Brazilian elderly. These covariates included the demographics of the individual, age, gender and race, and sector; urban-rural classification and major region.

Covariate	Health		Economic Well-being	
Age	Est.	(s.e.)	Est.	(s.e.)
60-64 years old	Ref.		Ref.	
65-69 years old	**-0.060**	(0.007)	**-0.046**	(0.010)
70-74 years old	**-0.142**	(0.008)	**-0.061**	(0.011)
75-79 years old	**-0.228**	(0.010)	**-0.088**	(0.013)
80-84 years old	**-0.320**	(0.013)	**-0.104**	(0.016)
≥85 years old	**-0.427**	(0.017)	**-0.198**	(0.021)
Gender				
Female	Ref.		Ref.	
Male	**0.160**	(0.007)	**-0.030**	(0.008)
Race				
White	Ref.		Ref.	
Non-white	**-0.024**	(0.006)	**-0.246**	(0.011)
Urban-rural				
Urban	Ref.		Ref.	
Rural	0.001	(0.007)	**-1.037**	(0.032)
Major Region				
South East	Ref.		Ref.	
South	**-0.045**	(0.008)	**-0.051**	(0.017)
Central West	**-0.048**	(0.010)	**-0.267**	(0.020)
North	**-0.034**	(0.010)	**-0.474**	(0.016)
North East	**-0.023**	(0.007)	**-0.449**	(0.011)

Bold highlight of regression coefficients significant at 5% level.

Table 5.19: Estimated regression coefficients from the SEM with covariates from the 2008 PNAD.

Health was best for white males, aged 60-64 years old, living in the South East region. In comparison to the South elderly in the East region had the best health with significantly poorer health in all other regions. The worst health was in the South and Central West regions. There was no significant difference between urban and rural areas.

Both health and economic well-being were worse with older age. The negative regression weights showed the youngest age group (the reference category) had the best health and economic well-being, while the increasing magnitude demonstrated the deteriorating health with the oldest age group (aged 85 or more years old) having the worst health.

Economic well-being was seen to be lower for individuals of male gender, non-white race and living in rural areas. The biggest difference was between urban and rural areas with a regression weight of -1.037 for rural areas, while the difference between genders had a regression weight of -0.030 for females. Residing in the north regions, additionally, had the lowest economic well-being, where all regions had significantly lower economic well-being than the South East region.

The standardised factor loading of functional status, from overall health, increased from 0.766 to 0.916, explaining 24.8% more of the variance in functional status. On the other hand, the standardised factor loadings of the other three health factors (Infectious and chronic health, metabolic health and musculoskeletal health) were smaller. The estimated λ factor loadings were relatively unchanged.

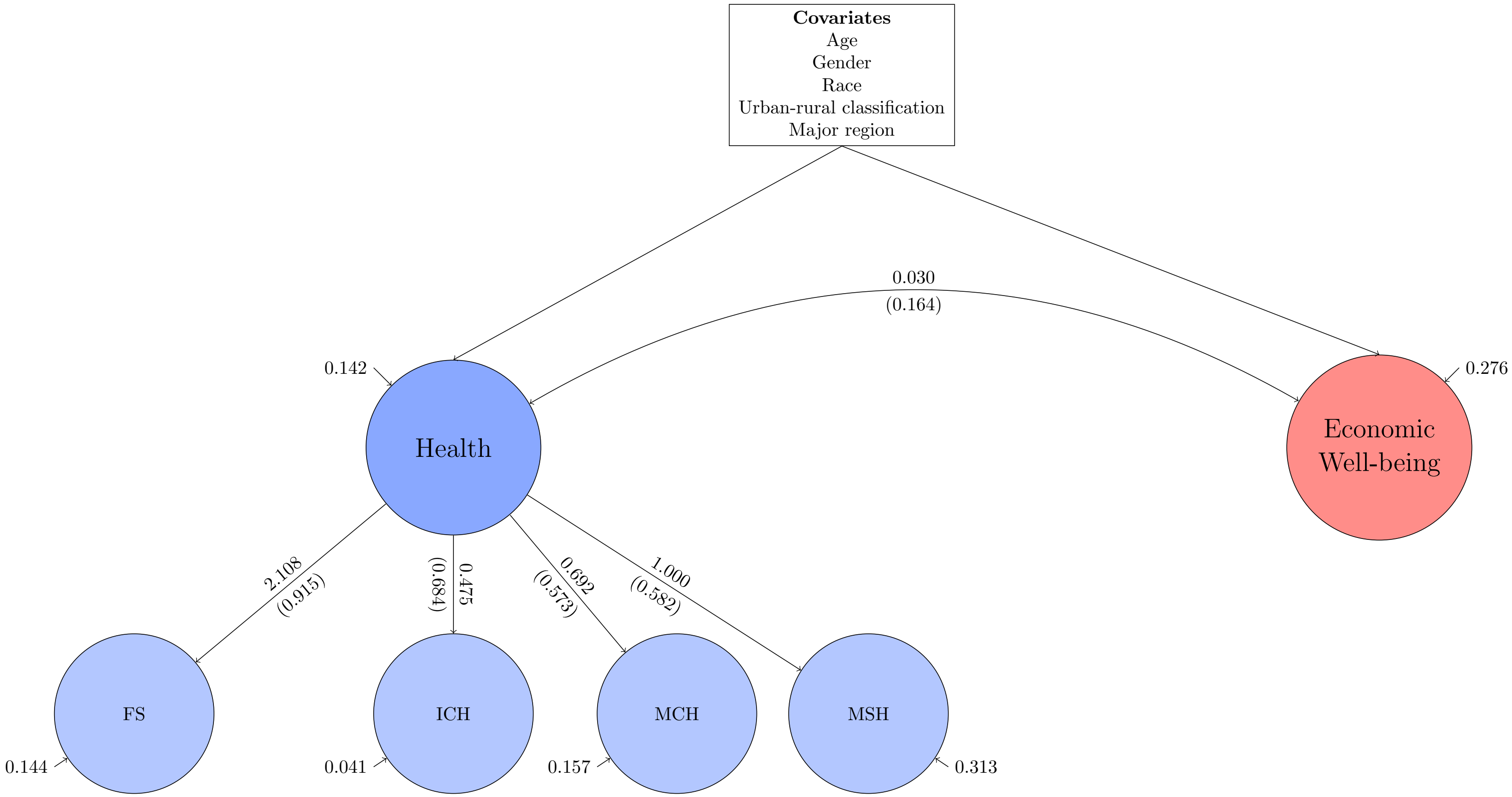


Figure 5.4: Path diagram for the estimated SEM with covariates, for the 2008 PNAD.
Abbreviations: FS, Functional status; ICH, Infectious and chronic health; MCH, Metabolic and cardiovascular health; and MSH, Musculoskeletal health.

5.3 Multilevel Structural Equation Model

The hierarchical structure present in the 2008 PNAD allowed for an approach with multilevel structural equation modelling (multilevel SEM). In the multilevel regression of SRH in Section 5.1, it was found that the household, sector and municipality clusters explained more than 5% of the variance in SRH between elderly individuals. In SEMs, the effect of clusters can inflate model fit indices (Muthén, 1994), thus leading to misinterpretation of the goodness-of-fit. Clustering by household, sector and municipality, were considered for a two level SEM.

The average cluster size of households, sectors and municipalities was 1.187, 4.121 and 35.927 elderly individuals. Hence, clustering by household was not reasonable as the average cluster size was too small and suggested a large proportion of households contained only one elderly individual.

To ascertain the amount of variation that was explained by either the sector or municipality, the intraclass correlation coefficients (ICCs) were estimated for each of the variables. The estimated ICCs are given in Table 5.20 for all the observed variables considered in the previous section. Clustering by municipality, the estimated ICCs were less than 0.100 (10%) for the health variables except the tendonitis and depression variables. Therefore, the municipality did not explain similarities in the health of elderly within same municipality or differences in health between municipalities. The economic well-being variables, however, did show variation between the municipalities.

Comparing the ICCs between the columns of Table 5.20, clustering individuals by sector explained a larger quantity of variation than clustering by municipality in all the variables. Hence, it was decided to proceed with sector as the higher level for clustering individuals in the multilevel SEM.

Over 10% of the variation in the majority of the health variables was explained by the sector; with the exception of the heart disease, diabetes, asthma and hypertension variables. These comorbidities were associated with the metabolic and cardiovascular health factor in the previous SEM, which suggested that comorbidities known to be related to metabolic syndrome and/or cardiovascular disease were primarily influenced by individual variations. The variables that the sector explained most were the economic well-being variables, in particular, variables dependent on local facilities and services such as the ownership of a washing machine, television and landline telephone, and the construction materials of the walls and roof variables. The provision of construction materials requires a local source of supply, due to the high costs involved in the logistics of transporting heavy materials, so were expected that similar within a sector and vary according to the availability of adequate materials locally. Also, as expected, tuberculosis also had a large ICC given the it is an infectious disease.

Aspect	Variable	Sector	Municipality
Self-reported morbidity			
	Disease of the back or column	0.113	0.059
	Heart disease	0.084	0.057
	Arthritis or rheumatism	0.111	0.062
	Depression	0.176	0.104
	Cancer	0.115	0.082
	Tuberculosis	0.246	0.051
	Diabetes	0.055	0.016
	Tendonitis or tenosynovitis	0.236	0.162
	Asthma or bronchitis	0.060	0.045
	Cirrhosis	0.219	0.085
	Hypertension	0.047	0.022
Functional status			
	Eat, bath & toilet	0.183	0.097
	Run, lift, play sport & heavy jobs	0.156	0.069
	Push furniture & home repairs	0.137	0.064
	Climb hill/stairs	0.116	0.051
	Stoop, kneel & bend	0.100	0.043
	Walk 1km	0.111	0.045
Economic well-being			
	Stove	0.364	0.175
	Freezer	0.495	0.340
	Fridge	0.597	0.397
	Washing machine	0.652	0.463
	Television	0.500	0.248
	Computer	0.480	0.249
	Radio	0.342	0.164
	Wall construction material	0.678	0.491
	Landline telephone	0.670	0.407
	Roof construction material	0.727	0.615
	Water dispenser	0.497	0.354

Table 5.20: Estimated intraclass correlation coefficients (ICCs) for the health and economic well-being variables, in the 2008 PNAD.

After estimating the between level variation, the following Sections 5.3.1 and 5.3.2 will now give the results from the multilevel EFA for steps 3 and 4 of the process, as described in Section 4.2.3. Then, Section 5.3.3 presents the multilevel SEMs that identified the interrelationships between the factors at the individual level and sector level, and the impact of individual and sector characteristics.

5.3.1 Multilevel Measurement Model of Health

Similarly to the previous section, all 12 self-reported morbidity and six functional status variables were considered in the multilevel EFA. As the latent structure of the individual level was defined with tuberculosis and cirrhosis included, they were also included whilst defining the latent structure of the sector level.

Self-reported Morbidity

The estimated tetrachoric correlation coefficients for the individual and sector level are shown in Table 5.21. Correlation was generally larger between variables at the sector level than the individual level. The individual level factor structure discovered in the previous, non-multilevel, SEM was clearly visible in the individual level tetrachoric correlation coefficients, while the same structure was not noticeable in the sector level correlation matrix. Alternatively, there was a very strong positive correlation between tuberculosis and cirrhosis, while there was also a strong positive association between cirrhosis and diabetes, cancer and heart disease, and asthma and depression. There was, however, negative correlations between tuberculosis and disease of the column or back, which could be explained by the progression of tuberculosis into the spine, discussed in Section 2.1.1; thus only one diagnosis being reported. Similarly, the negative correlation between hypertension and cirrhosis may be explained by those with hypertension becoming normotensive during the development of cirrhosis (Henriksen and Moller, 2006).

	Back or column disease	Arthritis	Cancer	Diabetes	Asthma or bronchitis	Hypertension	Heart disease	Chronic renal failure	Depression	Tuberculosis	Tendonitis or tenosynovitis	Cirrhosis
Individual level												
Back or column disease	1.000											
Arthritis	0.471	1.000										
Cancer	0.049	0.088	1.000									
Diabetes	0.065	0.141	-0.012	1.000								
Asthma or bronchitis	0.175	0.189	0.110	0.085	1.000							
Hypertension	0.152	0.195	0.043	0.342	0.074	1.000						
Heart disease	0.149	0.199	0.048	0.235	0.194	0.425	1.000					
Chronic renal failure	0.242	0.220	0.204	0.218	0.162	0.196	0.261	1.000				
Depression	0.278	0.279	0.103	0.133	0.224	0.225	0.265	0.215	1.000			
Tuberculosis	0.136	0.091	0.088	0.049	0.297	0.028	0.186	0.253	0.232	1.000		
Tendonitis or tenosynovitis	0.376	0.449	0.026	0.062	0.213	0.114	0.121	0.200	0.315	0.177	1.000	
Cirrhosis	0.083	0.056	0.257	0.113	0.166	-0.005	0.091	0.168	0.085	0.253	0.121	1.000
Sector level												
Back or column disease	1.000											
Arthritis	0.645	1.000										
Cancer	0.146	0.105	1.000									
Diabetes	0.173	0.159	0.465	1.000								
Asthma or bronchitis	0.510	0.400	0.466	0.396	1.000							
Hypertension	0.417	0.382	0.154	0.570	0.465	1.000						
Heart disease	0.432	0.381	0.723	0.415	0.489	0.606	1.000					
Chronic renal failure	0.343	0.274	0.260	0.189	0.699	0.293	0.448	1.000				
Depression	0.398	0.341	0.541	0.390	0.713	0.414	0.605	0.563	1.000			
Tuberculosis	-0.108	0.026	0.180	0.425	0.166	0.000	0.069	0.225	0.159	1.000		
Tendonitis or tenosynovitis	0.340	0.304	0.581	0.479	0.303	0.297	0.459	0.295	0.592	0.153	1.000	
Cirrhosis	0.280	0.184	0.121	0.731	0.598	-0.243	0.226	0.476	0.540	0.992	0.271	1.000

0.000 0.250 0.500 0.750 1.000

Table 5.21: Estimated tetrachoric correlation coefficients at the individual and sector levels for the self-reported morbidity variables in the 2008 PNAD.

Eigenvalues of the estimated within level correlation matrix and the model fit indices from the multilevel EFA for the Step 3, estimation of the within level latent structure, are given in first three rows of Table 5.22. The interpretation for number of factors required to obtain good model fit was mixed. A single factor was sufficient for good fit according to the RMSEA, but two factors were required good fit according to the CFI, and the individual level SRMR and TLI only showed good fit with at least three factors. The eigenvalues of the within level correlation matrix for the fourth factor was 1.004, indicating four factors were required. The factor loadings of the fourth factor highlighted separation of the chronic renal failure variable as the one variable with strong association with this factor. Hence it was decided to remain with three factors at the individual level, similar to the non-multilevel SEM previously.

Factor(s)		Eigenvalue		RMSEA	SRMR		CFI	TLI
Individual	Sector	Individual	Sector		Individual	Sector		
1	Unrestricted	3.007		0.033	0.081	0.000	0.803	0.519
2		1.409		0.011	0.062	0.000	0.981	0.941
3		1.350		0.007	0.032	0.000	0.995	0.978
3	1	1.350	5.171	0.007	0.032	0.175	0.985	0.978
	2		2.098	0.006		0.147	0.992	0.986
	3		1.366	0.005		0.077	0.993	0.987

Table 5.22: Eigenvalues and model fit indices from the multilevel EFA of self-reported morbidity variables, in the 2008 PNAD.

Estimation of the sector level latent structure then proceeded with the number of individual level factors fixed at three. The eigenvalues from the sector level correlation matrix and model fit indices from the multilevel EFA for step 4 are given in the last three rows of Table 5.22. All the model fit indices except the within level SRMR indicated only a single factor was required for good fit. Meanwhile, the SRMR within level did not indicate good fit until five factors were included and the eigenvalues did not drop below 1.000 until the fourth factor; that had an estimated value of 1.021. Above two factors at the between level revealed cross-loading with one strong negative factor loading, these may be due to strong regional variations where different factors of health are poor in different areas. It was, therefore, decided that a model with three within level factors and two between level factors provided a model with good fit that would be meaningful.

Previously, in the non-multilevel SEM of the total correlation matrix, three factors representing musculoskeletal health, metabolic and cardiovascular health, and infectious and chronic health were identified. This structure was also present in the same form at the individual level in the multilevel EFA. The estimated factor loadings from the multilevel SEM are given in Table 5.23.

The same structure was not present at the sector level. The asthma or bronchitis and heart

	Individual			Sector	
	1	2	3	1	2
Disease of the back or column	**0.653**	-0.023	-0.005	**0.856**	-0.001
Arthritis or rheumatism	**0.735**	0.031	-0.049	**0.739**	-0.001
Cancer	0.019	-0.006	0.240	-0.096	**0.794**
Diabetes	-0.010	**0.421**	0.076	-0.004	**0.639**
Asthma or bronchitis	0.139	0.010	**0.400**	**0.396**	**0.554**
Hypertension	0.000	**0.798**	-0.019	**0.416**	0.309
Heart disease	0.001	**0.498**	0.281	**0.331**	**0.575**
Depression	0.277	0.157	0.283	0.241	**0.719**
Chronic renal failure	0.164	0.197	0.300	0.257	**0.454**
Tuberculosis	-0.011	-0.046	**0.640**	-0.257	**0.455**
Tendonitis or tenosynovitis	**0.606**	-0.072	0.088	0.184	**0.564**
Cirrhosis	-0.023	-0.032	**0.422**	-0.047	**0.637**

Bold highlight of factors loadings greater than 0.316.

Table 5.23: Estimated factor loadings from the multilevel EFA, with three individual level factors and two sector level factors, of the self-reported morbidity variables, in the 2008 PNAD.

disease variables had cross-loadings with both factors, therefore the structure was defined such that these two variables were associated with the factor that had the largest factor loading. This was factor 2 at the sector level in Table 5.23. Other variables associated with the second factor were: cancer, diabetes, chronic renal failure, depression, tuberculosis, tendonitis or tenosynovitis and cirrhosis. Variables associated with the first factor 1 were: disease of the back or column, arthritis or rheumatism and hypertension. These are comorbidities are known to workplace related. For example, musculoskeletal comorbidities such as arthritis, rheumatism and back or column disease are associated with construction work, while hypertension is related to stress in the workplace (Schumann et al., 2011; Health and Safety Executive UK, 2017). The second factor, on the other hand, was associated with infectious and chronic comorbidities. The first and second factor were named 'Industry-related health' and 'Local health' respectively. The comorbidities relating to the local health factor were similar to those associated with the chronic and infectious health factor at the individual health, however was named local health to maintain clarity in the distinction between the two factors.

Functional Status

Estimated polychoric correlation coefficients for the functional status variables at the individual and sector level were reasonably high, although less at the sector level (see Table 5.24). At the sector level, the correlation coefficients were seen to be higher in the later variables containing less basic ADLs. The polychoric correlation coefficients for the first variable, 'eat, bath and toilet' were noticeably lower at the sector level; a distinction not made at the individual level.

	Eat, bath & toilet	Run, lift, play sport & heavy jobs	Push furniture & home repairs	Climb hill/stairs	Stoop, kneel & bend	Walk 1km
Individual level						
Eat, bath & toilet	1.000					
Run, lift, play sport & heavy jobs	0.806	1.000				
Push furniture & home repairs	0.842	0.876	1.000			
Climb hill/stairs	0.826	0.832	0.844	1.000		
Stoop, kneel & bend	0.814	0.795	0.804	0.847	1.000	
Walk 1km	0.833	0.806	0.819	0.871	0.835	1.000
Sector level						
Eat, bath & toilet	1.000					
Run, lift, play sport & heavy jobs	0.426	1.000				
Push furniture & home repairs	0.671	0.661	1.000			
Climb hill/stairs	0.603	0.795	0.766	1.000		
Stoop, kneel & bend	0.629	0.759	0.773	0.866	1.000	
Walk 1km	0.717	0.704	0.762	0.829	0.801	1.000

Table 5.24: Estimated polychoric correlation coefficients at the individual and sector levels for the functional status variables in the 2008 PNAD.

Estimation of the individual level latent structure revealed a single factor for the activities of daily living variables. The model fit indices for one within level factors and unrestricted structure between were; RMSEA=0.100, within level SRMR=0.016, CFI=0.995 and TLI=0.985. Excellent fit was therefore shown by all the model fit indices except the RMSEA, which may be indicating the poor fit at the between level. The eigenvalues for the first and second factor of the within level correlation matrix were 5.150 and 0.251 respectively, therefore agreed with a single factor at the individual level.

Estimation of the sector level latent structure had the same conclusions as the within level. The estimated model fit indices for a multilevel EFA model with one factor at the individual level and at the sector level were; RMSEA=0.084, within level SRMR=0.016, between level SRMR=0.043, CFI=0.994 and TLI=0.989. Therefore, a multilevel model with a single factor at both levels was concluded for functional status, which was confirmed by the estimated factor loadings shown in Table 5.25. The sector level factor was named 'Local functional status' referring to the average functional status in a locality.

Multilevel Health Measurement Model

The latent structure of individual health in the elderly therefore consisted of the four factors previously discussed in the non-multilevel SEM (musculoskeletal health, metabolic and

	Individual	Sector
Eat, bath & toilet	**0.901**	**0.684**
Run, lift, play sport & heavy jobs	**0.910**	**0.782**
Push furniture & home repairs	**0.924**	**0.853**
Climb hill/stairs	**0.932**	**0.934**
Stoop, kneel & bend	**0.895**	**0.922**
Walk 1km	**0.916**	**0.907**

Bold highlight of factors loadings greater than 0.316.

Table 5.25: Estimated factor loadings from the multilevel EFA, with one individual level factor and one sector level factor, of the functional status variables, in the 2008 PNAD.

cardiovascular health, chronic and infectious health and functional status), while the latent structure of the average health of the elderly by sector consisted of three factors, including industry related health, local health and local functional status. At the individual level the second order factor representing overall health of the elderly individual level was re-introduced, as previously defined in Section 5.2.1, given strong positive factor correlations. A factor of the overall health at the sector level was not appropriate as the correlation was low between industry-related health and local health. Thus, the sector level model was defined with correlations between the factors, with no direction of causation defined. Again, the tuberculosis and cirrhosis variables were dropped from the measurement model due to the low frequencies.

Figure 5.5 shows the estimated parameters of the multilevel measurement model, with the standard errors of the factors loadings, R-squared values and scale factors for each observed variable given in Table 5.26.

The fit of the model to the data was good with RMSEA = 0.019, CFI = 0.997, TLI = 0.997 and individual level SRMR = 0.038, however sector level SRMR was only 0.095. This may be the contribution of the self-reported morbidity variables at the sector level, which was attributed to potential regional disparities that are explored later in the multilevel SEM with covariates. Previously a larger number of factors was found to have better fit to the data although strong negative cross-loadings were present.

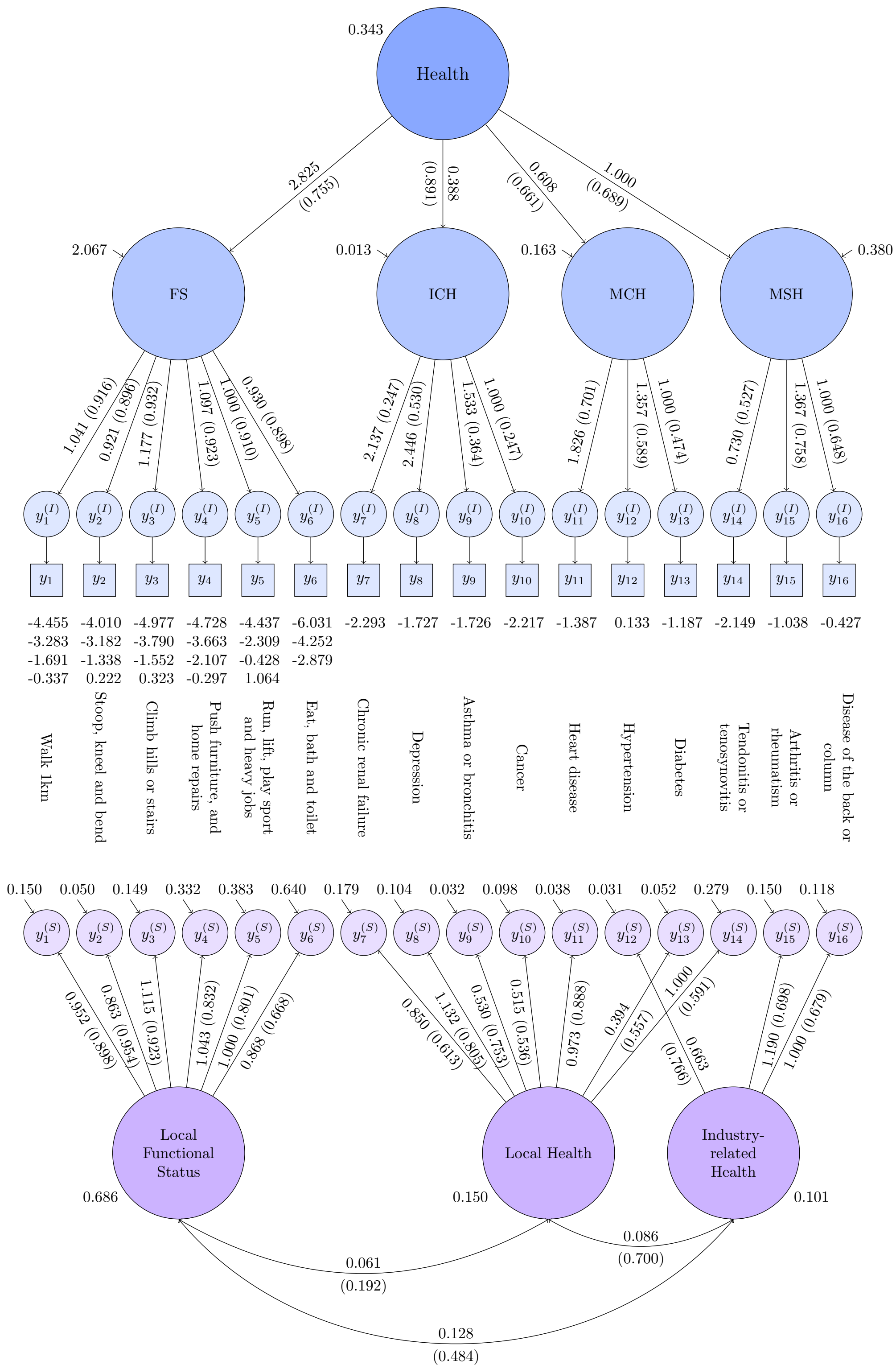


Figure 5.5: Path diagram for the estimated multilevel measurement model of health, at individual and sectors levels, in the 2008 PNAD.
Abbreviations: FS, Functional status; ICH, Infectious and chronic health; MCH, Metabolic and cardiovascular health; and MSH, Musculoskeletal health.

	Standard Error of Factor Loadings		R-squared		
	Individual level, $\lambda^{(I)}$	Sector level, $\lambda^{(S)}$	Individual level	Sector level	Scale Factor
Eat, bath & toilet	0.021	0.053	0.806	0.447	0.441
Run, lift, play sport & heavy jobs	-	-	0.828	0.642	0.415
Push furniture & home repairs	0.014	0.035	0.852	0.692	0.384
Climb hill/stairs	0.014	0.034	0.869	0.852	0.362
Stoop, kneel & bend	0.010	0.029	0.803	0.911	0.444
Walk 1km	0.012	0.035	0.839	0.806	0.402
Disease of the back or column	-	-	0.419	0.461	0.762
Arthritis or rheumatism	0.068	0.122	0.574	0.488	0.652
Cancer	-	0.118	0.061	0.287	0.969
Diabetes	-	0.073	0.224	0.310	0.881
Asthma or bronchitis	0.189	0.093	0.133	0.567	0.931
Hypertension	0.074	0.092	0.347	0.586	0.808
Heart disease	0.111	0.129	0.491	0.789	0.714
Depression	0.292	0.146	0.280	0.648	0.848
Chronic renal failure	0.267	0.147	0.229	0.376	0.878
Tendonitis or tenosynovitis	0.045	-	0.278	0.349	0.850

Table 5.26: Standard errors, R-squared estimates and scale factors for the multilevel measurement model of health, for the 2008 PNAD.

The importance of the functional status variables varied from the individual level to the sector level. By ordering the variables in relation to the standardised factor loadings it was seen that at the individual level the most important activities of daily living were climbing hills or stairs while stooping, kneeling and bending were the least important. However, at the sector level stooping, kneeling and bending has the highest standardised factor loading, and was therefore the most influential group of activities to the average functional status of a sector. The 'eat, bath and toilet' variable had relatively low association with the local functional status of the sector, compare to the other functional status variables.

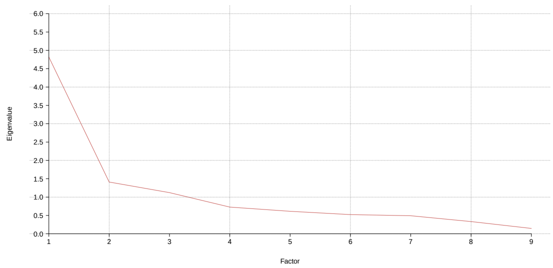
The ordering of the first order factors at the individual level can be seen to be different in terms of their association with overall health, compared to the ordering in the non-multilevel measurement model of health. Previously, functional status was the least associated with overall health of the elderly individual, however, in the multilevel measurement model, functional status has greater association with the second largest standardised factor loading.

At the sector level all the health factors were positively associated, with a significant correlation. Functional status had a stronger correlation with industry related health than with local health, while there was strong association between the two comorbidity factors; local health and industry related health.

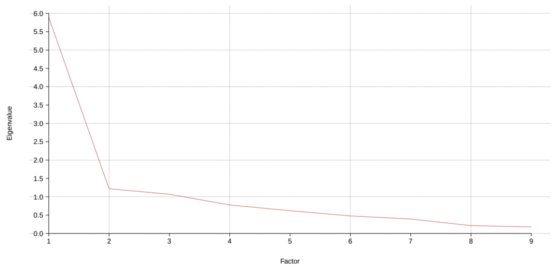
5.3.2 Multilevel Measurement Model of Economic Well-being

The separate correlations between the economic well-being variables for the individual and sector levels are given in Table 5.27. Generally, the correlations were higher at the sector level than the individual level, especially with the landline telephone, washing machine and computer variables, although the stove, water dispenser and freezer variables had lower correlations with the other variables at the sector level. The variables corresponding to services that were primarily provided by the sector included the sewage management system, rubbish disposal method and water supply, but with the occasional exception such as elderly individuals with a private water supply from a spring. The simultaneous modelling of these variables while clustering by sector was not possible, either as indicators to a latent variable due to the high correlation or as covariates due to the occasional exception.

In this instance the interpretation of the eigenvalues was taken by the alternative method of Cattell (1966), using the scree plots given in Figure 5.6, for both the individual and sector level. A sharp decline in eigenvalue can be seen from the first factor to the second which would suggest one factor as the appropriate number of factors. At both levels, the third factor was the last eigenvalue greater than one.



(a) Individual level



(b) Sector level

Figure 5.6: Scree plots of the eigenvalues for the individual level and sector level correlation matrices, from the 2008 PNAD.

The individual level had good fit with a single factor in the multilevel EFA with an unrestricted structure at the sector level. A single within level factor and unrestricted structure at the between level had the model fit indices; RMSEA=0.023, within level SRMR=0.133, CFI=0.960 and TLI =0.900. The RMSEA and CFI both indicated good fit, however, the within level SRMR and TLI did not. Unfortunately, the multilevel EFA failed to converge with more than one factor.

Estimation of the between level structure also revealed a single factor be most appropriate. The model fit indices for the multilevel EFA with one factor at both levels were; RMSEA=0.021, within level SRMR=0.133, between level SRMR=0.083, CFI=0.936 and TLI=0.920. Only the RMSEA and CFI indicated that model fit to be good. However, increasing the number of between level factors introduced strong negative cross-loadings, similarly to before. Variables that had strong factor loadings with the second factor were radio, water dispenser, freezer and washing machine, while the factor loading for water dispenser was negative. These variables can be dependent on the provision of local services such as laundrette and network connection, and the sanctity of the local water supply. Therefore, in areas where these needs are fulfilled by services outside the household, there may be higher affluence in the other variables.

The estimated parameters of the multilevel measurement model for economic well-being are shown on the path diagram in Figure 5.7, with the corresponding standard errors, R-squared and scale factors of the variables given in Table 5.28. At the individual level, the most influential variables were the possession of a fridge and stove, while variables with some requirement for a local service such as supply of construction materials and a telephone network were less important. In contrast, the possession of computers and landline telephones were the most influential indicators of economic well-being in the sector.

At the sector level, the factor loading of the waster dispenser variable was low and was comparatively lower at the individual level. The possession of a water dispenser can be considered a convenient luxury it locations where it is necessary, that provides safe drinking water on tap instead of requiring residents to boil their tap water before it is safe to drink. A water dispenser is also not always necessary, say if there is a nearby spring.

	Walls construction material	Roof construction material	Stove	Fridge	Television	Radio	Landline telephone	Water dispenser	Freezer	Washing machine	Computer
Individual level											
Walls construction material	1.000										
Roof construction material	0.737	1.000									
Stove	0.446	0.441	1.000								
Fridge	0.555	0.353	0.768	1.000							
Television	0.440	0.401	0.608	0.730	1.000						
Radio	0.134	-0.093	0.413	0.320	0.309	1.000					
Landline telephone	0.380	0.212	0.453	0.560	0.427	0.257	1.000				
Water dispenser	0.300	0.163	0.429	0.347	0.234	0.178	0.261	1.000			
Freezer	0.438	0.293	0.309	-0.004	0.486	0.235	0.442	0.167	1.000		
Washing machine	0.271	0.162	0.532	0.592	0.436	0.289	0.467	0.183	0.438	1.000	
Computer	0.285	0.326	0.371	0.463	0.458	0.349	0.479	0.152	0.362	0.473	1.000
Sector level											
Walls construction material	1.000										
Roof construction material	0.661	1.000									
Stove	0.497	0.109	1.000								
Fridge	0.627	0.569	0.547	1.000							
Television	0.591	0.594	0.433	0.790	1.000						
Radio	0.319	0.366	0.226	0.409	0.430	1.000					
Landline telephone	0.607	0.386	0.560	0.826	0.712	0.479	1.000				
Water dispenser	0.261	0.186	0.283	0.172	0.178	0.117	0.266	1.000			
Freezer	0.233	0.165	0.295	0.443	0.292	0.423	0.353	-0.068	1.000		
Washing machine	0.524	0.306	0.469	0.771	0.654	0.530	0.811	0.120	0.519	1.000	
Computer	0.675	0.338	0.482	0.769	0.685	0.539	0.857	0.231	0.476	0.796	1.000

0.000 0.250 0.500 0.750 1.000

Table 5.27: Estimated tetrachoric correlation coefficients for the individual and sector levels of economic well-being variables in the 2008 PNAD.

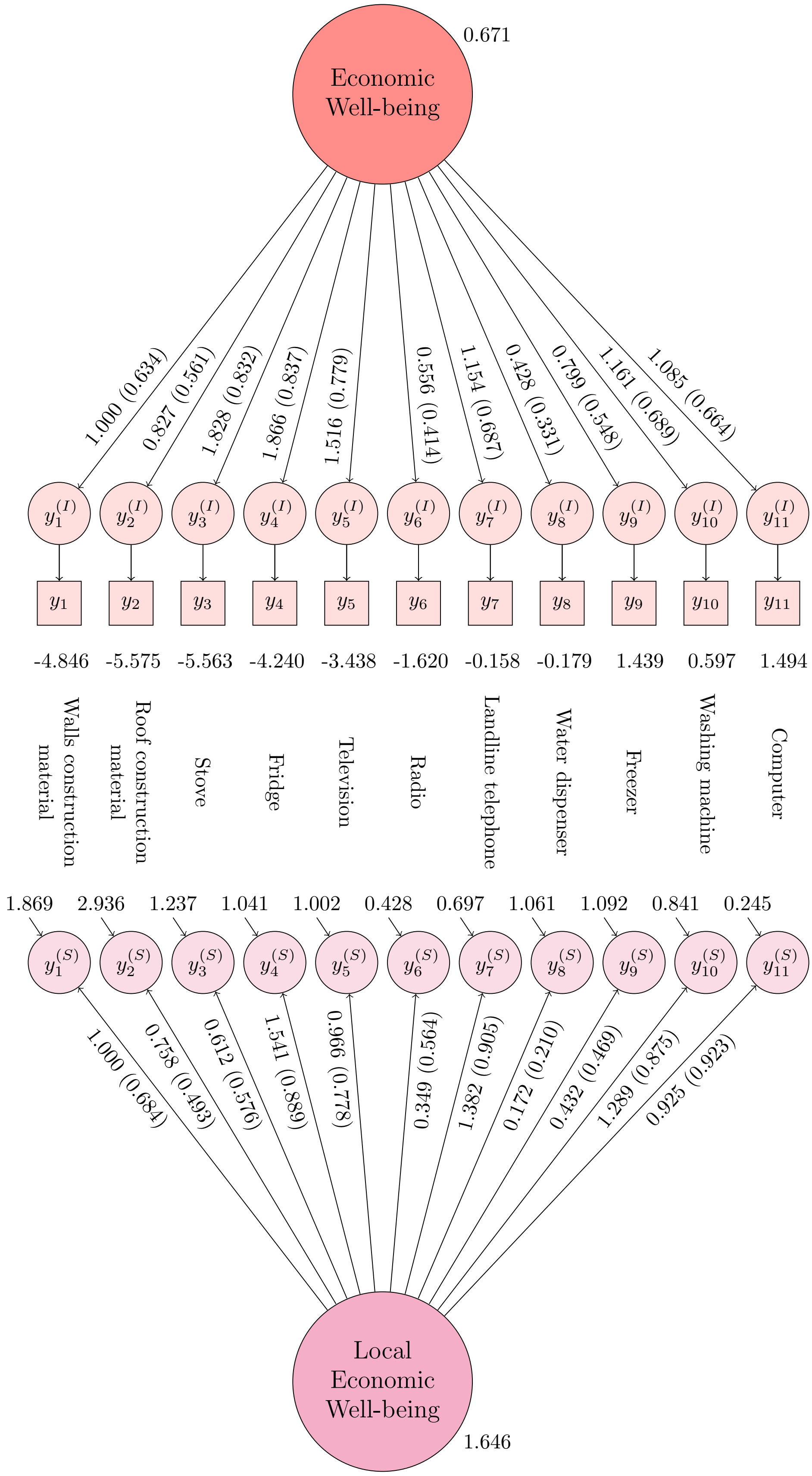


Figure 5.7: Path Diagram for the estimated multilevel measurement model, at the individual and sector levels, of economic well-being, in the 2008 PNAD.

	Standard Error of Factor Loadings		R-squared		
	Individual level, $\lambda^{(I)}$	Sector level, $\lambda^{(S)}$	Individual level	Sector level	Scale Factor
Computer	0.093	0.072	0.441	0.852	0.747
Washing machine	0.098	0.098	0.475	0.765	0.725
Freezer	0.068	0.036	0.300	0.220	0.837
Water dispenser	0.037	0.019	0.109	0.044	0.944
Landline telephone	0.097	0.104	0.472	0.818	0.727
Radio	0.049	0.029	0.172	0.318	0.910
Television	0.143	0.078	0.607	0.605	0.627
Fridge	0.188	0.130	0.700	0.790	0.547
Stove	0.278	0.100	0.692	0.332	0.555
Roof material	0.112	0.089	0.315	0.244	0.828
Wall material	-	-	0.402	0.468	0.774

Table 5.28: Standard errors of the unstandardised factor loadings, R-squared estimates and scale factors from the estimated multilevel measurement model of economic well-being from the 2008 PNAD.

5.3.3 Multilevel Structural Equation Model

The interrelationships between the individual level factors and between the sector levels factors, were then estimated in a multilevel SEM. The combination of the two multilevel measurement models allowed for the correlation between the four sector level factors: local functional status, local health, industry-related health and local economic well-being, and between the two individual level factors: health and economic well-being.

Correlations at the sector level between the functional status variables 'climbing hills or stairs' and the ownership of a washing machine or computer were 0.420 and 0.473 respectively. This high correlation across domains led to Heywood estimates in the model. Therefore, the 'climbing hills or stairs' variable was dropped from the model.

The fit of the to the data was good with RMSEA = 0.014, CFI = 0.993, TLI = 0.993 and individual level SRMR = 0.069, however the sector level SRMR = 0.115 remained inadequate. This could indicate that the overall model fit was good, but the structure at the sector level could be improved. This was consistent to the results found in the multilevel EFA, where additional factors seen with negative cross-loadings.

The estimated model parameters are shown on the path diagram in Figure 5.8 with the respective standard errors of the factor loadings, R-squared values and scale factors of the variables given in Table 5.29.

The health and economic well-being of elderly individuals remained positively associated, the same as in the non-multilevel SEM. Although statistically significant the correlation was not large. There was therefore slight association such that elderly individuals with poorer health were likewise in terms of their economic well-being, and vice versa with better economic well-being associated with better health.

At the sector level, there was no significant association between the industry related health and economic well-being of the elderly. Therefore, the presence of back or column disease, arthritis or rheumatism, or hypertension were not associated with the economic well-being of an elderly individual. Further, the local economic well-being was not associated with the prevalence of industry related comorbidities.

