Exploring the discordance between self-reported and performance-based measures of physical capability

Elizabeth Genowefa Wloch

Thesis submitted for the Degree of Doctor of Philosophy

MRC Unit for Lifelong Health and Ageing at UCL

University College London

2016

I, Elizabeth Genowefa Wloch, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Levels of physical capability, an individual's ability to undertake the physical tasks of daily living, are assessed using performance-based and self-reported measures. By examining discordance between these measures, two groups can be identified; one reports poor capability yet performs to higher levels than expected (underestimators), whilst the other report higher levels of capability than expected given their poor performance (overestimators). It was hypothesised that discordant groups would have different characteristics.

Data from the MRC National Survey of Health and Development were used to explore the concordance and discordance between self-reported and performance-based measures of physical capability at age 60-64. To identify two discordant groups, and distinguish these from a reference concordant group, summary scores of the self-reported and performance-based measures were produced and plotted against each other. The selection of measures used to produce the self-reported summary score was informed by analyses that explored the hierarchical order in which participants reported difficulty with the physical tasks of daily living. Multinomial logistic regression models were used to investigate the associations between discordance and factors selected a priori based on comprehensive literature reviews

Sex, education, occupational class, smoking history, pain and chronic depression, were associated with discordance at age 60-64. For example, those with higher socioeconomic position, who experienced pain or chronic depression were more likely to underestimate their physical capability, whilst women and ex-smokers were less likely to overestimate their physical capability.

Evidence of discordance suggests that there may be individuals in early old age who have unmet needs relating to their physical capability when assessment is based on either self-reported or performance based measures alone. Future policy should ensure these individuals are identified; factors associated with discordance can be used to identify target groups who may benefit from intervention that prevent or minimise age-related declines in physical capability.

Acknowledgments

I would first like to thank my supervisors, Dr Rachel Cooper and Prof Diana Kuh, for their continued guidance, feedback and encouragement throughout this project. I have learned a tremendous amount under your supervision and I hope that the skills I have developed during this time will continue to aid me throughout my career.

I would also like to thank my colleagues in the MRC Unit for Lifelong Health and Ageing at UCL, especially Andy Wong, Stephanie Pilling and Maria Popham, for their professional and friendly technical advice and support. I am indebted to the study members of the MRC National Survey of Health and Development (NSHD) for their longstanding, and ongoing, contribution to the survey; without them this project would not have been possible. For their continued funding of the NSHD and for providing financial support during my thesis I would also like to thank the Medical Research Council.

Finally, I would like to acknowledge the huge amount of support I have received from my family over the past few years, particularly from my husband and parents. The patience and kindness you have shown me has helped me to complete this thesis and get through one of the most challenging phases of my life; for this I will be forever grateful.

Contents

List of Tables

List of Figures

Abbreviations

WHO – World Health Organisation

ICF – International Classification of Functioning, Disability and Health

ICIDH – International Classification of Impairments, Disabilities and Handicaps

NSHD – MRC National Survey of Health and Development

CRFs - Clinical Research Facilities

TUG - Timed get up and go

OPCS – Office of Population Censuses and Surveys

BADL – Basic Activities of Daily Living

IADL – Instrumental Activities of Daily Living

1. Introduction

The overall aim of this thesis is to explore the concordance between self-reported and performance-based measures of physical capability in late midlife and early old age, and to identify which factors across life influence any observed discordance.

Physical capability can be defined as an individual's ability to perform the physical tasks of daily living (1;2). The concept of physical capability incorporates the complete spectrum of ability, rather than the narrow focus at the negative end of the scale associated with disability. The advantage of this approach is that it encourages researchers to investigate the wide variability between individuals in levels of physical capability observed within older populations in order to inform relevant policy and the direction of future research into interventions. A more detailed discussion of the conceptual definition of physical capability and the language used in this context will follow in Section 2.2.

1.1 Importance of physical capability in an ageing population

During the 20th century, life expectancy at birth rose by more than 30 years across Europe, from 42.7 years in 1900 to 76.8 years in 2001 (3), and reductions in mortality rates at older ages mean that life expectancy has steadily continued to rise in the 21st century (4). Combined with low fertility levels, this improvement in life expectancy has led to the ageing of populations across the world (5), especially in Western Europe. Population ageing occurs when the proportion of the population of older ages increases (old age usually taken as those over aged 65) (6). In the UK, the proportion of the population aged 65 and over rose from 15% in 1985 to 17% in 2010, which corresponds to an increase of 1.7 million people in this age group (7). Due to the current population age structure, together with further predicted improvements in mortality and life expectancy, it is forecast that by 2035 23% of the UK population will be over 65 (7).

In general there has been a focus on the negative consequences of population ageing in the literature and broader media (8;9). This negative perception of ageing was highlighted in a paper that reviewed geriatric medical textbooks published in the last 15 years, with a focus on cognitive ageing (10). The authors reported that, of the 40 books included in their review, not a single one framed population ageing in a positive manner. This is based on the fact that, as the proportion of older people rises, there is a fear that the economic, social and health care costs associated with old age will also rapidly increase (5;8;11). The population structure observed today is unprecedented, and the key societal institutions required to cope with this increased level of demand have not been appropriately equipped (9). It is important to note that population ageing is not a phenomenon occurring in isolation; there are other societal changes that are contributing to this challenge (11). For example, family structure has changed dramatically over the past century, with an increase in the number of single households. This means that the care provided in previous generations by family members, which helped to reduce the "burden" of old age, has become more limited (5), particularly when younger generations migrate away from home (11).

This negative outlook of population ageing is based on the assumption that chronological age is inextricably linked to health decline. For example, the economic dependency ratio, often used when forecasting the financial challenges of population ageing, assumes that everyone over the age of 65 is dependent and unable to contribute to society (8). The evidence collected in the literature challenges this assumption; although some individuals will experience a significant decline in health with increasing age, there is a heterogeneity observed in health trajectories, with some people maintaining good health into very old age (12;13). One study in the USA reported that 23% of study members who survived to the age of 100 did so without any major chronic disease, and 18% became centenarians with no disability (12). Throughout the course of their lives, older people have gained wisdom, developed their sense of logic and refined their strategic thinking (10;14), enabling them to adapt to their changing circumstances, and consequently many are able to maintain a relatively high level of function within society (8;9;11). On the basis of the evidence available in the literature, it is therefore inappropriate to extrapolate the experience of those with substantial age-related declines in health to the general population of older adults.