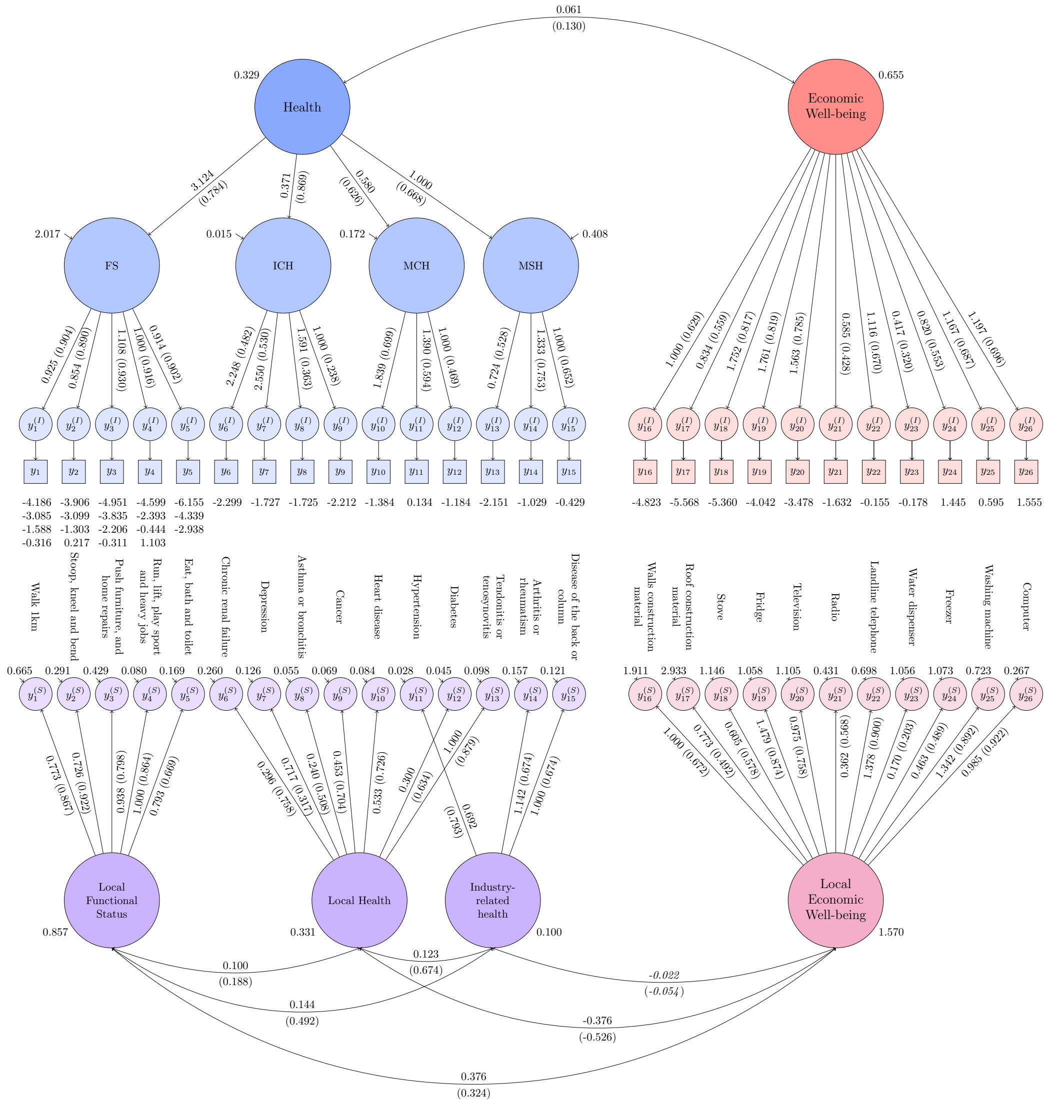


Figure 5.8: Path diagram for the estimated multilevel SEM, for the 2008 PNAD.
Abbreviations: FS, Functional status; ICH, Infectious and chronic health; MCH, Metabolic and cardiovascular health; and MSH, Musculoskeletal health.
Italics statistically insignificant estimates (at 5% level).

	Standard Error of Factor Loadings		R-squared		
	Individual level, $\lambda^{(I)}$ or $\beta^{(I)}$	Sector level, $\lambda^{(S)}$	Individual level	Sector level	Scale Factor
Functional status variables					
Eat, bath & toilet	0.022	0.048	0.814	0.448	0.432
Run, lift, play sport & heavy jobs	-	-	0.840	0.747	0.401
Push furniture & home repairs	0.017	0.034	0.865	0.638	0.367
Stoop, kneel & bend	0.011	0.026	0.792	0.849	0.456
Walk 1km	0.012	0.030	0.817	0.751	0.427
Self-reported morbidity variables					
Disease of the back or column	-	-	0.424	0.454	0.759
Arthritis or rheumatism	0.066	0.117	0.567	0.454	0.658
Cancer	-	0.063	0.057	0.496	0.971
Diabetes	-	0.036	0.220	0.401	0.883
Asthma or bronchitis	0.202	0.041	0.132	0.258	0.932
Hypertension	0.076	0.092	0.353	0.629	0.804
Heart disease	0.112	0.050	0.489	0.528	0.715
Depression	0.313	0.059	0.281	0.575	0.848
Chronic renal failure	0.289	0.058	0.233	0.101	0.876
Tendonitis or tenosynovitis	0.043	-	0.279	0.772	0.849
Economic well-being variables					
Computer	0.105	0.080	0.484	0.851	0.718
Washing machine	0.101	0.107	0.471	0.796	0.727
Freezer	0.071	0.040	0.306	0.239	0.833
Water dispenser	0.037	0.019	0.102	0.041	0.947
Landline telephone	0.096	0.108	0.449	0.810	0.742
Radio	0.052	0.031	0.183	0.323	0.904
Television	0.152	0.082	0.616	0.575	0.620
Fridge	0.174	0.125	0.670	0.765	0.574
Stove	0.259	0.097	0.668	0.334	0.576
Roof construction material	0.114	0.097	0.313	0.242	0.829
Walls construction material	-	-	0.396	0.451	0.777

Table 5.29: Standard errors, R-squared estimates and scale factors for the variables in the multilevel SEM, for the 2008 PNAD.

The functional status and economic well-being of the elderly in a sector were positively associated with significant correlation. In sectors where the elderly individuals had more functional limitations, the economic well-being was higher, and in sectors where economic well-being was lower the elderly had greater limitation to their functional status.

Controversially, the local health and local economic well-being had negative association. Therefore, areas where the economic well-being of the elderly was high it was more likely that the elderly individuals had at least one comorbidity of local health and conversely low economic well-being was associated with lower prevalence of comorbidities such as cancer, depression, chronic renal failure etc.

The relationship between the factors and the demographics of the individual and the sector were analysed in a multilevel SEM with covariates, where the covariates were modelled at both levels. At the individual level the covariates were age, gender and race, while at the sector level the covariates were the urban-rural classification and major region.

The estimated parameters of the structural equations are shown in Figure 5.9, with the estimated regression weights for the individual level shown in Table 5.30, and the estimated regression weights for the sector level covariates shown in Table 5.31.

A lack of fit at the sector was still present in the between level SRMR=0.112, and the within level SRMR=0.067 no longer gave good fit. However, the other model fit indices indicated very good fit. The other indices were RMSEA=0.017, CFI=0.982 and TLI=0.980. The lack of fit indicated by the SRMRs could be due to the more complicated factor structures indicated in the multilevel EFA, where the strong negative cross-loadings of additional factors show a complex relationship exists between the comorbidities such that the better health in one part is associated with poorer health in another. This was eluded to in the sector level correlation matrix in Table 5.21.

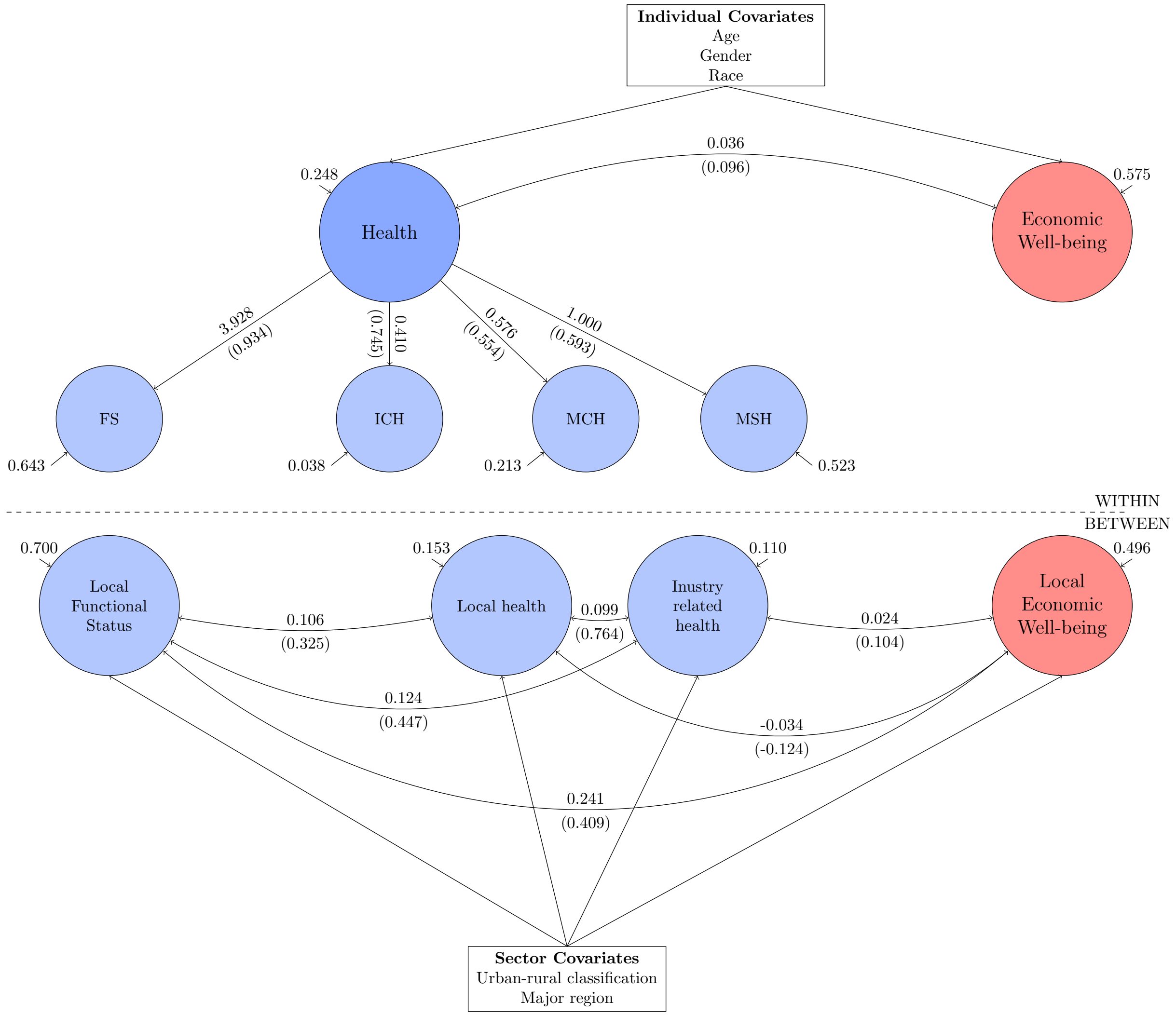


Figure 5.9: Path diagram for the multilevel SEM for the 2008 PNAD.

The effect of the individual level covariates had the same interpretation as in the non-multilevel model, in Section 5.2. The estimated regression weights were increasingly negative for the health and economic well-being for the elderly in the older age groups. Elderly men had significantly better health than elderly women, but the economic well-being was not significantly different between the genders. Meanwhile, both health and economic well-being were significantly worse for those of non-white race than white race. Overall health was therefore worst in non-white females aged 90 or more years old, while economic well-being was worst in the oldest non-white elderly.

Covariate	Health		Economic Well-being	
	Est.	s.e.	Est.	s.e.
Age				
60-64 years old	Ref.		Ref.	
65-69 years old	**-0.080**	(0.010)	**-0.099**	(0.019)
70-74 years old	**-0.196**	(0.013)	**-0.144**	(0.022)
75-79 years old	**-0.310**	(0.016)	**-0.202**	(0.027)
80-84 years old	**-0.438**	(0.022)	**-0.266**	(0.035)
85-89 years old	**-0.559**	(0.030)	**-0.395**	(0.052)
≥90 years old	**-0.644**	(0.042)	**-0.395**	(0.072)
Gender				
Female	Ref.		Ref.	
Male	**0.217**	(0.011)	-0.012	(0.016)
Race				
White	Ref.		Ref.	
Non-white	**-0.018**	(0.008)	**-0.368**	(0.032)

Bold highlight of regression coefficients significant at 5% level.

Table 5.30: Estimated regression weights of the individual level explanatory variable from the multilevel SEM with covariates of the 2008 PNAD.

In contrast, the effect of the sector level covariates, urban-rural classification and the major region, were altered in relation to their effect in the non-multilevel model.

Local functional status of the sector was significantly different between the South East and the other major regions. The functional status of the elderly in the South East region was significantly better than in all the other regions. Worst levels of limitation in functional status were in the North and Central West regions. The effect of urban-rural classification of the sector, however, had no significant impact on the functional status of the elderly.

Industry related health was worst in sectors located in the western regions of Brazil, while being better in rural sectors. Therefore, the presence of the comorbidities arthritis, disease of the back or column and hypertension was lower in the South East and North East regions of Brazil.

On the other hand, economic well-being was significantly worse in rural sectors, and in sectors located in the Central West, North and North East regions. Living in a rural sector had the greatest impact. Sectors where the elderly had the best economic well-being were the South East and South regions, where there was no significant difference in the economic well-being between the two regions. The worst economic well-being, however, was in the elderly located in the sectors of the North and North East.

A controversial result was the effects on local health. The regression weights indicated that local health was significantly better in rural sectors and in the North and North East region. This effect did not agree with the poorer economic well-being with the same demographic and the positive correlation in health and economic well-being at the individual level. A consideration may be the comorbidities associated with the local health factor. These include comorbidities such as cancer and diabetes which require more complex medical resources to diagnose and treat that were primarily available in state capitals and the major cities (de Albuquerque et al., 2017). This will be discussed further later in the conclusions Chapter 8.

Correcting for the effect of the covariates reduced the negative correlation between the economic well-being and local health at the sector level, and increased the correlation between economic well-being and industry related health. Previously the association was insignificant between economic well-being and industry related health, however, after controlling for the effect of urban-rural classification and the major region, there was a small but significant correlation between the two factors. Whereas, the correlation between economic well-being and local health was -0.124 which had been -0.526 without controlling for the covariates. Therefore, the local health, industry related health and economic well-being of a sector was strongly dependent on the urban-rural classification and the major region it was located in.

Covariate	Local Functional Status		Local Health		Industry Related Health		Economic Well-being	
	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.
Urban-rural								
Urban	Ref.		Ref.		Ref.		Ref.	
Rural	-0.064	(0.051)	**0.323**	(0.038)	**0.081**	(0.025)	**-1.275**	(0.098)
Major Region								
South East	Ref.		Ref.		Ref.		Ref.	
South	**-0.220**	(0.057)	**-0.260**	(0.036)	**-0.094**	(0.028)	-0.056	(0.033)
Central West	**-0.275**	(0.067)	0.007	(0.042)	**-0.085**	(0.033)	**-0.316**	(0.046)
North	**-0.325**	(0.072)	**0.440**	(0.054)	**-0.094**	(0.033)	**-0.761**	(0.068)
North East	**-0.236**	(0.049)	**0.409**	(0.040)	0.033	(0.024)	**-0.776**	(0.064)

Bold highlight of regression coefficients significant at 5% level.

Table 5.31: Estimated regression weights of the sector level explanatory variable from the multilevel SEM with covariates of the 2008 PNAD.

Summary of the Chapter

In this chapter we have presented the results of analyses using the data from 2008 PNAD. In the first section, multilevel cumulative logistic regression was used to determine the effects of a large range of covariates including characteristics of health and those relating to the wider setting such as the household and municipality, on an elderly person's self-rated health (SRH). Univariate analyses showed all the covariates in the large selection to be significant to SRH. Later multivariate analyses revealed all the considered individual characteristics to be significant, but at the household and municipality levels only the major region, ownership of durables, household income and status of living alone remained significant.

The perception of health was seen to improve with older age, despite the same status of health characteristics such as comorbidity and functional status, and socioeconomic characteristics such as education, housing quality and ownership of basic goods. This indicated that self-rated health did not provide a homogeneous measure of overall health in the elderly, as equal status was perceived differently.

Methods of structural equation modelling were then applied to the different aspects, self-reported morbidity and functional status, of health and aspects, housing quality and ownership of durables, of economic well-being. The latent structures were identified for each aspect to create measurement models for health and economic well-being. Firstly, the latent structures were ascertained in a non-multilevel analysis of the total correlation matrices, then in multilevel analyses we identified the separate structures for the elderly individuals and for sectors of elderly individuals. The latent structures identified in the non-multilevel analysis were found to agree with the structures found at the individual (within) level of the multilevel analysis; although the importance of functional status was greater in the multilevel analysis.

Health was found to have a different structure at the individual and sector levels. Health of elderly individuals was found to consist of the four factors; musculoskeletal health, metabolic and cardiovascular health, infectious and chronic health, and functional status. Meanwhile, the health of sectors of the elderly consisted of three factors; industry related health, local health and functional status. However, the multilevel EFA indicated a more complex structure including negative cross-loadings existed in the comorbidities.

Economic well-being was found to have a single factor structure at both levels. At the individual level the most influential variables were the ownership of the more basic durables such as a stove or fridge, while at the sector level the most influential variables required network and local services such as the ownership of a computer, telephone and the adequacy of construction materials.

Health and economic well-being of the individual were found to have a small but significantly positive association, while the interrelationships between the health factors and economic well-being at the sector level were mixed. Economic well-being of the sector had a strong positive association with the functional status and industry related health in the sector, while having a negative association with the local health factor. This was elaborated upon further in the multilevel SEM with covariates.

The multilevel SEM with covariates showed the effects of individual demographics and characteristics of the sector on the factor health and economic well-being at the respective level. Older age was associated with increasingly worse health and economic well-being, and increasingly worse. Both domains were also significantly worse for the elderly of non-white race compared to those of white race. Men had better health than women, however there was no significant difference in the economic well-being.

At the sector level, however, the effect of urban-rural classification and the major region revealed controversial results. As expected, worse levels of economic well-being in the North and North East region were reflected by poorer functional status but also better local health. This was counter to the positive association between health and economic well-being at the individual level. This result is discussed further later in Chapter 8.

In sectors of rural classification, the local health and industry related health were significantly better than in urban sectors. The effect of rural classification had the greatest impact on economic well-being which was significantly worse.

Chapter 6

Cross-sectional Analysis of England

The current chapter will present the results from the cross-sectional analysis of the English Longitudinal Study of Ageing. The first three sections describe the formation of the measurement model for each domain; health, economic well-being and subjective well-being. In each section exploratory factor analysis (EFA) identifies the factor structure of each aspect, that are then used to form the estimated measurement models for each domain. In the final section of this chapter the overall structural equation model (SEM) combines the domains to determine the relationships between the domains, then the effect of covariates was estimated.

Note, the direction of all observed variables has been set such that increasing values reflect better health, economic well-being or subjective well-being.

6.1 Health

The health domain consisted of three aspects; cardiovascular morbidity, chronic morbidity and functional ability. EFA was performed on each aspect, identifying the most appropriate number of factors for modelling the latent response variables and a simple latent structure that had interpretable meaning.

6.1.1 Cardiovascular Morbidity

The estimated tetrachoric correlation coefficients for the cardiovascular morbidity variables are shown in Table 6.2. The coefficients revealed some conditions that were related with strong correlations between pairs of variables, including angina and heart attack, heart failure and abnormal heart rhythm, and angina and heart failure. Weak correlations between heart murmur and diabetes, heart failure and high cholesterol, and abnormal heart rhythm and high cholesterol were also noticeable.

Eigenvalues of the correlation matrix and the model fit indices from the EFA (Table 6.1) highlighted two factors to be the model with good fit and the least number of factors. For reference the cut-off values for the indication of good fit were RMSEA< 0.05, SRMR< 0.06, CFI> 0.90 and TLI> 0.95, as aforementioned in Section 4.1.4. The RMSEA showed good fit with just one factor. Two factors were required in order for the eigenvalues, CFI and TLI to indicate good fit, and for the SRMR three factors were required. The three factor model highlighted the strong correlation between angina and myocardial infarction to be a third factor. Unfortunately, as there were fewer than two latent response variables the model lacked local identification.

Factor(s)	Eigenvalue	RMSEA	SRMR	CFI	TLI
1	3.206	0.046	0.094	0.881	0.842
2	1.348	0.033	0.067	0.959	0.922
3	0.985	0.014	0.040	0.995	0.985

Table 6.1: Eigenvalues and model fit indices for the EFA of cardiovascular morbidity variables in the ELSA (Wave 4).

	Hypertension	Angina	Heart attack	Heart failure	Heart murmur	Abnormal heart rhythm	Diabetes	Stroke	High cholesterol
Hypertension	1.000								
Angina	0.298	1.000							
Heart attack	0.261	0.674	1.000						
Heart failure	0.227	0.424	0.419	1.000					
Heart murmur	0.161	0.287	0.200	0.316	1.000				
Abnormal heart rhythm	0.184	0.283	0.235	0.458	0.373	1.000			
Diabetes	0.399	0.253	0.272	0.294	0.079	0.060	1.000		
Stroke	0.328	0.336	0.294	0.144	0.270	0.280	0.174	1.000	
High cholesterol	0.381	0.331	0.210	0.048	0.087	0.090	0.300	0.219	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.2: Estimated tetrachoric correlation coefficients for the cardiovascular morbidity variables in the ELSA (Wave 4).

A clear latent structure was presented by the estimated factor loadings from the EFA model with two factors; given in Table 6.3. Factor one had strong factor loading with hypertension, diabetes and high cholesterol that are conditions linked to metabolic disorder. Factor two loaded strongest onto the other conditions, including stroke that had no large factor loading. The latent structure was defined as such with factor one subsequently named metabolic health and Factor two named cardiovascular health.

	Factor	
	One (η_1)	Two (η_2)
Hypertension	**0.712**	-0.006
Angina	0.070	**0.777**
Heart attack	0.006	**0.756**
Heart failure	-0.064	**0.647**
Heart murmur	-0.060	**0.469**
Abnormal heart rhythm	-0.051	**0.474**
Diabetes	**0.551**	0.003
Stroke	0.240	**0.318**
High cholesterol	**0.528**	0.029

Bold highlight of factors loadings greater than 0.316.

Table 6.3: Estimated factor loadings from EFA with two factors for the cardiovascular morbidity variables in the ELSA (Wave 4).

6.1.2 Chronic Morbidity

The estimated tetrachoric correlation coefficients for the EFA for the chronic morbidity variables are shown in Table 6.4. Correlations were low with the strongest of 0.429 between chronic lung disease and asthma.

	Chronic lung disease	Asthma	Arthritis	Osteoporosis	Cancer	Psychiatric problems	Neurodegenerative disorder
Chronic lung disease	1.000						
Asthma	0.429	1.000					
Arthritis	0.154	0.098	1.000				
Osteoporosis	0.215	0.151	0.245	1.000			
Cancer	0.079	-0.020	0.059	0.114	1.000		
Psychiatric problems	0.136	0.177	0.151	0.096	-0.031	1.000	
Neurodegenerative conditions	0.097	0.058	0.071	0.050	0.034	0.108	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.4: Estimated tetrachoric correlation coefficients for the chronic morbidity variables in the ELSA (Wave 4).

Consequently, the EFA gave mixed interpretation for the factor structure. To ensure model identification (see Chapter 4 for details) the maximum number of factors considered was two. Eigenvalues of the matrix of correlation coefficients, however, indicated three factors were required with values less than 1.000 after this. Alternatively, interpretation of the scree plot showed a sharp decline after one factor suggesting a single factor was sufficient. The RMSEA and SRMR also show a single factor to obtain good model fit (see Table 6.5). However, the CFI and TLI did not and indicated two factors were required for good model fit.

The estimated factor loadings for a two factor model identified only two variables, chronic lung disease and asthma, that had a strong factor loading with the second factor.

In the single factor model the estimated factor loadings were; 0.669 chronic lung disease, 0.539 asthma, 0.307 arthritis, 0.392 osteoporosis, 0.094 cancer, 0.293 psychiatric problem and 0.162 neurodegenerative disorders. Hence, only chronic lung disease, asthma and osteoporosis had large factor loadings (greater than 0.316), and the factor loadings of cancer and neurodegenerative disorders were small. The low factor loadings were thought to be due to the lack of multiple morbidity, as 61.33% reported only one condition and the conditions may have higher relevance when accounting for other factors. The factor therefore has the alternative interpretation of multi-morbidity.

Factor(s)	Eigenvalue	RMSEA	SRMR	CFI	TLI
1	1.811	0.026	0.047	0.878	0.817
2	1.098	0.009	0.024	0.993	0.981

Table 6.5: Eigenvalues and model fit indices from the EFA of the chronic morbidity variables in the ELSA (Wave 4).

6.1.3 Functional Ability

Analysis of the functional ability variables revealed a two factor structure; one for activities requiring physical ability and the other requiring cognitive ability.

A two factor model was suggested by the EFA. The eigenvalues for factors one, two and three were 8.922, 1.833 and 0.769 respectively, and the model fit indices were RMSEA= 0.990, SRMR= 0.043, CFI= 0.990 and TLI= 0.986 for the two factor model. All demonstrating the two factor model therefore had good fit to the data. In terms of the RMSEA, CFI and TLI a single factor model also had good fit, however the SRMR was 0.104 suggesting the latent structure was not fully specified.

	Factor	
	One (η_1)	Two (η_2)
Dressing	**0.818**	-0.030
Walking across a room	**0.809**	0.111
Bathing	**0.842**	0.007
Eating	**0.551**	0.278
Getting in/out of bed	**0.910**	-0.113
Using the toilet	**0.853**	-0.078
Using a map	-0.013	**0.809**
Recognising danger	0.174	**0.736**
Preparing a meal	**0.614**	**0.412**
Shopping	**0.679**	**0.335**
Telephone calls	-0.031	**0.830**
Communication	0.003	**0.698**
Taking medications	0.218	**0.645**
Doing house/garden work	**0.805**	0.129
Managing money	0.191	**0.720**

Bold highlight of factors loadings greater than 0.316.

Table 6.6: Estimated factor loadings for the estimated two factor EFA of the functional ability variables in the ELSA (Wave 4).

The estimated factors loadings, shown in Table 6.6, and tetrachoric correlation coefficients (shown in Table 6.7) demonstrated a clear simple structure. The basic activities (first six variables) were together with association with factor along with the preparing a hot meal, shopping and doing house or garden work variables of the instrumental activities all of which require physical movement. Additionally, the remaining variables were associated with factor two and commonly required an element of cognitive ability. The factors were subsequently named physical function and cognitive function. Preparing a meal and shopping had cross-loading with large factor loadings from both factors, highlighting that some physical and cognitive ability was required, both variables were allocated to the factor with the larger loading (physical function in both instances).

	Dressing	Walking across a room	Bathing	Eating	Getting in/out of bed	Using the toilet	Using a map	Recognising danger	Preparing a meal	Shopping	Telephone calls	Communication	Taking medications	Doing house/garden work	Managing money
Dressing	1.000														
Walking across a room	0.682	1.000													
Bathing	0.701	0.738	1.000												
Eating	0.578	0.600	0.604	1.000											
Getting in/out of bed	0.723	0.737	0.668	0.584	1.000										
Using the toilet	0.679	0.732	0.688	0.512	0.754	1.000									
Using a map	0.289	0.440	0.336	0.422	0.324	0.280	1.000								
Recognising danger	0.462	0.564	0.460	0.553	0.448	0.492	0.769	1.000							
Preparing a meal	0.639	0.775	0.697	0.693	0.639	0.601	0.606	0.655	1.000						
Shopping	0.643	0.763	0.720	0.637	0.674	0.595	0.550	0.705	0.819	1.000					
Telephone calls	0.375	0.406	0.354	0.503	0.359	0.380	0.587	0.514	0.542	0.480	1.000				
Communication	0.345	0.429	0.338	0.396	0.277	0.352	0.501	0.527	0.442	0.435	0.711	1.000			
Taking medications	0.465	0.480	0.472	0.557	0.442	0.454	0.563	0.623	0.681	0.596	0.629	0.514	1.000		
Doing house/garden work	0.657	0.746	0.749	0.635	0.710	0.645	0.436	0.458	0.793	0.824	0.356	0.337	0.517	1.000	
Managing money	0.455	0.546	0.507	0.497	0.399	0.428	0.674	0.637	0.697	0.685	0.653	0.524	0.718	0.544	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.7: Estimated tetrachoric correlation coefficients for the functional ability variables in the ELSA (Wave 4).

6.1.4 Measurement Model of Health

The five factors cardiovascular health, metabolic health, chronic health, physical function and cognitive function were conglomerated into a measurement model of overall health, as a single second-order factor. Firstly, a confirmatory factor analysis (CFA) revealed the correlations between the factors were strong and positive; shown in Table 6.8. The correlations between chronic health and both the physical and cognitive function factors were strong, while cardiovascular health had moderate and positive correlations with all four other factors and metabolic health had low correlations with the other factors.

	Cardiovascular health	Metabolic health	Chronic health	Physical function	Cognitive function
Cardiovascular health	1.000				
Metabolic health	0.530	1.000			
Chronic health	0.536	0.291	1.000		
Physical function	0.432	0.199	0.876	1.000	
Cognitive function	0.437	0.090	0.804	0.862	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.8: Model estimated factor correlations from the CFA of the health factors in the ELSA (Wave 4).

Estimates of the model parameters, standard errors and r-squared values for the measurement model of health are given in Figure 6.1 and Table 6.9. Model fit was good with RMSEA=0.026, CFI=0.961 and TLI=0.958. For completeness, the chi-squared value of the model was $\chi^2 = 1,863.966$ with 429 degrees of freedom. In this section factor loadings refers to the standardised factor loadings and it will be specified when the unstandardised factor loadings are being referred to.

As expected, overall health had a strong positive association with chronic health, physical function and cognitive function with very high factor loadings, while metabolic health had low association. The factor loading for metabolic health was less than the conventional 0.300 cut-off point for determining an important factor, although statistically significant, thus had very little effect on overall health.

Both the physical function and the cognitive function factors had strong influences on all the corresponding latent response variables. Physical function had larger factor loadings for activities that required standing; walking across a room, preparing a meal and shopping. All activities that could be accomplished whilst sitting had lower factor loadings.

Cardiovascular health was most affected by heart failure and stroke, conversely to the EFA where stroke did not have a large factor loading. All the latent variables corresponding to cardiovascular health had large factor loadings, with heart murmur and abnormal heart rhythm had the smallest influence.

Similarly, chronic health was dominated by the neurodegenerative disorders with factor loading more than twice that of the other latent response variables. Previously, in the EFA, neurodegenerative disorders had the smallest factor loading while asthma had the second largest factor loading. Asthma and psychiatric problems had factor loadings less than 0.316 in the measurement model.

Hence, neither the cardiovascular health or chronic health factors reflect multi-morbidity in the factor loadings, instead reflect the impact of the conditions on overall health.

The bootstrapped estimates were, on average, 47.9% larger than the asymptotically estimated standard errors. The majority of this was contributed to by the functional status variables, where the bootstrapped standard errors were more than twice that of the original estimates. Otherwise, discounting the functional status variables, the bootstrapped standard errors were only 29.1% larger than the original. The reported standard error estimates for the parameters for functional status in the later SEM analysis should therefore be interpreted lightly as bootstrapped estimates were not possible due to the large number of variables and the low frequencies for some variables.

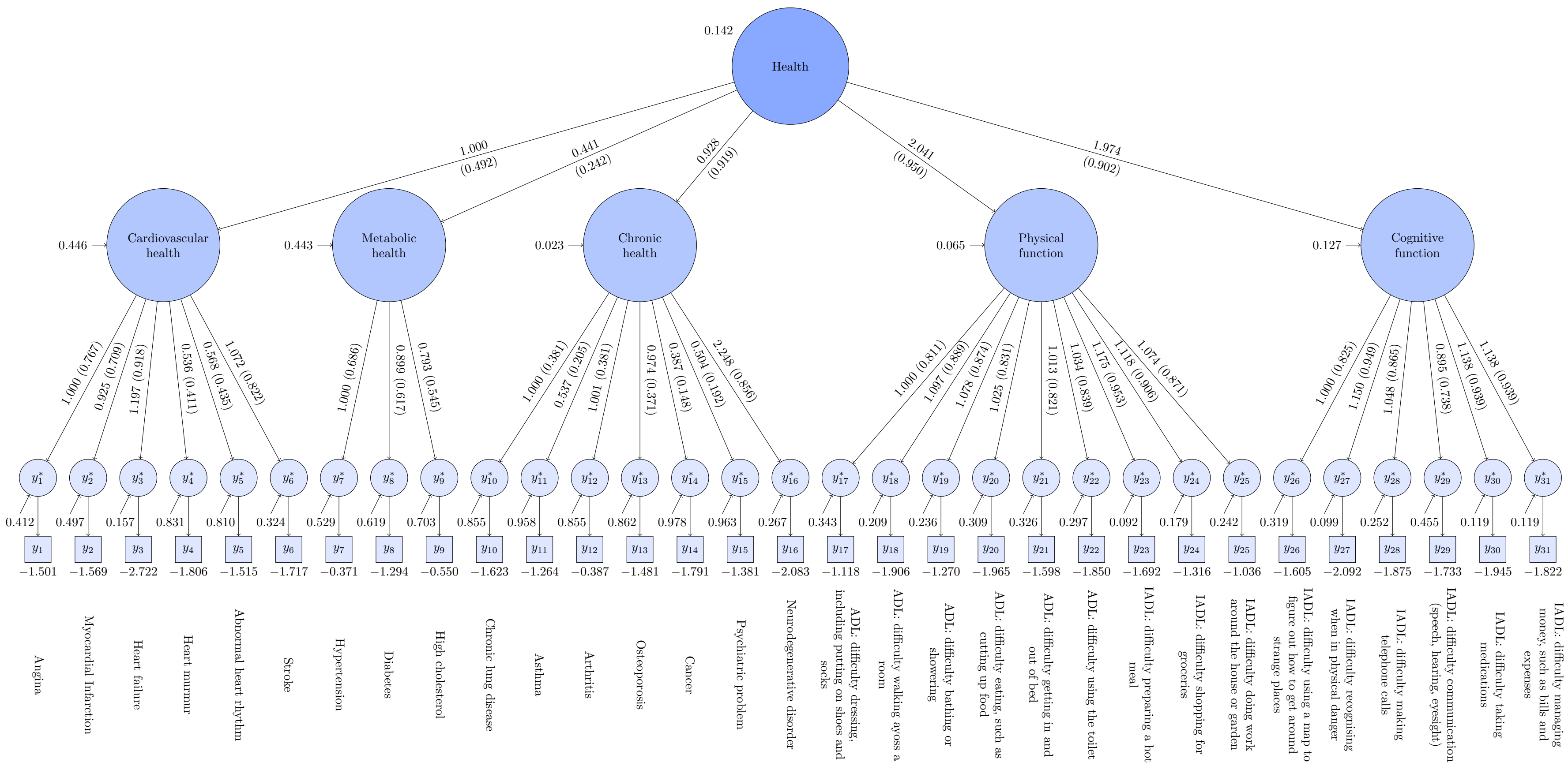


Figure 6.1: Path diagram for the estimated measurement model of health in the ELSA (Wave 4).

Variable	Standard Error Original	Standard Error Bootstrapped	R-squared
	s.e.($\hat{\lambda}_{m,h}$)		
Angina	-	-	0.611
Heart attack	0.046	0.046	0.506
Heart failure	0.079	0.115	0.703
Heart murmur	0.051	0.061	0.197
Abnormal heart rhythm	0.044	0.055	0.206
Stroke	0.057	0.081	0.675
Hypertension	-		0.499
Diabetes	0.073	0.075	0.390
High cholesterol	0.061	0.052	0.265
Chronic lung disease	-	-	0.146
Asthma	0.063	0.067	0.036
Arthritis	0.079	0.093	0.150
Osteoporosis	0.085	0.086	0.115
Cancer	0.089	0.077	0.022
Psychiatric problems	0.071	0.071	0.049
Neurodegenerative disorder	0.153	0.151	0.633
Dressing	-	-	0.681
Walking across a room	0.016	0.026	0.800
Bathing	0.013	0.021	0.786
Eating	0.021	0.039	0.683
Getting in/out of bed	0.015	0.023	0.715
Using the toilet	0.018	0.027	0.713
Preparing a meal	0.014	0.024	0.898
Shopping	0.013	0.023	0.843
Doing house/garden work	0.014	0.023	0.779
Using a map	-		0.657
Recognising danger	0.025	0.057	0.885
Telephone calls	0.022	0.057	0.781
Communication	0.025	0.051	0.558
Taking medications	0.023	0.061	0.871
Managing money	0.021	0.053	0.872
	s.e.($\hat{\beta}_m$)		
Cardiovascular health	-	-	0.276
Metabolic health	0.047	0.057	0.097
Chronic health	0.072	0.079	0.961
Physical function	0.099	0.120	0.862
Cognitive function	0.091	0.109	0.796

Table 6.9: Model estimated standard errors of the unstandardised factor loadings, original and bootstrapped, and R-squared estimates for the health measurement model, of the ELSA (Wave 4).

6.2 Economic Well-being

Economic well-being variables comprised of the two aspects; durables owned and accommodation problems. Similarly to above, the factor structure identified in the EFA is given for each aspect, then a measurement model of overall economic well-being is considered. Associations between economic well-being and net financial wealth and total wealth are considered as issues occurred from the due to the subjective variables.

6.2.1 Durables owned

There were 12 binary variables for the durables owned aspect, discussed in Section 3.2.1, that indicated the possession of the durable. Howeve,r the variables indicating possession of a television, fridge freezer or landline telephone were perfectly, or near perfectly, correlated, therefore, to include more than one of these variables would be redundant. Alternatively, all three durables were owned by 81.6% and the correlation with other the variables was too high to consider an aggregated variable. As a result, all three variables were removed from the analysis.

Estimated tetrachoric correlation coefficients between the remaining items are shown in Table 6.10. Higher correlations were generally with the CD player, washing machine and microwave oven variables, whereas the tumble dryer, dishwasher and digital TV variables had lower correlations. There was, however, no visibly obvious clustering of variables.

The EFA revealed a single factor model was most appropriate for the variables. The eigenvalues were 5.782 and 0.903 for one and two factors respectively. The model fit indices for one factor, SRMR=0.052, CFI=0.988 and TLI=0.984, demonstrated good fit, however the RMSEA=0.059 was marginal but indicated only adequate fit. Unfortunately a two factor model failed to converge.

The estimated factor loadings were; 0.839 video recorder, 0.948 CD player, 0.898 washing machine, 0.577 tumble dryer, 0.586 tumble dryer, 0.890 microwave oven, 0.748 computer, 0.527 digital tv and 0.901 DVD player. Supporting the conclusion for a single factor, all factor loadings were greater than 0.316, with a minimum of 0.527.

	Video recorder	CD player	Washing machine	Tumble dryer	Dishwasher	Microwave oven	Computer	Digital TV	DVD player
Video recorder	1.000								
CD player	0.813	1.000							
Washing machine	0.776	0.814	1.000						
Tumble dryer	0.485	0.538	0.523	1.000					
Dishwasher	0.345	0.536	0.580	0.404	1.000				
Microwave oven	0.776	0.799	0.850	0.562	0.482	1.000			
Computer	0.548	0.719	0.637	0.408	0.580	0.594	1.000		
Digital TV	0.370	0.528	0.505	0.233	0.155	0.507	0.443	1.000	
DVD player	0.761	0.881	0.748	0.493	0.502	0.763	0.701	0.516	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.10: Estimated tetrachoric correlation coefficients for the durables owned variables in the ELSA (Wave 4).

6.2.2 Accommodation problems

The estimated tetrachoric correlation coefficients for the 12 accommodation problem variables are given in Table 6.13, that were overall relatively low with some slightly larger values on the right-hand side of the table. Pollution and other street noise, however, had a much stronger positive correlation than the other variables. Variables concerned with water (water leaks, bad condensation and rising damp) also had notably higher correlation coefficients with each other.

Factor(s)	Eigenvalue	RMSEA	SRMR	CFI	TLI
1	4.513	0.028	0.104	0.860	0.829
2	1.452	0.014	0.055	0.974	0.960

Table 6.11: Eigenvalues and model fit indices from the EFA of the accommodation problem variables in the ELSA (Wave 4).

In the EFA, a two factor model was discovered to have good fit and factors that were interpreted as problems inside the accommodation and problems outside the accommodation, in the neighbourhood. Eigenvalues and model fit indices from the EFA are given in Table 6.11. The RMSEA indicated good fit was achieved with just one factor, while the SRMR, CFI and TLI indicated two factors were required. Conversely the eigenvalues dropped below 1.000 after three factors. The third factor was associated with the shortage of space and too dark variables, thus would have lacked model identification.

The estimated factor loadings determined the separation of variables into the two factors (shown in Table 6.12); one associated with the household and the other with the wider area. Factor one had large factor loading with the noise from neighbours, the other street noise and the pollution/grime variables, while the variables associated with problems with the household itself, such as shortage of space and bad condensation problems had large factor loadings with the second factor. The two factors were subsequently named neighbourhood quality and household quality respectively.

	Factor	
	One (η_1)	Two (η_2)
Shortage of space	0.125	**0.352**
Noise from neighbours	0.277	0.261
Other street noise	**0.945**	-0.005
Too dark/not enough light	0.140	**0.361**
Pollution/grime	**0.695**	0.195
Rising damp	0.049	**0.625**
Water leaks	-0.051	**0.591**
Bad condensation problems	-0.053	**0.751**
Electrical or plumbing problems	0.026	**0.625**
Rot and decay	-0.067	**0.820**
Insects, mice or rats	0.095	**0.541**
Too cold in winter	0.015	**0.647**

Bold highlight of factors loadings greater than 0.316.

Table 6.12: Estimated factor loadings from the EFA with two factors of the accommodation problem variables in the ELSA (Wave 4).

	Shortage of space	Noise from neighbours	Other street noise	Too dark	Pollution	Rising damp	Bad condensation	Water leaks	Electrical/ plumbing problems	Rot and decay	Insects, mice or rats	Too cold in winter
Shortage of space	1.000											
Noise from neighbours	0.227	1.000										
Other street noise	0.212	0.351	1.000									
Too dark	0.402	0.131	0.218	1.000								
Pollution	0.219	0.298	0.723	0.268	1.000							
Rising damp	0.240	0.199	0.272	0.143	0.327	1.000						
Water leaks	0.062	0.207	0.175	0.074	0.169	0.462	1.000					
Bad condensation	0.358	0.202	0.213	0.271	0.256	0.469	0.358	1.000				
Electrical/plumbing problem	0.298	0.258	0.188	0.206	0.360	0.368	0.386	0.422	1.000			
Rot and decay	0.164	0.248	0.173	0.247	0.396	0.544	0.452	0.612	0.520	1.000		
Insects, mice or rats	0.127	0.309	0.271	0.248	0.290	0.358	0.379	0.409	0.371	0.465	1.000	
Too cold in Winter	0.240	0.217	0.234	0.298	0.295	0.396	0.390	0.497	0.439	0.497	0.372	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.13: Estimated tetrachoric correlation coefficients for the accommodation problem variables in the ELSA (Wave 4).

6.2.3 Measurement Model of Economic Well-being

CFA of all three factors (ownership, household quality and neighbourhood quality) revealed small and negative correlations with the ownership factor; while the household and neighbourhood quality factors had a moderate positive correlation (shown in Table 6.14). Thus, one of the aspects did not reflect the economic well-being of the elderly.

	Neighbourhood quality	Household quality	Ownership
Neighbourhood quality	1.000		
Household quality	0.545	1.000	
Ownership	-0.045	-0.023	1.000

Table 6.14: Model estimated factor correlations from the CFA of the Economic Well-being factors in the ELSA (Wave 4).

The objective measures, financial wealth and total wealth, of economic status were modelled as covariates in separate SEMs in order to identify whether the three factors had the expected positive association (see Figure 6.2). To recap from Section 3.2.1, page 48, financial wealth was the combined value of investments such as savings, ISAs and shares, minus the value of debt from credit cards, private debt *etc.* Total wealth also included the value of businesses and property owned, and the primary housing wealth (value of the home minus the debt). Both variables were coded as their weighted deciles, with the first decile corresponding to the lowest value. The estimated regression coefficients, $\mathbf{\Gamma}$, for each model, are given in Table 6.15 and Table 6.16; with the respective standard errors and p-values.

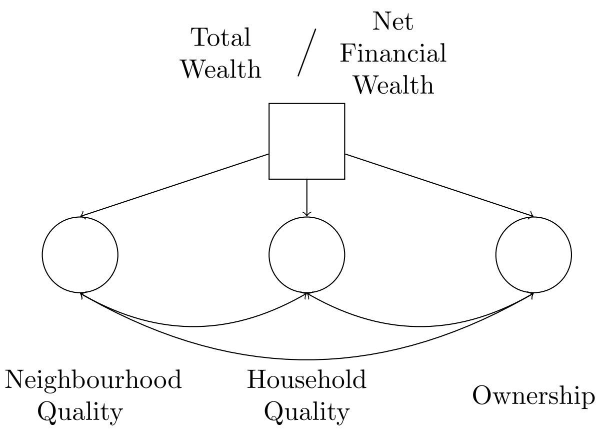


Figure 6.2: Path diagram for the MIMIC model including the net financial wealth or total wealth as a covariate of the economic well-being factors in the ELSA (Wave 4).

	Neighbourhood quality			Household quality			Ownership		
	Estimate	s.e.	p-value	Estimate	s.e.	p-value	Estimate	s.e.	p-value
Weighted Quantile									
First	-	-	-	-	-	-	-	-	-
Second	0.204	0.065	0.002	0.117	0.052	0.025	0.034	0.048	0.483
Third	0.035	0.062	0.572	0.136	0.054	0.011	0.214	0.053	< 0.0005
Fourth	0.331	0.066	< 0.0005	0.296	0.054	< 0.0005	0.176	0.051	< 0.0005
Fifth	0.212	0.060	< 0.0005	0.328	0.057	< 0.0005	0.226	0.051	< 0.0005
Sixth	0.210	0.063	0.001	0.326	0.058	< 0.0005	0.246	0.051	< 0.0005
Seventh	0.166	0.064	0.010	0.411	0.062	< 0.0005	0.138	0.053	0.009
Eighth	0.250	0.066	< 0.0005	0.433	0.063	< 0.0005	0.114	0.054	0.033
Ninth	0.230	0.064	< 0.0005	0.360	0.058	< 0.0005	0.128	0.054	0.017
Tenth	0.308	0.066	< 0.0005	0.303	0.056	< 0.0005	0.047	0.055	0.394

Table 6.15: Model estimated regression coefficients for the weighted total wealth deciles on the economic well-being factors in the ELSA (Wave 4).

	Neighbourhood quality			Household quality			Ownership		
	Estimate	s.e.	p-value	Estimate	s.e.	p-value	Estimate	s.e.	p-value
Weighted Quantile									
First	-	-	-	-	-	-	-	-	-
Second	0.053	0.066	0.421	0.194	0.057	0.001	-0.189	0.061	0.002
Third	0.167	0.068	0.014	0.304	0.056	< 0.0005	-0.139	0.059	0.018
Fourth	0.195	0.068	0.004	0.342	0.060	< 0.0005	-0.104	0.060	0.083
Fifth	0.178	0.066	0.007	0.415	0.062	< 0.0005	-0.019	0.060	0.747
Sixth	0.224	0.069	0.001	0.454	0.068	< 0.0005	0.027	0.060	0.654
Seventh	0.244	0.070	< 0.0005	0.394	0.059	< 0.0005	-0.056	0.062	0.361
Eighth	0.195	0.067	0.004	0.572	0.069	< 0.0005	-0.059	0.061	0.338
Ninth	0.323	0.071	< 0.0005	0.485	0.063	< 0.0005	-0.072	0.061	0.241
Tenth	0.225	0.068	0.001	0.453	0.062	< 0.0005	-0.056	0.062	0.363

Table 6.16: Model estimated regression coefficients for the weighted net financial wealth deciles on the economic well-being factors in the ELSA (Wave 4).

In comparison to the lowest decile of total wealth, higher total wealth was associated with higher levels of neighbourhood quality, household quality and ownership as excepted. Estimated regression weights for the effect of net financial wealth on ownership, however, were negative and/or insignificant. The second and third deciles of net financial wealth, however, were associated with lower levels of ownership, while the higher deciles had no significant association. In contrast, ownership had a significantly positive association with total wealth.

This indicated possession of durables was a decision of luxury which was contrary to our aim to measure economic well-being in relation to its impact on the everyday life and health of the elderly, rather than the direct economic position. Therefore, the ownership factor was not continued into the measurement model for economic well-being, resulting in only the two factors, neighbourhood quality and household quality, remaining.

As previously mentioned, a factor is required to load onto a minimum of three variables to ensure model identification. Thus, a single factor of economic well-being was not possible and was instead modelled with the two correlated factors; household and neighbourhood quality.

A path diagram of the resulting model is shown in Figure 6.3 with the estimated standard errors and R-squared given in Table 6.17. The model fit indices demonstrated good model fit (RMSEA=0.015, CFI=0.962 and TLI=0.953). The estimated chi-squared was $\chi^2 = 129.841$ on 53 degrees of freedom.

	Standard Error		R-squared
	Original	Bootstrapped	
Noise from neighbours	-	-	0.240
Other street noise	0.157	0.160	0.589
Pollution/grime	0.176	0.192	0.814
Shortage of space	-	-	0.185
Too dark	0.148	0.130	0.204
Rising damp	0.155	0.173	0.419
Water leaks	0.155	0.175	0.304
Bad condensation problem	0.170	0.167	0.506
Electrical/plumbing problem	0.171	0.176	0.408
Rot and decay	0.193	0.214	0.602
Infestation	0.144	0.159	0.347
Too cold in winter	0.151	0.163	0.418

Table 6.17: Model estimated standard errors of the unstandardised factor loadings, original and bootstrapped, and R-squared estimates for the economic well-being measurement model, of the ELSA (Wave 4).

Correlation between the two factors was positive but only moderate, demonstrating an overall factor of economic well-being was not appropriate. In other words, the quality of the household does not pertain to the quality of the neighbourhood. Therefore, economic well-being should

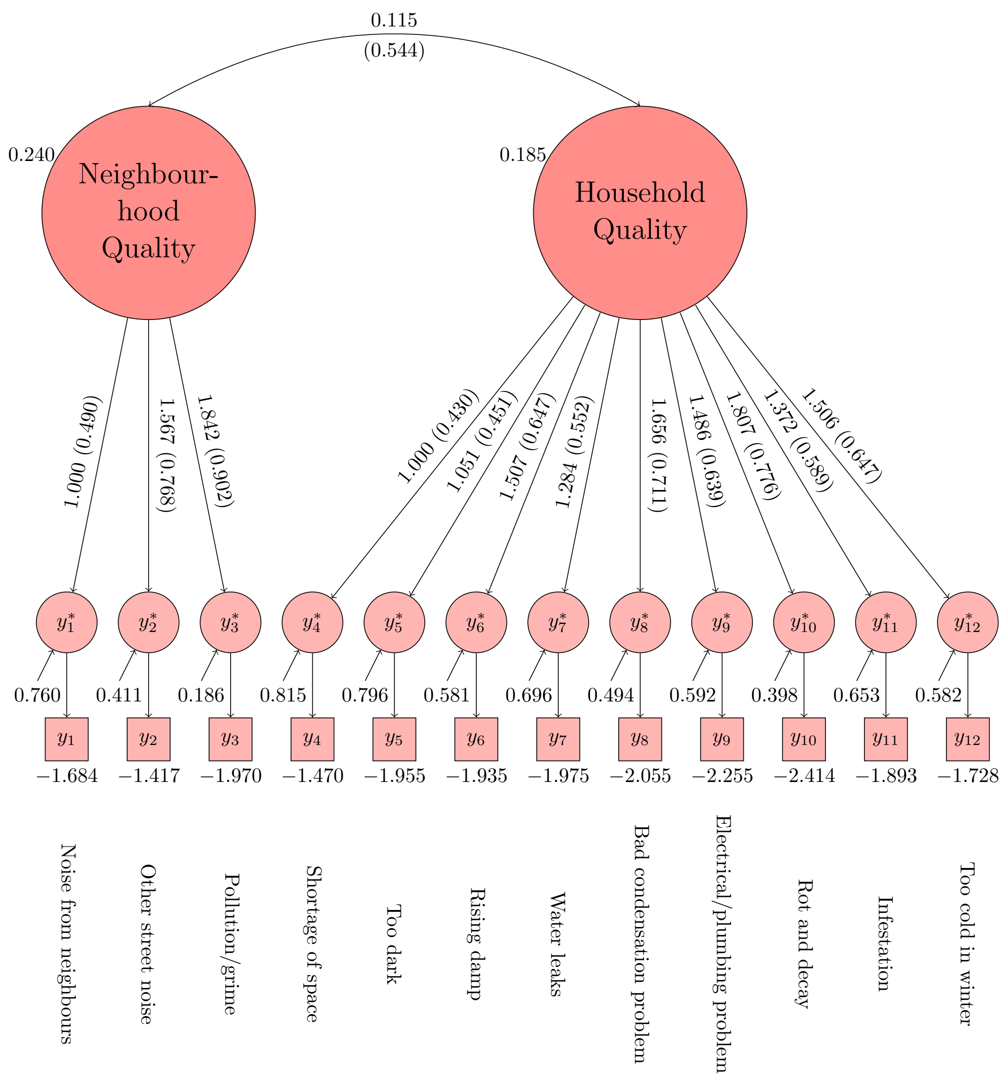


Figure 6.3: Path diagram for the estimated measurement model of economic well-being in the ELSA (Wave 4).

be interpreted separately in terms of the household in which the elderly person resides and the location of the household.

Pollution and grime was the most influential variable of the neighbourhood quality, and noise from neighbours the least. The R-squared estimate demonstrated that 81.4% of the variance in pollution/grime was explained by the factor. The lower factor loading of noise from neighbours may be due to the influence of the proximity of the neighbours, for example the noise level would be expected to have greater impact in a terraced house or a block of flats, where neighbours are within close proximity, than in a detached house in less proximity.

Standardised factor loadings of the household quality factor ranged from 0.430 (shortage of space) to 0.776 (rot and decay). The most influential variables, with higher factor loadings, were consequences of poor build quality and maintenance such as rot and decay, condensation problems, too cold *etc.*, whereas less influential variables were consequences of poor design in the layout of the building with the lack of space and light.

The bootstrapped standard errors were only marginally larger than the original asymptotically estimated standard errors, with an increase of 5.4% on average. The small difference indicates that the sample may be large enough for the asymptotic standard error to provide good estimation of the true standard errors.

6.3 Subjective Well-being

In addition to the measurement of health and economic well-being that were considered for the 2008 PNAD, the ELSA presented the opportunity to investigate the measurement of subjective well-being of the elderly. This section will analyse the latent structures of the variables concerning aspects of life satisfaction, loneliness, depression and quality of life. EFA was applied to the life satisfaction, loneliness and depression aspects, while the proposed multidimensional structure of the CASP-19 measure of quality of life was explored using item response theory (IRT) methods. In the IRT analysis the performance of the suggested test and items were analysed in terms of the discrimination ability and difficulty, particularly following repeated use of the measure in preceding waves of the ELSA. Finally, the measurement model for subjective well-being.

6.3.1 Life Satisfaction

The estimated polychoric correlation coefficients for the life satisfaction variables are given in Table 6.18, all of which were strong and positive. Notably the fifth variable 'If I could live my life over I would change almost nothing' had the lowest correlations of all the other variables.

Given there were only five variables only a single factor model was considered, which was supported by the EFA. Model fit was good with CFI=0.991 and TLI=0.981, for a single factor, and the eigenvalues were 3.740 to 0.531 for one and two factors respectively; thus also concluding a single factor. The RMSEA=0.169, however, did not indicate good fit, and neither did RMSEA=0.079 for two factors. In this case the RMSEA may be influenced by the large sample size and small number of latent response variables, as previously mentioned in Section 4.1.4.

	Close to ideal	Conditions are excellent	Satisfied with my life	Gotten the important things	Live my life over again
Close to ideal	1.000				
Conditions are excellent	0.822	1.000			
Satisfied with my life	0.791	0.808	1.000		
Gotten the important things	0.675	0.657	0.746	1.000	
Live my life over again	0.582	0.552	0.588	0.591	1.000

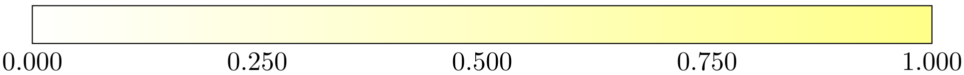


Table 6.18: Estimated polychoric correlation coefficients for the life satisfaction variables in the ELSA (Wave 4).

6.3.2 Loneliness

Similarly for the five loneliness variables, only a single factor model was considered appropriate and the RMSEA did not indicate good fit with either one or two factors. The estimated model fit indices were RMSEA=0.111, CFI=0.990 TLI=0.981 and the first two eigenvalues were 3.340 and 0.981.

The estimated polychoric correlation coefficients are given in Table 6.19, demonstrating very high correlation between all the variables except the fourth variable (In tune with the people around you) which had dramatically lower correlations with all the other variables. The estimated rotated factor loadings reflected the same structure: 0.902 lack companionship, 0.859 'left out', 0.867 'isolated from others', 0.131 'in tune with the people around me' and 0.924 'felt lonely'. The in tune with the people around me variable was seen to be poorly associated with the other variables. It is possible that, as the only question that was positively worded in the collection, that some of the elderly respondents did not recognise this and additional error occurred.

Item	Lack companionship	Left out	Isolated from others	In tune with the people around you	Felt lonely
Lack companionship	1.000				
Left out	0.719	1.000			
Isolated from others	0.736	0.810	1.000		
In tune with the people around you	0.071	0.139	0.143	1.000	
Felt lonely	0.864	0.746	0.757	0.119	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.19: Estimated polychoric correlation coefficients for loneliness variables in the ELSA (Wave 4).

6.3.3 Depression

The aspect of depression consisted of eight binary variables that had positive tetrachoric correlation shown in Table 6.20. As could be expected, stronger correlation coefficients; were between the happy and enjoy variables, depressed and sad variables and the depressed and effort variables.

A two factor structure was discovered in the EFA, one of which was the somatic factor, previously discussed in Section 3.2.1. The eight variables in the original 20 item scale (Radloff, 1977) belonged to three factors. The depressed, lonely and sad variables belonged to the depressed affect, the happy and enjoy variables belonged to the positive affect and the effort, sleep and get going variables belonged to the somatic factor. The estimated factor loadings with two factors are given in Table 6.22, where the somatic variables effort, restless and get going have strong associations with factor two whilst the remaining variables had stronger associations with factor one.

Table 6.21 gives the model fit indices and eigenvalues for the EFA with one and two factors. A

	Depressed	Effort	Restless	Happy	Lonely	Enjoy	Sad	Get going
Depressed	1.000							
Effort	0.726	1.000						
Restless	0.488	0.494	1.000					
Happy	0.712	0.519	0.362	1.000				
Lonely	0.620	0.490	0.321	0.583	1.000			
Enjoy	0.714	0.595	0.398	0.848	0.618	1.000		
Sad	0.728	0.504	0.403	0.664	0.670	0.674	1.000	
Get going	0.605	0.771	0.459	0.469	0.483	0.569	0.503	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.20: Estimated tetrachoric correlation coefficients for the CES-Depression scale variables in the ELSA (Wave 4).

Factor(s)	Eigenvalue	RMSEA	SRMR	CFI	TLI
1	5.049	0.081	0.074	0.962	0.947
2	0.917	0.045	0.030	0.992	0.983

Table 6.21: Eigenvalues and model fit indices from the EFA for the CES-Depression scale variables in the ELSA (Wave 4).

single factor model was supported by the eigenvalues and the CFI, however the RMSEA, SRMR and TLI required a two factor model to indicate good model fit.

	Factor	
	One (η_1)	Two (η_2)
Depressed	**0.587**	**0.365**
Effort	-0.006	**0.947**
Restless	0.161	**0.434**
Happy	**0.992**	-0.141
Lonely	**0.665**	0.092
Enjoy	**0.902**	< 0.0005
Sad	**0.756**	0.077
Get going	0.081	**0.758**

Bold highlight of factors loadings greater than 0.316.

Table 6.22: Estimated factors loadings from the EFA with two factors for the CES-Depression scale variables in the ELSA (Wave 4).

The eight item CES-Depression scale is intended to measure a single factor of depression, therefore to maintain consistency with previous studies and research (see Section 2.3.2) it was decided to retain a single factor model.

6.3.4 Quality of Life

Quality of life was measured using the 19 ordinal variables from the CASP-19, which has a predefined structure with four factors. This structure was suggested by Higgs et al. (2003) with the control, autonomy, self-realisation and pleasure dimensions as the factors. Control was the influencing factor of the first four variables, then the other factors influenced the remaining variables respectively in groups of five. More detailed descriptions were given in Sections 2.3.5 and 3.2.1.

It was therefore prudent to apply this structure so that any conclusions were comparable with previous research involving the CASP-19. However, the resulting confirmatory factor analysis was invalid with some Heywood estimates. Specifically, the estimated correlation between the control and autonomy factors was greater than one.

The matrix of estimated polychoric correlation coefficients shown in Table 6.23 suggested that the predetermined structure was not present in this data. For example, Item 8 from the Autonomy dimension had stronger correlations with Item 1 of the Control dimension and Item 15 of the Self-realisation dimension than with the other variables of its own dimension.

Hence, this section will consist of exploratory factor analysis as in previous sections in order to obtain a valid model. Multidimensional item response theory (IRT) then follows to evaluate the items' difficulty and discrimination ability. The model is equivalent to a standardised confirmatory factor analysis model; with the variance of the factors fixed to one and all factor loadings freely estimated. In the framework of IRT, however, the focus is the performance of the items, mainly how well they each relate to their construct and whether they together provide a balanced measure of the whole continuum.

	Items	Control 1	2	3	4	Autonomy 5	6	7	8	9	Pleasure 10	11	12	13	14	Self-realisation 15	16	17	18	19
Control	1	1.000																		
	2	0.512	1.000																	
	3	0.286	0.360	1.000																
	4	0.339	0.479	0.307	1.000															
Autonomy	5	0.364	0.411	0.594	0.341	1.000														
	6	0.090	0.218	0.161	0.241	0.116	1.000													
	7	0.111	0.205	0.411	0.211	0.512	0.405	1.000												
	8	0.678	0.482	0.326	0.329	0.431	0.057	0.112	1.000											
	9	0.214	0.267	0.244	0.256	0.238	0.228	0.169	0.263	1.000										
Pleasure	10	0.265	0.368	0.488	0.417	0.501	0.113	0.350	0.280	0.198	1.000									
	11	0.270	0.343	0.463	0.416	0.457	0.070	0.285	0.267	0.164	0.774	1.000								
	12	0.275	0.380	0.503	0.398	0.535	0.139	0.394	0.307	0.200	0.748	0.724	1.000							
	13	0.117	0.141	0.263	0.173	0.283	0.039	0.226	0.144	0.095	0.436	0.416	0.528	1.000						
	14	0.106	0.260	0.370	0.347	0.379	0.139	0.291	0.149	0.222	0.603	0.587	0.582	0.448	1.000					
Self-Realisation	15	0.525	0.427	0.453	0.351	0.518	0.025	0.255	0.662	0.186	0.565	0.505	0.534	0.329	0.409	1.000				
	16	0.327	0.252	0.337	0.160	0.357	-0.058	0.186	0.345	0.071	0.335	0.388	0.388	0.250	0.225	0.540	1.000			
	17	0.240	0.385	0.482	0.454	0.491	0.120	0.323	0.297	0.279	0.639	0.620	0.638	0.337	0.707	0.508	0.320	1.000		
	18	0.342	0.342	0.492	0.361	0.481	0.047	0.286	0.366	0.200	0.581	0.587	0.587	0.366	0.485	0.588	0.512	0.616	1.000	
	19	0.409	0.438	0.573	0.425	0.537	0.090	0.317	0.444	0.277	0.667	0.644	0.644	0.385	0.563	0.665	0.457	0.713	0.733	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.23: Estimated polychoric correlation coefficients for the quality of life (CASP-19) variables in the ELSA (Wave 4).

To obtain a valid model, a new latent structure for the quality of life variables was defined, applying the same EFA methods as in previous aspects. The EFA confirmed four factors was most appropriate. Eigenvalues for four and five factors were 1.066 and 0.818 respectively and the model fit indices with four factors were RMSEA=0.060, SRMR=0.027, CFI=0.977 and TLI=0.961; the latter three indicating good model fit. The RMSEA did not indicate good fit until at least six factors were included; while the fifth factor had large factor loadings with only two variables (Items 2 and 4).

The structure suggested by the estimated factor loadings (Table 6.24) supported the notion that the claimed structure was not present. The first four of items from the control factor of the claimed structure were separated among the four new factors. There was a clear structure except for several items having strong factor loadings from two factors and item 9 was not strongly associated with any of the factors. All items, except item 4, were allocated to the factor with greatest factor loading.

		Factor			
	Item	One (η_1)	Two (η_2)	Three (η_3)	Four (η_4)
1	age prevents me	**0.392**	-0.012	**0.643**	-0.034
2	out of my control	**0.480**	0.249	0.273	0.064
3	free to plan	-0.007	0.277	0.252	**0.379**
4	left out	**0.451**	**0.471**	0.007	0.002
5	can do things	0.022	0.183	**0.349**	**0.462**
6	family responsibilities	**0.353**	-0.014	-0.208	**0.460**
7	can please myself	-0.024	-0.023	-0.010	**0.845**
8	health stops me	**0.340**	-0.045	**0.780**	-0.030
9	shortage of money	0.304	0.174	0.030	0.130
10	look forward	-0.013	**0.846**	0.005	0.011
11	life has meaning	-0.024	**0.850**	0.006	-0.050
12	enjoy things	-0.035	**0.752**	0.052	0.112
13	company of others	-0.139	**0.495**	0.013	0.037
14	sense of happiness	0.043	**0.880**	-0.220	-0.034
15	full of energy	0.024	**0.317**	**0.651**	-0.006
16	never done before	-0.208	0.174	**0.527**	0.043
17	satisfied with life	0.106	**0.839**	-0.032	-0.007
18	full of opportunities	-0.110	**0.573**	**0.333**	0.025
19	future looks good	0.010	**0.663**	0.311	0.026

Bold highlight of factors loadings greater than 0.316.

Table 6.24: Estimated factor loadings from the EFA with four factors of the Quality of Life variables in the ELSA (Wave 4).

Factor one was seen as the statements related to the lack of control due to age, health or money, as well as feeling out of control. Increasing values of the factor indicate better quality of life, therefore these items indicated improved levels of control. Thus, the factor was named control. Allocating items to the factor with the largest loading, however, left only items 2 and 9.

The other factor item 4 was influenced by was factor two, that was associated with enjoying life, such as looking forward to the future, enjoying the company of others and being satisfied with life so far. Factor two was then named enjoyment. Item 4 was subsequently allocated to factor one for which feeling left out was more meaningful.

Factors three and four could be considered to represent opportunity and independence. Factor three was associated with doing new things, having opportunities and then ability to fulfil these with energy and without prevention from age and health. Factor four, on the other hand, related to the person's ability to be independent through freedom to plan and do things for themselves, and without other responsibilities.

The factors were therefore allocated as such: the Control factor consisted of items 2, 4 and 9, the Enjoyment factor of items 10-14 and 17-19, the Opportunity factor of items 1, 8, 15 and 16, and the Independence factor of items 3, 5, 6 and 7.

A multidimensional two parameter normal ogive model was then fitted with the newly structured quality of life variables. Model fit, as expected from the EFA, was bad with RMSEA=0.096, although the CFI=0.916 and TLI=0.902 did indicate adequate model fit. This was suspected to be due to poor performing items and the cross-loading seen in the EFA. Estimates of the discrimination and difficulty parameters for the items are given in Table 6.25.

Several items were located to the far left with negative estimates for $\hat{\tau}_{3,h}$. Thus, an individual with at least an average factor score had more than 50% chance of using the top most category (*i.e.* Often for positively worded items). These items provide more information about individuals whose factor scores are below average, but no information for individuals with scores above average.

The majority of items had good discrimination estimates, greater than 0.5, however several items were identified that had poor discriminatory ability. Items 2, 9 and 13 had low abilities to discriminate between individuals with different categorical responses and locations on the factor continuum. For example, seen below in the category response curve for item 6 (Figure 6.4), an individual located 3 standard deviations below the mean independence factor score, had 16.1% chance of responding with often, 40.4% with sometimes 28.8% with not often and 14.6% with never, thus were reasonably likely to have used any of the categories.

Item	Discrimination		Difficulty	
h	$\hat{\lambda}_h$	$\hat{\tau}_{1,h}$	$\hat{\tau}_{2,h}$	$\hat{\tau}_{3,h}$
Control				
2	0.712	-1.469	-0.292	0.857
4	0.662	-1.821	-0.667	0.267
9	0.396	-1.174	-0.242	0.598
Enjoyment				
10	0.837	-2.222	-1.669	-0.612
11	0.803	-1.993	-1.322	-0.300
12	0.827	-2.514	-2.087	-0.806
13	0.484	-2.557	-1.968	-0.604
14	0.698	-2.350	-1.622	-0.377
17	0.808	-1.858	-1.192	0.033
18	0.789	-1.633	-0.697	0.452
19	0.886	-1.687	-0.823	0.405
Opportunity				
1	0.677	-1.006	0.096	0.975
8	0.740	-0.933	-0.115	0.480
15	0.933	-1.424	-0.531	0.830
16	0.602	-1.169	-0.033	1.336
Independence				
3	0.781	-1.566	-0.831	0.196
5	0.831	-1.937	-1.246	-0.101
6	0.240	-1.680	-0.559	0.304
7	0.539	-1.973	-1.328	-0.100

Table 6.25: Estimated item discrimination and difficulty parameters from the unidimensional IRT model of the quality of life variables in the ELSA (Wave 4).

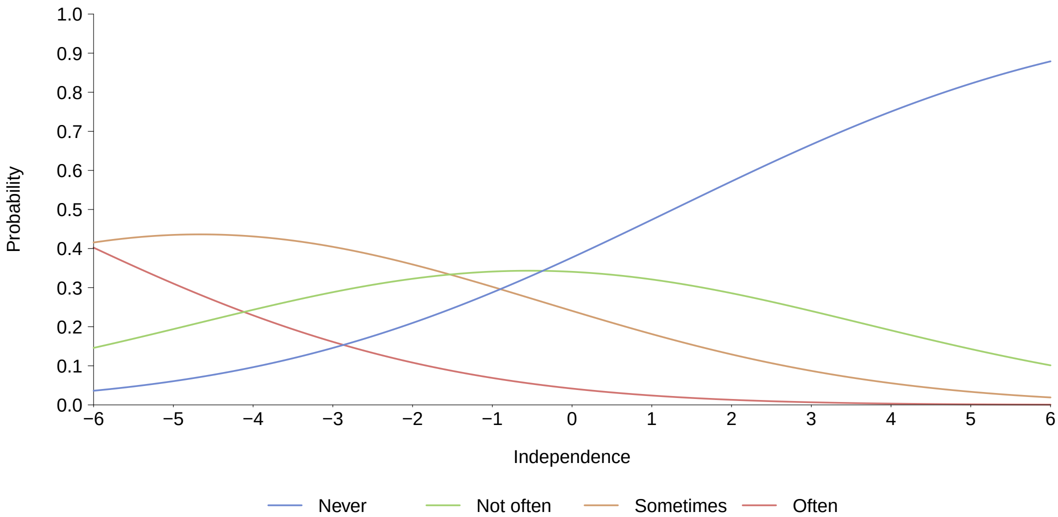
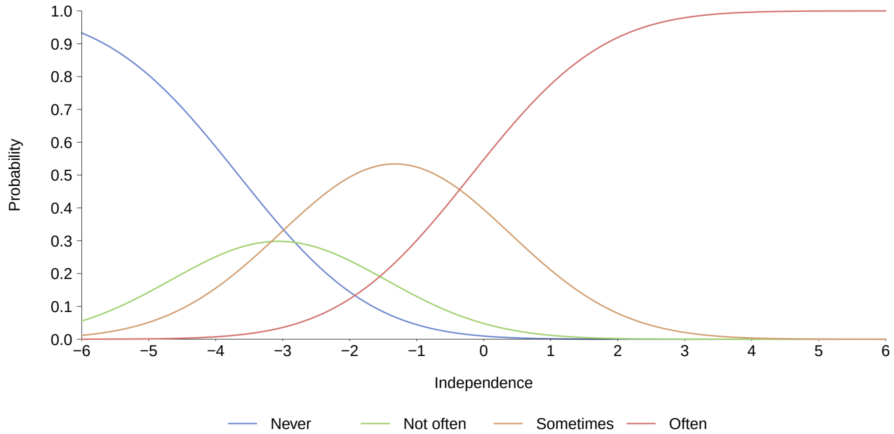
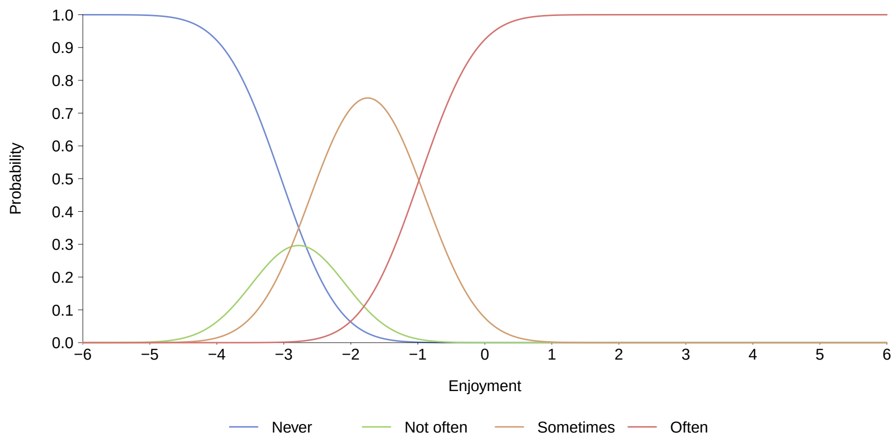


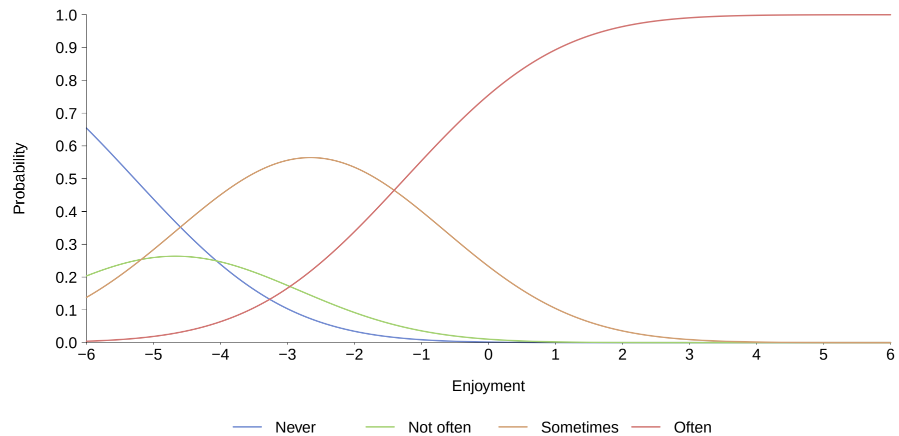
Figure 6.4: Category response curves for Item 6 from the unidimensional IRT of the quality of life variables in the ELSA (Wave 4).



(a) Item 7



(b) Item 12



(c) Item 13

Figure 6.5: Category response curves for items with unused categories in the unidimensional IRT of the quality of life variables in the ELSA (Wave 4).

The category response curves also highlighted that the not often category was unused by individuals. A category was classed as unused if, for any value on the continuum, it was not the most likely response. In other words, the curve on the category response curves did not appear as the top line at any point on the factor continuum, as seen in Figure 6.5. Low use of the not often category was seen in both positively and negatively worded items. The Likert scale was not properly utilised by the respondents.

For comparison the total information functions for all fours factor are shown on the same plot in Figure 6.6. Relatively the items provided little information about the control factor and the greatest amount of information about the enjoyment factor. Control reached a maximum of 0.440 ($\hat{\eta}_1 \approx -0.19$) whereas the enjoyment factor had a maximum of 2.281 ($\hat{\eta}_2 \approx -0.70$). The information for the enjoyment and independence was centred below the average factor score for both, which was indicated above by the negative difficulty parameters of some of the items.

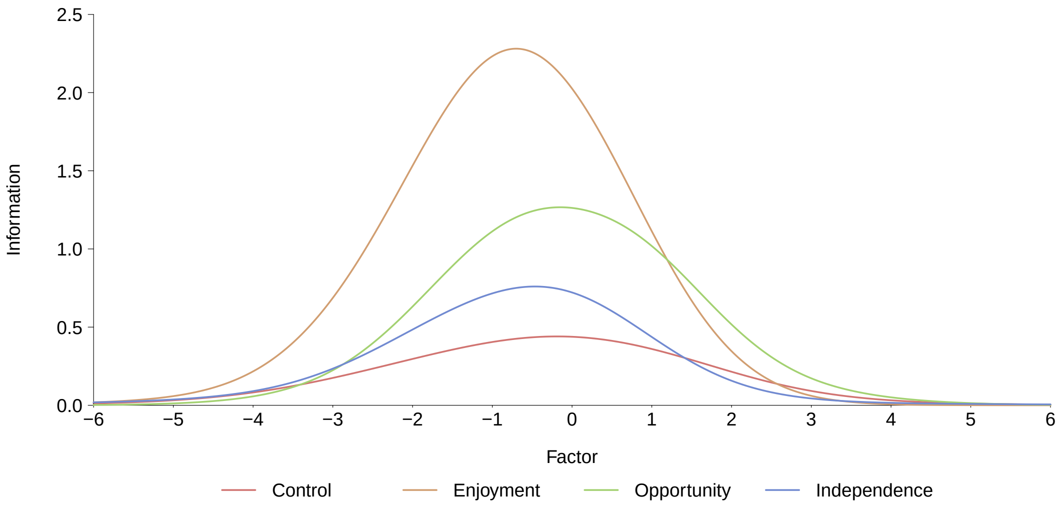


Figure 6.6: Total information curves from the unidimensional IRT of the quality of life variables in the ELSA (Wave 4).

In conclusion, there were some fundamental issues with the CASP-19 measure of quality of life, such as a lack of configural invariance (a different configuration of the relationship between factors and latent response variables from Wave 1 and Wave 4) and the inability for some items to discriminate between individuals with different levels of quality of life. In our analysis the suggested factor structure, although validated by Wiggins et al. (2008) using data from Wave 1 of the ELSA, was not present in our analysis using data from Wave 4. This demonstrated the configuration of the latent structure changed over time. The validity of the measure has not been proven in longitudinal studies with repeated use. Repeated exposure may change individuals' responses to the items with potential for memory bias and learning. There were concerns about the items, also, with issues such as poor discriminatory ability, the existence of an unused or under used category and the lack of information above the factor mean.

6.3.5 Measurement Model of Subjective Well-being

The seven factors identified in the previous sections (control, enjoyment, opportunity, independence, life satisfaction, loneliness and depression) were collected into a measurement model of overall subjective well-being. Due to repetition, the variables corresponding to 'how often do you feel left out?' and 'Whether felt lonely much of the time during past week' were removed from the loneliness and depression aspects respectively. The satisfied with life variables of the enjoyment and the life satisfaction aspects were considered to be distinct enough as the first corresponded to the frequency of feeling satisfied and the latter the level of agreement that they are satisfied. The correlations between the seven factors, given in Table 6.26, were all moderate and positive while loneliness had relatively high correlation with the control aspect of quality of life and depression.

	Control	Enjoyment	Opportunity	Independence	Life Satisfaction	Loneliness	Depression
Control	1.000						
Enjoyment	0.688	1.000					
Opportunity	0.714	0.670	1.000				
Independence	0.681	0.741	0.610	1.000			
Life Satisfaction	0.624	0.743	0.444	0.580	1.000		
Loneliness	0.853	0.624	0.434	0.478	0.559	1.000	
Depression	0.680	0.646	0.598	0.524	0.547	0.740	1.000

0.000 0.250 0.500 0.750 1.000

Table 6.26: Model estimated factor correlations of subjective well-being factors in the ELSA (Wave 4).

The estimated parameters of the measurement model are shown in Figure 6.7 with standard errors and R-squared values given in Table 6.27. The model fit indices, again, gave mixed interpretation with the RMSEA=0.067 that indicated poor fit, the CFI=0.925 that indicated good fit and the TLI=0.919 that indicated acceptable fit; which was attributed to the uncertain fit of quality of life aspects. The chi-squared value of the model was $\chi^2 = 17,044.187$ with 553 degrees of freedom.

Overall subjective well-being had a strong positive influence on each of the aspects. The most dominant aspects were enjoyment and control with the largest standardised factor loadings, where the overall factor explained 84.8% and 76.6% of the variance in each. Aspects with higher factor loadings were those which could be considered to be part of the elderly individuals' day-to-day life such as their independence to do what they want, to enjoy what they do and to be in the company of others (lack of loneliness). Longer term aspects such as satisfaction with life, depression and the ability to seize opportunities due to good health or age had lower factor loadings.

There were several variables that were noticeably weaker indicators of the respective factor, however only the family responsibilities variable of the independence factor and the 'in tune with the people around you' variable of the loneliness factor had low factor loadings (0.295 and 0.333). Control had less association with shortage of money than being left out or feeling out of control of one's own life, explaining only 16.0% of the variance in the shortage of money latent response variable. Enjoying the company of others was a weaker indicator of enjoyment.

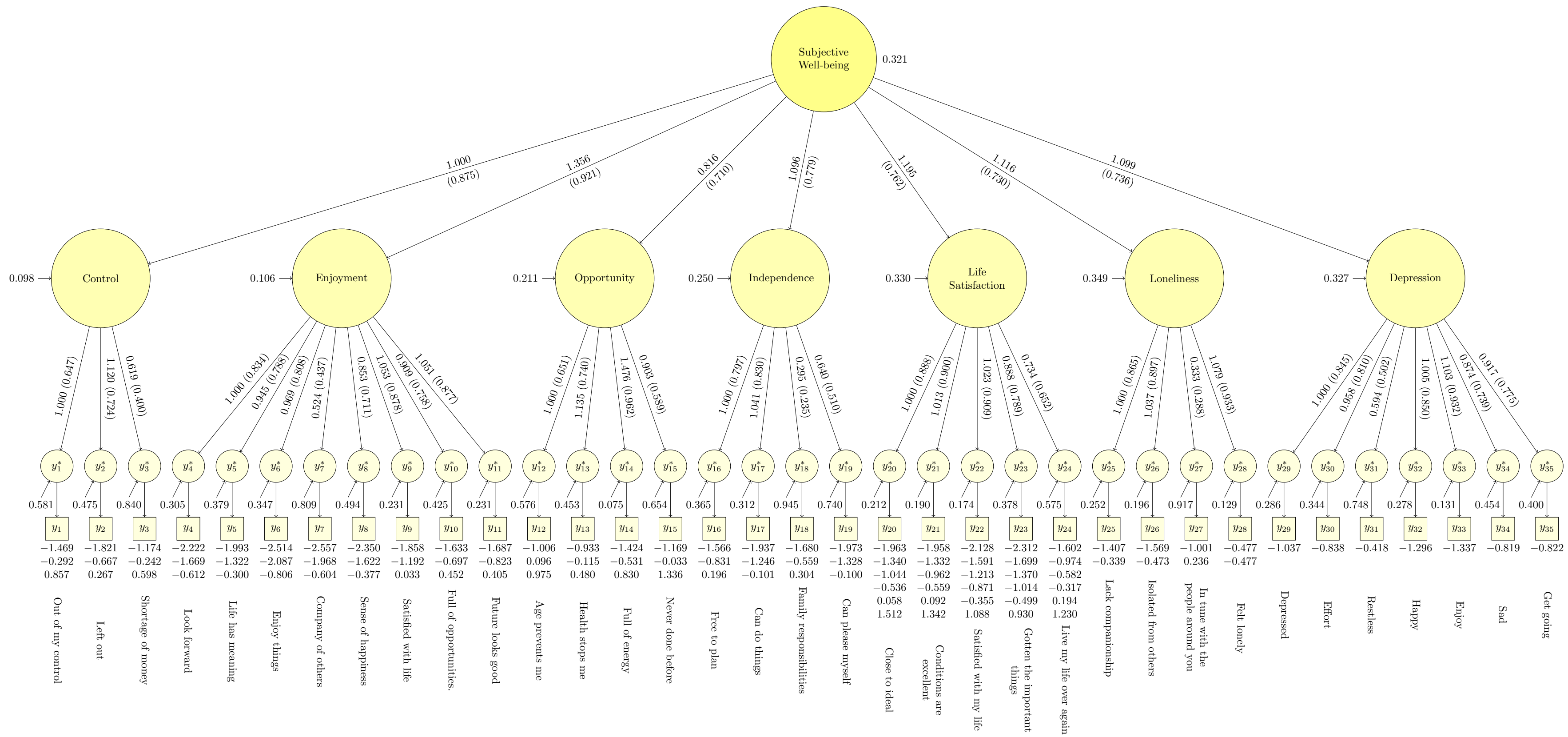


Figure 6.7: Path diagram for the estimated measurement model of subjective well-being in ELSA (Wave 4).

	Standard Errors		
	Original	Bootstrapped	R-squared
	s.e.($\lambda_{m,h}$)		
Control			
Out of my control	-	-	0.419
Left out	0.022	0.027	0.525
Shortage of money	0.023	0.027	0.160
Enjoyment			
Look forward	-	-	0.695
Life has meaning	0.011	0.013	0.621
Enjoy things	0.013	0.014	0.653
Company of others	0.018	0.019	0.191
Sense of happiness	0.013	0.016	0.505
Satisfied with life	0.011	0.014	0.771
Full of opportunities	0.011	0.014	0.575
Future looks good	0.011	0.013	0.769
Opportunity			
Age prevents me	-	-	0.424
Health stops me	0.022	0.024	0.547
Full of energy	0.027	0.037	0.925
Never done before	0.024	0.031	0.346
Independence			
Free to plan	-	-	0.635
Can do things	0.019	0.023	0.688
Family responsibilities	0.020	0.026	0.055
Can please myself	0.019	0.022	0.260
Life Satisfaction			
Close to ideal	-	-	0.788
Conditions are excellent	0.006	0.009	0.810
Satisfied with my life	0.005	0.008	0.826
Gotten the important things	0.007	0.011	0.622
Live my life over again	0.009	0.014	0.425
Loneliness			
Lack companionship	-	-	0.748
Isolated from others	0.013	0.015	0.804
In tune with the people around you	0.021	0.023	0.083
Lonely	0.013	0.014	0.871
Depression			
Depressed	-	-	0.714
Effort	0.020	0.021	0.656
Restless	0.021	0.022	0.252
Happy	0.021	0.021	0.722
Enjoy	0.021	0.023	0.869
Sad	0.020	0.020	0.546
Get going	0.021	0.022	0.600
	s.e.(β_m)		
Control	-	-	0.766
Enjoyment	0.027	0.031	0.848
Opportunity	0.019	0.022	0.503
Independence	0.024	0.029	0.607
Life Satisfaction	0.023	0.027	0.581
Loneliness	0.025	0.029	0.533
Depression	0.028	0.031	0.542

Table 6.27: Model estimated standard errors, original and bootstrapped, and R-squared values for the subjective well-being measurement model, of the ELSA (Wave 4).