The desired goal for policy makers and older people themselves is to create a compression of morbidity and disability at the end of life, to ensure that the extra years of life gained through increased life expectancy are spent in health rather than an elongated period of frailty. There are mixed findings in the literature as to whether this compression of morbidity is being achieved, depending on the measure of morbidity used to assess the theory. For example, when chronic disease was used as the proxy measure for morbidity, the evidence was less favourable, with none of the studies included in a recent review providing evidence in support of compression (15). However, when the more commonly used measure of disability was taken as a proxy indicator of morbidity, there were twice as many studies providing support for the compression of morbidity as there were against the theory (15). Across the world, with each passing generation, a greater proportion of younger cohorts are maintaining good health into older age (15-18). For example, one study that compared two cohorts of 75 year old Swedish adults born 30 years apart reported dependency in activities of daily living decreased from 13.9% to 5.6% between the two cohorts ($P<0.001$) (17). This suggests that any reduction in population average health caused by the increased lifespan of those with disabling chronic conditions (expansion of morbidity) has been negated by the effect of accumulated good health across the life course, resulting in progressively healthier cohorts (16). It is important to note however that this trend is not universal. A recent study observed that between 1991 and 2011 in England, whilst the proportion of life spent in severe disability decreased slightly, the period spent in mild disability rose during the same time period (19). It is plausible that the rise in the global obesity epidemic may contribute to this expansion of morbidity (19), and future generations will have to face this challenge.

If the compression of morbidity can be achieved and sustained then a greater proportion of older adults will be able to maintain good health, and consequently retain their independence. A recent Select Committee on Public Health and Demographic Change (20) emphasised the importance of maintaining independence. If independence is maintained then the impact of population ageing due to the economic implications associated with health and social care may be reduced, and quality of life for older adults will improve. In order to retain independence, older people need to maintain high levels of physical capability (defined on p.15) and cognition to complete the tasks of daily living. Although cognition is a significant factor in maintaining independence, this thesis focuses on physical capability as the primary outcome of interest.

From midlife onwards, with increasing age there is a significant decline in mean levels of physical capability at the population level (1;21;22). However, there are substantial variations in levels of physical capability found within populations of older people, indicating that age-related decline may not be experienced to the same extent by all individuals (21). It has been proposed that a quarter of the variability in physical capability at older ages is due to the genetic makeup of the individual but a large proportion of the observed heterogeneity is due to the consequences of inequalities in health accumulated across the life course (11). Preserving physical capability is considered to be one of the main components of healthy ageing by both researchers and older people themselves (5;23-26). Physical capability levels recorded at baseline in early old age have repeatedly been shown to predict the risk of adverse health outcome such as comorbidity, hospitalisation and premature mortality (27-31); individuals with higher initial levels of physical capability have a lower risk of adverse outcomes. For example, a 75 year old with low levels of physical capability has a life expectancy five years lower than that of an independent 75 year old with high levels of physical capability (8.2 years for women and 4.4 years for men, compared to 13.2 years and 9.4 years respectively) (32).

1.2 Measuring physical capability

Having established the importance of physical capability within the context of an ageing population, the following discussion will outline the two main measures used to assess physical capability.

There are two main approaches used to measure physical capability: performance-based and self-reported measures. Some researchers have shown a preference to move towards the use of performance-based measures over self-reported measures; however, the attributes of each measure ensure that, when both are used in conjunction, a more complete picture of physical capability is formed (33-35).

Self-reported measures of physical capability generally ask an individual to recall whether they have experienced any physical limitations in their ability to conduct a specific task within their daily lives, such as difficulty climbing stairs. Questions can be asked by an interviewer or self-administered in a survey, which makes the collection of these data relatively low in cost and convenient to administer because no specialist training is required (35;36). Many of the available self-reported measures focus on the lower end of the physical capability spectrum and detect the loss of function in ageing study populations (2;37). In most community-dwelling populations, loss of function in the tasks recalled for self-reported measures is uncommon, for example difficulty feeding oneself. This can result in ceiling effects and thus limits the ability of these measures to capture variation in individuals with high levels of physical capability (35). Although the limitations of ceiling effects are often associated with self-reported measures, it is important to note that performance-based measures may also have ceiling and floor effects (38). The limited ability of self-reported measures to capture the whole spectrum of physical capability also means there are concerns about the ability of self-reported measures to capture change over time (36;39). However, self-reported measures of physical capability capture the "lived experience" of an individual and this attribute has positive implications from a policy perspective, as the information collected can easily be translated to policymakers. Also the wording of questions in self-reported measures is such that, for those who report loss of function, the clinical significance and the impact on individuals' lives are clear. When recalling their levels of physical capability for self-reported measures, individuals reflect on their past experience (sometimes over a given time period defined in the question) to produce an overall estimate of their usual performance, rather than considering a single time point (39). This overall assessment may be influenced by the individual's environmental context, and may incorporate the effects of adaptation such as the use of aids or equipment (39;40).

There is an element of subjectivity to self-reported measures of physical capability, as the individual's perception of their ability (can do) may differ from their actual level of ability (34-36;39;41;42). Perceptions of physical capability levels may be influenced by a variety of factors, including culture and emotional state of mind (36;40;43). For example, the cultural context may dictate the significance an individual places on any experienced "difficulty". If an individual has difficulty walking more than half a mile, this may not register to the individual as a problem in a car-driven society where such distances are rarely covered. Another factor that should be considered when using self-reported measures is the impact of language and cognitive function on the ability of an individual to interpret the meaning of the question and respond appropriately (36;40;41;43). One key point of language is the difference between difficulty and dependency, as individuals will need to have lower levels of capability before reporting dependency.

In comparison, performance-based measures, such as walking speed, record an individual's physiological ability to conduct a specific task under observation, given standard conditions. In line with the conceptualisation of physical capability, performance-based measures record an individual's function on a continuum of low to high capability (2). Although this captures the full range of functioning within the population, some authors suggest that thresholds may need to be established to determine what level has clinical significance in terms of increased risk of future adverse outcomes (38;44). Performance-based measures are often described as objective measures of physical capability, because they capture levels of capability by observing individuals performing specific tasks rather than relying on subjective recall (34;36;44). For example, if an individual had not attempted a specific task for a considerable period of time due to personal circumstances, self-reported measures would rely on the presumed capability of the individual based on previous experience and current health, whereas performance-based measures would accurately record their level of capability at the time of data collection. This distinction between actual and presumed capability is one of the reasons that performance-based measures are often quoted as having better face validity than self-reported measures, because there is less ambiguity about the task under assessment (2;34;36;43;44).

Performance-based measures are also considered to be more reliable than self-reported measures because they are not so heavily influenced by perceptions and are conducted under standard conditions (35;44). However, precise replication of standard conditions is not always possible given the variation in practical details between assessment centres. For example, in the chair rise performance-based measure it is possible that chair heights may differ, increasing the difficulty of the performance test (34). Furthermore, it is not only the physical environment that should be standardised during the collection of performance-based measures. Procedural conduct, such as the level of encouragement given throughout assessment and the number of repetitions, also needs to be standardised as both have been shown to influence recorded levels of physical capability (44). The time of day when individuals are asked to perform may also affect the levels of recorded physical capability due to fatigue, which may change over the course of the day. Individuals are more motivated to perform well during a one off clinic visit than they are in their daily lives, so performance-based measures tend to capture maximal capacity (36), regardless of whether the protocol aimed to elicit maximum response. It is important to note that not all individuals will be able to complete the tasks required to assess physical capability using performance-based measures. Whilst the inability to complete a task for health reasons may provide useful information, some researchers do not incorporate these data into their analyses and recode such individuals as missing, potentially introducing bias.