6.4 Structural Equation Model

Next, the relationship between the domains, and the effects of an array of covariates on the domains, were analysed in a structural equation model. Firstly, the unconditional SEM was fitted to provide an overview of the relationship between domains on average in the elderly population. The model was then extended into a multiple indicator and multiple causes model (MIMIC), modelling the domains conditional on covariate variables.

The overall SEM combined the measurement models for each of health, economic well-being and subjective well-being. Health and economic well-being factors were modelled with a direct effect on subjective well-being, where subjective well-being is the endogenous variable. On the other hand, health and economic well-being were thought to have a bidirectional relationship; reflecting the premise of the National Health Service that ill-health inhibits economic well-being and poor economic well-being restricts the ability to prevent and treat ill-health.

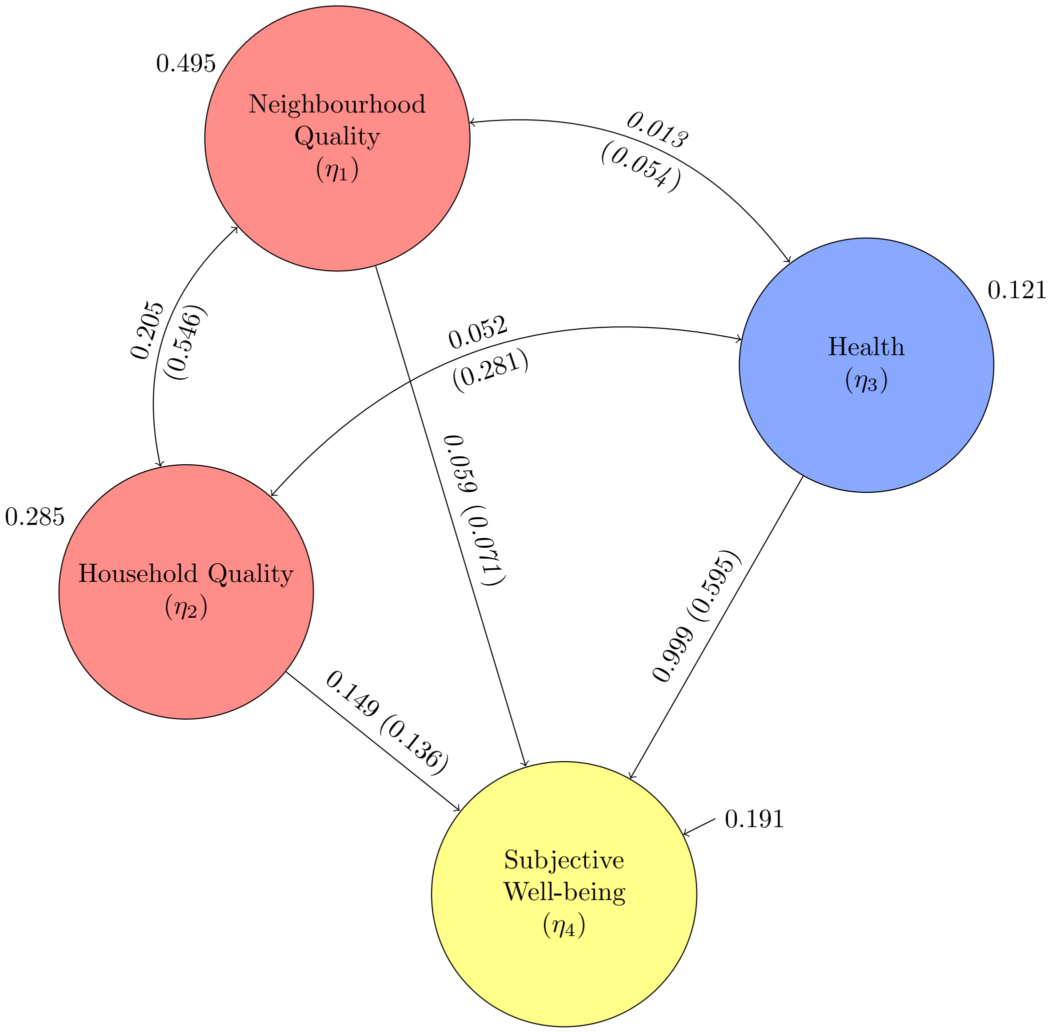
The relationships between the three domains were first seen in the unconditional SEM, with no covariates considered. This gave an overview of the population means for the elderly population, however inferences may be incorrect if covariates are influential to the missingness of data, such as age.

Low frequencies led to the removal of the two variables heart failure and difficulty recognising when in physical danger, from the health measurement model.

Path diagrams of the structural and measurement models that comprised the overall SEM are given in Figure 6.8 on pages 188, 189 and 190. The structural model between the three domains in contained in Figure 6.8a, and the measurement model for each domain, economic well-being, health and subjective well-being, are given in Figures 6.8b to 6.8d respectively.

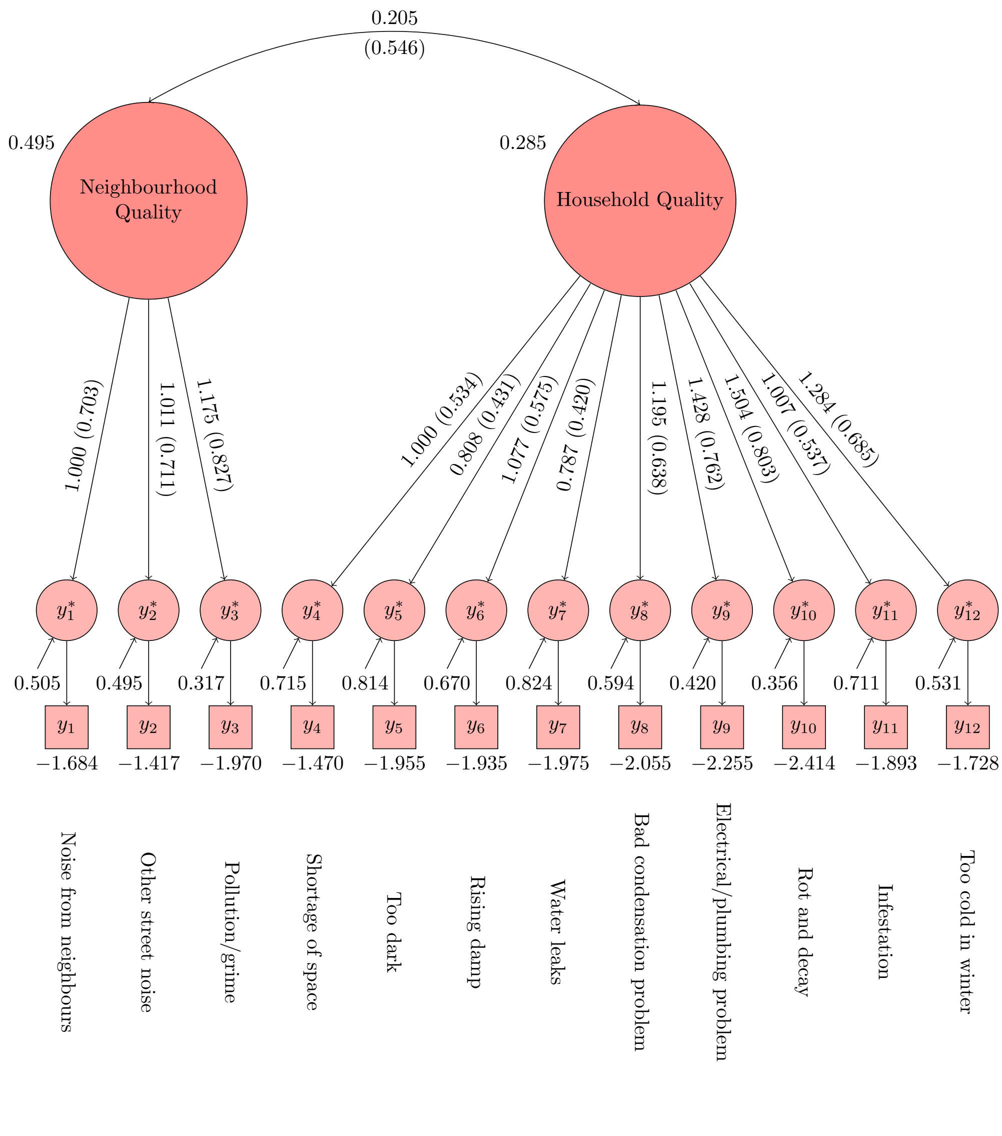
Model fit of the unconditional SEM was good with RMSEA=0.033, CFI=0.932 and TLI=0.930. The chi-squared value was $\chi^2 = 23,368.586$ on 2,756 degrees of freedom.

Figure 6.8: Path diagrams for the estimated unconditional SEM for the ELSA (Wave 4). Continued on the following pages.



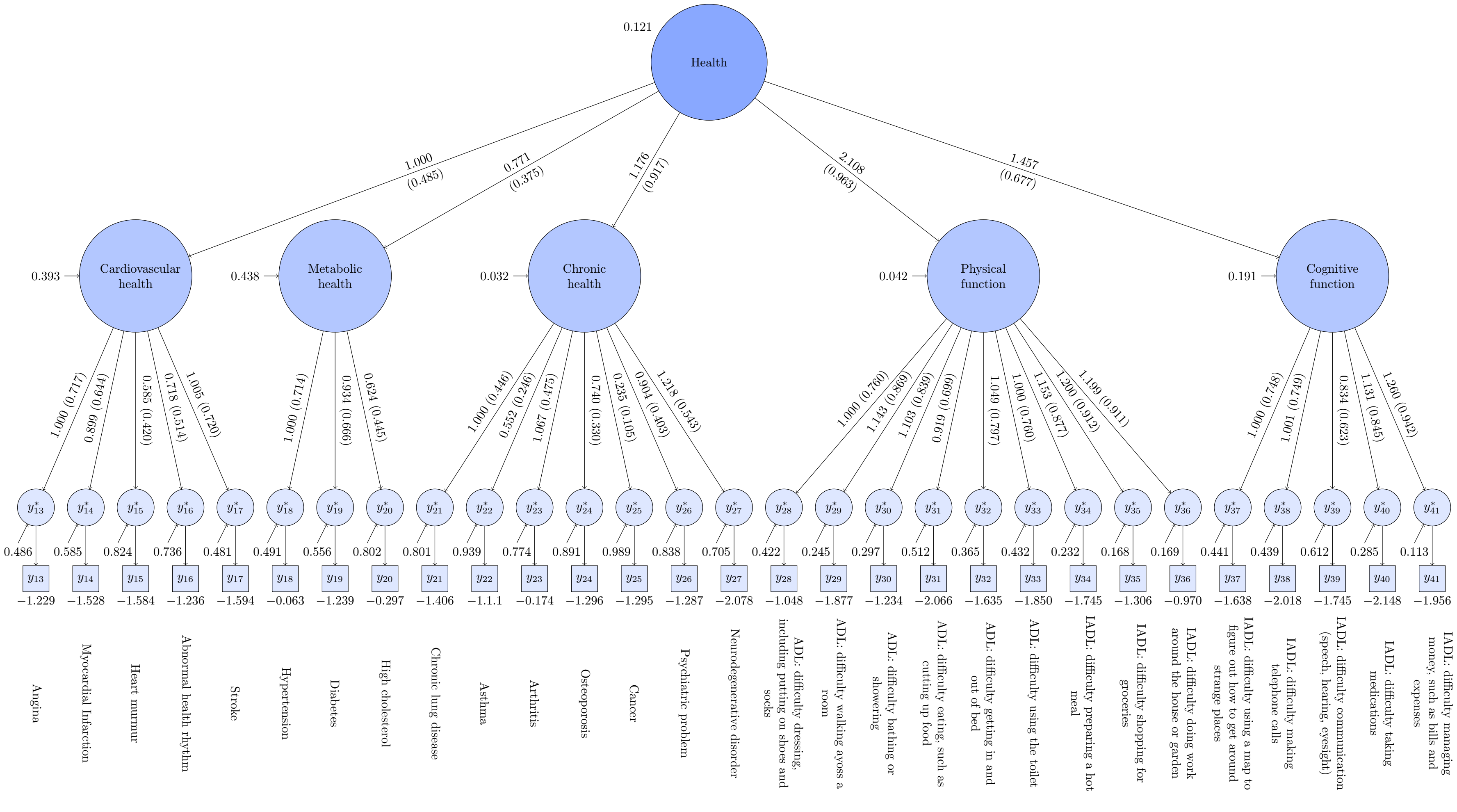
(a) Structural Model.
Italics used to highlight insignificant estimates, at 5% level.

Figure 6.8: Continued from previous page
Path diagrams for the estimated unconditional SEM for the ELSA (Wave 4).



(b) Economic Well-being Measurement Model

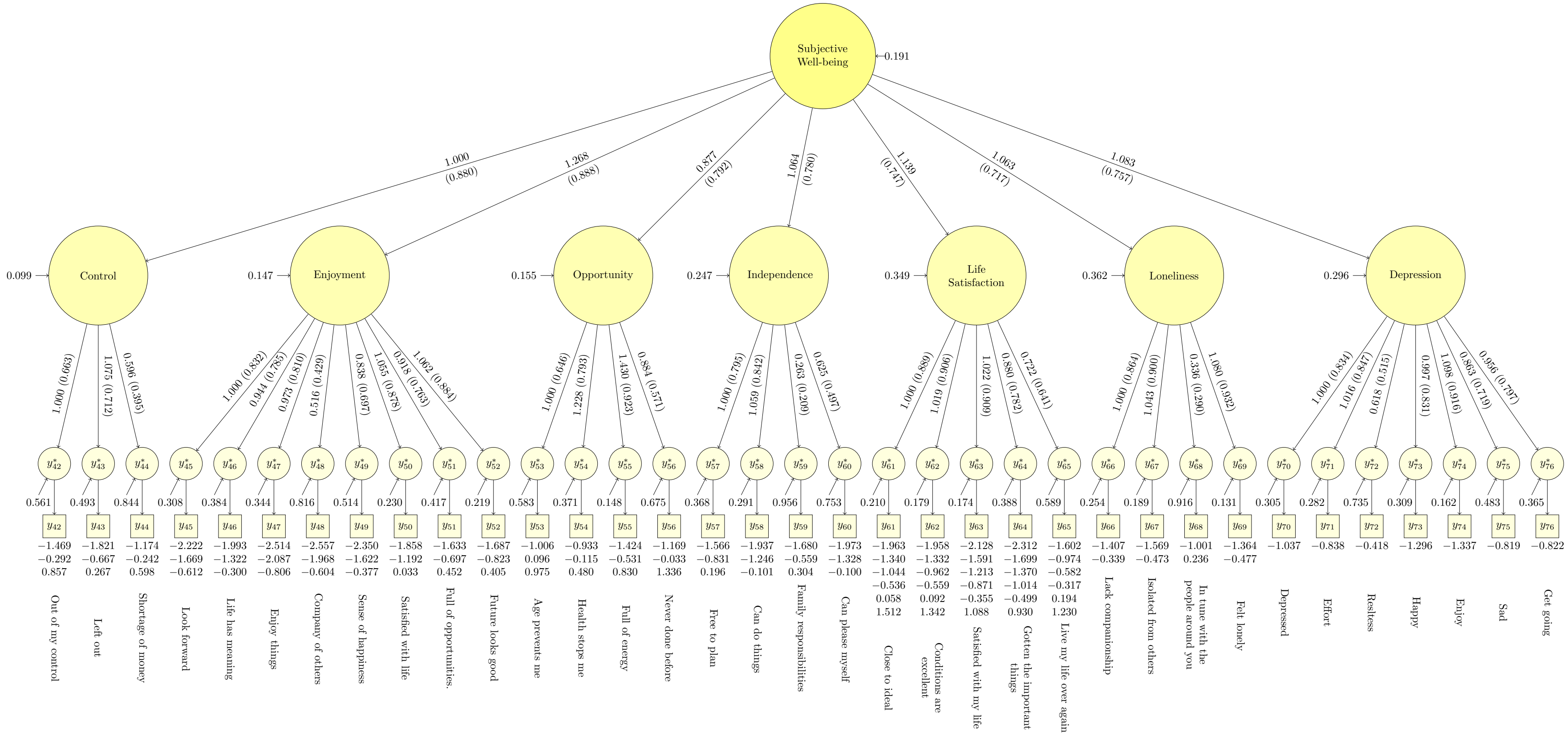
Figure 6.8: Continued from previous page
Path diagrams for the estimated unconditional SEM for the ELSA (Wave 4).



(c) Health Measurement Model

Figure 6.8: Continued from previous page

Path diagrams for the estimated unconditional SEM for the ELSA (Wave 4).



(d) Subjective Well-being Measurement Model

Subjective well-being was influenced by health and household quality but not neighbourhood quality. Health had the greatest effect with a standardised factor loading of 0.595 onto subjective well-being, and household quality had a smaller, but significant, effect. Neighbourhood quality was insignificant to subjective well-being. Therefore, fewer health problems and problems within the accommodation were associated with greater subjective well-being; including improved life satisfaction and enjoyment as well as lower levels of depression and loneliness.

Neighbourhood quality only had a strong association with household quality, but no significant association with either health or subjective well-being. Hence, the level of noise and pollution in an elderly person's neighbourhood had no impact on their health or their feeling of satisfaction with life, loneliness or depression.

The correlation between household quality and health was positive and significant demonstrating that better household quality was associated with better health and *vice versa*. However, a correlation of 0.281 was not large.

Overall, the majority of the measurement models in the unconditional SEM remained unchanged from the independent measurement models shown in Figure 6.1, Figure 6.3 and Figure 6.7.

The measurement models for both economic well-being and subjective well-being were very similar between the independent models (Figures 6.3 and 6.7) and the measurement models within the unconditional SEM (Figures 6.8b and 6.8d). The structure of the variables within the household quality factor had slight differences, for example the standardised factor loading increased for the shortage of space and decreased for the water leaks variable. The regression slope estimates in the subjective well-being model increased for the opportunity factor from 0.710 to 0.792, whilst the other regression slopes had the same interpretations.

The measurement model for health in the unconditional SEM was expected to have some differences from the independent measurement model following the removal of the two variables; heart failure from the cardiovascular health factor and difficulty recognising physical danger from the cognitive function factor. Comparing the regression slopes, the association between overall health and cardiovascular health was relatively unchanged whereas the association with cognitive function was reduced from 0.902 to 0.677. Elsewhere, there was little difference in the standardised regression slopes except for metabolic health which increased from 0.242 to 0.375.

In comparing the standardised factor loadings, there were no large changes to the structure of the cardiovascular and metabolic health variables. Additionally, several variables of the chronic health factor had noticeable changes, for example, psychiatric problems increased from 0.192 to

0.403 and neurodegenerative disorders decreased from 0.856 to 0.543. In general, the standardised factor loadings of the variables for both physical and cognitive function were lower, with some minor differences in the comparison between the variables of physical function.

6.4.1 Conditional Model

To determine the effect of individual characteristics on each of the domains, economic well-being, health and subjective well-being, covariates were included in the structural model. Covariates considered were age, gender, race, education, marital status and Government office region.

An illustration of the model is shown in Figure 6.9 containing the estimated regression slope parameters and the residual variances of the factors. Correlation between the health factor and the household and neighbourhood quality factors was replaced by the association with the covariates. The effect of health, household quality and neighbourhood quality on subjective well-being was larger than previous, in the unconditional SEM, whilst the residual variances were slightly smaller.

The estimated regression coefficients for the covariates are given in Table 6.28, where significant coefficients are highlighted in bold.

There were significant differences in the neighbourhood quality for the different age groups and regions. No significant differences were present between the genders, races, marital statuses and levels of educations. The estimated regression coefficients for the effect of age showed improved neighbourhood quality with increasing age. Compared with those aged 60-64 years old, those aged $\geq$70 years old had significantly better neighbourhood quality. The effect of region showed that in comparison to the South East region, elderly individuals in the North West, Yorkshire and the Humber, the East Midlands, and the South West had significantly better neighbourhood quality, whereas those in the London region had significantly worse neighbourhood quality. The effect of government office region on neighbourhood quality is illustrated on the map in Figure 6.10a, where darker colour indicated higher estimated regression coefficients thus better neighbourhood quality. It was clear to see that comparatively, elderly individuals in the South East and North East regions had the lowest neighbourhood quality and those in the East Midlands had the highest.

Household quality was significantly higher for the oldest elderly individuals (aged $\geq$85 years old) and for those that were either married, in a civil partnership or were widowed. The estimated

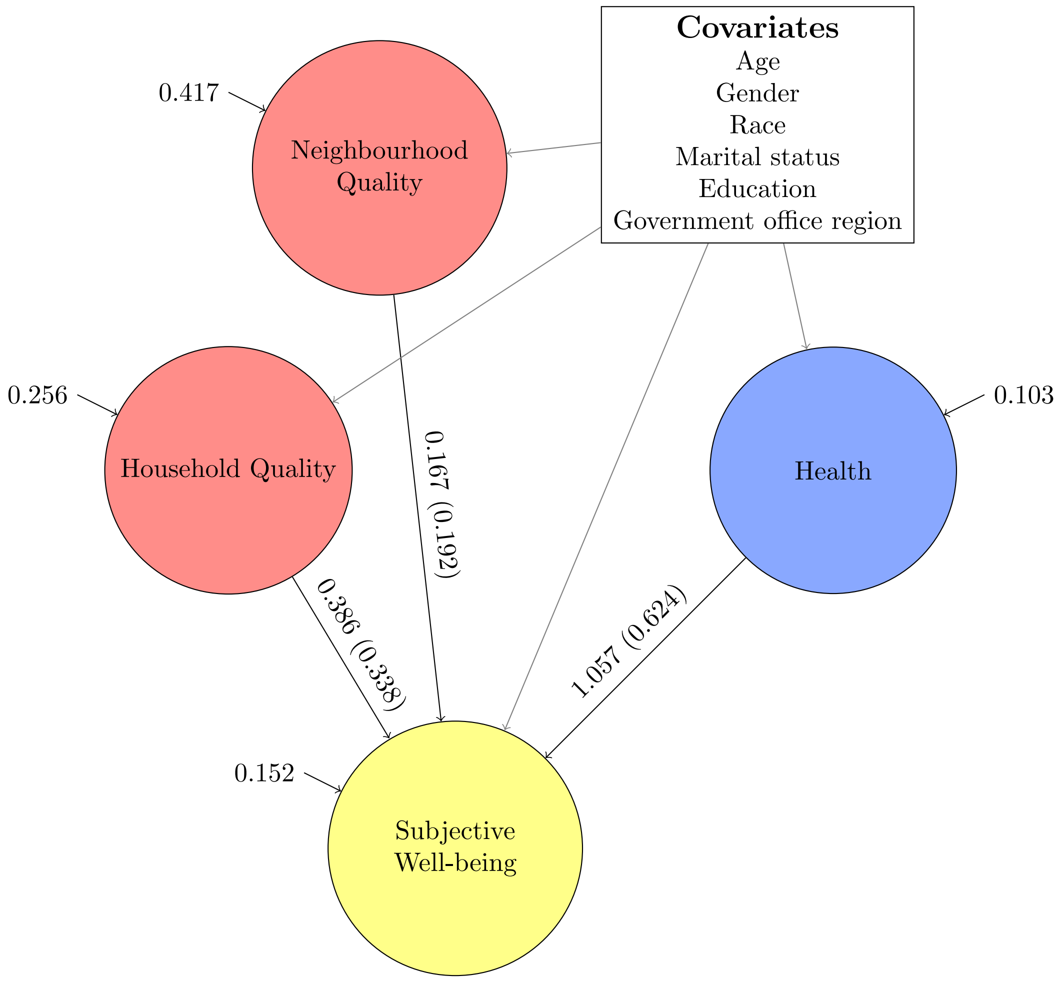
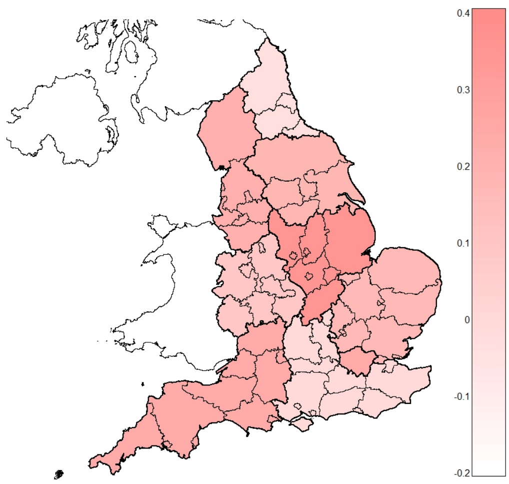
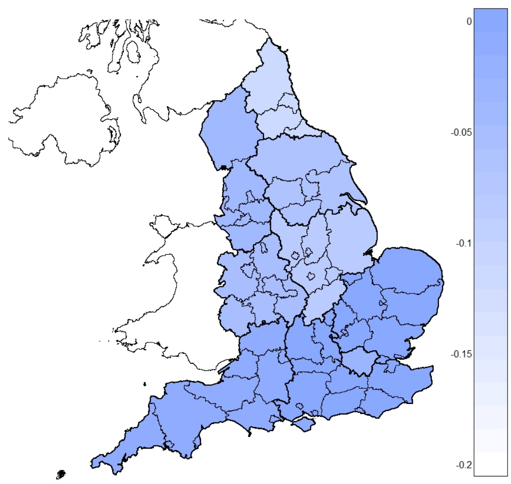


Figure 6.9: Path diagram for the estimated conditional SEM of the ELSA (Wave 4).



(a) Neighbourhood Quality



(b) Health

Figure 6.10: Map illustrations of the regression coefficients for the regression coefficients for the Government Office Region on the neighbourhood quality and health factors.

Covariate		Neighbourhood Quality	Household Quality	Health	Subjective Well-being
Age					
	60-64 years old	-	-	-	-
	65-69 years old	0.056 (0.045)	-0.054 (0.036)	**-0.054** (0.018)	**0.103** (0.029)
	70-74 years old	**0.173** (0.050)	0.032 (0.036)	**-0.100** (0.019)	**0.072** (0.029)
	75-79 years old	**0.211** (0.061)	0.077 (0.045)	**-0.185** (0.023)	**0.091** (0.034)
	80-84 years old	**0.250** (0.075)	0.079 (0.052)	**-0.244** (0.027)	0.057 (0.040)
	≥85 years old	**0.577** (0.116)	**0.146** (0.064)	**-0.340** (0.034)	0.037 (0.050)
Gender					
	Female	-	-	-	-
	Male	-0.018 (0.035)	<0.0005 (0.027)	**0.040** (0.013)	**-0.043** (0.021)
Race					
	White	-	-	-	-
	Non-white	0.199 (0.119)	-0.123 (0.104)	**-0.117** (0.042)	-0.034 (0.079)
Marital status					
	Single	-	-	-	-
	Married or civil partnership	0.006 (0.073)	**0.259** (0.057)	0.050 (0.028)	0.086 (0.046)
	Separated or divorced	-0.157 (0.085)	0.032 (0.061)	-0.055 (0.032)	-0.010 (0.052)
	Widowed	0.027 (0.082)	**0.251** (0.063)	0.011 (0.029)	**-0.098** (0.049)
Education					
	Degree	0.170 (0.054)	0.054 (0.040)	**0.134** (0.022)	0.045 (0.033)
	Higher education	0.138 (0.055)	0.054 (0.041)	**0.107** (0.020)	0.001 (0.032)
	NVQ3	0.032 (0.066)	0.048 (0.055)	0.035 (0.026)	0.066 (0.042)
	NVQ2	0.089 (0.050)	0.028 (0.037)	**0.119** (0.019)	-0.053 (0.030)
	NVQ1	0.049 (0.082)	0.045 (0.288)	**0.053** (0.026)	-0.007 (0.119)
	Foreign	0.097 (0.066)	0.085 (0.050)	**0.044** (0.021)	-0.049 (0.037)
	None	-	-	-	-
Government office region					
	North East	-0.029 (0.071)	-0.091 (0.059)	**-0.112** (0.027)	0.087 (0.044)
	North West	**0.217** (0.066)	-0.040 (0.047)	-0.037 (0.021)	0.014 (0.037)
	Yorkshire and The Humber	**0.182** (0.064)	0.037 (0.050)	**-0.056** (0.023)	-0.043 (0.039)
	East Midlands	**0.350** (0.074)	0.008 (0.050)	**-0.075** (0.024)	-0.021 (0.040)
	West Midlands	0.093 (0.062)	-0.005 (0.050)	**-0.049** (0.022)	-0.011 (0.037)
	East	**0.167** (0.062)	0.021 (0.048)	0.001 (0.022)	-0.029 (0.036)
	London	**-0.196** (0.065)	-0.042 (0.051)	-0.028 (0.026)	0.055 (0.041)
	South East	-	-	-	-
	South West	**0.236** (0.065)	-0.062 (0.047)	-0.010 (0.022)	-0.026 (0.037)

Bold text to highlight significant regression coefficient estimates, at 5% level.

Table 6.28: Model estimated regression coefficients for the covariates in the conditional SEM of the ELSA (Wave 4).

regression coefficients from the age of 70 years old were positive and increasing with age, although the difference with those aged 60-64 years old was not significant until the oldest age group, suggesting that household quality improved slowly as age increased. Marriage was also seen to be associated with better household quality. The effect of currently being in a marriage, civil partnership or being widowed was significantly positive compared to being single, whilst the status of being separated or divorced had no significant difference.

All the covariates except marital status had a significant effect on health. Health deteriorated significantly between those aged 60-64 years and those aged 65-69 years old, then continued to deteriorate with increasingly negative regression coefficients with increasing age. Health was also significantly worse for those of non-white race, while in terms of gender, health was significantly better from men than women.

The effect of level of education was mixed, although two distinct groups where health improved as the level of education improved could be identified. Firstly, health improved and education increased from no further education to NVQ level 2, or equivalent, then NVQ level 3 had no significant difference from those with no further education but health improved, again, as education increased up to degree level.

The estimated regression coefficients from the regional differences showed a north-south divide in the health of the elderly, which was highlighted by the illustration in Figure 6.10b. As previously mentioned, darker colour indicates better health. Elderly individuals in the northern regions, except the North West, (North East, Yorkshire and the Humber, East Midlands and West Midlands) had significantly worse health than those in the South East region, but the south regions (South West, London, East of England) did not.

Comparison of the two maps in Figure 6.10 identified elderly individuals living in the North East region had both poorer neighbourhood quality and health. The North East had the lowest health coefficient and only those in the London region had a lower coefficient for the neighbourhood quality. Elsewhere those in the East Midlands had better neighbourhood quality but poorer health, and for those in the South East neighbourhood quality was poorer whilst health was better.

Men and widowed individuals had significantly lower subjective well-being. Widowed marital status was the only significant regression coefficient of the covariate, indicating these individuals had significantly lower subjective well-being than single individuals.

All the age groups had positive regression coefficients, therefore subjective well-being was lowest in those aged 60-64 years old. Subjective well-being increased to its highest in those aged 65-69 years old, then as age increased the effect lessened with decreasing regression coefficients; but remaining positive compared to those aged 60-64 years old.

Summary of the Chapter

This chapter started by defining the underlying structures for each of the domains, using exploratory factor analysis on each collection of items. It was found that the cardiovascular medical conditions variables, activities of daily living variables and the accommodation problems variables each had a latent structure containing multiple factors. Underlying the cardiovascular comorbidities, we found a separation of variables related more specifically to metabolic health rather than cardiovascular health. Activities of daily living contrarily did not separate into their formally defined basic and instrumental activities, but into activities related to physical function and activities related to cognitive function. The accommodation problems belonged to two factors, the first relating to problems in the household and the second to problems related to outside the household such as the surrounding neighbourhood.

Further investigation was conducted into the association between the economic well-being factors and objective measures including net financial wealth. It was found that the factor underlying the durables owned variables, named ownership, was negatively associated with increased levels of net financial wealth.

Item response theory (IRT) was also applied to the CASP-19 variables, where it was discovered that a different factor structure was present to the one previously stated by the authors. A new four factor structure was proposed from exploratory factor analysis of the variables. Modelling the new structure using IRT, several issues were found within the variables. Firstly, several items had poor discrimination ability with respondents being likely to use any of the category response given extreme values on the latent continuum. The category response of 'not often' was also either under used or un-used by the elderly individuals.

Measurement models were then created for each of the domains using the factors structure identified previously, then these were combined into a single SEM that allowed for the relationships between the domains to be analysed and the effect of covariates to be determined. Neighbourhood quality was seen to have no effect on health or subjective well-being, whereas greater health and household quality influenced higher levels of subjective well-being. Age was found to significant to all four factors with differing effects such as better household and neighbourhood quality with older age but deteriorating health and subjective well-being. Men had significantly better health than women, but worse subjective well-being. Health was lower for elderly individuals of non-white race. The status of being married (or in a civil partnership) or widowed were associated with greater household quality, while being widowed was also associated with lower subjective well-being. Regional disparities in the levels of neighbourhood quality and health were also identified with elderly individuals in the North East having comparatively poorer levels in both domains. Education had a complex relationship with health, with increasingly better health with higher levels of education firstly from no further education to an NVQ level 2 or equivalent, then NVQ level 3, or equivalent, having no significant difference from no further education, and finally health improving again into the higher levels of education.

Subjective well-being was therefore a strong component of the model that was influenced by both health and economic well-being, and had differences in the age groups, genders and with marital status. In the next chapter we will extend the analysis of subjective well-being to include the longitudinal properties of the ELSA, focusing on the changes to aspects of subjective well-being over time, and how this is influenced by the afore considered covariates.

Chapter 7

Longitudinal Analysis of Subjective Well-being

This chapter will outline the results from longitudinal analyses of subjective well-being in the English Longitudinal Study of Ageing (ELSA). The repeated measures of life satisfaction, depression and loneliness were taken from three time points; Waves 2, 4 and 6. Invariance in the measurement of each aspect was tested in a confirmatory factor analysis (CFA) that contained the measurement model repeated at three time points. Multivariate latent growth models were applied to jointly model the growth processes of the three aspects. Firstly, an unconditional model was applied to the elderly individuals present at all three time points, to obtain the associations between the processes and population trends within each process given the ability to participate was retained. A conditional model with time-invariant covariates was then applied to allow for the inclusion of elderly individuals that dropped out of the study between Waves 2 and 6, under the missing at random with respect to covariates assumption (MARX). The effect of the age, gender, race and marital status on the growth processes was then seen for each aspect. In addition, a further model was applied with change in marital status as a time varying covariate to determine the effect of an event such as divorce from a marriage or death of a spouse on subjective well-being. Again, only the elderly individuals that did not drop out were included in the analysis, due to the requirement for complete covariate observations with the MARX assumption.

7.1 Confirmatory Factor Analysis and Measurement Invariance

Repeated measurements were available at Waves 2, 4 and 6 for the subjective well-being aspects. The first aim was to model a linear growth process for each aspect. In order to ensure inferences are drawn for the changes in the aspect rather than the scale of the measurement model, invariance of the measurement model was paramount. Strong measurement invariance was imposed alongside a covariance structure of the residuals of repeated latent response variables. Strong measurement invariance requires the configuration of the factor structure, the factor loadings and a subset of thresholds to be equal for each time point (Millsap and Yun-Tein, 2004).

Previously, a single factor model was concluded for the life satisfaction, loneliness and depression aspects, while the quality of life was measured by four factors (see Chapter 6). The quality of life factors lacked configural invariance, with a different configuration of factors and latent response variables defined at Wave 4 than previously confirmed by Wiggins et al. (2008) at Wave 2. Thus, the quality of life measurement model lacked configural invariance and could not be considered in the latent growth models. The 'felt lonely' variable of the loneliness aspect was not available at Wave 2 so was not included in the loneliness growth model.

The aspects were modelled in a separate CFA with the measurement model for the latent variable repeated at each time point. The covariance structure of the residuals for repeated latent response variables, was included in the model without the equality constraint between residuals of lag-1. The equality of the factor loadings and thresholds across time were each tested using a Wald test, in order to determine whether invariance was present in the model. Firstly, the presence of invariance in the unstandardised factor loadings was tested, then the loadings were fixed to be equal, while the presence of invariance in the thresholds was determined. Results of the Wald test are given in Table 7.1. The equality of the unstandardised factor loadings, over the three time points, was tested such that for each variable, h, $\Lambda_h^{(w)} = \Lambda_h$, for $w = 2, 4, 6$. For model identification one factor loading of each aspect was fixed to 1.000 at every time point, hence the respective factor loading was selected to correspond to the same latent response variable at all time points and was not part of the test for invariance.

Factor loadings were invariant for the loneliness and depression aspects, but not life satisfaction. This suggested that there were changes in the relationship between life satisfaction and the observed variables over time, either from their repeated use or as an effect of ageing. Differences in

Variable		Life Satisfaction	Loneliness	Depression
	Wald test value	97.073	6.376	22.644
Factor loadings	df	8	6	14
	p-value	< 0.00005	0.3824	0.0663
	Wald test value	122.266	166.909	198.329
Thresholds	df	20	16	16
	p-value	< 0.00005	< 0.00005	< 0.00005

Table 7.1: Wald test estimates of measurement invariance in the life satisfaction, loneliness and depression in the ELSA (Waves 2,4 & 6); excluding drop-out.

the unstandardised factor loadings were practically small, with the only noticeable large difference being in the third, 'Satisfied with my life', variable; where freely estimated unstandardised factors loadings were 1.066, 1.026 and 0.995 at waves 2, 4 and 6 respectively. The unstandardised factor loadings were then constrained to be equal, for each aspect, at all time points.

Invariance of the thresholds was then tested, given that the factor loadings were constrained to be equal. For the binary variables, of the depression aspect, the equality of the single threshold was tested. While, the equality of first two thresholds over time was tested for variables of life satisfaction and loneliness.

The Wald test for the equality of the thresholds concluded a significant difference in the thresholds over time, shown in Table 7.1. This emphasised the need for a latent growth model with significant changes in the distribution of the responses. Applying equality constraints to the threshold parameters ensured the trajectory was only captured by the random intercept and slope.

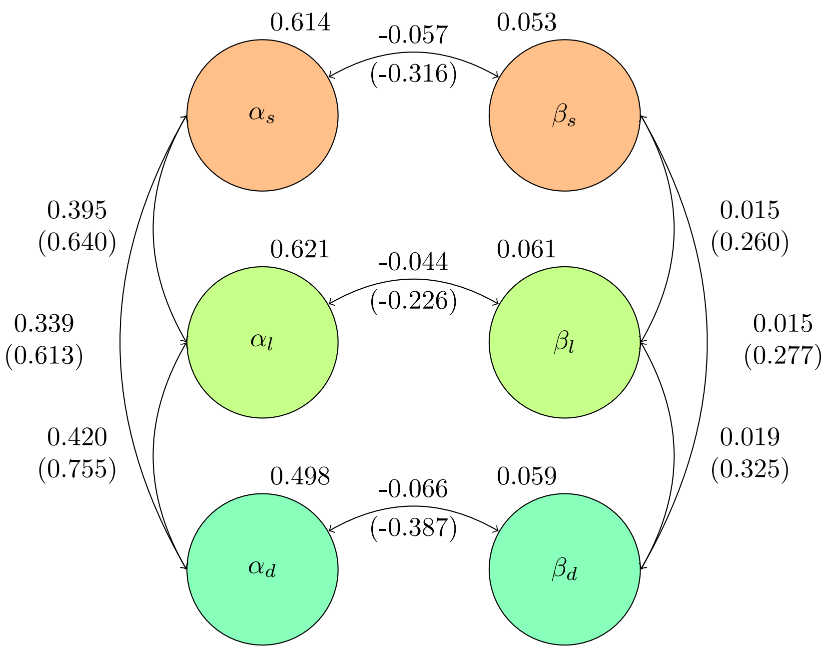
Equality constraints were added to the covariances between repeated latent response residuals of equal lag (lag-1) such that, the covariances between the latent response residuals at Waves two and four were constrained to be equal to those between Waves four and six; while the covariances between Waves two and six were free to be different.

7.2 Unconditional Multivariate Latent Growth Model

The first multivariate latent growth model contained the growth processes for the three aspects, modelling each aspect with a random intercept and random slope. The random intercepts were correlated between the aspects, as well as correlation between the random slopes. The aim of the analysis was to determine the overall trajectories of life satisfaction, loneliness and depression for the elderly over the eight years covered by the three time points. The data for this model was restricted to the elderly individuals present at all three time points, but included all age groups.

The estimated model is displayed in Figure 7.1, across pages 202 to 207. Figure 7.1a contains the relationships between the growth processes, while Figures 7.1b to 7.1d contain the growth processes for life satisfaction, loneliness and depression respectively. The subscripts s, l, d denote the aspects, life satisfaction, loneliness and depression respectively, and the superscripts (w), $w = 2, 4, 6$, denote the corresponding wave of time-varying variable; that correspond to time points t, $t = 0, 1, 2$. Random intercepts and slopes are denoted by α_a and β_a respectively for aspect $a \in s, l, d$.

Figure 7.1: Unconditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 & 6); excluding drop-out.
Continued on the following pages.



(a) Estimated structural equations of from the unconditional multivariate latent growth model, from the ELSA (Waves 2, 4 & 6).

Model fit was very good with RMSEA=0.032, CFI = 0.974, and TLI=0.974, with the estimated $\chi^2 = 5,192.374$ on df $= 1,252$.

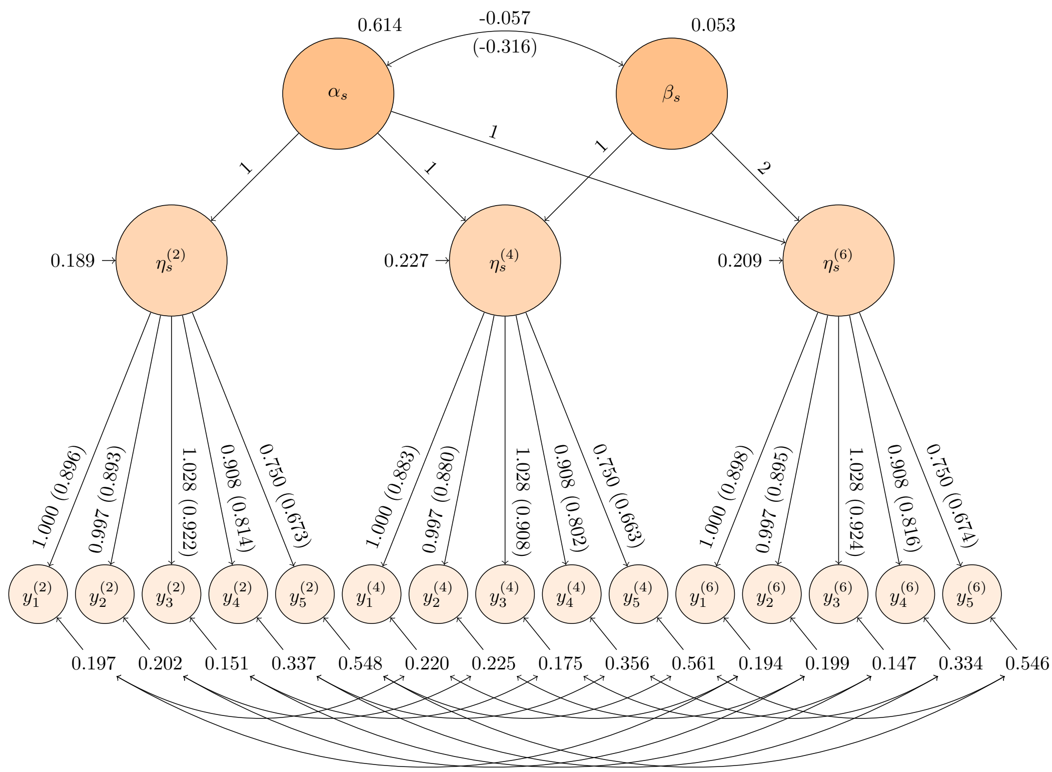
The slope means for life satisfaction and loneliness were significant with $\hat{\mu}_{\beta_s} = -0.114$ and $\hat{\mu}_{\beta_l} = -0.079$ respectively, while depression had insignificant slope mean ($\hat{\mu}_{\beta_d} = 0.001$). Therefore, over a period of eight years, life satisfaction and loneliness, on average, deteriorated over time, while depression remained unchanged.