Performance-based measures capture levels of physical capability at a single time point, so it is equally conceivable that a one-off measurement may capture the temporary effects of some ailment or injury (45). However, when performance-based measures are repeated longitudinally, they appear to be more sensitive to change than self-reported measures (36;39;44). It should be acknowledged that this increased sensitivity is due to the underlying design of performance-based measures, which assess physical capability on a continuous scale, and can therefore detect smaller changes than can be detected in the categorical responses used in self-reported measures. It has been suggested that performance-based measures are more sensitive to early decline in function amongst individuals with high levels of physical capability, and may be able to detect loss of function before it is noticed by the individuals themselves (35;36;46). Performance-based measures can require specialist equipment and training, which makes them more costly and time consuming than their self-reported counterparts (34;36;44).

Despite the different characteristics of each approach, there is a role for both sets of physical capability measures, as they provide additional, complementary information which would not be captured through the use of a single type of measure.

1.3 Concordance of physical capability measures

As part of the discussion surrounding self-reported and performance-based measures of physical capability, numerous studies have measured concordance between the two sets of measures. Coman and Richardson (36) conducted a review of 18 such studies and noted a wide variability in the level of concordance observed between the two sets of measures. The authors reported that the correlation coefficients ranged from -0.72 to 0.6 within the studies reviewed; however, it is not clear from their methods how these correlation coefficients were derived. The negative correlation coefficients could represent a negative association between self-reported and performance-based measures; it is more likely that these values are negative simply because of inconsistencies in the way the two types of measures are coded and scaled.

If the sample of papers included in Coman and Richardson's review is limited to those where the same tasks were assessed using both types of physical capability measures, the variation in correlation coefficients between self-reported and performance-based measures lessens to a range of 0.6 to 0.86. Although this suggests a strong correlation, there is also evidence of some discordance between the two sets of measures which it is important to investigate. To help visualise this discordance, a conceptual diagram has been produced (see Figure 1.1). By examining the discordance between self-reported and performance-based measures of physical capability, researchers can identify two distinct groups of individuals with different characteristics. One group performs to higher levels of capability yet reports poor capability (coloured orange in Figure 1.1), whilst the other group suggests more resilience, with higher levels of reported capability than might be expected given their poor performance (coloured green in Figure 1.1). When only one type of measure is used to assess physical capability, it is possible that within each of these discordant groups there may be some individuals whose physical capability needs may go undetected. Further investigation into the two discordant groups, and the factors associated with each, may enable researchers to identify some of the determinants of this discordance. If determinants are established, these can be used to predict discordant individuals who may prove to be useful target groups for interventions that prevent or minimise age-related functional decline.

There are however several limitations to the existing evidence base, as summarised by Coman and Richardson, that need to be addressed in order to identify potential determinants of discordance. Firstly, most of the previous studies have focused on concordance, merely inferring the presence of discordance. Secondly, there is a lack of consistency in the underlying theoretical constructs, particularly in the definition of physical capability, which may influence the level of observed discordance. Thirdly, most of the previous studies have been cross-sectional in nature, not accounting for life course factors or longitudinal trends that may affect levels of discordance. In addition, of the 18 papers included in the review by Coman and Richardson, most were based on small sample sizes, and several had sex-specific samples, thus limiting the generalisability of the findings. As a result of the limitations discussed, further research is required to explore the discordance between self-reported and performance-based measures of physical capability in more detail, to assist in the development of suitable interventions to minimise the impact of a declining trajectory of physical capability.

1.4 Aims and objectives

The purpose of this thesis is therefore to strengthen the existing evidence base by addressing some of the gaps highlighted in the literature. The aim is to explore the concordance and discordance between self-reported and performance-based measures of physical capability in early old age using data from the MRC National Survey of Health and Development (NSHD). The thesis will also aim to investigate the association between discordance in measures of physical capability and socio-demographic, behavioural risk and health factors.

1.4.1 Objectives

In order to achieve these aims, specific objectives were set as outlined below:

1. Review the literature to conceptualise physical capability within the existing disability frameworks, and identify the appropriate terminology associated with the construct.
2. Describe the levels of physical capability experienced by participants in the study population using self-reported and performance-based measures of physical

capability, and explore the hierarchical order in which participants report difficulty with the tasks of daily living.

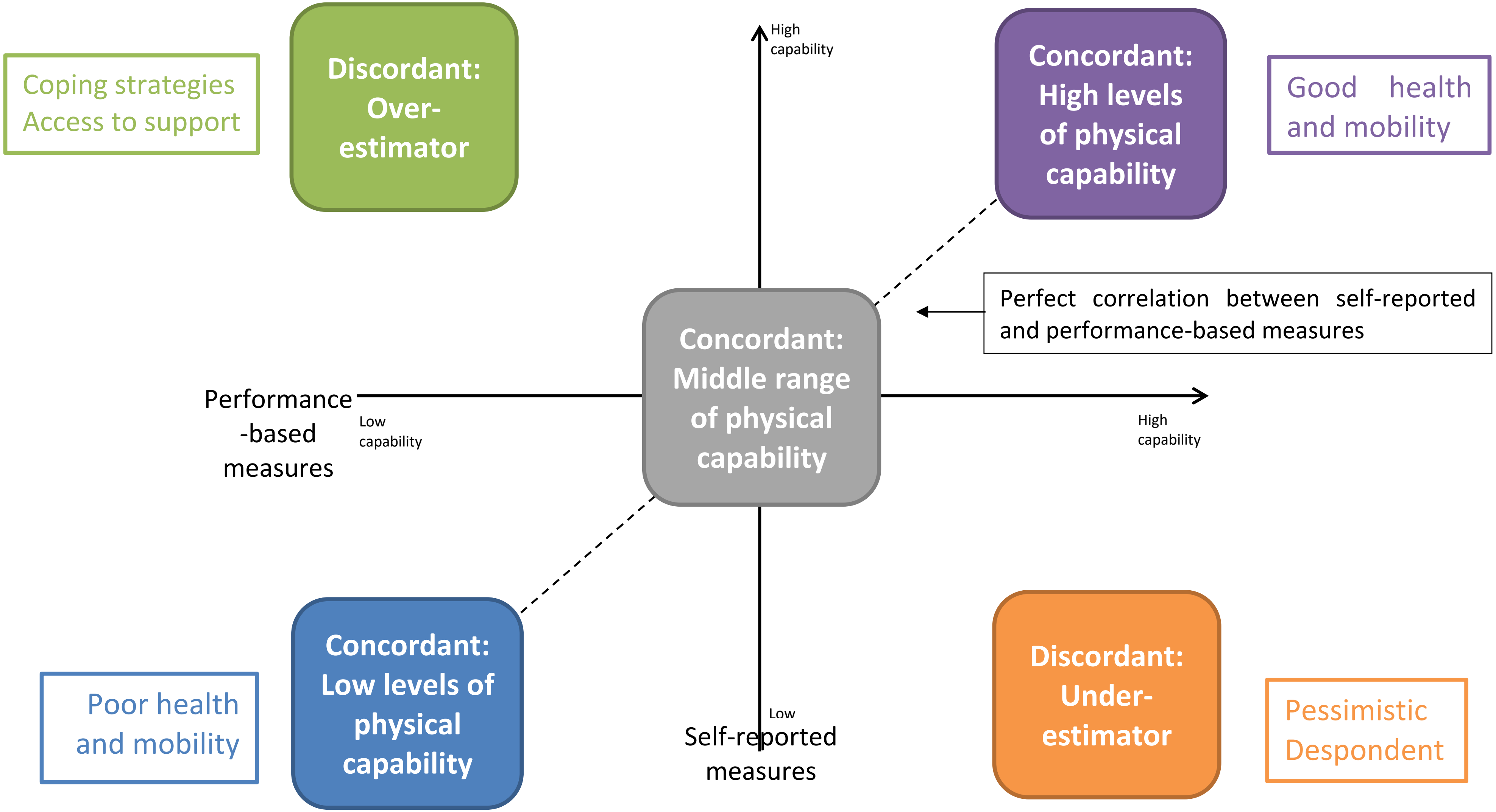
3. Conduct two further literature reviews to summarise and critique the existing evidence base for a) the concordance and b) the discordance of self-reported and performance-based measures of physical capability.
4. Establish the level of concordance and discordance observed between the measures of physical capability within the study population and identify five distinct groups of concordant and discordant individuals (Figure 1.1).
5. Explore the association between discordance and potential risk factors, identified a priori, by ascertaining whether the two discordant groups are distinctly different from the concordant group with respect to these factors.
6. Incorporate data collected prospectively to extend the existing discordance evidence base beyond cross sectional analyses by analysing repeated measures and exploring change between time points.