The correlation between the intercept and slope within each aspect was negative, demonstrating a regression to the mean. This demonstrated that initially better levels of life satisfaction, loneliness or depression initially were associated with a greater deterioration over time, while on the converse those with initially worse levels had the greatest improvement over time.

Life satisfaction, loneliness and depression were strongly associated, with significantly positive correlations between both the random intercepts and random slopes. The random intercepts had strong correlations between the aspects while slightly weaker, but still significant, correlations existed between the random slopes. Correlations of both intercepts and slopes were both strongest between the loneliness and depression processes however both aspects had similar associations with life satisfaction.

Figure 7.1: Continued from previous page
Unconditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 & 6); excluding drop-out.

$y_1^{(w)}$: 'Close to ideal'
$y_2^{(w)}$: 'Conditions are excellent'
$y_3^{(w)}$: 'Satisfied with my life'
$y_4^{(w)}$: 'Gotten the important things'
$y_5^{(w)}$: 'Live my life over again'



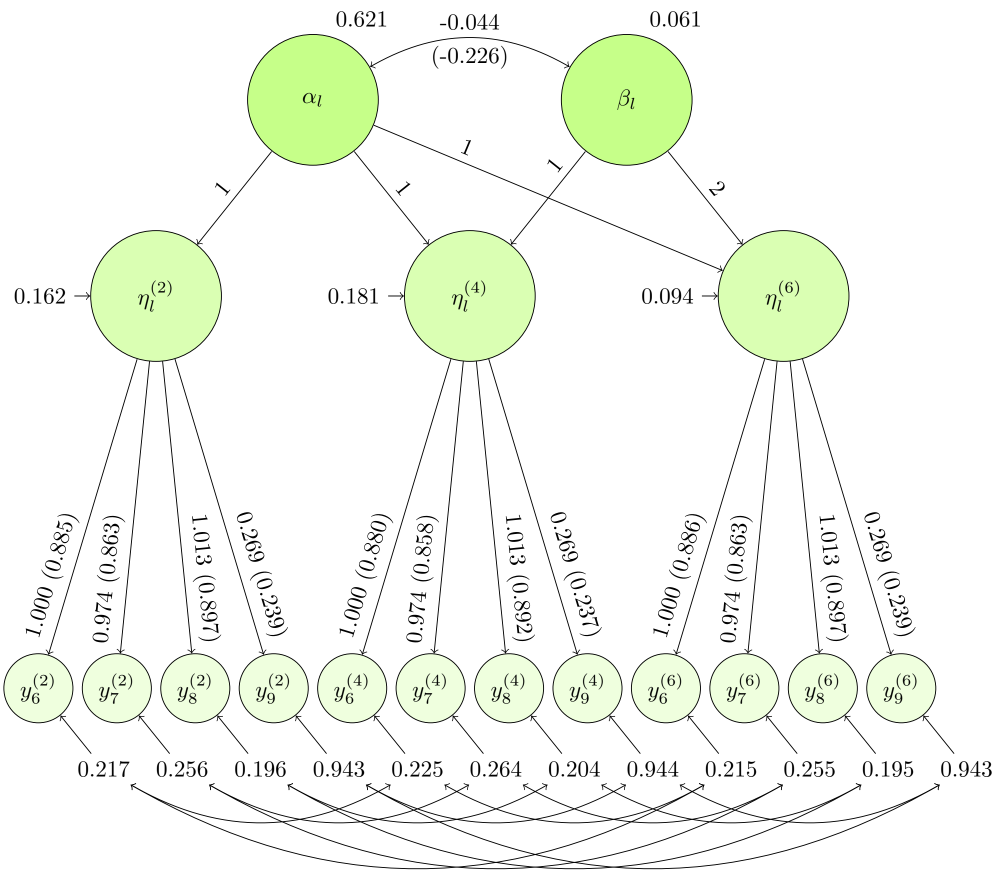
Covariance and correlation of the residuals of the repeated latent response variables:

h	$\text{Cov}(\epsilon_h^{(2)}, \epsilon_h^{(4)})$	$\text{Cov}(\epsilon_h^{(2)}, \epsilon_h^{(6)})$	$\text{Corr}(\epsilon_h^{(2)}, \epsilon_h^{(4)})$	$\text{Corr}(\epsilon_h^{(4)}, \epsilon_h^{(6)})$	$\text{Corr}(\epsilon_h^{(2)}, \epsilon_h^{(6)})$
1	0.040	0.007	0.194	0.196	0.038
2	0.049	0.031	0.228	0.230	0.153
3	-0.004	-0.033	-0.025	-0.025	-0.225
4	0.081	0.085	0.235	0.236	0.255
5	0.283	0.299	0.511	0.512	0.546

(b) Estimated growth model for life satisfaction.

Figure 7.1: Continued from previous page
Unconditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 & 6); excluding drop-out.

$y_6^{(w)}$: 'Lack companionship' $y_7^{(w)}$: 'Isolated from others'
$y_8^{(w)}$: 'Feel left out' $y_9^{(w)}$: 'In tune with the people around you'



Covariance and correlation of the residuals of the repeated latent response variables:

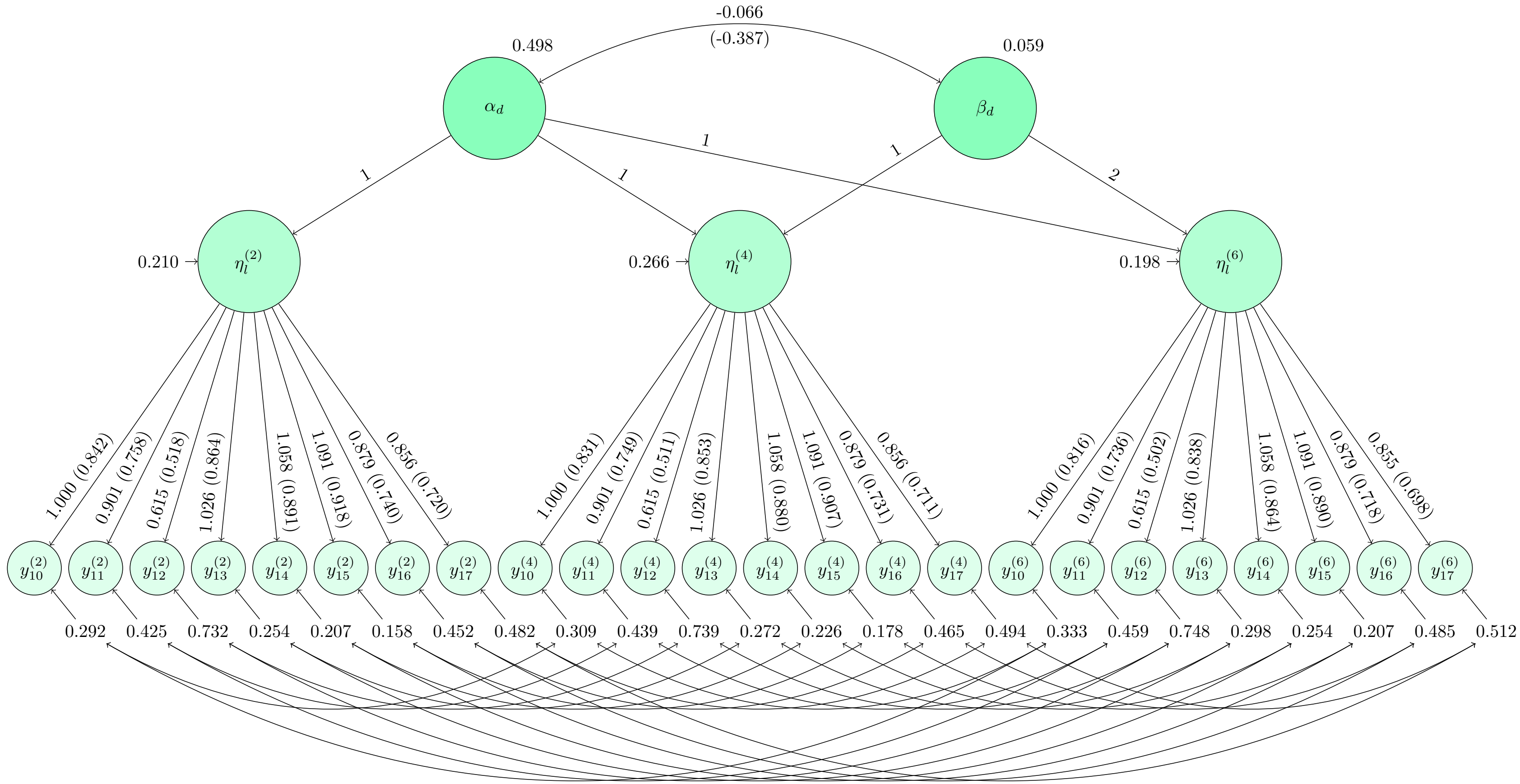
h	$\text{Cov}(\epsilon_h^{(2)}, \epsilon_h^{(4)})$	$\text{Cov}(\epsilon_h^{(2)}, \epsilon_h^{(6)})$	$\text{Corr}(\epsilon_h^{(2)}, \epsilon_h^{(4)})$	$\text{Corr}(\epsilon_h^{(4)}, \epsilon_h^{(6)})$	$\text{Corr}(\epsilon_h^{(2)}, \epsilon_h^{(6)})$
1	0.117	0.078	0.528	0.530	0.363
2	0.100	0.081	0.386	0.387	0.316
3	0.044	0.016	0.218	0.218	0.083
4	0.293	0.258	0.310	0.310	0.274

(c) Estimated growth model for loneliness.

Figure 7.1: Continued from previous page

Unconditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 & 6); excluding drop-out.

$y_{10}^{(w)}$: 'Depressed' $y_{11}^{(w)}$: 'Effort' $y_{12}^{(w)}$: 'Restless' $y_{13}^{(w)}$: 'Happy' $y_{14}^{(w)}$: 'Lonely' $y_{15}^{(w)}$: 'Enjoy' $y_{16}^{(w)}$: 'Sad' $y_{17}^{(w)}$: 'Get going'



(d) Estimated growth model for depression.

Figure 7.1: Continued from previous page
Unconditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 & 6); excluding drop-out.

h	$\text{Cov}(\epsilon_h^{(2)}, \epsilon_h^{(4)})$	$\text{Cov}(\epsilon_h^{(2)}, \epsilon_h^{(6)})$	$\text{Corr}(\epsilon_h^{(2)}, \epsilon_h^{(4)})$	$\text{Corr}(\epsilon_h^{(4)}, \epsilon_h^{(6)})$	$\text{Corr}(\epsilon_h^{(2)}, \epsilon_h^{(6)})$
1	0.106	0.085	0.352	0.330	0.273
2	0.237	0.171	0.549	0.528	0.387
3	0.431	0.366	0.585	0.579	0.494
4	0.059	0.010	0.226	0.208	0.035
5	0.189	0.155	0.874	0.790	0.678
6	0.042	0.051	0.252	0.219	0.282
7	0.152	0.132	0.332	0.321	0.282
8	0.224	0.208	0.459	0.445	0.419

Estimated growth model for depression: covariance and correlation of the latent response variable residuals.

7.3 Conditional Multivariate Latent Growth Model with Time Invariant Covariates

Drop-out is a key feature of ageing research as drop-out can indicate the ability to participate was lost, possibly due to deterioration in health or even death. It was therefore of interest to include those that dropped out in the latent growth analyses. It was assumed that the probability of drop out was not dependent on life satisfaction, loneliness or depression, however, under the assumption of missing at random with respect to covariates (MARX) could be associated with age, gender, race or marital status as the covariates.

To test sensitivity to the drop-out, the previous model was estimated both including and excluding individuals that did drop-out. This indicated there was some robustness in the presence of drop-out, up to the MARX assumption with little difference seen in the parameter estimation.

To minimise the effect of drop out, the analysis was restricted to those aged 60-74 years old the majority (over 50%) of those aged 75 or more years old dropped out between Waves 2 and 6. Per five year age group the percentage of drop out was; 30.51% aged 60-64 years old, 33.08% aged 65-69 years old, 41.64% aged 70-74 years old, 51.37% aged 75-79 years old, 67.14% aged 80-84 years old and 81.76% aged 85 or more years old.

Incorporating the drop out in the model would have complicated the model, although may be possible using similar methods to those described by Muthén et al. (2011). Further complication arose from the estimation using a weighted least squares estimator rather than maximum likelihood.

The path diagram for the estimated conditional multivariate latent growth model is given in Figure 7.2 and the estimated regression coefficients in Table 7.2.

Model fit was good with RMSEA= 0.024, CFI= 0.979 and TLI= 0.978. For completeness, the $\chi^2 = 5,108.313$ was highly significant on $1,567$ degrees of freedom. The relationships between the growth processes were similar to those previously seen in the unconditional model.

The estimated intercepts of the random slopes for life satisfaction, loneliness and for depression were $\hat{\nu}_{\beta_s} = 0.160$, $\hat{\nu}_{\beta_l} = 0.201$ and $\hat{\nu}_{\beta_d} = -0.030$, all of which were statistically insignificant. However, the positive estimates for life satisfaction and loneliness showed improving trends that were exaggerated by covariate effects. The intercepts referred to the estimated change over time for the reference profile of white elderly women aged 60-64 years old of single marital status.

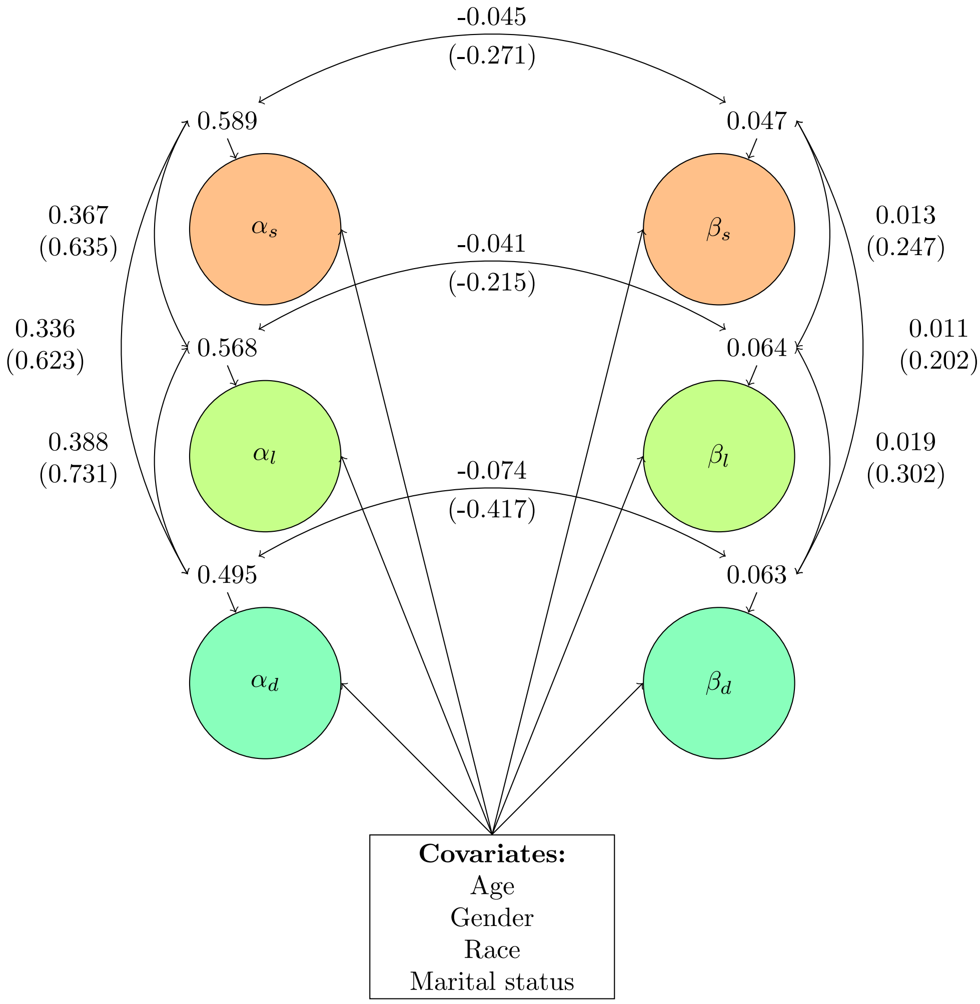


Figure 7.2: Estimated structural equations from the conditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 &6); including drop-out.

Life satisfaction was initially better for those in the older age groups, however the oldest age group had slower improvement over time, while depression significantly deteriorated over time and loneliness was unaffected by age. The estimated regression coefficients (Table 7.2) showed those in the 65-69 years old and 70-74 years old age groups had significantly positive regression coefficients, therefore life satisfaction was initially higher. The improvement in life satisfaction was slower in older age with a significant difference for those in the oldest age group, aged 70-74 years old, than those in the youngest age group. The model estimated trends are shown in Figures 7.3b, 7.4b and 7.5b which illustrate how the effects transpire over time. Figure 7.3b highlights the differences in life satisfaction at the first time point, $t = 0$ but the similarity at $t = 2$, whereas conversely, loneliness and depression were similar at $t = 0$ but those aged 70-74 years old were visibly worse by $t = 2$.

The effect of gender showed men were significantly less lonely and depressed initially and had greater improvement over time in all aspects than women. For all aspects the regression coefficient on the random slope was significantly positive for men, therefore the improvement in life satisfaction and loneliness was greater, while the change in depression over time was marginally positive rather than negative as it was for women.

Race only affected the initial level of depression with no significant effect on life satisfaction or loneliness. Those of non-white race had significantly worse depression initially and had a similar change over time to those of white race.

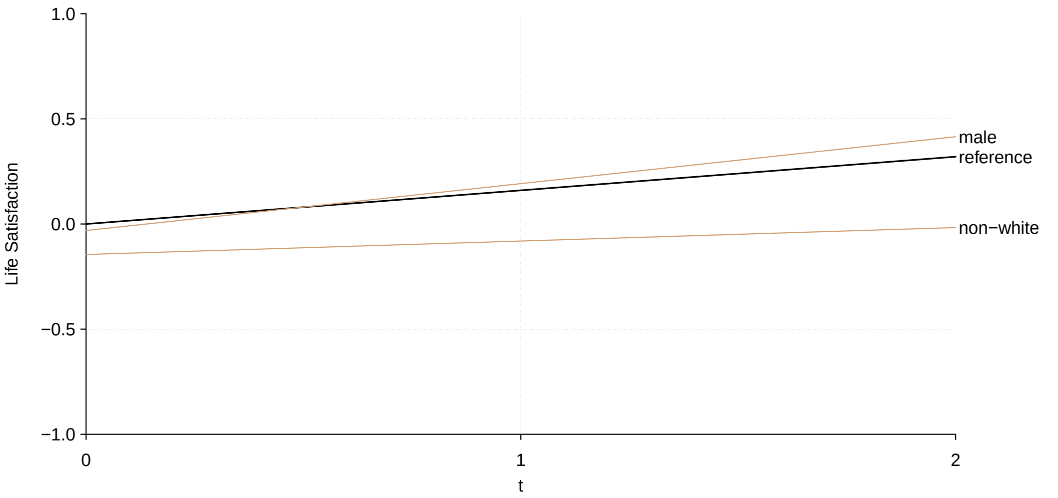
Figures 7.3a, 7.4a and 7.5a show the effects of gender and race over time. Those of non-white race were seen to be persistently worse off in all three aspects over all time points. Similarly, males had persistently lower levels of loneliness and depression. Life satisfaction, however, was initially better for women but, again, was better for men by the third time point ($t = 2$).

The effect of baseline marital status, at Wave 2, was varied across the marital statuses. Those that were married or in a civil partnership, at Wave 2, were persistently better in all aspects with no change over time. Divorced or separated elderly individuals initially had worse life satisfaction and depression, but the change over time was similar to single elderly individuals. Those who were widowed, initially had worse depression except there was significant improvement over time. The results of this improvement can be seen in Figure 7.5c where at $t = 0$ both depression and loneliness in the widowed elderly were worse than those who were divorced at baseline, however the recovery over time resulted in better loneliness and depression for those who were widowed than those who were divorced; and was even similar to those that were single by the third time point.

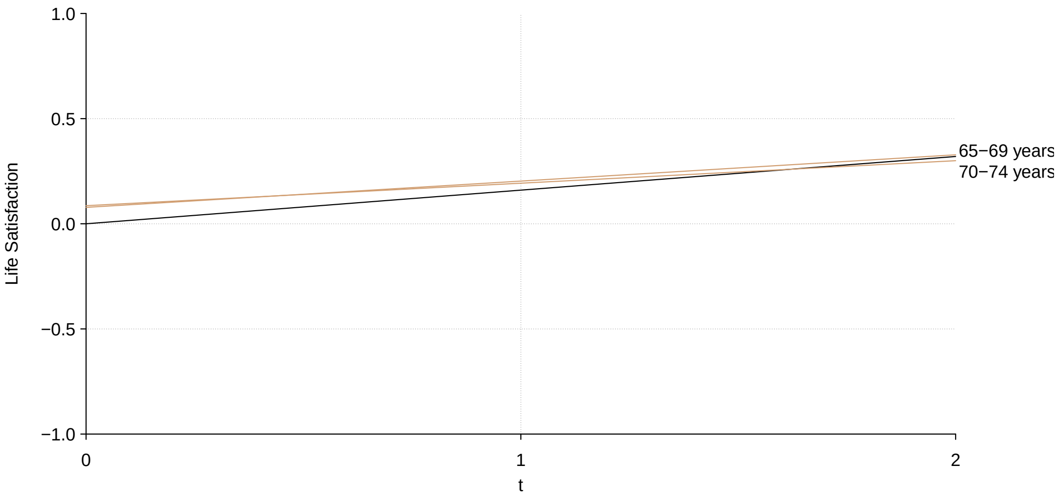
	Intercept, α		Slope, β	
	Estimate	s.e.	Estimate	s.e.
Life Satisfaction				
Age				
60-64 years old	-	-	-	-
65-69 years old	**0.078**	0.036	-0.035	0.023
70-74 years old	**0.086**	0.039	**-0.053**	0.026
Gender				
Female	-	-	-	-
Male	-0.031	0.031	**0.063**	0.020
Race				
White	-	-	-	-
Non-white	-0.145	0.118	-0.096	0.088
Marital Status				
Single	-	-	-	
Married or civil partnership	**0.401**	0.078	**-0.104**	0.044
Divorced or separated	**-0.269**	0.090	0.012	0.051
Widowed	-0.050	0.087	0.015	0.051
Loneliness				
Age				
60-64 years old	-	-	-	-
65-69 years old	0.032	0.041	-0.027	0.026
70-74 years old	0.010	0.043	-0.060	0.029
Gender				
Female	-	-	-	-
Male	**0.109**	0.035	**0.068**	0.023
Race				
White	-	-	-	-
Non-white	-0.241	0.127	-0.085	0.085
Marital Status				
Single	-	-	-	
Married or civil partnership	**0.545**	0.083	**-0.208**	0.054
Divorced or separated	-0.048	0.096	-0.117	0.062
Widowed	-0.151	0.093	0.016	0.061
Depression				
Age				
60-64 years old	-	-	-	-
65-69 years old	0.030	0.042	-0.032	0.028
70-74 years old	-0.003	0.044	**-0.071**	0.030
Gender				
Female	-	-	-	-
Male	**0.217**	0.036	**0.095**	0.025
Race				
White	-	-	-	-
Non-white	**-0.406**	0.124	0.057	0.077
Marital Status				
Single	-	-	-	
Married or civil partnership	**0.221**	0.081	-0.033	0.048
Divorced or separated	**-0.187**	0.094	0.042	0.056
Widowed	**-0.238**	0.090	**0.121**	0.055

Bold highlights significant regression coefficients at 5% level.

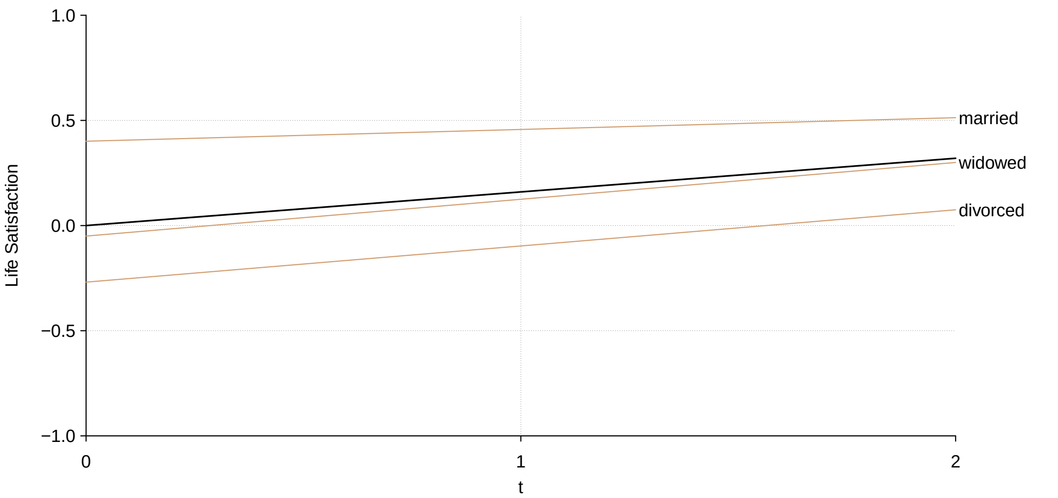
Table 7.2: Model estimated regression coefficients of the time-invariant covariate effects on life satisfaction, loneliness and depression in the conditional multivariate latent growth model in the ELSA (Waves 2, 4 & 6). For reference, the estimated intercepts of the random slopes were $\hat{\nu}_{\beta_s} = 0.160$, $\hat{\nu}_{\beta_l} = 0.201$ and $\hat{\nu}_{\beta_d} = -0.030$.



(a) Demographics

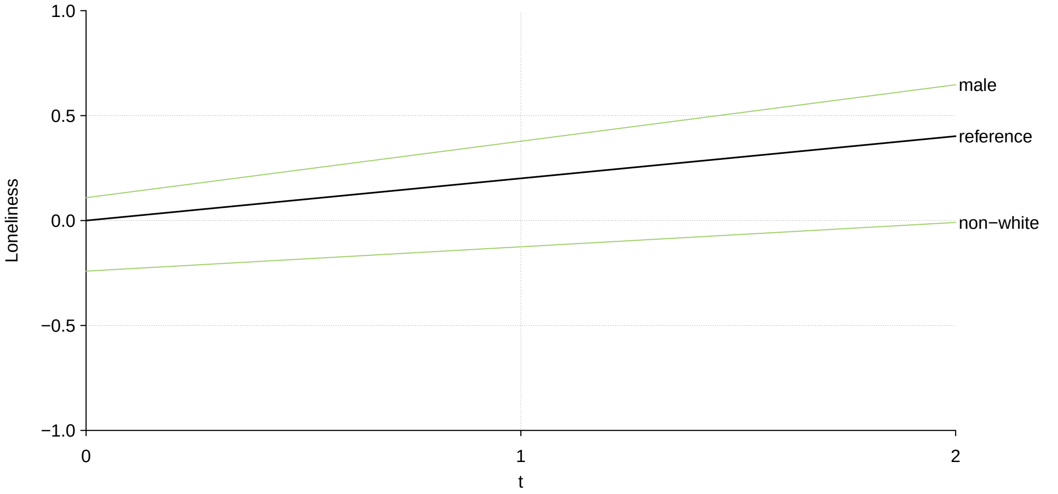


(b) Age

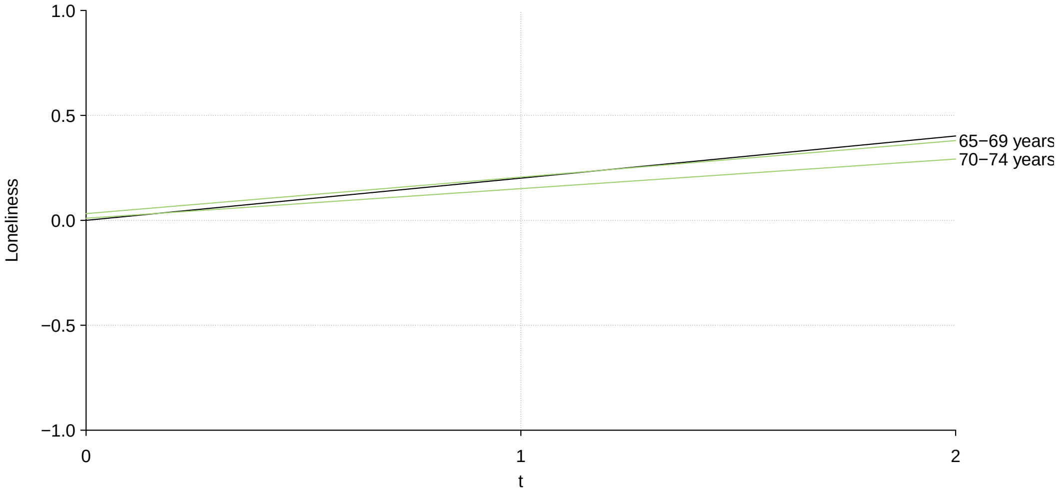


(c) Marital Status

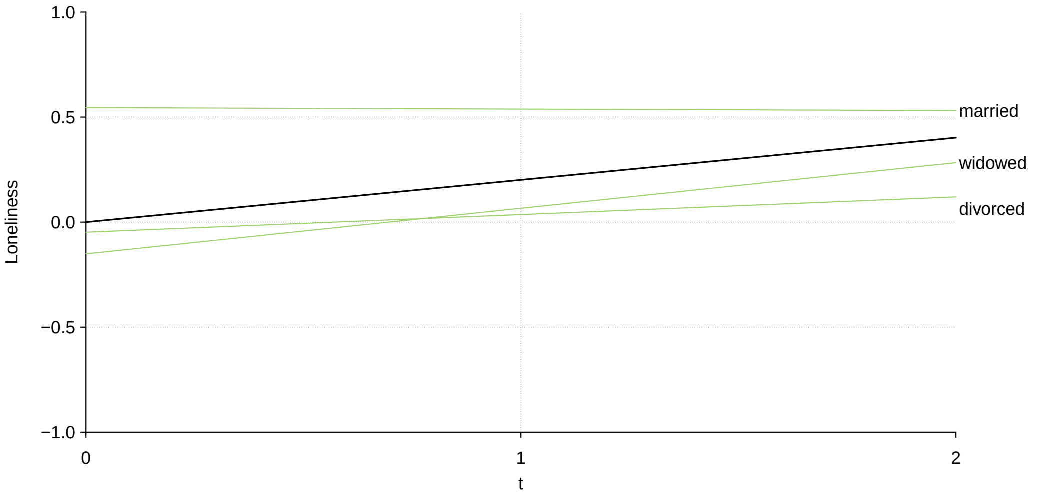
Figure 7.3: Model predicted growth process of Life Satisfaction and the effect of covariates from the conditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 &6); including drop-out. Black line represents the reference profile (white, single, female, aged 60-64 years old).



(a) Demographics

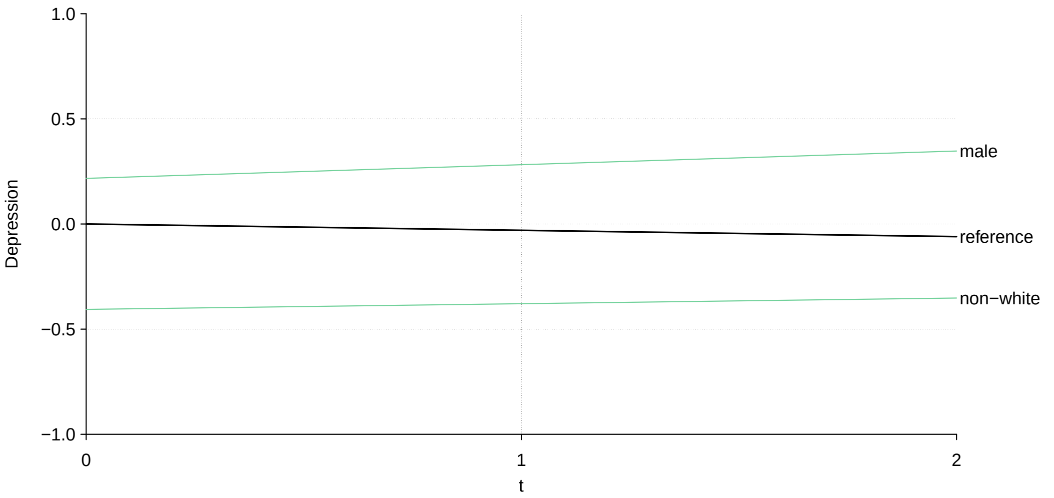


(b) Age

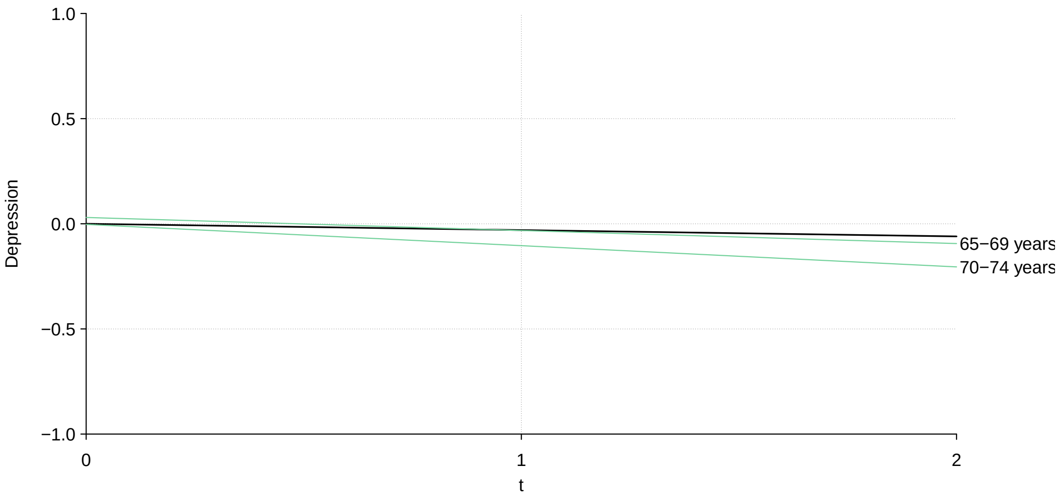


(c) Marital Status

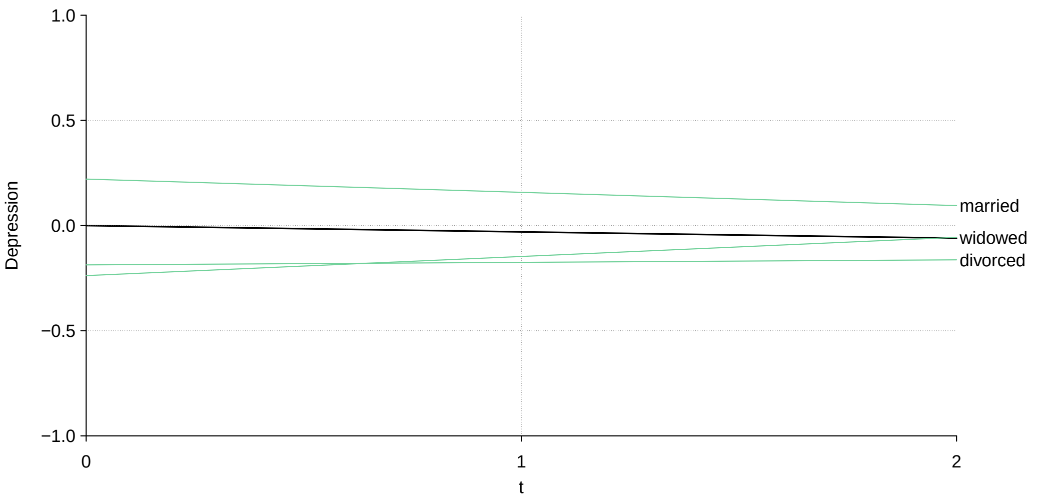
Figure 7.4: Model predicted growth process of Loneliness and the effect of covariates from the conditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 &6); including drop-out. Black line represents the reference profile (white, single, female, aged 60-64 years old).



(a) Demographics



(b) Age



(c) Marital Status

Figure 7.5: Model predicted growth process of Depression and the effect of covariates from the conditional multivariate latent growth model of subjective well-being in the ELSA (Waves 2, 4 &6); including drop-out. Black line represents the reference profile (white, single, female, aged 60-64 years old).

7.4 Conditional Multivariate Latent Growth Model with Time-varying Covariate

The dynamic effect of marital status suggested that the impact of widowhood may have been a temporary reaction to a change in marital status rather than an impact that was sustained over time. To investigate this hypothesis a second conditional multivariate latent growth model was fitted, which included changes in marital status as a time-varying covariate as well as the initial marital status at wave 0 (from the Health Survey for England) as a time invariant covariate. This provided an estimate of the difference in life satisfaction, loneliness and depression at the time proceeding the change in marital status compared to the expected growth if the change in marital status had no occurred.

To accommodate the time varying covariate the analysis was restricted to those present at all three time points and aged 60-74 years old. The MARX assumption required that the marital status variable was observed at all three time points, therefore those that drop-out, and 29 individuals that had missingness on the marital status variable, were excluded from the analysis. Single marital status was defined as "never married" and changing to single status was impossible, hence those whose marital status changed to single were also excluded. Baseline marital status was taken as the marital status variable in Wave 0 of the ELSA; that contained the respective Health Survey for England (1998, 1999 or 2001) data from which the elderly individual was sampled. Change in marital status at Wave 2 therefore corresponded to change from Wave 0, baseline marital status.

Time invariant covariates age, gender, race and baseline marital status were modelled to have an effect on the growth process; both the random intercept and random slope. Besides, time-varying covariates were modelled to affect the latent variable of the aspect at the respective time point, as illustrated in Figure 7.7. The time-varying covariates consisted of three indicators, the first indicated a change from married or in a civil partnership at the previous time point $t-1$ to divorced or separated at time point t, similarly the second indicated a change from married or in a civil partnership at the previous time point $t-1$ to widowed at time point t and finally the third indicated a change from either divorced, separated or widowed at the previous time point $t-1$ to married at time point t. Therefore, for each aspect and baseline marital status the model consisted of the predicted growth process given the baseline marital status and then the different level of the aspect, at time point t, given one of the previously mentioned events occurred in the interval $(t-1, t)$.

Overall, model fit was very good with RMSEA= 0.026, CFI= 0.973 and TLI= 0.972 ($\chi^2 = 4,863.963$ on 1,864 degrees of freedom) and the relationships between aspects had similar form to previously (Figure 7.6). On comparison to previous, the random slopes had marginally smaller residual variances and larger correlations between them, while the random intercepts had less correlation to previously.

The estimated intercepts of the random slopes were $\hat{\nu}_{\beta_s} = 0.042$, $\hat{\nu}_{\beta_l} = 0.062$ and $\hat{\nu}_{\beta_d} = -0.018$, and corresponded to model growth process for the reference profile of a white, female, aged 60-64 years old with single and never married baseline marital status (Wave 0). None of these estimates were significantly different from zero, however the positive values of $\hat{\nu}_{\beta_s}$ and $\hat{\nu}_{\beta_l}$ indicated marginal improvement in life satisfaction and loneliness over time, while the negative estimate of $\hat{\nu}_{\beta_d}$ indicated marginal deterioration in depression over time.

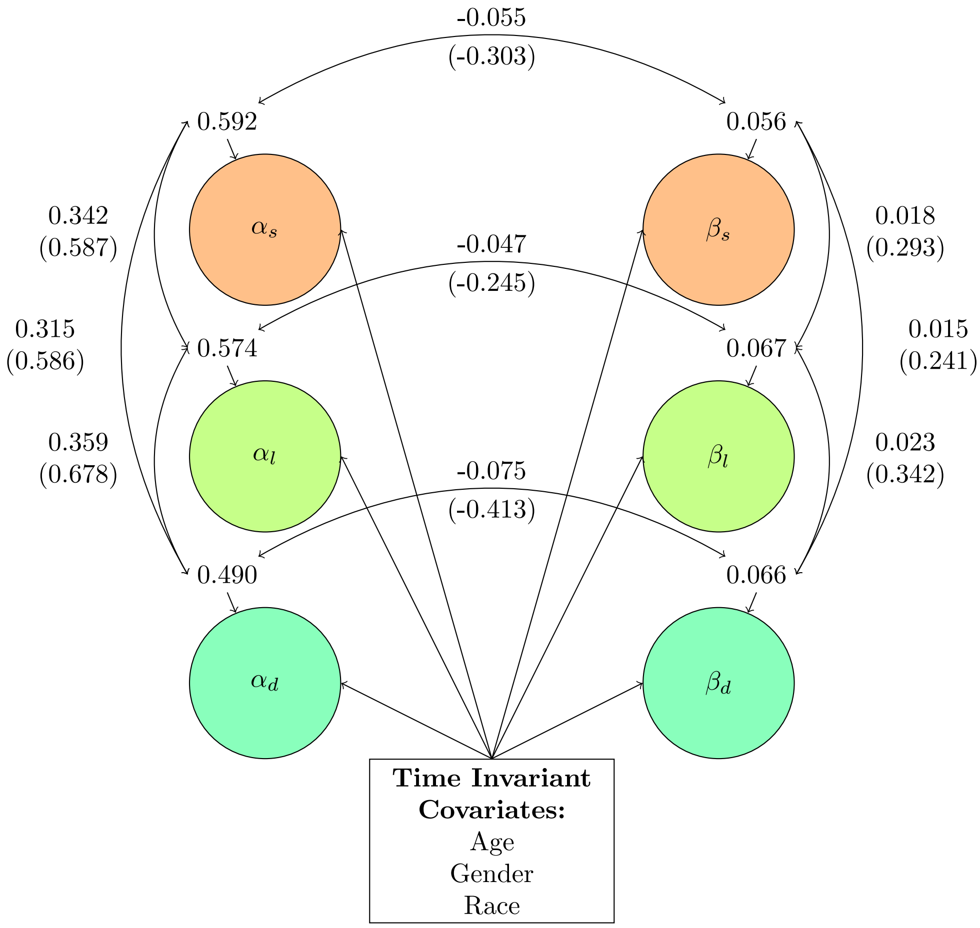


Figure 7.6: Estimated structural equations from the conditional multivariate latent growth model, with time varying marital status, of subjective well-being, for those aged 60-74 years old, in the ELSA (Waves 2, 4 & 6); excluding drop-out.

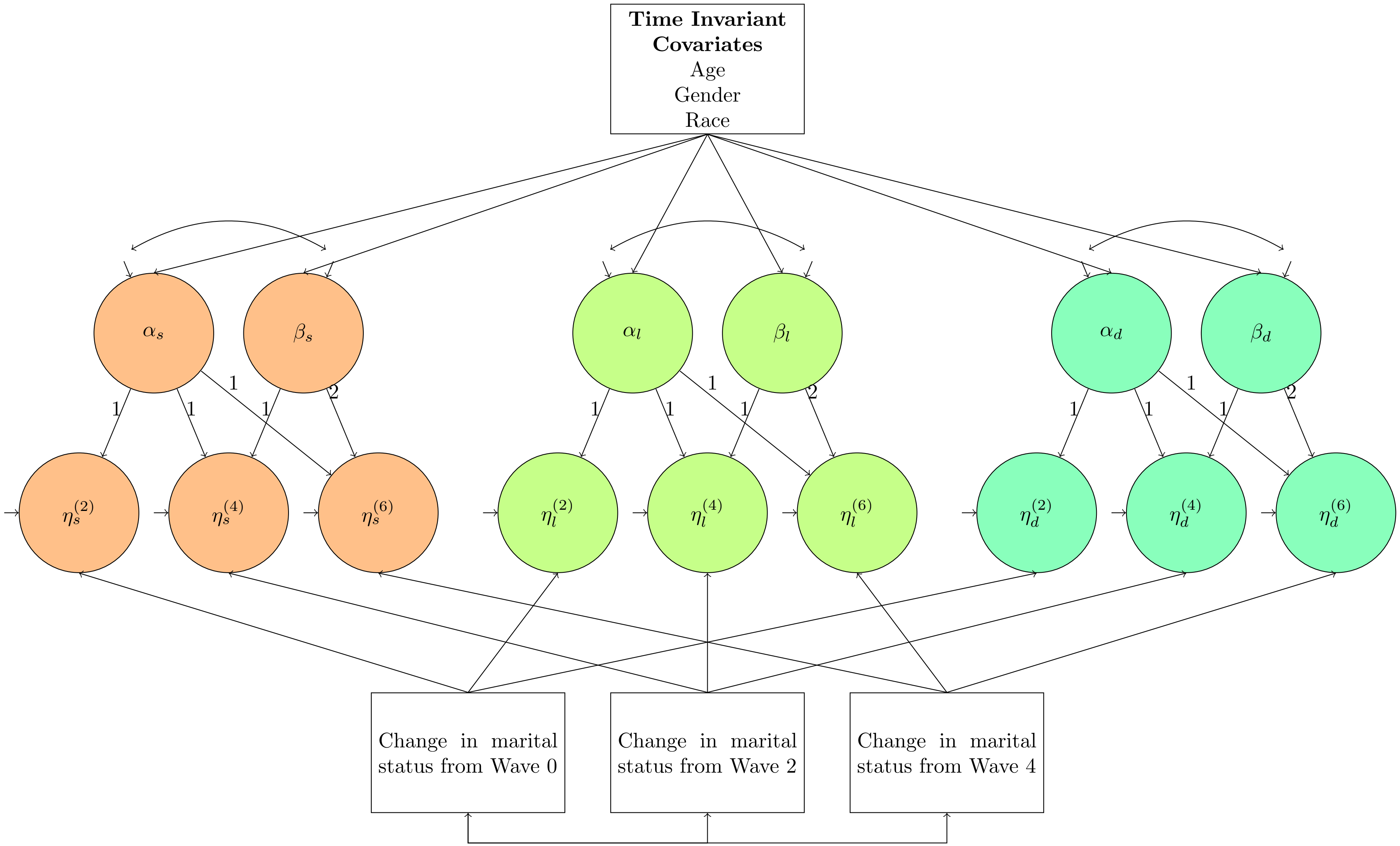


Figure 7.7: Path diagram of the growth process of the conditional latent growth model with a time-varying covariate.

The estimated regression coefficients for the time invariant covariates are given in Table 7.3. The effects on the random intercepts showed that initially life satisfaction was greatest for those aged 65-69 years old, and married at baseline but worse for those that were divorced; compared to those that were single. Gender and race, however, had no significant effect on the random intercept of life satisfaction, and there was no difference between single and widowed baseline marital statuses. Similarly for both, initial loneliness and depression were less, and therefore better, for men that were married at baseline, but were worse for non-white elderly individuals of widowed marital status at baseline. Age had no effect on either random intercept for loneliness or depression.

The random slopes of the aspects were influenced by fewer covariates, in particular no covariate had a significant effect on the change in loneliness over time. Race also had no effect on any of the random slopes. Life satisfaction was seen to improve for men more than for women, however there was marginal deterioration for those aged 65-69 years old compared to those aged 60-64 years old and for those married at baseline compared to those that were single. Men also had more positive changes in depression than women. Depression, however, deteriorated for all age groups with increasing magnitude with older age.

Estimated regression coefficients for the time-varying marital status are given in Table 7.4. These show the modelled difference from the modelled growth curve given either an individual was married and became widowed or divorced or was divorced or widowed and got re-married between the previous and current time point.

At every time point the event of changing to divorced or widowed from married significantly decreased life satisfaction and worsened loneliness and depression, with the greatest impact from becoming divorced. The effect of becoming divorced on life satisfaction and depression, however, lessened with time, while the effect of becoming widowed on loneliness increased with time. All three aspects were significantly improved with a change to married from either divorced or widowed, with increasingly positive influences over time. There were, however, insignificant regression coefficients for becoming divorced or married at the second two time points which was partially due to large standard errors. This could be due to either the low numbers of observations with only 16 cases becoming married between Waves 4 and 6 or, alternatively, there could be a larger variation in the effect of these events on depression for the different ages, as was seen by the time invariant covariates.

Figure 7.8 contains three plots that illustrate the respective growth models for each baseline marital status and the respective value for the time-varying covariates. In each of Figures 7.8a

Covariate	Intercept, α		Slope, β	
	Estimate	s.e.	Estimate	s.e.
Life Satisfaction				
Age				
60-64 years old	-	-	-	-
65-69 years old	**0.135**	0.047	**-0.053**	0.024
70-74 years old	0.082	0.052	-0.040	0.026
Gender				
Female	-	-	-	-
Male	-0.070	0.042	**0.060**	0.021
Race				
White	-	-		
Non-white	-0.246	0.154	-0.052	0.096
Baseline Marital Status				
Single	-	-	-	-
Married or civil partnership	**0.410**	0.101	**-0.093**	0.043
Divorced or separated	**-0.432**	0.117	0.036	0.052
Widowed	-0.084	0.116	-0.006	0.052
Loneliness				
Age				
60-64 years old	-	-	-	-
65-69 years old	0.031	0.052	-0.021	0.027
70-74 years old	0.017	0.058	-0.044	0.031
Gender				
Female	-	-	-	-
Male	**0.094**	0.047	0.038	0.025
Race				
White	-	-	-	-
Non-white	**-0.350**	0.173	-0.040	0.095
Baseline Marital Status				
Single	-	-	-	-
Married or civil partnership	**0.457**	0.104	-0.072	0.056
Divorced or separated	-0.214	0.121	-0.037	0.063
Widowed	**-0.237**	0.120	0.071	0.064
Depression				
Age				
60-64 years old	-	-	-	-
65-69 years old	0.093	0.053	-0.049	0.031
70-74 years old	0.042	0.057	**-0.071**	0.033
Gender				
Female	-	-	-	-
Male	**0.226**	0.049	**0.063**	0.027
Race				
White	-	-	-	-
Non-white	**-0.452**	0.158	0.047	0.085
Baseline Marital Status				
Single	-	-	-	-
Married or civil partnership	**0.210**	0.099	0.021	0.046
Divorced or separated	-0.231	0.119	0.049	0.056
Widowed	**-0.259**	0.116	0.102	0.058

Table 7.3: Model estimated regression coefficients for the time invariant covariates in the conditional multivariate latent growth model with time-varying covariates, of subjective well-being, for those aged 60-74 years old in the ELSA (Waves 2, 4 & 6); excluding drop-out. For reference, the estimated intercepts of the random slopes were $\hat{\nu}_{\beta_s} = 0.042$, $\hat{\nu}\beta_l = 0.062$ and $\hat{\nu}_{\beta_d} = -0.018$.

to 7.8c the lines represent the model estimated growth process given the baseline marital status and that the events of the time-varying covariates do not occur. The points then represent where an individual would be predicted to be on the latent scale at that time point, given their baseline marital status, and that there was a change from married to divorced or widowed, or from divorced to married or from widowed to married; between the previous and current time point. The black line represents the growth process for the reference profile; female, white, aged 60-64 years old and single baseline marital status.

	Wave 2, $\eta^{(2)}$		Wave 4, $\eta^{(4)}$		Wave 6, $\eta^{(6)}$	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Life Satisfaction						
Married or civil partnership to...						
Divorced or separated	**-0.906**	0.308	**-0.756**	0.332	**-0.730**	0.354
Widowed	**-0.462**	0.097	**-0.631**	0.110	**-0.587**	0.107
Divorced, separated or widowed to...						
Married or civil partnership	**0.727**	0.202	**0.745**	0.255	**1.108**	0.287
Loneliness						
Married or civil partnership to...						
Divorced or separated	**-0.914**	0.280	**-1.003**	0.354	**-0.963**	0.484
Widowed	**-0.879**	0.109	**-0.885**	0.122	**-0.993**	0.107
Divorced, separated or widowed to...						
Married or civil partnership	**0.567**	0.272	**0.729**	0.250	**0.914**	0.287
Depression						
Married or civil partnership to...						
Divorced or separated	**-0.912**	0.335	-0.617	1.768	-0.595	0.437
Widowed	**-0.695**	0.101	**-0.743**	0.098	**-0.678**	0.105
Divorced, separated or widowed to...						
Married or civil partnership	**0.422**	0.196	0.431	0.455	-0.026	13.497

Bold highlight for estimates significant at 5% level

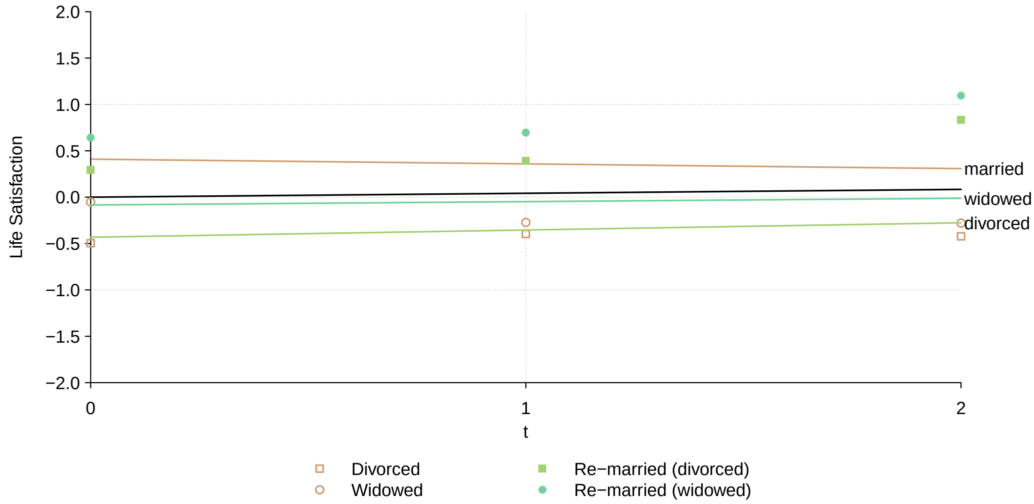
Table 7.4: Model estimated regression coefficients for the time-varying covariates in the conditional multivariate latent growth model, with time-varying covariates, of subjective well-being, for those aged 60-74 years old in the ELSA (Waves 2, 4 & 6); excluding drop-out.

Figure 7.8a demonstrates the relatively little change in life satisfaction over time for each of the baseline marital statuses except the decline for married at baseline. The black line represents the growth process for single marital status at baseline. At the first time point, $t = 0$, the life satisfaction of those divorced or widowed between Waves 0 and 2 was the same as for those divorced or widowed at baseline. The effect of becoming divorced then appears to remain constant and to be generally near the growth process for being divorced at baseline. Comparatively, those that become widowed also had similar life satisfaction as those that were divorced at baseline, and had worse life satisfaction than those widowed at baseline. Conversely, becoming married from either divorced or widowed showed greater life satisfaction than those that were married from baseline; except at the first time point when becoming divorced had slightly lower life satisfaction.

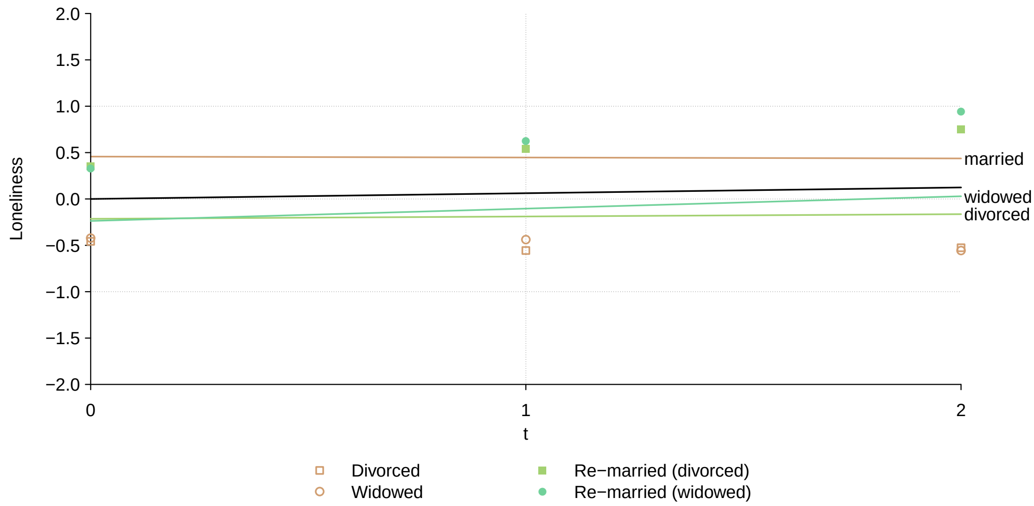
Figure 7.8b highlighted the recovery for those that were widowed at baseline compared to those that were divorced. There was relatively little difference between becoming widowed or divorced or between getting married from either widowhood or divorce, at each time point. Similarly to with life satisfaction, those that remarried were less lonely than those that were married from baseline except at the first time point. Becoming divorced or widowed also resulted in similar levels of loneliness, that were much worse than from any baseline marital status.

Depression had very little change over time or different between the baseline marital statuses, shown in Figure 7.8c. Those that remarried from being divorced or widowed at baseline had similar levels of depression to those that were married from baseline, while those that became divorced or widowed had worse levels of depression than those divorced or widowed at baseline. Alternatively to the other aspects, those that became widowed had worse depression than those that became divorced.

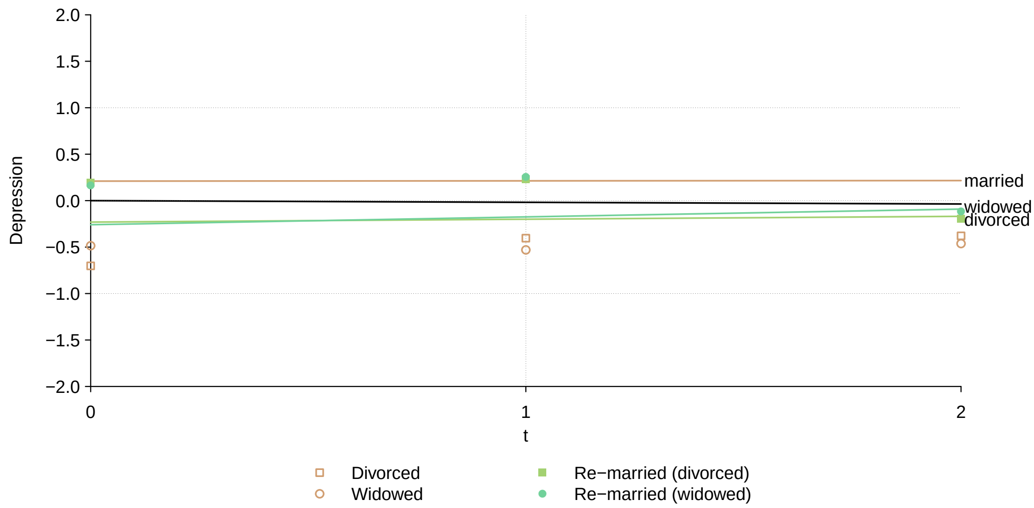
Ideally the model would include the continued effect of an event, such as divorce, widowhood or re-marriage, on the subsequent growth process. In future, this could be achieved by allowing for a lagged time-varying covariate, that is to model the location of the latent scale at time t given an event occurred in the periods between $t-2$ and $t-1$. However, to date there were only three time points with the subjective well-being variable available from the ELSA, but this could be accomplished in future.



(a) Life Satisfaction



(b) Loneliness



(c) Depression

Figure 7.8: Model estimated growth curves and time-vary points for the conditional multivariate latent growth model, with time-varying covariates, of subjective well-being, for those aged 60-74 years old in the ELSA (Waves 2, 4 & 6); excluding drop-out. Black line represents the reference profile (white, single, female, aged 60-64 years old).

Summary of the Chapter

In this chapter we have modelled the growth of the three aspects, life satisfaction, loneliness and depression, of subjective well-being using Wave 2, 4 and 6 of the ELSA.

Firstly, the measurement invariance of each aspect was explored in relation to the factor loadings and thresholds, that are required for strong measurement invariance. As was expected all aspects showed some level of invariance, which required equality constraints for the factor loadings and a subset of thresholds at all the time points.

The three growth processes were then modelled in a multivariate model, with a random intercept and random slope for each aspect, in order to determine the relationship between the growth processes. Alongside, the estimated trend in each aspect for the elderly; given they retained ability to participate. On average, life satisfaction and loneliness were both seen to deteriorate over time whilst depression remain unchanged. However, loneliness and depression had a strong relationship, and were also strongly connected to life satisfaction.

Next, the model was extended to include those that drop-out, but restricted to the younger elderly population, aged 60-74 years old. Conversely to the previous model, there was a positive trend in life satisfaction and in loneliness, while depression had a negative trend. Although the changes over time were not statistically significant, the differences may be significant in practical terms.

Considering the time invariant covariates, it was found that being male or married were protective and were persistently better in all three aspects. Besides there was little difference between the age groups except at the third time point, where the oldest were marginally worse. Marital status also revealed an element of recovery in the loneliness and depression in those that were widowed, that was not present in those that were divorced. This was then investigated further, owing to the inclusion of a change in marital status as a time-varying covariate.

The time-varying covariate demonstrated significant changes in life satisfaction, loneliness and depression following a change in marital status. Those that were either divorced or widowed at baseline and became married had better levels of life satisfaction and loneliness than those that were married at baseline. Depression also improved and was generally similar to those that were married at baseline. Alternatively, those that became divorced or widowed from being married at baseline had worse loneliness and depression than those that were divorced or widowed at baseline. Life satisfaction was lower for those that got divorced compared to those that became widowed, however depression was worse for those that became widowed than got divorced.

Chapter 8

Comparison and Conclusions

In this chapter, we outline the results from the separate analyses to create a comparison of the elderly in England and Brazil. Firstly, the latent structures identified for health and economic well-being are compared, highlighting some common features before discussing the more specific results that apply to the separate nations. Cross-nation comparisons are also made in terms of the effect of covariates on the health and economic well-being of the elderly. A specific section on the results concerning subjective well-being then follows. Some limitations of the data and the analysis are then discussed, followed by a discussion of possible future work.

8.1 Latent structures

Comparison of the latent structures highlighted several common features of health in the elderly between England and Brazil; despite the differences in the available variables from the surveys.

In both nations, the comorbidities associated with metabolic syndrome in the literature were indicators of a separate factor. In England, these comorbidities were diabetes, hypertension and high cholesterol, while in Brazil the comorbidities were diabetes, hypertension and heart disease. Previously discussed in Chapter 2, metabolic syndrome is a cluster of risk factors of cardiovascular health, which included hypertension and high cholesterol as known components of metabolic syndrome; while diabetes is strongly associated with increased risk of metabolic syndrome (He et al., 2006; Maggi et al., 2006).

Contrarily, Bo et al. (2009) applied structural equation modelling (SEM) to investigate the presence of a single factor of metabolic syndrome in the elderly, finding low internal consistency among the components of metabolic syndrome. However, this study was conducted on a small restrictive sample of 152 non-diabetic older men (aged 70 or more years) in the metropolitan area of Turin, Italy; therefore, lacks representation and may unintentionally exclude those with insulin resistance, which could be mistaken for diabetes.

It is clear that risk factors associated with metabolic syndrome are an important area of morbidity in the elderly in England and Brazil that requires specific attention in an elderly population.

Alternatively, in the analysis of Brazil, a factor representing the musculoskeletal disorders was identified. This may also be present in the elderly in England, however, the ELSA was limited with only two variables concerning musculoskeletal disorders, arthritis or rheumatism and osteoporosis, so a similar factor could not be identified. Although, the ELSA contained more detail in relation to the cardiovascular disorders with a specific group of variables, while the PNAD contained a single variable of heart disease.

One main difference in the latent structures between England and Brazil were in the representation of economic well-being. The ownership of basic goods such as fridge or TV were not associated with higher wealth in England, therefore did not represent greater economic well-being but were conveying the lack of necessity with increased wealth. Conversely, in Brazil, the ownership of even the most basic durables, such as a stove, were strong indicators of economic well-being in the elderly. The housing quality variables also differed between the nations with Brazilian variable focussed on the basic structure and amenities rather than problems with the accommodation such as repairs that are required or the building not fitting the needs of the occupants.

In England, problems within the household, such as electrical or plumbing problems, and those outside the household such as noise and pollution. This separation did not occur in the PNAD, however, there was some similar but small clustering visible in the tetrachoric correlation coefficients. This was particularly visible in the multilevel SEM.

Functional status was found to have a latent structure with two distinct factors in the ELSA. These factors did not represent the basic and instrumental activities of daily living, as one might expect, instead separated the activities according to their requirement for physical or cognitive function. Activities requiring cognitive function were not included in the functional status variables of the 2008 PNAD, so the distinction could not be analysed here. This is not found to feature in the previous analysis of the elderly during the literature review (Section 2.4), therefore, more research is required into the relationship between physical and cognitive function to better understand the concept and role of functional status in the health of the elderly.

Interrelationships between economic well-being and health were consistent between the two countries, with positive association between the two domains. In England this was more specifically between household quality and health, and not with neighbourhood quality. Therefore, in the elderly, a poorer environment within the household was associated with poorer health and vice versa. It is hence important that policies aimed at improving the health and well-being of the elderly should also provide support in relation to housing.

8.2 Multilevel latent structure

The multilevel SEM revealed different latent structures for the health and economic well-being of neighbourhoods of elderly individuals than of the elderly individuals. Once the sector was considered in the model, the importance of functional status to the overall health of the elderly individual was much greater. Prior to the multilevel analysis, functional status was the least influential factor of overall health, however was the second greatest influence in the multilevel analysis. The importance of each group of activities of daily living was comparably different to the elderly individual than to the sector. Limitations to stooping, kneeling and bending had the least influence on the individual functional status but was the most influential variable of functional status in the sector. The basic activities of daily living, eating, bathing and toileting, were more important to an elderly individual's functional status than the functional status in the sector. This highlights the strong importance of sector based provisions, such as healthcare and local public services, to reducing the impact of health problems on the functional status of the elderly.

The associations between the factors at the sector level revealed functional status to be more strongly associated with industry related health, such as arthritis, disease of the column or back and hypertension, than with chronic disorders and infectious diseases. The comorbidities of the industry related health factor are known to be associated with the workplace (Schumann et al., 2011; Health and Safety Executive UK, 2017), it is therefore important that care for the elderly starts by caring for the health of working adults, that may be at increased risk of functional limitations later in life.

A persistent issue in the multilevel modelling of the health variables occurred, was the poor fit at the sector level; indicated by the standardized root mean square error (SRMR) of the between level correlation matrix. A larger number of factors were identified to provide better fit to the data but at the compromise of allowing strong negative cross-loadings. This highlighted that some variables may indicate greater health in one aspect but poorer health in another. An example of this seen in the correlation between cirrhosis and hypertension, where the development of cirrhosis in hypertensive individuals leads to the individual becoming normotensive (Henriksen and Moller, 2006). Further work is needed to expand and create understanding on these more complicated relationships; which was beyond the focus of this thesis.

It was in the multilevel analyses that the importance of local services to the economic well-being was discerned. To the economic well-being of the elderly individual the possession of a stove and fridge were the two most influential variables, however the economic well-being of localities of the elderly was mostly indicated by the possession of a computer or telephone that require access to a network.

Further to this analysis, it would be desirable in future work to extend the model with three levels approach, potentially with municipality as another level. In the intraclass correlation coefficients, we identified that the municipality explained a large percentage of the variation in some of the variables, in particular of economic well-being. Municipality was also found to be informative as the third level in the multilevel cumulative logistic regression analysis of self-rated health.

8.3 Population characteristics

The effects of age, gender and race on the health of the elderly were consistent between England and Brazil. Older age, female gender and non-white race were associated with significantly poorer health, while significant regional variations in health were also present.

Health of the elderly was better in the southern regions of England (East of England, South West and London) than in the northern regions of England (West Midlands, Yorkshire and the Humber, East Midlands and North East). These were the same as the inequalities in health presented by Public Health England (2017). This included a higher life expectancy and healthy life expectancy in the south of England (South East, South West, the East of England and London), while in the north of the England there was a greater proportion of the local authority districts in the 10% most deprived in England (Department of Communities and Local Government, 2015).

In Brazil, health of the elderly was better in the South East region of Brazil. In all regions of

Brazil, the health of the elderly was significantly worse compared to the South East, while the worst health was seen in the South and Central West regions. Functional status was worst in the northern regions of Brazil, while industry related health was worst in the South region and North region. Functional status and comorbidities related to industry, which included mobility related comorbidities such as arthritis, were best in the South East region where the two largest cities, São Paulo and Rio de Janeiro, are located. Controversially, local health was worst in the southern regions of Brazil. However, this could be constituted to disparity in the diagnosis of comorbidities that required more complex medical facilities, hence under-diagnosis appearing as better health. Health inequalities in Brazil include a concentration of medium and complex facilities in the state capitals and larger cities (de Albuquerque et al., 2017) and fewer that one physician per 1,000 people in some municipalities (Santos et al., 2017). Access to basic health facilities, on the other hand, have been improving with the introduction of the Family Health Program (de Albuquerque et al., 2017). The Family Health Program provides health care to families by establishing professional medical teams, composed of a general practitioner, generalist nurse auxiliary or nursing technician and community health agents, in the more remote areas of Brazil (Ministério da Saúde Departamento de Atenção Básica, 2017).

As discussed in Section 2.2.1, the right to health was given to the people by the 1988 Constitution, since then the Brazilian Government has been taking steps towards providing primary healthcare to all with the Unified Healthcare System and the Family Health Programme. This is Brazil's main strategy to provide primary healthcare at the same time as encouraging poorer families to use the services (World Health Organisation [WHO], 2008).

Alternatively, functional status was poorest in the North and Central West regions and was indifferent between urban and rural areas. The functional status of the elderly was significantly better in the South East region than in any other.

Regional disparities existed in the economic well-being in both England and Brazil. As mentioned above economic well-being was significantly worse in the northern and Central West regions of Brazil, with additionally worse economic well-being in rural areas. The disparities in England, however, only existed in the quality of the neighbourhood and not the conditions inside the household. There were the least problems with noise and pollution in the East Midlands, South West, North West, however the worst problems were in the London and North East regions.

Brazil and England differ in the effect of age, gender and race on the economic well-being of the elderly. Firstly, in Brazil, economic well-being deteriorated with older age, and was significantly worse for men and the elderly of non-white race. Meanwhile, in England, both household and neighbourhood quality improved with older age, and there was no difference in either factors by gender or race.

Additionally, the effect of education and marital status were determined in the analysis for England. Problems inside the household were fewer for those that were married, in a civil partnership or widowed, however subjective well-being was significantly worse for those that were widowed. Meanwhile, education only showed significant difference in the health of the elderly in what appears to be two separate groups, distinguishing between those with the equivalent of an NVQ level 3 or higher and those with less than NVQ level 3. Firstly, there was no difference health of the elderly with NVQ level 3 or equivalent education compared to those with no qualifications. However, as the level of education improved from no qualifications to NVQ level 2 or equivalent health was better then, again, health improved from NVQ level 3 or equivalent to degree level.

8.4 Subjective Well-being

Subjective well-being was a domain that was only available in the ELSA, that contributed to the cross-sectional analysis, and was the focus of longitudinal analyses. Each aspect, of subjective well-being, consisted of a selection of variables from previously validated psychometric measure of either quality of life, satisfaction with life, loneliness or depression. The latent structure of the variables given by the authors was confirmed, for the satisfaction with life, loneliness and depression aspects, with a single factor underlying each, as prescribed. The suggested latent structure of the CASP-19 measure of quality of life, however, provided an invalid model with Heywood estimates; although the existence of four factors was confirmed by the exploratory factor analysis (EFA). The structure between the factors and latent response variables, however, was different. Previously, the authors had also tested the CASP-19 from the ELSA, using Wave 1 (Wiggins et al., 2008) from which they state that "as far as model fit is concerned we could do better!". The CASP-19 therefore requires further research to confirm and reassure about the reliability and validity of its latent structure, particularly the consistency of the measure in a test re-test scenario. Due to the lack of consistency in the latent structure of the CASP-19, it was not included in the longitudinal analyses of subjective well-being.

Item response analysis of the CASP-19 variables revealed several areas highlighted a few issues with the performance of the items. Several of the items had poor discriminatory ability and were unable to distinguish between the responses of elderly individuals located at far extremes of the trait score. Additionally, the category response scale presented a problem with the under use of the 'not often' category and in some cases, it was unused.

Health had the largest effect on subjective well-being, with better health associated with better subjective well-being, while better household quality had the same effect. Neighbourhood quality did not have a significant effect. Supporting the elderly with the maintenance of their homes, such as with repairs or providing suitable housing can help decrease loneliness and depression as well as increasing their quality of life and satisfaction. This relationship means that the deterioration of health that occurs with ageing causes a deterioration of subjective well-being in older age, thus it is important to support the elderly in terms of their mental health and social support as their health deteriorates.

The effect of individual characteristics was also significant with differences between age groups, genders and marital statuses. From the age of 65-69 years old subjective well-being was worse with older age, subjective well-being was lower for men and those that were widowed. Education, race and government office region had no effect on the subjective well-being of the elderly.

Trajectories in life satisfaction, loneliness and depression were found to be strongly associated. For the elderly, in all age groups, that were present over all the time points, the trajectories showed significant deterioration in satisfaction with life and loneliness over time, while depression remained unchanged. Hence, the elderly that maintain the ability to participate in the survey are still at risk of deterioration in well-being as they age. The majority of the older age groups (75 or more years old) did drop out.

In the younger ages (65-74 years old), there was no significant change in either of the three aspects over time. Although statistically insignificant, life satisfaction and loneliness showed small improvements over time, which for men and for the elderly that were widowed at the initial time point, was significantly larger. We discovered that despite subjective well-being being the worst for those that were widowed at the first time point, there appeared to be marginally greater improvement over time, and a significant improvement in depression. In comparison, those that were divorced had significantly lower subjective well-being which had no change over time.

Marital status was the most determinant factor of subjective well-being that we considered. Over time, subjective well-being was found to be consistently better for men and for those that were married or in a civil partnership, whilst being consistently worse for those of non-white race and for those that were divorced or separated. The effect of a change in marital status further emphasised the influence of marital status on subjective well-being. Becoming divorced, separated or widowed severely reduced the subjective well-being of the elderly, with loneliness and depression worse than that expected level of those that had been divorced, separated or widowed initially. Conversely, getting re-married significantly improved subjective well-being, such that life satisfaction and loneliness were better than was expected of those that were married at the start.

8.5 Limitations

In the analysis of self-rated health of the elderly in Brazil, we discerned a lack of homogeneity in the perception of health by the elderly with older age. When controlling for characteristics such as comorbidity, functional status and healthcare usage the elderly in the older age groups had a more positive rating of their health, despite an equal status in terms of health characteristics. Self-rated health provides good indication of the subjective perception of health by the elderly, allowing us to determine where that support may be required, such as the youngest age groups as they adjust to the deterioration of health, and that education would be beneficial especially in relation to health behaviours; as shown in our analysis using multilevel cumulative logistic regression. It does not, however, provide an unbiased measure of either overall health or subjective well-being, rather it is a measure of the combination but the effect of health on the well-being of the elderly can not be deciphered.

Another disadvantage from the use of multilevel cumulative logistic regression to model the ordinal observed variable, self-rated health, was the inability to determine different effects for each of the comorbidities, activities of daily living, the elements of housing quality and the different durables. It would be unreasonable to assume that these variables were independent, thus could not be modelled without aggregation, due to the potential collinearity.

The percentages of self-reported morbidities in both datasets under-estimated the prevalence found in the literature. The comorbidity with the most under-representation was arthritis. The prevalence of arthritis in the elderly is estimated to be around 70% of the elderly (Kee and Epps, 2001), however in the ELSA only 43.08% reported a diagnosis of arthritis and in the PNAD only 26.20% reported a diagnosis of arthritis or rheumatism. The samples from both survey only contained individuals within private households, rather than institutions or hospital. From both surveys, we excluded those that gave interview by proxy in the ELSA and those that had proxy response to the health section of the PNAD. This may have excluded some of those with comorbidities due to their ill-health and incapable of providing responses on their own. The under-representation in the PNAD may also be a result of under-diagnosis rather than just under-reporting, as discussed above.

The four UCLA Loneliness Scale items used in the ELSA did not correspond to the four items recommended by Russell et al. (1980) as the shortened 4-item version. The items in the ELSA do not follow the recommended equal distribution of items between positively and negatively worded items that was a key aim of both the UCLA Loneliness Scale Russell et al. (1980) and the Social and Emotional Loneliness Scale (DiTommaso et al., 2004). The additional fifth item, that was not from the UCLA Loneliness Scale, explicitly mentions feeling lonely, a feature that was keenly avoided by the authors of the UCLA Loneliness Scale. It is, therefore, noted that the reliability and validity properties of these items as a measure of loneliness should be held with caution, especially when a summative score is used.

Longitudinal analysis of subjective well-being in the ELSA was restricted by the limited number of time points, and large periods in between measurements. Waves of the ELSA are conducted every two years with the subjective well-being variables only included in evenly numbered waves. Hence, the latent growth modelling was restricted to the analysis of a linear trend. The difference in the trajectories modelled from all age groups and the younger age groups hinted a change in the direction of the trajectory may occur, it would therefore be beneficial to investigate another trend such as a quadratic curve; for which a minimum number of four time points are required. A larger frequency of time points could also allow for the inclusion of lagged time-varying covariates that would enable us to determine any lasting effect from a change in marital status.

In the multilevel analysis of the PNAD, it would have been preferable to include the sampling probabilities for each stage of the complex sampling plan. Inclusion of the sampling weights would restrict selection bias in the analysis. However, identifiers for the exact clustering of sectors was not available (see Section 3.1.1) so the current analysis was limited to only the inclusion of weights in non-multilevel analysis. The provided weights were only suitable for calculation of means and not for multilevel analysis (Asparouhov, 2006).

Currently, structural equation modelling with a large number of categorical variables is based on the limited information approach of multi-stage methods with weighted least squares approximation. Computation of the full information likelihood function is difficult due to high dimensional integration required for the estimation of the large quantity of parameters; including thresholds, polychoric or tetrachoric correction coefficients and parameters of the structural model.

The multi-stage method of WLSMV also has no satisfactory method for the evaluation of the goodness of fit of the model to the data (Maydeu-Olivares, 2006; Muthén, 1993). This is due to the additional level of goodness of fit required by categorical variables given the multivariate normality assumption made for the estimation of the tetrachoric or polychoric correlation coefficients (Muthén, 1993).

8.6 Considerations

Measurement equivalence was an obvious problem with the cross-national comparison. The SEMs did not have direct equivalence, with the inclusion of different sets of variables, and so lack configural or metric invariance. However, there are economic and cultural differences between the nations, that alter the understanding and relevance of the items (Davidov et al., 2014).

It is possible that some equivalence exists where variables were irrelevant between the nations. For example, due to its rarity in England, tuberculosis may be considered an indicator of health that was negligible. Similarly, the variables of housing quality considered in the Brazilian analysis may also be negligible in England. Conversely, problems of cold and damp or the ownership of a dishwasher within a Brazilian household may also be negligible, given the differences in climate and culture. This is area that requires greater research, to determine possibility of obtaining measurement equivalence in future.

8.7 Further Work

Some areas where further work would be possible that have been mentioned above include: further analysis of the complex structure underlying the comorbidity variables, that were particularly noticeable at the sector level; more detailed analysis of the separate physical and cognitive dimensions in functional status; and the extension of the multilevel structural equation model to a three level model, with municipality as the potential higher level.

Health and economic well-being were not included in the longitudinal analysis of subjective well-being, however the measurement models for health and for economic well-being could be included either as covariates for the baseline health/economic well-being or included as another parallel growth process. This will require a larger dataset with more time points to ensure model identification.

Drop-out was a major feature of the longitudinal analysis, with over 50% of those aged over 75 years old dropping out of the ELSA. It would, therefore, be of interest to include an element of survival or time-to-event analysis in the structural equation model; treating drop-out as a time-to-event which is correlated to the latent variables of interest. This would have the potential to identify trajectories of subjective well-being and health that lead to such an event. Unfortunately, at this time, data on the drop-out was limited with the outcome of many respondents unknown.

Progression to a higher order model for the multivariate latent growth model would be natural. Firstly, as was discussed in Chapter 2, the inclusion of a higher order categorical latent variable for a growth mixture model would categorise the elderly with similar trajectories. This would help identify if there was a profile where the elderly were most at risk of deterioration in subjective well-being. Alternatively, a higher order factor of curves model could be considered. This would include a common intercept and common slope of all the growth process that could be perceived as overall growth process for subjective well-being (Duncan et al., 1999).

Bibliography

Abrantes, J. D. d. A. P. à. S. d. S. (2015). Secretaria de Estado de Saúde do Distrito Federal - Programa Saúde da Família.

Ahto, M., Isoaho, R., Puolijoki, H., Laippala, P., Romo, M., and Kivelä, S. L. (1998). Prevalence of coronary heart disease, associated manifestations and electrocardiographic findings in elderly Finns. *Age and Ageing*, 27(6):729–738.

Akinbami, L. J. and Liu, X. (2011). Chronic obstructive pulmonary disease among adults aged 18 and over in the United States, 1998-2009. *NCHS Data Brief*, (63):1–8.

Alexander, C. M., Landsman, P. B., Teutsch, S. M., and Haffner, S. M. (2003). NCEP-Defined Metabolic Syndrome, Diabetes, and Prevalence of Coronary Heart Disease Among NHANES III Participants Age 50 Years and Older. *Diabetes*, 52(5):1210–1214.

Almeida, O. P. (2014). Prevention of depression in older age. *Maturitas*, 79(2):136–141.

Alpass, F. M. and Neville, S. (2003). Loneliness, health and depression in older males. *Aging & Mental Health*, 7(3):212–216.

Alves, L. C. and Rodrigues, R. N. (2005). Determinants of self-perception of health among elderly people in the municipality of Sao Paulo, Brazil - Determinantes da autopercepção de saúde entre idosos do Município de São Paulo, Brasil. *Revista Panamericana de Salud Pública*, 17(1):333–341.

Andersson, L. (1998). Loneliness research and interventions: A review of the literature. *Aging & Mental Health*, 2(4):264–274.

Arnadottir, S. A., Gunnarsdottir, E. D., Stenlund, H., and Lundin-Olsson, L. (2011). Determinants of self-rated health in old age: a population-based, cross-sectional study using the International Classification of Functioning. *BMC public health*, 11:670.

Asparouhov and Muthén (2010). Weighted least squares estimation with missing data. *MplusTechnical Appendix*, pages 1–10.

Asparouhov, T. (2006). General multi-level modeling with sampling weights. *Communications in Statistics: Theory & Methods*, 35(3):439–460(22).

Asparouhov, T. and Muthén, B. (2016). IRT in Mplus, Version 2. Technical report.

Asparouhov, T. and Muthén, B. O. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. *Proceedings of the 2007 Joint Statistical Meetings*, pages 2531–2535.

Aylaz, R., Aktürk, Ü., Erci, B., Öztürk, H., and Aslan, H. (2012). Relationship between depression and loneliness in elderly and examination of influential factors. *Archives of Gerontology and Geriatrics*, 55(3):548–554.

Barbosa, A. T. F., Carneiro, J. A., Ramos, G. C. F., Leite, M. T., and Caldeira, A. P. (2017). Factors associated with Chronic Obstructive Pulmonary Disease among the elderly. *Ciencia & Saude Coletiva*, 22(1):63–73.

Bartholomew, D. J., Knott, M., and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. Wiley, 3rd edition.

Battaglia, S., Basile, M., Scichilone, N., and Bellia, V. (2015). Prevalence of Co-morbidities and severity of COPD. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 12(4):390–394.

Beilby, J. (2004). Definition of Metabolic Syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association Conference on Scientific Issues Related to Definition. *The Clinical Biochemist Reviews*, 25(3):195–198.

Bellia, V., Battaglia, S., Catalano, F., Scichilone, N., Incalzi, R. A., Imperiale, C., and Rengo, F. (2003). Aging and Disability Affect Misdiagnosis of COPD in Elderly Asthmatics. *Chest*, 123(4):1066–1072.

Bentayeb, M., Simoni, M., Norback, D., Baldacci, S., Maio, S., Viegi, G., and Annesi-Maesano, I. (2013). Indoor air pollution and respiratory health in the elderly. *Journal of environmental science and health. Part A, Toxic/hazardous substances & environmental engineering*, 48(14):1783–9.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2):238–246.

Blackwell, D. L., Lucas, J. W., and Clarke, T. C. (2014). Summary health statistics for U.S. adults: national health interview survey, 2012. *Vital And Health Statistics. Series 10, Data From The National Health Survey*, (260):1–161.

Blazer, D. G. (2003). Depression in Late Life: Review and Commentary. *The Journals of Gerontology: Series A*, 58(3):M249–M265.

Blazer, D. G. and Hybels, C. F. (2005). Origins of depression in later life. *Psychological Medicine*, 35(9):1241–1252.

Bo, M., Sona, A., Astengo, M., Fiandra, U., Quagliotti, E., Brescianini, A., and Fonte, G. (2009). Metabolic syndrome in older subjects: Coincidence or clustering? *Archives of Gerontology and Geriatrics*, 48(2):146–150.

Bollen, K. A. and Curran, P. J. (2006). *Latent curves models: a structural equation perspective.* John Wiley & Sons, Inc., 1st edition.

Borim, F. S. A., Barros, M. B. d. A., and Neri, A. L. (2012). Self-rated health in the elderly: a population-based study in Campinas, São Paulo, Brazil - Autoavaliação da saúde em idosos: pesquisa de base populacional no Município de Campinas, São Paulo, Brasil. *Cadernos de Saúde Pública*, 28(4):769–780.