Each of these objectives has been expanded and refined based on reviews of the existing literature in the corresponding chapters.

1.5 Structure of thesis

The following chapter will outline the development of the existing conceptual frameworks and critically assess the constructs and terminology used in the literature. Chapter 3 will then introduce the data set and provide the context to the thesis by exploring levels of physical capability in the study population in early old age (60-64 years old). The descriptive analysis is extended in Chapter 4 by exploring the hierarchical order with which participants first report the experience of difficulty with the physical tasks of daily living. Chapter 5, updates the literature review conducted by Coman and Richardson (36), focusing on papers that explore the concordance between self-reported and performance-based measures of physical capability as conceptualised in this thesis. The second part of the chapter will ascertain the levels of concordance and discordance to be found in the study population, before producing appropriate variables in order to identify five distinct concordant and discordant groups of individuals. Chapter 6 will start with the second literature review, using a subsample of papers from the concordance review, focusing on discordance to identify potential factors of interest that are associated with discordance. A selection of factors will then be chosen based on the findings of this literature review, and the second part of Chapter 6 and Chapter 7 will use appropriate analyses to determine the extent of the association between the factors of interest and discordance. In Chapter 6 the focus will be on sociodemographic factors and health behaviours, whereas Chapter 7 will focus on markers of health status, before extending the analysis to explore the association of specific factors across life. Finally, Chapter 8 will conclude the thesis, reviewing key themes and findings, addressing the limitations of the methods used and formulating policy implications.

Figure 1.1: Conceptual diagram of the relationship between self-reported and performance-based measures of physical capability



2. Conceptualising physical capability

One of the challenges faced when researching physical capability is the absence of a common language within the literature relating to this concept. The choice of terminology depends in part on the underlying conceptual framework used by the authors. Much of the existing research has focused on the negative end of the spectrum of physical capability (i.e. loss of function), and therefore to fully engage with the literature it is necessary to place the concept of physical capability within the context of the existing functional ability frameworks. Functional ability is used in this chapter as an umbrella term to describe physical ability at all conceptual levels. This chapter aims to review the existing conceptual frameworks and critically assess the constructs and terminology used within them, before clarifying the definition of physical capability used within this thesis.

There are three main conceptual approaches to functional ability that underpin the existing frameworks: the medical, social and biopsychosocial models of disability. The medical model of disability conceptualises the loss of functional ability as a characteristic of the individual, as the consequence of a specific health condition that can be "treated" through intervention by medical professionals (47;48). In contrast to this medical approach, which places the onus on the individual, the social model of disability perceives the loss of functional ability as a problem generated by society. The wider context is crucial to the experience of functional ability, as inhospitable features of the social and physical environment create the difficulties experienced, and political and social reforms are required to remove these barriers (47;48). The biopsychosocial model integrates components from both the medical and social approaches, with functional ability considered to be the result of biological, personal and social factors (47). It is the biopsychosocial approach that underpins the two most commonly used functional ability frameworks within the literature: "The Disablement Process" based on Nagi's seminal work (49-51) and the World Health Organisation's (WHO) International Classification of Functioning, Disability and Health (ICF) (48). The following section will describe the history behind the two frameworks and outline the conceptual features of each.

2.1 Conceptual frameworks used within the literature

2.1.1 Nagi's Disablement Model

In 1965, Nagi first proposed a framework of disability consisting of four key concepts: active pathology, impairment, functional limitations and disability (50). These four concepts form a sequential theoretical pathway (38) from a physiological or biochemical abnormality (pathology) to activity restrictions of an individual within a wider social environment (disability) (see Figure 2.1). The concepts depicted at each stage of the pathway operate from the cellular to societal level under the umbrella term of "disablement". The first stage in the pathway has been termed "active pathology", and operates at the lowest level. Active pathology can be defined as the disruption of normal cellular processes combined with the loss of homeostasis, which would ordinarily have enabled the organism to recover (31;47). The continued disruption of these processes causes structural and functional abnormalities within certain body systems, and Nagi identified these abnormalities as "impairments" (47). These impairments in turn contribute to "functional limitations" which are conceptualised by Nagi as the difficulties experienced when individuals attempt tasks considered to be essential for daily living (31;38). The distinction between the final two stages of the pathway (functional limitations and disability) is the scale at which they have been conceptualised. Functional limitations are focused at the level of the "whole organism or person", and therefore consider the ability of an individual in a situation free environment, whereas disability incorporates the broader physical and social environment experienced in the "real world".

Figure 2.1: Nagi's Disablement Model

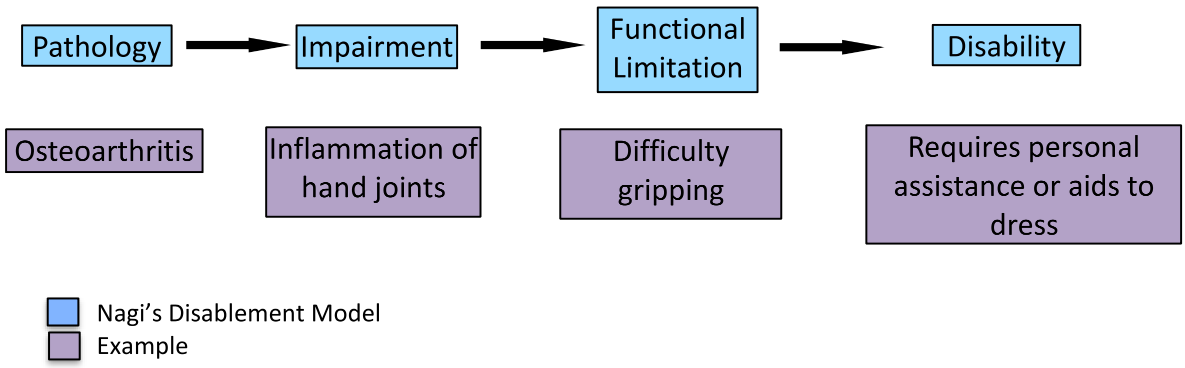


Figure adapted from Guralnik and Ferrucci 2003 (38)