Boss, L., Kang, D.-H., and Branson, S. (2015). Loneliness and cognitive function in the older adult: a systematic review. *International Psychogeriatrics*, 27(4):541–553.

Brasil (1974). Lei N^{o} 6.179, de 11 de Dezembro de 1974. Institui amparo previdenciário para maiores de setenta anos de idade e para inválidos, e dá outras providências.

Brasil (1975). Lei N^{o} 6.260, de 6 de Novembro de 1975. Institui benefícios de previdência e assistência social em favor dos empregadores rurais e seus dependentes, e dá outras providências.

Brasil (1977). Lei N^{o} 6.439, de 1^{o} de Setembro de 1977. Institui o sistema Nacional de Previdência e Assistência Social e dá outras providências.

Brasil (1988). Constituição da República Federativa do Brasil de 1988.

Brasil (1993). Lei N^{o} 8.742, de 7 de Dezembro de 1993. Dispõe sobre a organização da Assistência Social e dá outras providências.

Brasil (1994). Lei N^{o} 8.842, de 4 de Janeiro de 1994. Dispõe sobre a política nacional do idoso, cria o Conselho Nacional do Idoso e dá outras providências.

Brasil (1996). Decreto N^{o} 1.948, de 3 de Julho de 1996. Regulamenta a Lei n 8.842, de 4 de janeiro de 1994, que dispõe sobre a Política Nacional do Idoso, e dá outras providências.

Brasil (2003). Lei No 10.741, de 1^{o} de Outubro de 2003. Dispõe sobre o Estatuto do Idoso e dá outras providências.

Brasil. Ministério de Saúde. (1999). Portaria N^o 1.395, de 10 de Dezembro de 1999. a Política Nacional de Saúde do Idoso.

Brasil. Ministério de Saúde. (2006). Portaria N^o 2.528 de 19 de Outubro de 2006 - Política Nacional de Saúde da Pessoa Idosa.

Brasil. Ministério do Desenvolvimento Social e Combate à Fome Secretaria Nacional de Assistência Social. (2004). Resolução N^o 145, de 15 de Outubro de 2004 (dou 28/10/2004) - Política Nacional de Assistência Social.

Brenes-Camacho, G. (2011). Favourable changes in economic well-being and self-rated health among the elderly. *Social science & medicine (1982)*, 72(8):1228–35.

Cacciatore, F., Gaudiosi, C., Mazzella, F., Scognamiglio, A., Mattucci, I., Carone, M., Ferrara, N., and Abete, P. (2017). Pneumonia and hospitalizations in the elderly. *Geriatric Care*, 3(1).

Camargos, M. C. S., Rodrigues, R. d. N., and Machado, C. J. (2009). Healthy life expectancy in Brazilian elders, 2003, Expectativa de vida saudável para idosos brasileiros, 2003. *Ciência & Saúde Coletiva*, 14(5):1903–1909.

Care Act (2014). c. 23.

Carmel, S., Raveis, V. H., O'Rourke, N., and Tovel, H. (2017). Health, coping and subjective well-being: results of a longitudinal study of elderly Israelis. *Aging & Mental Health*, 21(6):616–623.

Casey, D. A. (2012). Depression in the elderly: A review and update. *Asia-Pacific Psychiatry*, 4(3):160–167.

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2):245–276.

Chang, K. P., Center, J. R., Nguyen, T. V., and Eisman, J. A. (2004). Incidence of Hip and Other Osteoporotic Fractures in Elderly Men and Women: Dubbo Osteoporosis Epidemiology Study. *Journal of Bone and Mineral Research*, 19(4):532–536.

Chang, S. S., Chen, S., McAvay, G. J., and Tinetti, M. E. (2012). Effect of Coexisting Chronic Obstructive Pulmonary Disease and Cognitive Impairment on Health Outcomes in Older Adults. *Journal of the American Geriatrics Society*, 60(10):1839–1846.

Chen, C.-M., Lee, I.-C., Su, Y.-Y., Mullan, J., and Chiu, H.-C. (2017). The longitudinal relationship between mental health disorders and chronic disease for older adults: a population-based study. *International Journal of Geriatric Psychiatry*, 32(9):1017–1026.

Chiavegatto Filho, A. D. P., Lebrão, M. L., and Kawachi, I. (2012). Income inequality and elderly self-rated health in São Paulo, Brazil. *Annals of epidemiology*, 22(12):863–7.

Christensen, R. H. B. (2015). ordinal—Regression Models for Ordinal Data.

Cole, M. G. and Dendukuri, N. (2003). Risk Factors for Depression Among Elderly Community Subjects: A Systematic Review and Meta-Analysis. *American Journal of Psychiatry*, 160(6):1147–1156.

Coletta, E. M. and Murphy, J. B. (1994). Physical and functional assessment of the elderly stroke patient. *American Family Physician*, 49(8):1777–1785.

Corriere, M., Rooparinesingh, N., and Kalyani, R. (2013). Epidemiology of Diabetes and Diabetes Complications in the Elderly: An Emerging Public Health Burden. *Current Diabetes Reports*, 13(6):805–813.

Craigs, C. L., Twiddy, M., Parker, S. G., and West, R. M. (2014). Understanding causal associations between self-rated health and personal relationships in older adults: A review of evidence from longitudinal studies. *Archives of gerontology and geriatrics*, 59(2):211–26.

Crewdson, J. A. (2016). The Effect of Loneliness in the Elderly Population: A Review. *Healthy Aging & Clinical Care in the Elderly*, 8(5439-HACCE-The-Effect-of-Loneliness-in-the-Elderly-Population:-A-Review.pdf):1–8.

Crisafulli, E., Morandi, A., Olivini, A., Malerba, M., and Clini, E. M. (2014). Rehabilitation and supportive therapy in elderly patients with chronic obstructive pulmonary disease. *European Journal of Internal Medicine*, 25(4):329–335.

Dahlberg, L. and McKee, K. J. (2014). Correlates of social and emotional loneliness in older people: evidence from an English community study. *Aging & Mental Health*, 18(4):504–514.

Dang, T. T., Majumdar, S. R., Marrie, T. J., and Eurich, D. T. (2015). Recurrent Pneumonia: A Review with Focus on Clinical Epidemiology and Modifiable Risk Factors in Elderly Patients. *Drugs & Aging*, 32(1):13–19.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., and Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*, 40(1):55–75.

de Albuquerque, M. V., d'Ávila Viana, A. L., de Lima, L. D., Ferreira, M. P., Fusaro, E. R., Iozzi, F. L., de Albuquerque, M. V., d'Ávila Viana, A. L., de Lima, L. D., Ferreira, M. P., Fusaro, E. R., and Iozzi, F. L. (2017). Desigualdades regionais na saúde: mudanças observadas no Brasil de 2000 a 2016. *Ciência & Saúde Coletiva*, 22(4):1055–1064.

de Guzman, A. B., Jurado, J. B. N., and Juson, A. J. A. (2015a). Examining the Structural Relationship of Chronic Illness, Physical Function, Life Satisfaction, and Social Support in the Development of Depression Among Filipino Elderly in Institutionalized Settings. *Educational Gerontology*, 41(3):193–206.

de Guzman, A. B., Lacampuenga, P. E. U., and Lagunsad, A. P. V. (2015b). Examining the Structural Relationship of Physical Activity, Cognition, Fear of Falling, and Mobility Limitation of Filipino in Nursing Homes. *Educational Gerontology*, 41(7):527–542.

de Moraes, J. R., Luiz, J. P. d. L. M., and Raggio, R. (2011). Associação entre o estado de saúde autorreferido de adultos e a área de localização do domicílio: uma análise de regressão logística ordinal usando a PNAD 2008 - Association between self-reported state of health among adults and the location of the home:. *Ciência & Saúde Coletiva*, 16(9):3769–3780.

de Rijk, M., Breteler, M., Graveland, G., Ott, A., Grobbee, D., van der Meche, F., and Hofman, A. (1995). Prevalence of Parkinson's disease in the elderly: The Rotterdam Study. *Neurology*, 45(12):2143–2146.

Deaton, A. (2007). Income, Aging, Health and Wellbeing Around the World: Evidence from the Gallup World Poll. Technical report, National Bureau of Economic Research, Cambridge, MA.

DeFronzo, R. A. and Ferrannini, E. (1991). Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. *Diabetes Care*, 14(3):173–194.

Delhom, I., Gutierrez, M., Mayordomo, T., and Melendez, J. C. (2017). Does Emotional Intelligence Predict Depressed Mood? A Structural Equation Model with Elderly People. *Journal of Happiness Studies*.

Department of Communities and Local Government (2015). English Indices of Deprivation 2015: File 10 Local Authority District Summaries.

DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., and Muntner, P. (2006). Mortality prediction with a single general self-rated health question. A meta-analysis. *Journal of general internal medicine*, 21(3):267–75.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55(1):34–43.

Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49(1):71.

DiTommaso, E., Brannen, C., and Best, L. A. (2004). Measurement and Validity Characteristics of the Short Version of the Social and Emotional Loneliness Scale for Adults. *Educational and Psychological Measurement*, 64(1):99–119.

DiTommaso, E. and Spinner, B. (1993). The development and initial validation of the Social and Emotional Loneliness Scale for Adults (SELSA). *Personality and Individual Differences*, 14(1):127–134.

Divgi, D. R. (1979). Calculation of the Tetrachoric Correlation-Coefficient. *Psychometrika*, 44(2):169–172.

Dontsov, A. V. and Vasilyeva, L. V. (2013). Gender features of metabolic syndrome in old age: Literature review. *Advances in Gerontology*, 3(4):309–313.

dos Santos, N. F. and do Rosário de Fátima e Silva, M. (2013). AS POLÍTICAS PÚBLICAS VOLTADAS AO IDOSO: MELHORIA DA QUALIDADE DE VIDA OU REPRIVATIZAÇÃO DA VELHICE - PUBLIC POLICIES DIRECTED TO THE ELDERLY: IMPROVED QUALITY OF LIFE OR PRIVATIZATION OF OLD AGE. *Revista FSA*, 10(2):pp. 358–371.

Douthitt, R. A., Macdonald, M., and Mullis, R. (1992). The relationship between measures of subjective and economic well-being: A new look. *Social Indicators Research*, 26(4):407–422.

Drennan, J., Treacy, M., Butler, M., Byrne, A., Fealy, G., Frazer, K., and Irving, K. (2008). The experience of social and emotional loneliness among older people in Ireland. *Ageing & Society*, 28(08):1113–1132.

Dugan, E. and Kivett, V. R. (1994). The Importance of Emotional and Social Isolation to Loneliness Among Very Old Rural Adults. *The Gerontologist*, 34(3):340–346.

Dujardin, C., Lorant, V., and Thomas, I. (2014). Self-assessed health of elderly people in Brussels: does the built environment matter? *Health & place*, 27:59–67.

Duncan, R., Francis, R. M., Collerton, J., Davies, K., Jagger, C., Kingston, A., Kirkwood, T., Robinson, L., and Birrell, F. (2011). Prevalence of arthritis and joint pain in the oldest old: findings from the Newcastle 85+ Study. *Age and Ageing*, 40(6):752–755.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., and Alpert, A. (1999). *An Introduction to Latent Variable Growth Curve Modeling*. Lawrence Erlbaum Associates, Inc., 1st edition.

Egol, K. A., Koval, K. J., and Zuckerman, J. D. (1997). Functional recovery following hip fracture in the elderly. *Journal of orthopaedic trauma*, 11(8):594–9.

Ekström, J. (2011). A Generalized Definition of the Polychoric Correlation Coefficient. *Department of Statistics, UCLA*.

Elder, T. E. (2013). The predictive validity of subjective mortality expectations: evidence from the Health and Retirement Study. *Demography*, 50(2):569–89.

Farooqui, T. and Farooqui, A. A. (2009). Aging: An important factor for the pathogenesis of neurodegenerative diseases. *Mechanisms of Ageing and Development*, 130(4):203–215.

Faverio, P., Aliberti, S., Bellelli, G., Suigo, G., Lonni, S., Pesci, A., and Restrepo, M. I. (2014). The management of community-acquired pneumonia in the elderly. *European Journal of Internal Medicine*, 25(4):312–319.

Fees, B. S., Martin, P., and Poon, L. W. (1999). A Model of Loneliness in Older Adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 54B(4):P231–P239.

Fernandes, M. T. D. O. and Soares, S. M. (2012). The development of public policies for elderly care in Brazil. *Revista da Escola de Enfermagem da U S P*, 46(6):1494–502.

Ferrer, E. and Ghisletta, P. (2011). Methodological and Analytical Issues in the Psychology of Aging. In Schaie, K. W. and Willis, S. L., editors, *Handbook of the Psychology of Aging*, chapter 2, pages 25–39. Elsevier, 7th edition.

Foulke, B. A., Kendal, A. R., Murray, D. W., and Pandit, H. (2016). Fracture healing in the elderly: A review. *Maturitas*, 92:49–55.

Franks, P., Gold, M. R., and Fiscella, K. (2003). Sociodemographics, self-rated health, and mortality in the US. *Social Science & Medicine*, 56(12):2505–2514.

Gabriel, R., Alonso, M., Reviriego, B., Muñiz, J., Vega, S., López, I., Novella, B., Suárez, C., and Rodríguez-Salvanés, F. (2009). Ten-year fatal and non-fatal myocardial infarction incidence in elderly populations in Spain: the EPICARDIAN cohort study. *BMC Public Health*, 9(1):360.

Gheno, R., Cepparo, J. M., Rosca, C. E., and Cotten, A. (2012). Musculoskeletal Disorders in the Elderly. *Journal of Clinical Imaging Science*, 2:39.

Giltay, E. J., Vollaard, A. M., and Kromhout, D. (2011). Self-rated health and physician-rated health as independent predictors of mortality in elderly men. *Age and Ageing*, 41(2):165–171.

Given, B. and Given, C. W. (2008). Older Adults and Cancer Treatment. *Cancer*, 113(12 Suppl):3505–3511.

Goldstein, H., Browne, W., and Rasbash, J. (2002). Partitioning Variation in Multilevel Models. *Understanding Statistics*, 1(4):223–231.

Golini, N. and Egidi, V. (2016). The Latent Dimensions of Poor Self-Rated Health: How Chronic Diseases, Functional and Emotional Dimensions Interact Influencing Self-Rated Health in Italian Elderly. *Social Indicators Research*, 128(1):321–339.

Government Office for Science (2016). Future of an Ageing Population. *FORESIGHT (Government Office for Science, UK).*

Grilli, L. and Rampichini, C. (2007). Structural Equation Modeling: A Multilevel Factor Models for Ordinal Variables Multilevel Factor Models for Ordinal Variables. *A Multidisciplinary Journal*, 14(1):37–41.

Gruber, R., Koch, H., Doll, B. A., Tegtmeier, F., Einhorn, T. A., and Hollinger, J. O. (2006). Fracture healing in the elderly patient. *Experimental Gerontology*, 41(11):1080–1093.

Haaksma, M. L., Vilela, L. R., Marengoni, A., Calderón-Larrañaga, A., Leoutsakos, J.-M. S., Olde Rikkert, M. G. M., and Melis, R. J. F. (2017). Comorbidity and progression of late onset Alzheimer's disease: A systematic review. *Plos One*, 12(5):e0177044–e0177044.

Halbert, R. J., Natoli, J. L., Gano, A., Badamgarav, E., Buist, A. S., and Mannino, D. M. (2006). Global burden of COPD: systematic review and meta-analysis. *European Respiratory Journal*, 28(3):523 LP – 532.

Hallquist, M. and Wiley, J. (2016). *MplusAutomation: Automating Mplus Model Estimation and Interpretation.*

Han, L., Allore, H., Murphy, T., Gill, T., Peduzzi, P., and Lin, H. (2013). Dynamics of functional aging based on latent-class trajectories of activities of daily living. *Annals of Epidemiology*, 23(2):87–92.

Hansen, T., Slagsvold, B., and Moum, T. (2008). Financial Satisfaction in Old Age: A Satisfaction Paradox or a Result of Accumulated Wealth?

Haq, I., Murphy, E., and Dacre, J. (2003). Osteoarthritis. *Postgraduate Medical Journal*, 79(933):377–383.

He, Y., Jiang, B., Wang, J., Feng, K., Chang, Q., Fan, L., Li, X., and Hu, F. B. (2006). Prevalence of the Metabolic Syndrome and its Relation to Cardiovascular Disease in an Elderly Chinese Population. *Journal of the American College of Cardiology*, 47(8):1588–1594.

Health and Safety Executive UK (2017). Controlling physical health risks - Managing occupational ill health risks in construction.

Hedges, L. V. and Hedberg, E. C. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis*, 29(1):60–87.

Helmer, C., BarbergerGateau, P., and Dartigues, J. F. (1999). Subjective health and mortality in French elderly women and men. *J.Gerontol.Ser.B Psychol.Sci.*, 54(2):S84–S92.

Henriksen, J. H. and Moller, S. (2006). Liver cirrhosis and arterial hypertension. *World Journal of Gastroenterology : WJG*, 12(5):678–685.

Higgs, P., Hyde, M., Wiggins, R., and Blane, D. (2003). Researching Quality of Life in Early Old Age: The Importance of the Sociological Dimension. *Social Policy & Administration*, 37(3):239–252.

Hindle, J. V. (2010). Ageing, neurodegeneration and Parkinson's disease. *Age and Ageing*, 39(2):156–161.

Hirve, S., Oud, J. H. L., Sambhudas, S., Juvekar, S., Blomstedt, Y., Tollman, S., Wall, S., and Ng, N. (2014). Unpacking Self-Rated Health and Quality of Life in Older Adults and Elderly in India: A Structural Equation Modelling Approach. *Social Indicators Research*, 117(1):105–119.

Höfelmann, D. A., Garcia, L. P., and de Freitas, L. R. S. (2014). Self-rated health in Brazilian adults and elderly: Data from the National Household Sample Survey 2008 .

Hollander, M., Koudstaal, P. J., Bots, M. L., Grobbee, D. E., Hofman, A., and Breteler, M. M. B. (2003). Incidence, risk, and case fatality of first ever stroke in the elderly population. The Rotterdam Study.

Holmén, K., Ericsson, K., Andersson, L., and Winblad, B. (1992). Loneliness among elderly people living in Stockholm: a population study. *Journal of Advanced Nursing*, 17(1):43–51.

Holmén, K., Ericsson, K., and Winblad, B. (2000). Social and emotional loneliness among non-demented and demented elderly people. *Archives of Gerontology and Geriatrics*, 31(3):177–192.

Holroyd, C., Cooper, C., and Dennison, E. (2008). Epidemiology of osteoporosis. *Best Practice & Research Clinical Endocrinology & Metabolism*, 22(5):671–685.

Hooper, D., Coughlan, J., and Mullen, M. (2008). Structural Equation Modelling : Guidelines for Determining Model Fit Structural equation modelling : guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1):53–60.

Hu, L. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1):1–55.

Huang, P. L. (2009). A comprehensive definition for metabolic syndrome. *Disease Models & Mechanisms*, 2(5-6):231–237.

Hung, C.-W., Chen, Y.-C., Hsieh, W.-L., Chiou, S.-H., and Kao, C.-L. (2010). Ageing and neurodegenerative diseases. *Ageing Research Reviews*, 9(SUPPL.):S36–S46.

Hyde, M., Wiggins, R. D., Higgs, P., and Blane, D. B. (2003). A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19). *Aging & Mental Health*, 7(3):186–194.

Idler, E. L. and Benyamini, Y. (1997). Self-Rated Health and Mortality : A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior*, 38(1):21–37.

Iecovich, E. and Biderman, A. (2012). Attendance in adult day care centers and its relation to loneliness among frail older adults. *International Psychogeriatrics*, 24(3):439–448.

Instituto Brasileiro de Geografia e Estatística (IBGE) (1999). Pesquisa sobre Padrões de Vidas 1996-1997. Technical report, Rio de Janeiro.

Instituto Brasileiro de Geografia e Estatística (IBGE) (2008). Pesquisa Nacional por Amostra de Domicílios 2008.

Instituto Brasileiro de Geografia e Estatística (IBGE) (2010). *Síntese de Indicadores Sociais uma Amálise das Condições de Vida da População Brasileira.*

Jackson, M. L., Neuzil, K. M., Thompson, W. W., Shay, D. K., Yu, O., Hanson, C. A., and Jackson, L. A. (2004). The Burden of Community-Acquired Pneumonia in Seniors: Results of a Population-Based Study. *Clinical Infectious Diseases*, 39(11):1642–1650.

Janssens, J.-P. and Krause, K.-H. (2004). Pneumonia in the very old. *The Lancet Infectious Diseases*, 4(2):112–124.

Johansson, P., Riegel, B., Svensson, E., Broström, A., Alehagen, U., Dahlström, U., and Jaarsma, T. (2014). Sickness Behavior in Community-Dwelling Elderly: Associations With Impaired Cardiac Function and Inflammation. *Biological Research For Nursing*, 16(1):105–113.

Johnell, O. and Kanis, J. a. (2006). An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporosis International*, 17(12):1726–1733.

Jones, L. D., Pandit, H., and Lavy, C. (2014). Back pain in the elderly: A review. *Maturitas*, 78(4):258–262.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1):141–151.

Kanis, J. A., Johnell, O., Oden, A., Jonsson, B., De Laet, C., and Dawson, A. (2000). Risk of hip fracture according to the World Health Organization criteria for osteopenia and osteoporosis. *Bone*, 27(5):585–590.

Kaplan, V., Angus, D. C., Griffin, M. F., Clermont, G., Scott Watson, R., and Linde-Zwirble, W. T. (2002). Hospitalized community-acquired pneumonia in the elderly. Age-and sex-related patterns of care and outcome in the United States. *American Journal of Respiratory and Critical Care Medicine*, 165(6):766–772.

Karim, J., Weisz, R., Bibi, Z., and ur Rehman, S. (2014). Validation of the Eight-Item Center for Epidemiologic Studies Depression Scale (CES-D) Among Older Adults. *Current Psychology*.

Kee, C. C. and Epps, C. D. (2001). Pain Management Practices of Nurses Caring for Older Patients with Osteoarthritis. *Western Journal of Nursing Research*, 23(2):195–210.

Kimura, D., Nakatani, K., Takeda, T., Fujita, T., Sunahara, N., Inoue, K., and Notoya, M. (2015). Analysis of causal relationships by structural equation modeling to determine the factors influencing cognitive function in elderly people in Japan. *PLoS ONE*, 10(2):1–13.

Kirkman, M. S., Briscoe, V. J., Clark, N., Florez, H., Haas, L. B., Halter, J. B., Huang, E. S., Korytkowski, M. T., Munshi, M. N., Odegard, P. S., Pratley, R. E., and Swift, C. S. (2012). Diabetes in older adults: a consensus report. *Journal of the American Geriatrics Society*, 60(12):2342–56.

Kitch, B. T., Levy, B. D., and Fanta, C. H. (2000). Late onset asthma: epidemiology, diagnosis and treatment. *Drugs & Aging*, 17(5):385–397.

Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling.* Guilford Press, 3rd edition.

Koenig, H. G. and Studenski, S. (1988). Post-stroke depression in the elderly. *Journal of General Internal Medicine*, 3(5):508–517.

Kukull, W. A., Higdon, R., Bowen, J. D., McCormick, W. C., Teri, L., Schellenberg, G. D., van Belle, G., Jolley, L., and Larson, E. B. (2002). Dementia and Alzheimer disease incidence: a prospective cohort study. *Archives Of Neurology*, 59(11):1737–1746.

Landeiro, F., Barrows, P., Nuttall Musson, E., Gray, A. M., and Leal, J. (2017). Reducing social isolation and loneliness in older people: a systematic review protocol. *BMJ Open*, 7(5).

LaRue, A., Bank, L., Jarvik, U., and Hetland, M. (1979). Health in Old Age: How Do Physicians' Ratings and Self-ratings Compare ? *Journal of Gerontology*, 34(5):687–691.

Lechleitner, M. (2008). Obesity and the metabolic syndrome in the elderly - A mini-review. *Gerontology*, 54(5):253–259.

Lee, J., Choi, M., Jung, D., Sohn, Y. H., and Hong, J. (2015). A Structural Model of Health-Related Quality of Life in Parkinson's Disease Patients. *Western Journal of Nursing Research*, 37(8):1062–1080.

Lee, S.-Y., Poon, W.-Y., and Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, 57(1):89–105.

Lima, F. M., Hyde, M., Chungkham, H. S., Correia, C., Campos, A. S., Campos, M., Novaes, M., Laks, J., and Petribu, K. (2014). Quality of Life amongst Older Brazilians: A Cross-Cultural Validation of the CASP-19 into Brazilian-Portuguese. *PLoS ONE*, 9(4):1–8.

Lima-Costa, M. F., Steptoe, A., Cesar, C. C., De Oliveira, C., Proietti, F. a., and Marmot, M. (2012). The influence of socioeconomic status on the predictive power of self-rated health for 6-year mortality in English and Brazilian older adults: The ELSA and Bambui cohort studies. *Annals of Epidemiology*, 22(9):644–648.

Lindblad, U. (2001). Metabolic Syndrome and Ischemic Heart Disease in Elderly Men and Women. *American Journal of Epidemiology*, 153(5):481–489.

Liu, L., Gou, Z., and Zuo, J. (2016). Social support mediates loneliness and depression in elderly people. *Journal of Health Psychology*, 21(5):750–758.

Lord, F. M. and Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, Mass., Addison-Wesley Pub. Co.

Lu, C., Miclau, T., Hu, D., Hansen, E., Tsui, K., Puttlitz, C., and Marcucio, R. S. (2005). Cellular basis for age-related changes in fracture repair. *Journal of Orthopaedic Research*, 23(6):1300–1307.

Luanaigh, C. Ó. and Lawlor, B. A. (2008). Loneliness and the health of older people. *International Journal of Geriatric Psychiatry*, 23(12):1213–1221.

Maggi, S., Noale, M., Gallina, P., Bianchi, D., Marzari, C., Limongi, F., and Crepaldi, G. (2006). Metabolic Syndrome, Diabetes, and Cardiovascular Disease in an Elderly Caucasian Cohort:

The Italian Longitudinal Study on Aging. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 61(5):505–510.

Mai, L. L., Young Owl, M., and Karsting, P. (2005). musculoskeletal disorder. In *Cambridge Dictionary of Human Biology and Evolution*, number Cambridge, UK. Cambridge University Press.

Marmot, M., Oldfield, Z., Clemens, S., Blake, M., Phelps, A., Nazroo, J., Steptoe, A., Rogers, N., Banks, J., and Oskala, A. (2016). English Longitudinal Study of Ageing: Waves 0-7, 1998-2015. [data collection] (Usage Number 102242).

Marosi, C. and Köller, M. (2016). Challenge of cancer in the elderly. *ESMO Open*, 1(3).

Masumori, N., Homma, D., and Tsukamoto, T. (2005). Web-based research of lower urinary tract symptoms that affect quality of life in elderly Japanese men: analysis using a structural equation model. *BJU International*, 95(7):1013–1022.

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, 71(1):57–77.

McKay, A. J., Mahesh, P. A., Fordham, J. Z., and Majeed, A. (2012). Prevalence of COPD in India: a systematic review. *Primary Care Respiratory Journal: Journal Of The General Practice Airways Group*, 21(3):313–321.

Mehta, R. H., Rathore, S. S., Radford, M. J., Wang, Y., Wang, Y., and Krumholz, H. M. (2001). Acute myocardial infarction in the elderly: differences by age. *Journal of the American College of Cardiology*, 38(3):736–741.

Meléndez, J. C., Tomás, J. M., Oliver, A., and Navarro, E. (2009). Psychological and physical dimensions explaining life satisfaction among the elderly: A structural model examination. *Archives of Gerontology and Geriatrics*, 48(3):291–295.

Menec, V. H., Chipperfield, J. G., and Perry, R. P. (1999). Self-Perceptions of Health: A Prospective Analysis of Mortality, Control, and Health. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 54B(2):P85–P93.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4):525–543.

Millsap, R. E. and Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research*, 39(3):479–515.

Ministério da Saúde Departamento de Atenção Básica (2017). Estratégia Saúde da Família [Health Strategy of the Family].

Miranda, V. S., DeCarvalho, V. B. F., Machado, L. A. C., and Dias, J. M. D. J. M. (2012). Prevalence of chronic musculoskeletal disorders in elderly Brazilians: a systematic review of the literature. *BMC Musculoskeletal Disorders*, 13(1):82.

Moore, S. L. and Rafii, M. (2001). Imaging of Musculoskeletal and Spinal Tuberculosis. *Radiologic Clinics of North America*, 39(2):329–342.

Mullis, R. J. (1992). Measures of economic well-being as predictors of psychological well-being. *Social Indicators Research*, 26(2):119–135.

Muniz-Terrera, G., Matthews, F., Dening, T., Huppert, F. A., and Brayne, C. (2009). Education and trajectories of cognitive decline over 9 years in very old people: methods and risk analysis. *Age and Ageing*, 38(3):277–282.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1):115–132.

Muthén, B., Asparouhov, T., Hunter, A. M., and Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: alternative analyses of the STAR*D antidepressant trial. *Psychological methods*, 16(1):17–33.

Muthén, B. O. (1993). Fit With Categorical and Other Nonnormal Variables. In Bollen, K. A. and Long, J. S., editors, *Testing Structural Equation Models*, chapter 9, pages 205–234. SAGE Publications.

Muthén, B. O. (1994). Multilevel Covariance Structure Analysis. *Sociological Methods & Research*, 22(3):376–398.

Muthén, B. O. (1998). (1998-2004) Mplus Technical Appendices. Los Angeles, CA: Muthén & Muthén.

Muthén, B. O. and Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Note No. 4. Technical report.

Muthén, B. O., du Toit, S. H., and Spisic, D. (1997). Robust Inference using Weighted Least Squares and Quadratic Estimating Equations in Latent Variable Modeling with Categorical and Continuous Outcomes.

Muthén, L. K. and Muthén, B. O. (1998). Mplus User's Guide. Sixth Edition. (1998-2011).

NIH Consensus Development Conference (1992). Diagnosis and treatment of depression in late life. *JAMA*, 268(8):1018–1024.

Nikolaou, V. S., Efstathopoulos, N., Kontakis, G., Kanakaris, N. K., and Giannoudis, P. V. (2009). The influence of osteoporosis in femoral fracture healing time. *Injury*, 40(6):663–668.

Nishi, A., Kawachi, I., Shirai, K., Hirai, H., Jeong, S., and Kondo, K. (2012). Sex/gender and socioeconomic differences in the predictive ability of self-rated health for mortality. *PloS one*, 7(1):e30179.

Nitrini, R., Bottino, C. M. C., Albala, C., Custodio Capuñay, N. S., Ketzoian, C., Llibre Rodriguez, J. J., Maestre, G. E., Ramos-Cerqueira, A. T. A., and Caramelli, P. (2009). Prevalence of dementia in Latin America: a collaborative study of population-based cohorts. *International Psychogeriatrics*, 21(4):622–630.

Nitrini, R., Caramelli, P., Herrera, E., de Castro, I., Bahia, V. S., Anghinah, R., Caixeta, L. F., Radanovic, M., Charchat-Fichman, H., Porto, C. S., Teresa Carthery, M., Hartmann, A. P. J., Huang, N., Smid, J., Lima, E. P., Takahashi, D. Y., and Takada, L. T. (2005). Mortality from dementia in a community-dwelling Brazilian population. *International Journal of Geriatric Psychiatry*, 20(3):247–253.

Nummela, O., Raivio, R., and Uutela, A. (2012). Trust, self-rated health and mortality: a longitudinal study among ageing people in Southern Finland. *Social science & medicine (1982)*, 74(10):1639–43.

Nummela, O., Seppänen, M., and Uutela, A. (2011). The effect of loneliness and change in loneliness on self-rated health (SRH): a longitudinal study among aging people. *Archives of gerontology and geriatrics*, 53(2):163–7.

Ocampo, J. M. (2010). Self-rated health: Importance of use in elderly adults. *Colombia Médica*.

Office for National Statistics (2015). Statistical bulletin: Cancer Registration Statistics, England: 2015 Cancer diagnoses and age-standardised incidence rates for all cancer sites by age, sex and region.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460.

Pandey, M. K. and Jha, A. K. (2012). Widowhood and health of elderly in India: examining the role of economic factors using structural equation modeling. *International Review of Applied Economics*, 26(1):111–124.

Parkinson's UK (2009). Parkinson's prevalence in the United Kingdom (2009). Technical report, Parkinson's UK.

Pavot, W. and Diener, E. (2008). The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, 3(2):137–152.

Pavot, W., Diener, E., and Suh, E. (1998). The Temporal Satisfaction With Life Scale. *Journal of Personality Assessment*, 70(2):340–354.

Pensions Act (2014). c. 19.

Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., and Rasbash, J. (1998). Weighting for Unequal Selection Probabilities in Multilevel Models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(1):23–40.

Pirani, E. and Salvini, S. (2012). Socioeconomic Inequalities and Self-Rated Health: A Multilevel Study of Italian Elderly. *Population Research and Policy Review*, 31(1):97–117.

Pluhar, J. (2014). Metabolic Syndrome - Lifestyle Changes Are Key. In Harvard Health Publications, editor, *Harvard Medical School Commentaries on Health*, number Boston, MA, USA. Harvard Health Publications.

Pongiglione, B., Ploubidis, G. B., and De Stavola, B. L. (2017). Levels of disability in the older population of England: Comparing binary and ordinal classifications. *Disability and Health Journal*.

Preacher, K. J., Zyphur, M. J., and Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological methods*, 15(3):209–233.

Prince, M., Kapp, M., Guerchet, M., McCrone, P., Prina, M., Comas-Herrera, A., Wittenberg, R., Adelaja, B., Hu, B., King, D., Rehill, A., and Salimkumar, D. (2014). Dementia UK Update. Technical report.

Public Health England (2017). Inequality in health. In *Research and analysis*, chapter 5.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3):385–401.

Rajabally, Y. A. and Martey, J. (2011). Neuropathy in Parkinson disease: Prevalence and determinants. *Neurology*, 77(22):1947–1950.

Reckase, M. (2009). *Multidimensional Item Response Theory*. Springer New York, New York, NY.

Reed, C. E. (2010). Asthma in the elderly: Diagnosis and management. *Journal of Allergy and Clinical Immunology*, 126(4):681–687.

Russell, D., Peplau, L. A., and Cutrona, C. E. (1980). The revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*, 39(3):472–480.

Russell, D., Peplau, L. A., and Ferguson, M. L. (1978). Developing a Measure of Loneliness. *Journal of Personality Assessment*, 42(3):290–294.

Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure. *Journal of Personality Assessment*, 66(1):20.

Rzewuska, M., Mallen, C. D., Strauss, V. Y., Belcher, J., and Peat, G. (2015). One-year trajectories of depression and anxiety symptoms in older patients presenting in general practice with musculoskeletal pain: A latent class growth analysis. *Journal of Psychosomatic Research*, 79(3):195–201.

Sabbagh, M. N., Lahti, T., Connor, D. J., Caviness, J. N., Shill, H., Vedders, L., Mahant, P., Samanta, J., Burns, R. S., Evidente, V. G. H., Driver-Dunckley, E., Reisberg, B., Bircea, S., and Adler, C. H. (2007). Functional ability correlates with cognitive impairment in Parkinson's disease and Alzheimer's disease. *Dementia And Geriatric Cognitive Disorders*, 24(5):327–334.

Salter, D. M. (2002). Degenerative joint disease. *Current Diagnostic Pathology*, 8(1):11–18.

Sánchez Martínez, M., Blanco, A., Castell, M. V., Gutiérrez Misis, A., González Montalvo, J. I., Zunzunegui, M. V., and Otero, Á. (2014). Diabetes in older people: Prevalence, incidence and its association with medium- and long-term mortality from all causes. *Atención Primaria*, 46(7):376–384.

Santos, L. M. P., Oliveira, A., Trindade, J. S., Barreto, I. C., Palmeira, P. A., Comes, Y., Santos, F. O., Santos, W., Oliveira, J. P. A., Pessoa, V. M., and Shimizu, H. E. (2017). Implementation research: towards universal health coverage with more doctors in Brazil. *Bulletin of the World Health Organization*, 95(2):103–112.

Schulz, R. (2006). Activities of Daily Living. In *The Encyclopedia of Aging*, number New York, NY, USA. Springer Publishing Company.

Schumann, B., Seidler, A., Kluttig, A., Werdan, K., Haerting, J., and Greiser, K. H. (2011). Association of occupation with prevalent hypertension in an elderly East German population: an exploratory cross-sectional analysis. *International Archives of Occupational and Environmental Health*, 84(4):361–369.

Scichilone, N., Pedone, C., Battaglia, S., Sorino, C., and Bellia, V. (2014). Diagnosis and management of asthma in the elderly. *European Journal of Internal Medicine*, 25(4):336–342.

Sexton, E., King-Kallimanis, B. L., Conroy, R. M., and Hickey, A. (2013). Psychometric evaluation of the CASP-19 quality of life scale in an older Irish cohort. *Quality of Life Research*, 22(9):2549–2559.

Sheikh, J. I. and Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. In Brink, T. L., editor, *Clnical Gerontology: A Guide to Assessment and Intervention*, pages 165–173. New York, Hawton Press.

Shen, C., Schooling, C. M., Chan, W. M., Zhou, J. X., Johnston, J. M., Lee, S. Y., and Lam, T. H. (2014). Self-rated health and mortality in a prospective Chinese elderly cohort study in Hong Kong. *Preventive medicine*, 67:112–8.

Sim, J., Bartlam, B., and Bernard, M. (2011). The CASP-19 as a measure of quality of life in old age: evaluation of its use in a retirement community. *Quality of Life Research*, 20(7):997–1004.

Simonetti, A. F., Viasus, D., Garcia-Vidal, C., and Carratalà, J. (2014). Management of community-acquired pneumonia in older adults. *Therapeutic Advances in Infectious Disease*, 2(1):3–16.

Smith, T. O., Purdy, R., Latham, S. K., Kingsbury, S. R., Mulley, G., and Conaghan, P. G. (2016). The prevalence, impact and management of musculoskeletal disorders in older people living in care homes: a systematic review. *Rheumatology International*, 36(1):55–64.

Stupka, E. and DeShazo, R. (2009). Asthma in Seniors: Part 1. Evidence for Underdiagnosis, Undertreatment, and Increasing Morbidity and Mortality. *The American Journal of Medicine*, 122(1):6–11.

Stupka, J. E., Mortensen, E. M., Anzueto, A., and Restrepo, M. I. (2009). Community-acquired pneumonia in elderly patients. *Aging health*, 5(6):763–774.

Tabachnick, B. G. (2001). *Using multivariate statistics.* Allyn and Bacon, Boston, Mass., 4th ed. edition.

Taffet, G. E., Donohue, J. F., and Altman, P. R. (2014). Considerations for managing chronic obstructive pulmonary disease in the elderly. *Clinical Interventions in Aging*, 9:23–30.

Tian, Q. (2016). Intergeneration social support affects the subjective well-being of the elderly: Mediator roles of self-esteem and loneliness. *Journal of Health Psychology*, 21(6):1137–1144.

Titman, A. C., Lancaster, G. A., and Colver, A. F. (2016). Item response theory and structural equation modelling for ordinal data: Describing the relationship between KIDSCREEN and Life-H. *Statistical Methods in Medical Research*, 25(5):1892–1924.

Torres, A., Peetermans, W. E., Viegi, G., and Blasi, F. (2013). Risk factors for community-acquired pneumonia in adults in Europe: a literature review. *Thorax*, 68(11):1057 LP – 1065.

Trade Economics (2017). Brazil Minimum Monthly Wages.

Tran, T. V., Wright, R., and Chatters, L. (1991). Health, stress, psychological resources, and subjective well-being among older Blacks. *Psychology and Aging*, 6(1):100–108.

Tucker, L. R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1):1–10.

Tuomilehto, J., Nissinen, A., Kivelä, S.-L., Pekkanen, J., Kaarsalo, E., Wolf, E., Aro, A., Punsar, S., and Karvonens, M. J. (1986). Prevalence of diabetes mellitus in elderly men aged 65 to 84 years in eastern and western Finland. *Diabetologia*, 29(9):611–615.

Turvey, C. L., Wallace, R. B., and Herzog, R. (1999). A Revised CES-D Measure of Depressive Symptoms and a DSM-Based Measure of Major Depressive Episodes in the Elderly. *International Psychogeriatrics*, 11(2):139–148.

United Nations: Department of Economic and Social Affairs. Population Division (2013). World Population Prospects: The 2012 Revision, DVD Edition.

van 't Land, H., Verdurmen, J., Ten Have, M., van Dorsselaer, S., Beekman, A., and de Graaf, R. (2010). The association between arthritis and psychiatric disorders; results from a longitudinal population-based study. *Journal of Psychosomatic Research*, 68(2):187–193.

Vink, D., Aartsen, M. J., and Schoevers, R. A. (2008). Risk factors for anxiety and depression in the elderly: A review. *Journal of Affective Disorders*, 106(1):29–44.

Vogelsang, E. M. (2014). Self-rated health changes and oldest-old mortality. *The journals of gerontology. Series B, Psychological sciences and social sciences*, 69(4):612–21.

Wang, J. and Wang, X. (2013). *Structural equation modeling: applications using Mplus*. Wiley.

Wang, X. (2014). Subjective well-being associated with size of social network and social support of elderly. *Journal of Health Psychology*, 21(6):1037–1042.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070.

Weiner, D. K., Sakamoto, S., Perera, S., and Breuer, P. (2006). Chronic Low Back Pain in Older Adults: Prevalence, Reliability, and Validity of Physical Examination Findings. *Journal of the American Geriatrics Society*, 54(1):11–20.

Wiggins, R. D., Netuveli, G., Hyde, M., Higgs, P., and Blane, D. (2008). The Evaluation of a Self-enumerated Scale of Quality of Life (CASP-19) in the Context of Research on Ageing: A Combination of Exploratory and Confirmatory Approaches. *Social Indicators Research*, 89(1):61–77.

Wilson-Genderson, M. and Pruchno, R. (2015). Functional Limitations and Gender Differences. *The International Journal of Aging and Human Development*, 81(1-2):83–100.

Woolf, A. D. and Pfleger, B. (2003). Burden of major musculoskeletal conditions. *Bulletin Of The World Health Organization*, 81(9):646–656.

World Health Organisation [WHO] (2008). WHO — Flawed but fair: Brazil's health system reaches out to the poor. *WHO*, 28(4):241–320.

World Health Organization [WHO] (2013). Cardiovascular diseases (CVDs) Fact sheet N317.

World Health Organization [WHO] (2014). What is "active ageing"?

World Health Organization [WHO] (2016). Global Tuberculosis Report. Technical report.

Yan, B., Gao, X., and Lyon, M. (2014). Modeling satisfaction amongst the elderly in different Chinese urban neighborhoods. *Social Science & Medicine*, 118(C):127–134.

Yáñez, A., Cho, S.-H., Soriano, J. B., Rosenwasser, L. J., Rodrigo, G. J., Rabe, K. F., Peters, S., Niimi, A., Ledford, D. K., Katial, R., Fabbri, L. M., Celedón, J. C., Canonica, G. W., Busse, P., Boulet, L.-P., Baena-Cagnani, C. E., Hamid, Q., Bachert, C., Pawankar, R., and Holgate, S. T. (2014). Asthma in the elderly: what we know and what we have yet to know. *World Allergy Organization Journal*, 7(1):8.