To help clarify the separate stages depicted in Nagi's Disablement Model, it is useful to apply the framework to a hypothetical scenario or example (see Figure 2.1). For example, consider a woman in older age who has osteoarthritis. The osteoarthritis is the active pathology within the Disablement Model framework. As the condition progresses, the osteoarthritis causes inflammation of the hand joints, a structural abnormality, and this is considered to be an impairment. The woman will have difficulty gripping, defined as a functional limitation, due to the impairment, and this will affect her ability to complete activities of daily living. At an individual level, the woman will have limited functional ability to dress herself due to her restricted dexterity. Therefore, without suitable adaptations such as aids or personal help, she may no longer be able to maintain independent living; this would be considered a disability.

2.1.2 International Classification of Impairments, Disabilities and Handicaps

Nagi's disablement model was not widely used until the late 1980s (31), by which time a second framework had been proposed within the literature: the International Classification of Impairments, Disabilities and Handicaps (ICIDH) (52). Whilst Nagi's Disablement model had been based on sociological theory (53), the ICIDH was created with an underlying interest in taxonomy (47). The ICIDH was developed as a corresponding framework to the WHO's International Classification of Diseases, producing a classification system of function and disability that results from disease (47). In a similar manner to Nagi's model, the ICIDH framework has a linear causal pathway from disease to "handicap" (see Figure 2.2). In this framework, disease leads to impairments, which are defined as the loss or abnormality of bodily structure and functions (54) and have been conceptualised at the organ level. Disability, the next stage in the pathway, has been defined as the inability to perform activities considered normal for everyday life (54). Unlike Nagi's model, where the term "disability" was used to conceptualise the final stage in the pathway, the ICIDH perceived disability at the individual level and places "handicap" at the end of the pathway. Handicaps are conceptualised as limitations in the ability to fulfil socially and culturally defined 'normal' roles (31).

Figure 2.2: International Classification of Impairments, Disabilities, and Handicaps

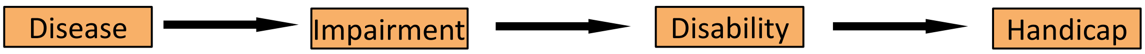


Figure adapted from WHO 1980 (52)

2.1.3 The Disablement Process

As the literature around disability and functional limitations developed, there was a growing appreciation that the linear pathways proposed within the existing frameworks had limitations. One of the key concerns was that the process of "disablement" does not always occur in a linear, uniform manner, because there are sociocultural and personal factors that can modify the course and pace of the pathway (31). To address this concern, Verbrugge and Jette (31) extended Nagi's framework to incorporate these factors, and named their model "The Disablement Process" (see Figure 2.3). The additional sociocultural and personal modifying factors were split into three categories: risk factors, extra-personal factors and intra-personal factors. Risk factors were described as the predisposing, longstanding personal characteristics and attributes of an individual which were present before the start of the disablement process, and which could increase the chance of developing functional limitations or disability (31). During the disablement process, intra-individual factors (those focused at the individual level) and extra-individual factors (the physical environment and social context) operate either as exacerbating factors or interventions (47). Interventions are factors introduced during the disablement process which have the effect of reducing, or sometimes reversing, the negative consequences of the process (31). For example, behavioural changes as part of a coping strategy or assistive devices such as a walking stick. Exacerbating factors work in the opposite direction, accelerating the rate of functional decline or stimulating the onset of a further decline, for example the negative side effects of medical surgery (31). It should be acknowledged that some factors, such as obesity, could be categorised as both risk factors and intra-individual/exacerbating factors depending on the timing within the disablement process.

Figure 2.3: The Disablement Process

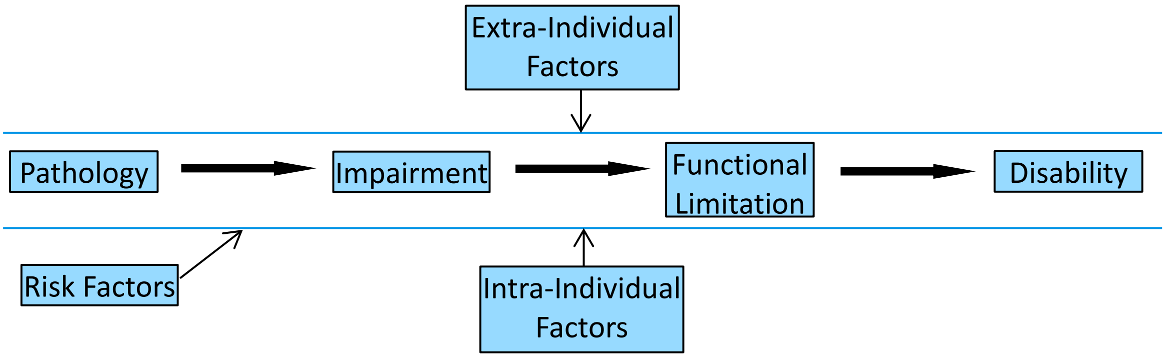


Figure adapted from Verbrugge and Jette 1994 (31)

2.1.4 International Classification of Functioning, Disability and Health

Verbrugge and Jette were not the only authors to develop the existing frameworks. In 2001, the WHO released the International Classification of Functioning, Disability and Health (ICF) (48) to address criticisms raised within the literature of the ICIDH. The ICF responded to three main failings of the ICIDH (55). Firstly, the ICF highlighted the importance of the environment and the interaction it has with the key concepts of function and disability. Secondly, the ICF recognised the dynamic nature of function and disability, choosing bi-directional relationships to depict the complex interrelations between the key concepts (see Figure 2.4). The third change was the development of a new language, using more positive terms to describe function and disability. This new language was created in part because terms such as “handicap” were no longer considered politically appropriate (56), but it also emphasised the value of including higher functioning individuals within studies and the significance of the variation across the full spectrum of function (37). Within the ICF framework, functioning is conceptually perceived as a continuum (57), and this model, combined with the new language, encourages research to move away from the traditional focus on the negative end of the functioning spectrum towards a focus on maintaining maximum levels of physical ability (58;59).

Figure 2.4: International Classification of Functioning, Disability and Health

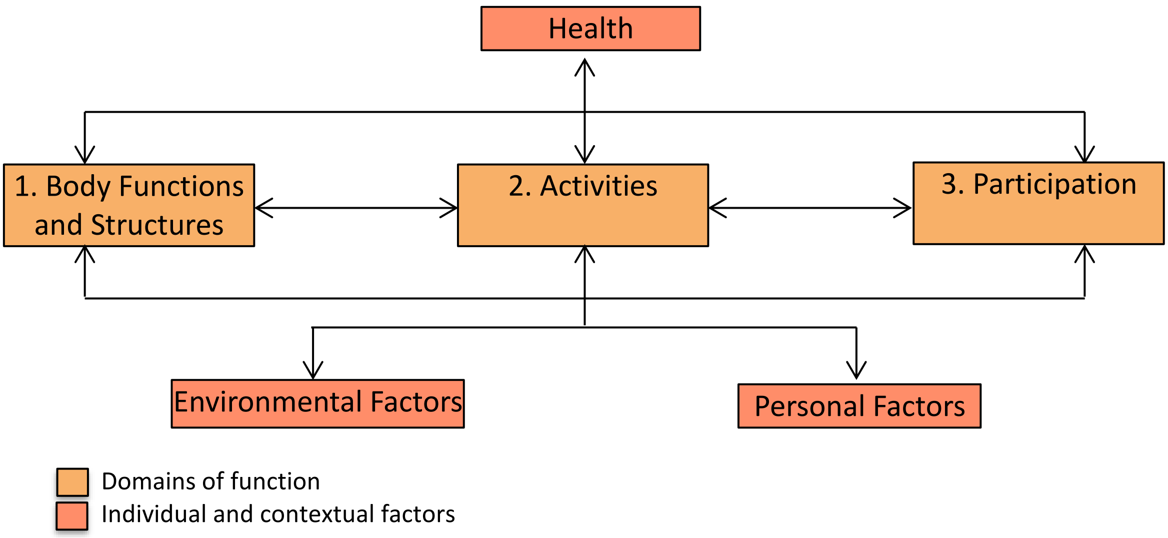


Figure adapted from WHO 2001 (48)

The ICF framework consists of three main components: 1) body functions and structures, 2) activities and 3) participation (see Figure 2.4). In line with other frameworks, each of these components or domains of function have been conceptualised at a different level, starting with body functions and structures at the body system level, through activities at the level of the whole person, up to participation, which incorporates the complete social and physical environment (47). An individual's functional ability, considered from any of the three domains, is determined by the complex interactions between the health condition of the individual and contextual factors (environmental and personal factors) (48). The relationships between the domains and these factors are multidirectional, as are the relationships between the three domains, reflecting the potential feedback effects within the system.

Qualifiers are used within the framework to assess the presence and extent of functional decline in each domain (47;60). In the domain of body functions and structures, the qualifier is based on the presence and severity of any specified impairments (47). For the domains of activities and participation, two qualifiers are provided: capacity and performance (60). The first qualifier, capacity, assesses an individual's ability to carry out a specified task or action. The aim is to capture the highest potential functioning level of an individual within a controlled environment at a given time point, and consequently indicate the ability of the individual adjusted for their environment. In comparison the second qualifier, performance, assesses what an individual can achieve in their usual environment, and therefore reflects the "lived experience" (48). The WHO suggest that the domains of activities and participation should be considered in unison when evaluating the functional ability of an individual, because each task or action has the potential to be classified in either domain, and the two qualifiers should be used to discriminate the data collected (60).

2.1.5 Comparison and critique of frameworks

The evolution of Nagi's Disablement Model and the ICIDH framework into the Disablement Process and ICF frameworks provides the theoretical context for the concept of physical capability. The current frameworks however need to be critically analysed and evaluated before the conceptual definition of physical capability used in this thesis can be clarified.

Although the Disablement Process and ICF frameworks use different vocabulary, the underlying constructs are similar (47;57). When the two frameworks are presented next to each other (as shown in Figure 2.5), it is possible to see how the two models align conceptually. Both frameworks conceptualise functional ability at three corresponding levels (61), and it is at each of these levels that the constructs have been matched. The lowest level at which the two frameworks can be aligned is the body systems level; the concept of 'impairments' from the Disablement Process model and the 'body functions and structures' domain of the ICF model. Compared to other domains within the frameworks, these are closer to the medical model of disability. In both frameworks, functioning has been conceptualised at the individual or whole person level, where an individual's ability is considered within a standard environment. Accordingly, Nagi's functional limitations component and the ICF's activity domain have been matched within the literature. At the highest level, disability and participation are aligned.

Compared to the other domains used in the two frameworks, these components have been most heavily influenced by the social model of disability, with each item conceptualised at the societal level, reflecting how people operate in the lived environment.

Figure 2.5: Alignment of conceptual frameworks

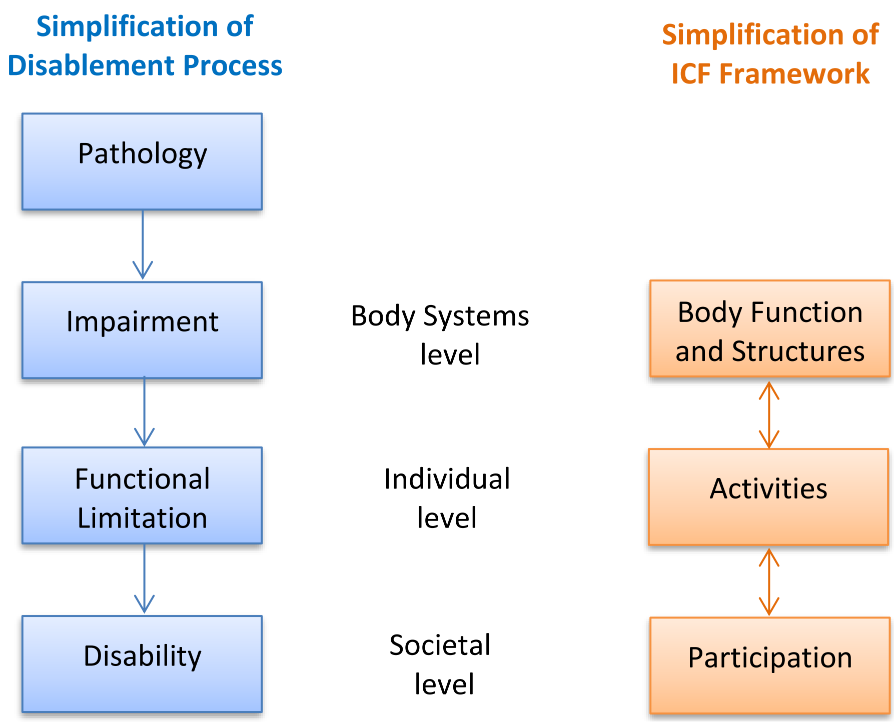


Figure adapted from Freedman 2009 (58)

Both frameworks have evolved in response to developments within the literature, with one of the key areas of discussion focusing on the dynamic nature of functional ability. There has been widespread recognition that the traditionally held view of disablement as a linear trajectory is too simplistic (47;56;62). The original versions of both frameworks assumed a causal linear progression from one domain to the next, implying that each domain was the direct consequence of the preceding domain. Within both frameworks however, the three key domains have been conceptualised at different levels, meaning that it is plausible for them to operate simultaneously (53), and that the interactions between the domains are likely to be bi-directional (43). The linear model of disablement was also heavily criticised because it was believed to discriminate against those with poor functional ability, as the causal connection between domains implied that individuals were responsible for their own disablement at the societal level (62). Instead, disablement should be considered as a result of the complex interactions between the individual and environmental factors in the past, present and future. In recognition of this, the two frameworks have developed from the original uni-directional models, with later models incorporating sociocultural, personal and environmental factors.