Yang, L., Martikainen, P., Silventoinen, K., and Konttinen, H. (2016). Association of socioeconomic status and cognitive functioning change among elderly Chinese people. *Age and Ageing*, 45(5):674–680.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., and Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*, 17(1):37–49.

Zein, J. G. and Erzurum, S. C. (2015). Asthma in the Elderly. *Current Geriatrics Reports*, 4(2):174–182.

Zhang, Y. (2002). Prevalence of Symptomatic Hand Osteoarthritis and Its Impact on Functional Status among the Elderly: The Framingham Study. *American Journal of Epidemiology*, 156(11):1021–1027.

Zheng, H. and Thomas, P. A. (2013). Marital status, self-rated health, and mortality: overestimation of health or diminishing protection of marriage? *Journal of health and social behavior*, 54(1):128–43.

Appendix A

PNAD, 2008

Mplus Input Code

The Mplus input code for all structural equation modelling (SEM) and multilevel structural equation modelling (MLSEM) of the 2008 PNAD are given, within this appendix, to enable replication of analyses. The code is presented in the same order that the results are presented in Chapter 5. Appendices A.1 to A.3 present the analysis of the total covariance matrices, including the exploratory factor analysis (EFA) of the variables within each aspect and the confirmatory factor analysis (CFA) of the measurement models for each domain, then finally, in Appendix A.3, the SEMs for the interrelationship of the domains and the effect of covariate are given. This is then repeated for the equivalent multilevel structural equation analyses in Appendices A.4 to A.6.

In all analyses the data file `pnad.dat` was used which contains all the identification variables, dependent variables and covariates. Each input code therefore began with the same preceding lines, which declared the names of all the variables in the data and the expression used to indicate a missing value;

```
DATA: FILE = "pnad.dat";
VARIABLE:
  NAMES = hhid secid munid fs1-fs6 srm1-srm12 hq1-hq5 dur1-dur9
    age_60 age_65 age_70 age_75 age_80 age_85 age_85p
    male nonwhite rural
    region_ne region_n region_cw region_s region_se;
  MISSING=.;
```

A.1 Exploratory Factor Analysis

The first exploratory factor analysis, in Section 5.2, was of the 12 binary self-reported morbidity variables. These corresponded to the diagnosis of back or column disease, arthritis or rheumatism, cancer, diabetes, asthma or bronchitis, hypertension, heart disease, chronic renal failure, depression, tuberculosis, tendonitis or tenosynovitis, and cirrhosis respectively. The variables were named `srm1-srm12` respectively, and each had the possible values $1, 2$. To provide the intuitive interpretation of greater values as better health reverse scoring was used, such that a value of 1 indicated the comorbidity was diagnosed, *i.e.* poorer health. The variables are then modelled incrementing the number of factors from one to five. The estimator requested was the weighted least squares mean and variance adjusted (WLSMV) estimator, which was used in analyses. The input code for this was:

```
  USEVARIABLES = srm1-srm12;
  CATEGORICAL = srm1-srm12;
ANALYSIS: TYPE = EFA 1 5;
  ESTIMATOR = WLSMV;
```

Followed by the EFA of the ordinal functional status variables, named `fs1-fs6`. The first variable, `fs1`, had the possible values $1, 2, 3$ and 4 representing the four categorical responses 'can not', 'moderate limitation', 'mild limitation' and 'no limitation' respectively. From the PNAD a value of 1 or 2 on `fs1` meant the respondent was ineligible to respond to the later items, `fs2-fs6`. The assumption was that functional status was poorer for these respondents so an additional fifth category was created for `fs2-fs6` that represented that indicated this poorer functional status in these variables. The new category took the value 1, thus the original four categories ('can not', 'moderate limitation', 'mild limitation' and 'no limitation') had the values $2, 3, 4$ and 5 respectively.

In the EFA, the number of factors was incremented from one through to three factors.

```
  USEVARIABLES = fs1-fs6;
  CATEGORICAL = fs1-fs6;
ANALYSIS: TYPE = EFA 1 3;
  ESTIMATOR = WLSMV;
```

The EFA of the housing quality variables contained the five binary variables named `hq1-hq5`, where each variable indicated adequate status of; the construction material of the walls, the construction material of the roof, the water supply, the system of sewage management and the system for disposal of rubbish respectively. The number of factors requested in the EFA was one and two factors:

```
  USEVARIABLES = hq1-hq5;
  CATEGORICAL = hq1-hq5;
ANALYSIS: TYPE = EFA 1 2;
  ESTIMATOR = WLSMV;
```

The variables `dur1-dur9` indicated the ownership of nine durables including a stove, fridge, television, radio, landline telephone, water dispenser, freezer, washing machine or computer within the household. The number of factors in the EFA was incremented from one to four factors:

```
  USEVARIABLES = dur1-dur9;
  CATEGORICAL = dur1-dur9;
ANALYSIS: TYPE = EFA 1 4;
  ESTIMATOR = WLSMV;
```

A.2 Confirmatory Factor Analysis

Confirmatory factor analysis for the health measurement model in Section 5.2.1 was composed of the self-reported morbidity variables `srm1-srm12` and functional status variables `fs1-fs6`. The factors were `msh` (musculoskeletal health), `mch` (metabolic and cardiovascular health), `ich` (infectious and chronic health), `fs` (functional status) and `hf` (overall health). The model was defined in the `MODEL` command, with the factor on the left-hand side of the `BY` statement and the respective dependent variables on the right-hand side. The code for the health measurement model was:

```
  USEVARIABLES = fs1-fs6 srm1-srm12;
  CATEGORICAL = fs1-fs6 srm1-srm12;
ANALYSIS: ESTIMATOR = WLSMV;
MODEL:
  msh BY srm1 srm2 srm11;
  mch BY srm4 srm6 srm7;
  ich BY srm3 srm5 srm8 srm9 srm10 srm12;
  fs  BY fs1* fs2@1 fs3-fs6;
  hf  BY msh mch ich fs;
OUTPUT: SAMPSTAT STDYX TECH4 RES;
```

The additional outputs requested were the sample statistics (`SAMPSTAT`), standardised parameters (`STDYX`), the estimated means, covariances and correlations for the latent variables (`TECH4`), and the residuals for the observed variables (`RES`). As with all analyses in Mplus involving categorical dependent variables the WLSMV estimator was used.

To ensure a reasonable scale for the functional status factor, `fs`, the factor loading of `fs2` was fixed to 1 in the health measurement model. This was coded by `fs2@1`, while `fs1*` allowed the factor loading of `fs1` to be freely estimated.

Similarly, the economic well-being measurement model, in Section 5.2.2, was coded as below, with a single factor `ef` (economic well-being) loading onto all the relevant housing quality and durables owned dependent variables:

```
  USEVARIABLES = hq1-hq5 dur1-dur9;
  CATEGORICAL = hq1-hq5 dur1-dur9;
ANALYSIS: ESTIMATOR = WLSMV;
MODEL:
  ef BY hq1-hq5 dur1-dur9;
OUTPUT: SAMPSTAT STDYX TECH4 RES;
```

A.3 Structural Equation Models

This section gives the input code for the structural equation models presented in Section 5.2.3. These models combined the measurement model for health and for economic well-being to ascertain the interrelationship between them and their relationships with covariates. The measurement models were the same as in the previous section.

The first SEM determined the interrelationship between health and economic well-being, modelling the covariance between `hf`, the overall health factor, and `ef`, the economic well-being factor, by the `WITH` command line:

```
  USEVARIABLES = fs1-fs6 srm1-srm12 hq1-hq5 dur1-dur9;
  CATEGORICAL = fs1-fs6 srm1-srm12 hq1-hq5 dur1-dur9;
ANALYSIS: ESTIMATOR = WLSMV;
MODEL:
  !measurement model of health
    msh BY srm1 srm2 srm11;
    mch BY srm4 srm6 srm7;
    ich BY srm3 srm5 srm8 srm9 srm10 srm12;
    fs BY fs1* fs2@1 fs3-fs6;
    hf BY msh mch ich fs;
  !measurement model of economic well-being
    ef BY hq1-hq5 dur1-dur9;
  !structural model
    hf WITH ef;
OUTPUT: SAMPSTAT STDYX TECH4 RES;
```

The conditional SEM modelled used this same code with the additional covariate variables and the modelling of their effect on health, `hf`, and economic well-being, `ef`. The covariates were the dummy variables for age (`age_65`, `age_70`, `age_75`, `age_80`, `age_85p`), gender (`male`), race (`nonwhite`), urban-rural classification (`rural`) and major region (`region_ne`, `region_n`, `region_cw`, `region_s`). The `ON` command line defined both factors `hf` and `ef` to be affected by all covariates. These additional lines of code were:

```
    hf ef ON age_65 age_70 age_75 age_80 age_85p male nonwhite rural
      region_ne region_n region_cw region_s;
```

A.4 Multilevel Exploratory Factor Analysis

In this section, the input code for the multilevel structural equation modelling, including multilevel exploratory factor analysis and confirmatory factor analysis, in Section 5.3 are presented. The identity variables for the clusters were `hhid`, `secid` and `munid` corresponding to the household, sector and municipality respectively; used to define the clustering of individuals in the analysis.

The intraclass correlation coefficients (ICCs) of all the dependents variables (`srm1-srm12`, `fs1-fs6`, `dur1-dur9` and `hq1-hq5`) were estimated; firstly with clustering by sector then again with clustering by municipality. The code used to calculate the ICCs for the clustering by sector was:

```
  USEVARIABLES = secid srm1-srm12 fs1-fs6 dur1-dur9 hq1-hq5;
  CATEGORICAL = srm1-srm12 fs1-fs6 dur1-dur9 hq1-hq5;
  CLUSTER = secid;
ANALYSIS:
  TYPE = TWOLEVEL;
  ESTIMATOR = WLSMV;
OUTPUT: SAMPSTAT;
```

where `CLUSTER = secid` was replaced with `CLUSTER = munid` to calculate the amount of variation explained by the municipality clustering.

The multilevel EFA was then as follows, with the number of factors at the within level and at the between level incremented in turn. The unrestricted within and between level factors were requested by `UW` and `UB` respectively. The code for multilevel EFA of the self-reported morbidity variables was:

```
  USEVARIABLES = secid srm1-srm12;
  CATEGORICAL = srm1-srm12;
  CLUSTER = secid;
ANALYSIS:
  TYPE = TWOLEVEL EFA 1 5 UW 1 5 UB;
  ESTIMATOR = WLSMV;
OUTPUT: SAMPSTAT;
```

For the functional status variables was:

```
  USEVARIABLES = secid fs1-fs6;
  CATEGORICAL =  fs1-fs6;
  CLUSTER = secid;
```

```
ANALYSIS:
  TYPE = TWOLEVEL EFA 1 3 UW 1 3 UB;
  ESTIMATOR = WLSMV;
OUTPUT: SAMPSTAT;
```

And, for the economic well-being variables was:

```
  USEVARIABLES = secid hq1 hq2 dur1-dur9;
  CATEGORICAL = hq1 hq2 dur1-dur9;
  CLUSTER = secid;
ANALYSIS:
  TYPE = TWOLEVEL EFA 1 4 UW 1 4 UB;
  ESTIMATOR = WLSMV;
OUTPUT: SAMPSTAT;
```

The housing quality variables and the durables owned variables were combined in the multilevel EFA, to form a multilevel EFA of economic well-being. This was due to only two of the housing quality variables (`hq1` and `hq2`) not relating to services provided by the sector. Variables that related to services provided by the authority of the sector were too similar, with very little variation, within a sector to be considered as dependent variables but with exceptions that prevented their inclusion as covariates.

A.5 Multilevel Confirmatory Factor Analysis

The multilevel measurement model for health was presented in Section 5.3.1. This model did not contain the variables `srm10` and `srm12` that represented tuberculosis and cirrhosis respectively. The factors `msh`, `mch`, `ich`, `fs` and `hf` at the within level as defined previously in the non-multilevel analysis; defined in the code under the `%WITHIN%` command. At the between level, under the `%BETWEEN%` command, the factors `irh`, `lh` and `lfs` were industry-related health, local health and local functional status respectively. Industry-related health was measured by the binary variables indicating the lack of: disease of the back or column (`srm1`), arthritis, (`srm2`) and hypertension, (`srm6`), while local health was measured by the lack of: cancer (`srm3`), diabetes (`srm4`), asthma (`srm5`), heart disease (`srm7`), chronic renal failure (`srm8`), depression (`srm9`) and tendonitis or tenosynovitis (`srm11`).

```
  USEVARIABLES = secid fs1-fs6 srm1-srm9 srm11;
  CATEGORICAL = fs1-fs6 srm1-srm9 srm11;
  CLUSTER = secid;
ANALYSIS:
  TYPE = TWOLEVEL;
  ESTIMATOR = WLSMV;
MODEL:
  %WITHIN%
    msh BY srm1 srm2 srm11;
    mch BY srm4 srm6 srm7;
    ich BY srm3 srm5 srm8 srm9;
    fs  BY fs1* fs2@1 fs3-fs6;
    hf  BY msh mch ich fs;
  %BETWEEN%
    irh BY srm1 srm2 srm6;
    lh  BY srm3 srm4 srm5 srm8 srm7 srm9 srm11;
    lfs BY fs1* fs2@1 fs3-fs6;
OUTPUT: SAMPSTAT STDYX TECH1 TECH8 RES;
```

The multilevel measurement model for economic well-being contained a subset of the dependent variables considered in the non-multilevel model. Again, there was only one factor of overall economic well-being, this time with one factor of the individual economic well-being `ewf` and another single factor at the sector level, `ebf`, corresponding to the economic well-being of the sector, each measured by the same housing quality and durables owned variables.

```
MODEL:
  %WITHIN%
    ewf BY hq1 hq2 dur1-dur9;
  %BETWEEN%
    ebf BY hq1 hq2 dur1-dur9;
OUTPUT: SAMPSTAT STDYX TECH1 TECH8 RES;
```

A.6 Multilevel Structural Equation Model

The multilevel structural equation model was then the culmination of these two measurement models, as presented in Section 5.3.3, with the association between health and economic well-being estimated. This was defined in the `MODEL` command in the `WITH` statements, whereby all associations between the sector level factors was modelled and the association between the health and economic well-being of older individuals.

```
  USEVARIABLES = secid fs1-fs6 srm1-srm9 srm11 hq1 hq2 dur1-dur9;
  CATEGORICAL = fs1-fs6 srm1-srm9 srm11 hq1 hq2 dur1-dur9;
  CLUSTER = secid;
ANALYSIS:
  TYPE = TWOLEVEL;
  ESTIMATOR = WLSMV;
MODEL:
  %WITHIN%
    !measurement model of health
      msh BY srm1 srm2 srm11;
      mch BY srm4 srm6 srm7;
      ich BY srm3 srm5 srm8 srm9;
      fs  BY fs1* fs2@1 fs3-fs6;
      hf  BY msh mch ich fs;
    !measurement model of economic well-being
      ewf BY hq1 hq2 dur1-dur9;
    !structural model
      hw  WITH ew;
  %BETWEEN%
    !measurement model of health
      irh BY srm1 srm2 srm6;
      lh  BY srm3 srm4 srm5 srm8 srm7 srm9 srm11;
      lfs BY fs1* fs2@1 fs3-fs6;
    !measurement model of economic well-being
      ebf BY hq1 hq2 dur1-dur9;
    !structural model
      irh WITH lh lfs ebf;
      lh  WITH lfs ebf;
      lfs WITH ebf;
OUTPUT: SAMPSTAT STDYX TECH1 TECH8 RES;
```

The covariate effects were then estimated, with the respect `ON` commands added to the model in the `%WITHIN%` and `%BETWEEN%`. At the within level were the dummy variables for age (`age_65`, `age_70`, `age_75`, `age_80`, `age_85p`), gender (`male`) and race (`nonwhite`) and at the between level were the dummy variables for the region (`region_ne`, `region_n`, `region_cw`, `region_s`) and urban-rural classification (`rural`).

```
MODEL:
  %WITHIN%
    !measurement model of health
      msh BY srm1 srm2 srm11;
      mch BY srm4 srm6 srm7;
      ich BY srm3 srm5 srm8 srm9;
      fs  BY fs1* fs2@1 fs3-fs6;
      hf  BY msh mch ich fs;
    !measurement model of economic well-being
      ewf BY hq1 hq2 dur1-dur9;
    !structural model
      hw    WITH ew;
      hf ewf  ON age_65 age_70 age_75 age_80 age_85 age_90 male nonwhite;
  %BETWEEN%
    !measurement model of health
      irh BY srm1 srm2 srm6;
      lh  BY srm3 srm4 srm5 srm8 srm7 srm9 srm11;
      lfs BY fs1* fs2@1 fs3-fs6;
    !measurement model of economic well-being
      ebf BY hq1 hq2 dur1-dur9;
    !structural model
      irh WITH lh lfs ebf;
      lh  WITH lfs ebf;
      lfs WITH ebf;
      irh lh lfs ebf ON rural region_ne region_n region_cw region_s;
OUTPUT: SAMPSTAT STDYX TECH1 TECH8  RES;
```

Appendix B

ELSA, Wave 4 Mplus Input Code

In this appendix, the Mplus input code for the cross-sectional structural equation modelling and item response analyses, of Wave 4 of the English Longitudinal Study of Ageing (ELSA), in Chapter 6 are presented.

Each analysis is based on data in the file `elsa_wave4.dat`, which contained all the dependents variables for health, economic well-being and subjective well-being, and the covariates (age, gender, race, education, region and marital status). The variables for wealth were also available in the data file, including `netfw1-netfw10` for the deciles of net financial wealth and `totw1-totw10` for the deciles of total wealth. The preceding code in each of the Mplus inputs was:

```
DATA: FILE = "elsa_wave4.dat";
VARIABLE:
  NAMES = casp1-casp19 swls1-swls5 lone1-lone5 cesd1-cesd8 srm1-srm9
    srm10-srm16 adl1-adl6 iadl1-iadl9 dur1-dur9 ap1-ap12
    netfw1-netfw10 totw1-totw10
    age60 age65 age70 age75 age80 age85
    male nonwhite edu1-edu7
    regne regnw regyh regem regwm regne regl regse regsw
    maritsi maritma maritdi maritwi;
  MISSING=.;
```

B.1 Exploratory Factor Analysis

The exploratory factor analysis (EFA) presented in Section 6.1.1 contained the nine binary variables indicating the lack of diagnosis for the cardiovascular comorbidities `srm1-srm9`. These corresponded to hypertension, angina, heart attack, heart failure, heart murmur, abnormal heart rhythm, diabetes, stroke and high cholesterol respectively. The EFA incremented the number of factors from one to four:

```
  USEVARIABLES = srm1-srm9;
  CATEGORICAL = srm1-srm9;
ANALYSIS:
  TYPE = EFA 1 4;
  ESTIMATOR = WLSMV;
```

Next the EFA for the chronic comorbidities was for the seven binary variables `srm10-srm16` that indicated the lack of diagnosis of: chronic lung disease, asthma, arthritis, osteoporosis, cancer, psychiatric disorders and neurodegenerative disorders (dementia, Alzheimer's disease or Parkinson's disease). The number of factors was incremented from one to four:

```
  USEVARIABLES = srm10-srm16;
  CATEGORICAL = srm10-srm16;
ANALYSIS:
  TYPE = EFA 1 4;
  ESTIMATOR = WLSMV;
```

EFA of functional status contained the 15 binary variables indicating the lack of difficulty performing the activities of daily living. The first six variables `adl1-adl6` contained the basic activities of daily living, while the latter variables, `iadl1-iadl9`, contained the instrumental activities of daily living. Models with one to six factors were explored in the EFA:

```
  USEVARIABLES = adl1-adl6 iadl1-iadl9;
  CATEGORICAL = adl1-adl6 iadl1-iadl9;
ANALYSIS:
  TYPE = EFA 1 6;
  ESTIMATOR = WLSMV;
```

In the analysis of economic well-being, the accommodation problems and durables owned were analysed separately as in the analysis of the 2008 PNAD. Accommodation variables `ap1-ap12` were binary indicators of the lack of the problem with the household. One to five factors were considered in the EFA:

```
  USEVARIABLES = ap1-ap12;
  CATEGORICAL = ap1-ap12;
ANALYSIS:
  TYPE = EFA 1 5;
  ESTIMATOR = WLSMV;
```

Nine binary variables `dur1-dur9` indicated the ownership of the durables within the household, and the EFA incremented the number of factors from one to four:

```
  USEVARIABLES = dur1-dur9;
  CATEGORICAL = dur1-dur9;
ANALYSIS:
  TYPE = EFA 1 4;
  ESTIMATOR = WLSMV;
```

The subjective well-being consisted of the four measures; Satisfaction With Life Scale, Loneliness, CES-Depression Scale and the CASP-19 (see Section 6.3 for details). The items of each measure were analysed in separate EFA. Firstly, life satisfaction was measured by five ordinal dependent variables `swls1-swls5`, which were analysed in EFA with one to three factors:

```
  USEVARIABLES = swls1-swls5;
  CATEGORICAL = swls1-swls5;
ANALYSIS:
  TYPE = EFA 1 3;
      ESTIMATOR = WLSMV;
```

There were the five ordinal variables `lone1-lone5` observed for loneliness, with higher scores relating to lower levels of loneliness. The EFA exploring the underlying constructs incremented the number of factors from one to three factors:

```
  USEVARIABLES = lone1-lone5;
  CATEGORICAL = lone1-lone5;
ANALYSIS:
  TYPE = EFA 1 3;
  ESTIMATOR = WLSMV;
```

Depression was indicated by the eight binary variables `cesd1-cesd8`, each indicating a lack of depression. These were analysed in EFA with the number of factors incremented from one to four:

```
  USEVARIABLES = cesd1-cesd8;
  CATEGORICAL = cesd1-cesd8;
ANALYSIS:
  TYPE = EFA 1 4;
  ESTIMATOR = WLSMV;
```

B.1.1 CASP-19

The 19 ordinal variables, `casp1-casp19`, of the CASP-19 were first analysed in a confirmatory factor analysis using the structure defined in Higgs et al. (2003). The input code for this model was:

```
  USEVARIABLES = casp1-casp19;
  CATEGORICAL = casp1-casp19;
MODEL:
  con BY casp1-casp4;
  aut BY casp5-casp9;
  sel BY casp10-casp14;
  ple BY casp15-casp19;
OUTPUT: SAMPSTAT TECH1 TECH4 STDYX RESIDUAL;
```

However, as discussed in Section 6.3.4, this produced a model with invalid, Heywood estimates. The variables were therefore analysed in EFA, which incremented the number of factors from one to six:

```
  USEVARIABLES = casp1-casp19;
  CATEGORICAL = casp1-casp19;
ANALYSIS:
  TYPE = EFA 1 6;
  ESTIMATOR = WLSMV;
```

Item Response Analysis

The code for the item response analysis of the CASP-19 variables was as follows. The fours factors control, Enjoyment, Opportunity and Independence were named `con`, `enj`, `opp` and `ind` respectively in the code, with the variance of these factors fixed at 1 by `@1`. The category response curves were obtained under the `PLOT` command with type `PLOT3`; which returns histograms and scatterplots of the estimated factor scores, the item characteristic curves and the information curves.

```
  USEVARIABLES = casp1-casp19;
  CATEGORICAL = casp1-casp19;
MODEL:
  con BY casp2* casp4 casp9;
  enj BY casp10* casp11 casp12 casp13 casp14 casp17 casp18 casp19;
  opp BY casp1* casp8 casp15 casp16;
  ind BY casp3* casp5 casp6 casp7;
  con@1;
  enj@1;
  opp@1;
```

```
  ind@1;
OUTPUT: SAMPSTAT TECH1 TECH4 STDYX RESIDUAL;
PLOT: TYPE=PLOT3;
```

B.2 Confirmatory Factor Analysis

The measurement model of health presented in Section 6.1 was estimated in a confirmatory factor analysis. The dependent variables measured the five factors of Cardiovascular health `car`, Metabolic health `met` and Chronic health `chr` from the binary variables of self-reported morbidity, and Physical function `phy` and Cognitive function `cog` from the binary variables of functional status. These five factors then in turn measured an overall factor of health `hea`. Cardiovascular health was measured by angina `srm2`, heart attack `srm3`, heart failure `srm4`, heart murmur `srm5`, abnormal heart rhythm `srm6` and stroke `srm8`. Metabolic health was measured by hypertension `srm1`, diabetes `srm7` and high cholesterol `srm9` Chronic health was measured by chronic lung disease, asthma, arthritis, osteoporosis, cancer, psychiatric problems and neurodegenerative disorders (`srm10-srm16`). Physical function was measured by the basic activities of daily `adl1-adl6` and the instrumental activities that involved preparing a meal `iadl3`, shopping `iadl4` and doing house or garden work `iadl8`. Finally, cognitive function was measured by the instrumental activities; using a map `iadl1`, recognising danger `iadl2`, telephone calls `iadl5`, communication `iadl6`, taking medication `iadl7` and managing money `iadl9`. The Mplus input code for the estimation of this was model was then:

```
  USEVARIABLES = srm1-srm9 srm10-srm16 adl1-adl6 iadl1-iadl9;
  CATEGORICAL = srm1-srm9 srm10-srm16 adl1-adl6 iadl1-iadl9;
ANALYSIS:
  ESTIMATOR = WLSMV;
MODEL:
  car BY srm2-srm6 srm8;
  met BY srm1 srm7 srm9;
  chr BY srm10-srm16;
  phy BY adl1-adl6 iadl3 iadl4 iadl8;
  cog BY iadl1 iadl2 iadl5-iadl7 iadl9;
  hea BY car met chr phy cog;
OUTPUT: SAMPSTAT PATTERNS TECH1 TECH3 TECH4 STDYX RES;
```

In Section 6.2.3 it was discussed that the ownership of durables was negatively associated with the two factors underlying the observed accommodation problems. In the Mplus code, the factor of the neighbourhood quality was `nei`, household quality was `hou` and the ownership of durables factor was named `dur`. Neighbourhood quality measured by the noise from the neighbours `apnz`, other street noise `apsn` and pollution/grime `appo`. Household quality was measured by variables of: the shortage of space `apsp`, too dark `apdk`, rising damp `aprd`, water leaks `apwa`, condensation problems `apco`, electrical or plumbing problems `apep`, rot and decay `apro`, infestation `apin` and too cold `apco`. The effect of wealth on these three factors was estimated in a confirmatory factor analysis, with the deciles of either net financial wealth (`netfw1-netfw10`) or total wealth (`totw1-totw10`) modelled as covariates. The `MODEL` command in the Mplus input code for this was:

```
MODEL:
  nei BY apnz apsn appo;
  hou BY apsp apdk aprd apwa apcp apep apro apin apco;
  dur by dovr docd dowm dowd dodw domo dopc dodt dodv;
  nei hou dur ON netfw2 netfw3 netfw4 netfw5 netfw6 netfw7 netfw8 netfw9 netfw10;
```

with net financial wealth as the covariate, where `netfw1` and `netfw10` represented low and high wealth respectively. Similarly, for the model with total wealth as the covariate with `totw2-totw10` in place of `netf21-netfw10`.

The measurement model of economic well-being defined in Section 6.2, therefore, consisted of the neighbourhood quality factor `nei` and the household quality factor `hou`, such that the model was defined in the code as:

```
  USEVARIABLES = ap1-ap12;
  CATEGORICAL = ap1-ap12;
ANALYSIS: ESTIMATOR = WLSMV;
MODEL:
  nei BY ap2 ap3 ap5;
  hou BY ap1 ap4 ap6-ap12;
OUTPUT: SAMPSTAT PATTERNS TECH1 TECH3 TECH4 STDYX RES;
```

The measurement model of subjective well-being defined in Section 6.3, which combined the single factors of life satisfaction `sat`, loneliness `lon` and depression `dep` with the four factors of quality of life (`con`, `enj`, `opp` and `ind`) defined from the CASP-19 variables, as measures of the overall subjective well-being factor `sub`. At this point, `lone2` was dropped from the analysis as a repeat of `casp4`, where both items asked how often the respondent felt left out. The Mplus input code for the definition of the confirmatory factor analysis was:

```
  USEVARIABLES =  casp1-casp19 swls1-swls5 lone1 lone3-lone5 cesd1-cesd4
          cesd6-cesd8;
  CATEGORICAL =   casp1-casp19 swls1-swls5 lone1 lone3-lone5 cesd1-cesd4
          cesd6-cesd8;
ANALYSIS:
  ESTIMATOR = WLSMV;
MODEL:
  !factors from the CASP-19, quality of life
    con BY casp1 casp8 casp15 casp16;
    enj BY casp2 casp4 casp9;
    opp BY casp10-casp14 casp17-casp19;
    ind BY casp3 casp5 casp6 casp7;
  !SWLS, loneliness, CES-D
    sat BY swls1-swls5;
    lon BY lone1-lone5;
    dep BY cesd1-cesd8;
  !structural model
    sub BY con enj opp ind sat lon dep;
OUTPUT: SAMPSTAT PATTERNS TECH1 TECH3 TECH4 STDYX RES;
```

B.3 Structural Equation Models

The Mplus input code for the structural equation model presented in Figure 6.8 was as follows. The measurement models for health, economic well-being and subjective well-being were as defined above in the confirmatory factor analysis. The association between health, `hea`, and the economic well-being factors, `nei` and `hou` was modelled in the `WITH` command, then the effect of these factors on subjective well-being were defined in the `ON` command:

```
  USEVARIABLES = casp1-casp19 swls1-swls5 lone1 lone3-lone5 cesd1-cesd4
  cesd6-cesd8 srm1-srm9 srm10-srm16 adl1-adl6 iadl1-iadl9 ap1-ap12;
  CATEGORICAL = casp1-casp19 swls1-swls5 lone1 lone3-lone5 cesd1-cesd4
  cesd6-cesd8 srm1-srm9 srm10-srm16 adl1-adl6 iadl1-iadl9 ap1-ap12;
ANALYSIS:
  ESTIMATOR = WLSMV;
MODEL:
  !economic well-being measurement model
    nei BY ap2 ap3 ap5;
    hou BY ap1 ap4 ap6-ap12;
  !health measurement model
    car BY srm2-srm6 srm8;
    met BY srm1 srm7 srm9;
    chr BY srm10-srm16;
    phy BY adl1-adl6 iadl3 iadl4 iadl8;
    cog BY iadl1 iadl2 iadl5-iadl7 iadl9;
    hea BY car met chr phy cog;
  !subjective well-being measurement model
    con BY casp1 casp8 casp15 casp16;
    enj BY casp2 casp4 casp9;
    opp BY casp10-casp14 casp17-casp19;
    ind BY casp3 casp5 casp6 casp7;
    sat BY swls1-swls5;
    lon BY lone1-lone5;
    dep BY cesd1-cesd8;
    sub BY con enj opp ind sat lon dep;
  !structural model
    hea WITH  nei hou;
    sub ON    nei hou hea;
```

Covariates were then included in the model with the same code as above with the additional lines of code;

```
  !covariates
    hea nei hou sub ON  age60 age65 age70 age75 age80 age85
      male nonwhite edu1-edu7 netfw1-netfw10 totw1-totw10
      regne regnw regyh regem regwm regne regl regse regsw
      maritsi maritma maritdi maritwi;
```

in the `MODEL` section. The results of this model were presented in Section 6.4.1.

Appendix C

ELSA, Waves 2, 4 & 6 Mplus Input Code

In this appendix, the Mplus input code for analyses in Chapter 7 are presented, including the analysis of measurement invariance and the latent growth modelling. The datafiles used in these analyses were `elsa_wave246_inc_dropout.dat` and `elsa_wave246_exc_dropout.dat`. The variables within in file was identical, containing an identification variable for each individual the observed variables of subjective well-being at waves 2, 4 and 6 of the English Longitudinal Study of Ageing (ELSA) and the covariates. Dependent variable names were of the form '*aspect*'-'*wave*'-'*item number*'. The code at the start of each input file was:

```
DATA: FILE = "elsa_wave246_inc_dropout.dat";
VARIABLE:
  NAMES = casp201-casp219 swls21-swls25 lone21-lone25 cesd21-cesd28
      casp401-casp419 swls41-swls45 lone41-lone45 cesd41-cesd48
      casp601-casp619 swls61-swls65 lone61-lone65 cesd61-cesd68
      male nonwhite age60 age65 age70 age75 age80 age85
      mssi_0 msma_0 msdi_0 mswi_0
      mssi_2 msma_2 msdi_2 mswi_2
      chno2 madi2 mawi2 dwma2
      chno4 madi4 mawi4 dwma4
      chno6 madi6 mawi6 dwma6;
  MISSING IS .;
```

C.1 Measurement Invariance

To test the measurement invariance of the repeated measures of life satisfaction, loneliness and depression with time a confirmatory factor analysis was applied to each aspect separately. This analysis used the `elsa_wave246_inc_dropout.dat` file. The input code provided below is given for the life satisfaction process with variables `swls21-swls25` from Wave 2, `swls41-swls45` from Wave 4 and `swls61-swls65` from Wave 6. In this the equality of the factor loadings and thresholds for each repeated dependent variable was tested. For example, the factors loadings and thresholds of `swls21`, `swls41` and `swls61` were compared as the repeated measures of the 'Close to ideal' variable.

Due to the complexity of the later models, the number of iterations in the estimation of the models was increased to 50,000 (from the default 10,000) to ensure convergence, which was kept consist across all the analyses in Chapter 7.

The confirmatory factor analysis contained the measurement model for life satisfaction repeated at each wave, such that the factors `satf2`, `satf4` and `satf6` were the measured life satisfaction at waves 2, 4 and 6 respectively. The factor loadings were of dependent variables were then labelled by the parentheses following the name of the variable in the model. For example, the factor loading of `satf2` on `swls22` was labelled `sat22`. The invariance of the factor loadings was then tested in the MODEL TEST— section of the model, by testing whether the difference in the factor loadings was equal to zero. The covariance between the residuals of the repeat variables was modelled using the `WITH` command, as seen below:

```
  USEVARIABLES =  swls21-swls25 swls41-swls45 swls61-swls65;
  CATEGORICAL = swls21-swls25 swls41-swls45 swls61-swls65;
ANALYSIS:
  ESTIMATOR=WLSMV;
  ITERATIONS=50000;
MODEL:
!life satisfaction Wave 2 measurement model
  satf2 BY swls21 (sat21)
    swls22 (sat22)
    swls23 (sat23)
    swls24 (sat24)
    swls25 (sat25);
!life satisfaction Wave 4 measurement model
  satf4 BY swls41 (sat41)
    swls42 (sat42)
    swls43 (sat43)
    swls44 (sat44)
    swls45 (sat45);
!life satisfaction Wave 6 measurement model
  satf6 BY swls61 (sat61)
```

```
    swls62 (sat62)
    swls63 (sat63)
    swls64 (sat64)
    swls65 (sat65);
!Residual covariance structure
  swls21 WITH swls41 swls61;
  swls41 WITH swls61;
  [...]
  swls25 WITH swls45 swls65;
  swls45 WITH swls65;
MODEL TEST:
  !equality of loadings
    0 = sat22 - sat42; 0 = sat42 - sat62;
    0 = sat23 - sat43; 0 = sat43 - sat63;
    0 = sat24 - sat44; 0 = sat44 - sat64;
    0 = sat25 - sat45; 0 = sat45 - sat65;
```

The factor loadings were then constrained to be equal by giving the same label to the respective parameters, so that the code for the measurement models was:

```
  USEVARIABLES = swls21-swls25 swls41-swls45 swls61-swls65;
  CATEGORICAL = swls21-swls25 swls41-swls45 swls61-swls65;
ANALYSIS:
  ESTIMATOR=WLSMV;
  ITERATIONS=50000;
MODEL:
!satisfaction Wave 2 measurement model
  satf2 BY swls21
    swls22 (sat2)
    [...]
    swls25 (sat5);
!satisfaction Wave 4 measurement model
  satf4 BY swls41
    swls42 (sat2)
    [...]
    swls45 (sat5);
!satisfaction Wave 6 measurement model
  satf6 BY swls61
    swls62 (sat2)
    [...]
    swls65 (sat5);
!Residual covariance structure
  swls21 WITH swls41 swls61;
  swls41 WITH swls61;
  [...]
  swls25 WITH swls45 swls65;
  swls485WITH swls65;
```

Similarly, for the testing of the invariance of the thresholds. Thresholds were specified within square parentheses, for example `[swls21$1]` specifies the first threshold of the variable `swls21`, and these were labelled by the round parentheses that followed. The equality of the specified thresholds was then tested in the `MODEL TEST` command as before:

```
!threshold structure
  [swls21$1](i211);
  [swls22$1](i212);
  [swls23$1](i213);
  [swls24$1](i214);
  [swls25$1](i215);

  [swls41$1](i411);
  [swls42$1](i412);
  [swls43$1](i413);
  [swls44$1](i414);
  [swls45$1](i415);

  [swls61$1](i611);
  [swls62$1](i612);
  [swls63$1](i613);
  [swls64$1](i614);
  [swls64$1](i615);

  [swls21$2](i221);
  [swls22$2](i222);
  [swls23$2](i223);
  [swls24$2](i224);
  [swls25$2](i225);

  [swls41$2](i421);
  [swls42$2](i422);
  [swls43$2](i423);
  [swls44$2](i424);
  [swls45$2](i425);

  [swls61$2](i621);
  [swls62$2](i622);
  [swls63$2](i623);
  [swls64$2](i624);
  [swls64$2](i625);
  MODEL TEST:
  !equality of thresholds
    0 = i211 - i411; 0 = i411 - i611;
    0 = i212 - i412; 0 = i412 - i612;
    0 = i213 - i413; 0 = i413 - i613;
    0 = i214 - i414; 0 = i414 - i614;
    0 = i215 - i415; 0 = i415 - i615;
    0 = i221 - i421; 0 = i421 - i621;
    0 = i222 - i422; 0 = i422 - i622;
    0 = i223 - i423; 0 = i423 - i623;
    0 = i224 - i424; 0 = i424 - i624;
    0 = i225 - i425; 0 = i425 - i625;
```

This was then repeated for the loneliness and depression aspects, with their respective dependent variables.

C.2 Latent Growth Model, without covariates

The Mplus input code for the multivariate latent growth model presented in Section 7.2 was then as follows. This was repeated for both datafiles, `elsa_wave246_inc_dropout.dat` and `elsa_wave246_exc_dropout.dat`, but only the results from `elsa_wave246_inc_dropout.dat` were presented in Section 7.2. For the growth process life satisfaction, the random intercept and random slope were named `ins` and `sls` respectively, for loneliness the random intercept and random slope were named `inl` and `sll` respectively, and for depression the random intercept and random slope were named `ind` and `sld` respectively. The growth model was coded by the random intercept and random slope followed by — and the names of required factors. The time scale of the growth process was set by the `@0`, `@1` and `@2` following the name of factor.

```
  USEVARIABLES =  casp201-casp219 swls21-swls25 lone21-lone24 cesd21-cesd28
          casp401-casp419 swls41-swls45 lone41-lone44 cesd41-cesd48
          casp601-casp619 swls61-swls65 lone61-lone64 cesd61-cesd68;
  CATEGORICAL =
          casp201-casp219 swls21-swls25 lone21-lone24 cesd21-cesd28
          casp401-casp419 swls41-swls45 lone41-lone44 cesd41-cesd48
          casp601-casp619 swls61-swls65 lone61-lone64 cesd61-cesd68;
ANALYSIS:
  ESTIMATOR=WLSMV;
  ITERATIONS=50000;
MODEL:
  !life satisfaction measurement models
    satf2 BY swls21 (sat1)
      swls22 (sat2)
      swls23 (sat3)
      swls24 (sat4)
      swls25 (sat5);
    satf4 BY swls41 (sat1)
      swls42 (sat2)
      swls43 (sat3)
      swls44 (sat4)
      swls45 (sat5);
    satf6 BY swls61 (sat1)
      swls62 (sat2)
      swls63 (sat3)
      swls64 (sat4)
      swls65 (sat5);
  !loneliness measurement models
    lonf2 BY lone21 (lon1)
      lone22 (lon2)
      lone23 (lon3)
      lone24 (lon4);
    lonf4 BY lone41 (lon1)
      lone42 (lon2)
      lone43 (lon3)
      lone44 (lon4);
    lonf6 BY lone61 (lon1)
      lone62 (lon2)
      lone63 (lon3)
```

```
        lone64 (lon4);
    !depression measurement models
      depf2 BY cesd21 (dep1)
        cesd22 (dep2)
        cesd23 (dep3)
        cesd24 (dep4)
        cesd25 (dep5)
        cesd26 (dep6)
        cesd27 (dep7)
        cesd28 (dep8);
      depf4 BY cesd41 (dep1)
        cesd42 (dep2)
        cesd43 (dep3)
        cesd44 (dep4)
        cesd45 (dep5)
        cesd46 (dep6)
        cesd47 (dep7)
        cesd48 (dep8);
      depf6 BY cesd61 (dep1)
        cesd62 (dep2)
        cesd63 (dep3)
        cesd64 (dep4)
        cesd65 (dep5)
        cesd66 (dep6)
        cesd67 (dep7)
        cesd68 (dep8);
    !latent growth model - random intercept and slope
      ins sls | s2@0 s4@1 s6@2;
      ind sld | c2@0 c4@1 c6@2;
      inl sll | l2@0 l4@1 l6@2;
    !factor covariance structure
      ins WITH sld@0 sll@0;
      ind WITH sls@0 sll@0;
      inl WITH sls@0 sld@0;
    !residual covariance structure of the latent response variables
      swls_1_2 WITH swls_1_4 (satl1)
      swls_1_6;
      swls_1_4 WITH swls_1_6 (satl1);
      [...]
      lonely_1_2 WITH lonely_1_4 (lonl1)
      lonely_1_6;
      lonely_1_4 WITH lonely_1_6 (lonl1);
      [...]
      cesd_1_2 WITH cesd_1_4 (depl1)
      cesd_1_6;
      cesd_1_4 WITH cesd_1_6 (depl1);
      [...]
OUTPUT: SAMPSTAT TECH1 TECH4 STDYX RES;
```

The conditional models presented in Sections 7.3 and 7.4 both used the datafile `elsa_wave246_exc_dropout.dat`.

The time invariant covariates, for the model presented in Section 7.3, the above code was used with the addition of:

```
  !structural model
    ins ind inl sls sld sld ON male nonwhite age_65_2 age_70_2
     msma_2 msdi_2 mswi_2;
```

in the `MODEL` section.

The model presented in Section 7.4 was then similar, except the covariates were specified by the code:

```
!time varying covariates
  s2 l2 c2 ON madi2 mawi2 dwma2;
  s4 l4 c4 ON madi4 mawi4 dwma4;
  s6 l6 c6 ON madi6 mawi6 dwma6;
!time invariant covariates
  ins sls ind sld inl sll ON sex race
  age_D65_2 age_D70_2 msma_0 msdi_0 mswi_0;
```

where the variables `madi2`, `madi4` and `madi6` were indicators of a change in marital status from married to divorced between Waves 0 and 2, 2 and 4, and 4 and 6 respectively, similarly `mawi2`, `mawi4` and `mawi6` indicated a change from married to widowed, and `dwma2`, `dwma4` and `dwma6` a change from divorced or widowed to married. The variables `msma_0`, `msdi_0` and `mswi_0` indicated the marital status of married, divorced or widowed at Wave 0.