The interaction between individual and environmental factors, and current functional domains within the two frameworks, will also influence the life course trajectory of physical capability. The interaction between the functional components of the framework may create feedback loops, with either a positive or negative outcome. Feedback can lead to a downward spiral of limitations in physical capability associated with a specific health condition, or in some circumstances it can stimulate a new secondary pathway of disablement based on a related health condition (31). However, feedback can also have positive effects, potentially leading to limited levels of recovery. Interventions can be based on personal factors such as coping strategies and peer support, or on environmental factors such as the introduction of aids and specialised equipment. Such adaptations have been used successfully by people with low levels of physical capability (31); in fact, it has been suggested that the improvement observed in the prevalence of disability at older ages in the United States (US) is partially due to successful adaptation through the increased use of assistive devices (4;63).

There are limitations in trying to produce a “one model fits all” framework for such a dynamic and complex construct as functional ability. The framework has to account for both progressive and catastrophic disability. Those who experience progressive disability experience a steady decline in levels of physical capability over a substantial period of time, usually in mid to late life. In contrast individuals with catastrophic disability experience a rapid and severe decline in physical capability as a result of an accident or illness. Cases of catastrophic disability are more often recorded at younger ages, however evidence of catastrophic disability has also been observed in older populations (64). Without frequent observations, those with catastrophic disability in old age may not be captured as they are more likely to have died and been lost to follow-up before any relevant data on their disability status has been collected, but they remain a relevant group of interest within any theoretical framework. Levels of recovery are also different between the two categories of functional decline, and consequently the two groups will have very different disablement pathways which may require different theoretical approaches (59).

The two frameworks have been accepted and used to a different extent globally and within specific areas of research. In the US, the gerontological research community have generally favoured Nagi's Disablement Process model (59), whilst European researchers have been more open to the ICF framework (61). The Disablement Process framework is a useful tool in gerontological research given the progressive nature of the model and the typical physical capability pathway experienced by people in older ages. The pathway depicted within Nagi's framework has been tested empirically, and a large body of evidence has accumulated over the years since publication to support the theoretical model (38;59). However, given the acceptance of the dynamic nature of physical capability, it is important to note the bi-directional relationship between the functional domains of the disablement process. Whilst functional limitations often lead to disabilities, evidence of the reverse has also been recorded in the literature (43). As a more complex picture emerges, especially when incorporating the interaction of personal and environmental factors, the ICF framework has also presented itself as a viable framework for use in gerontological research. The ICF framework was developed through a lengthy process of revisions, in consultation with relevant experts from across the globe, and aims to provide the research community with a common language to enable communication and stimulate further discussion (47;56;57).

Despite the US focus on Nagi, there has been a gradual rise in the use of the ICF framework within the North American literature in the last ten years, with the Institute of Medicine (65) endorsing the framework in 2006. A systematic review of papers relating to the ICF framework (66) reported that a total of 672 papers were published in the first eight years after the framework was released, with most published in the last year of the review period (2009). There is however some concern in the wider literature about the extent to which the conceptual ICF framework has been successfully operationalised in a research or clinical setting (67). For example, over 30% of the papers included in the systematic review were conceptual discussions of the framework, and a further 10% only mentioned the ICF (66). The North American literature has tended to produce more theoretically focused publications, whilst researchers in Europe have moved towards the operationalisation of the ICF, using the relevant sections of the classification system particularly from 2007 onwards. ICF-based tools were developed or validated in 45 of the papers included in the review (7% of the total), with the majority focused on neurological and musculoskeletal diseases (66). For example, one study in Germany created a 58-item questionnaire using the mobility, self-care and domestic life sections of the ICF classification system, and reported that the tool demonstrated good reliability and high construct validity when used in rehabilitation for patients with musculoskeletal, cardiac or neurology disease (68). Having established that it may be possible to operationalise the ICF framework, the current challenge is to establish a shorter and more effective tool for use in the wider research community (69) which can be applied across a range of clinical and general population samples, and across a variety of ages.

One of the remaining obstacles in the operationalisation of the ICF framework is the lack of conceptual clarity between the domains of activity and participation. Within the theoretical framework, activity and participation are presented as distinct concepts with clear definitions; however, in the classification scheme used to operationalise the framework, the two domains are combined into one component (70). Activity is conceptualised at the individual level, where intra-individual psychological and biological factors are the predominant influence on functional ability (55). Therefore, within biopsychosocial theory, the concept of activity has strong influences from the medical model of disability (56). Whereas the concept of participation incorporates social factors (55), and is conceptualised at the societal level to encapsulate the interaction between these social factors and the personal factors, which will influence functional ability. Participation captures an individual’s “lived experience” of physical capability within their usual environment, and consequently is linked to perceived quality of life (56). When developing the ICF framework, WHO chose not to separate activity and participation on the basis of individual and societal level (48) “given the international variation and differences in the approaches of professionals and theoretical frameworks” (p.16). However, in the context of physical capability, the distinction between the individual level and the societal level of functional ability may help researchers distinguish between activity and participation.

It has been suggested by some that the role of the environment could be used as a way of separating the concepts of activity and participation. This seems inappropriate given environmental factors are a key concept within the ICF framework, and thus can be applied to all functional domains. It is often perceived that the environment has a larger influence on the concept of participation (56), as it incorporates the wider social and political environment when assessing functional ability. However, all examples of physical functioning must take place in an environmental context (54). Within the domain of activity, there is an implicit assumption of a "standard environment". Caution should be applied to the notion of a "standard environment" as this is likely to vary across different study settings (54). Consequently, it has been suggested that the definition of activity should be altered so that it considers functional ability as an action or task completed by an individual in a specified context (61).

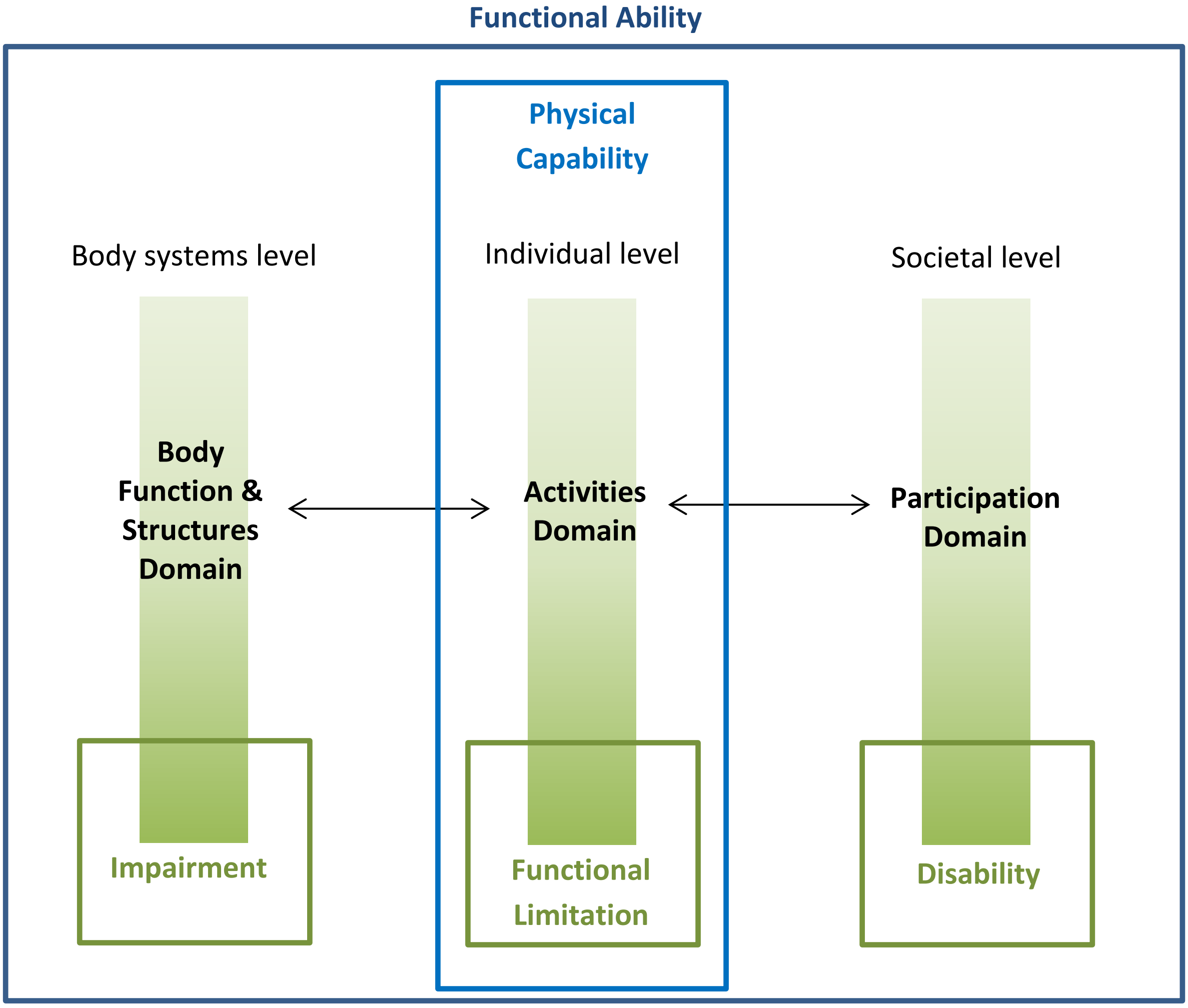
It should be acknowledged that sometimes it may be difficult to distinguish between the different concepts within a functional ability framework, as there is overlap (38). It is also worth noting that this problem is not unique to the ICF framework. Within the Nagi framework, the distinction between concepts on the pathway may not always be apparent, particularly between the concepts of functional limitation and disability (38), which correspond to the ICF's activity and participation components. Verbrugge and Jette (31) have suggested the use of specific vocabulary to help distinguish between the two concepts, with the words "task" or "action" relating to the concept of functional limitations, and the words "role" or "activity" associated with the concept of disability.

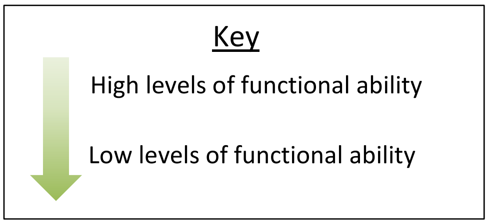
2.2 Conceptualising physical capability within this thesis

2.2.1 Conceptual definition of physical capability

Having critically assessed the existing theoretical frameworks, the following section outlines how physical capability has been conceptualised in this thesis. A conceptual diagram has been produced to summarise the relationship between the different components incorporated within the Disablement Process and the ICF frameworks, and how these relate to the concept of physical capability (see Figure 2.6).

Figure 2.6: Conceptualising the individual components of functional ability





Within this thesis, physical capability is perceived to be a spectrum of ability, with high function at one end and functional limitations at the other (see Figure 2.6). A spectrum encapsulates all levels of ability found within the population, not just those identified as being "disabled". When placing the concept of physical capability within the context of the disability frameworks previously discussed, it is important to note that the ICF framework is applicable across the whole spectrum of ability, because the constructs have been specifically developed for universal application (48). Within the ICF framework, the activity domain and the associated capacity qualifier fit conceptually with how physical capability is perceived in this thesis. Whereas, in the Disablement Process framework the functional limitations construct conceptually fits within the lower part of the spectrum of physical capability (see Figure 2.6). Both of the constructs, activity and functional limitations, focus on the capability of an individual to complete tasks of daily life rather than the lived experience (31;48).

The lived experience of functioning conceptually incorporates the social environment, and is linked to the ICF's concept of participation and Nagi's concept of disability. It can be useful to consider the terms suggested by Verbrugge and Jette to help distinguish the concepts of participation and disability from physical capability. Functioning is perceived within the concepts of participation and disability as an "activity" or "role", set within a broader social environment which influences levels of functioning through cultural, political and socioeconomic factors. When conceptualising physical capability, the environment provides the context to the "action" or "task" in question, without influencing the levels of function measured. Ideally the environment should be explicitly stated, rather than implying standard conditions within the protocol.

2.2.2 Defining the vocabulary to be used throughout this thesis

As highlighted throughout this chapter, the choice of vocabulary used in the context of physical capability research is highly important, as it can enhance conceptual clarity. The terminology used in the conceptual diagram shown in Figure 2.6 will be used throughout this thesis and Table 2.1 has been constructed to provide a brief synopsis of the key concepts. It is important to note the difference between the activity domain of the ICF framework and the term "activity", used to indicate functioning within the wider context of society, akin to disability. To distinguish between the two terms, the functional domain of the ICF framework will always be referred to as the "activity domain".

Table 2.1: Glossary of terms to be used throughout thesis

Term	Definition/Detail
Functional Ability	Umbrella term used to describe physical ability at all conceptual levels
Physical Capability	The complete spectrum of physical ability conceptualised at the individual level • Akin to the activity domain of the ICF framework • Denoted with the terms "task" and "action"
Disability	A social construct akin to the negative end of the participation domain, which incorporates the social and political context • Denoted with the terms "role" and "activity"

2.2.3 Applying the conceptual definition of physical capability to the appropriate measures

Having established how the concept of physical capability applies to the existing frameworks and defined the language to be used throughout this thesis, it is important to apply this knowledge to the two types of measures used to assess physical capability: performance-based measures and self-reported measures.

As previously outlined in Section 1.2, the two approaches used to measure physical capability have different characteristics, which it has been suggested result in the two sets of measures being conceptually linked to different theoretical constructs (39;43;71-73). Performance-based measures are perceived to measure physical capability, because they assess an individual's ability to perform a task in a standard environment. In contrast, some authors suggest that self-reported measures reflect the concept of disability rather than physical capability, because they assess an individual's physical ability within the context of their daily lives. It could be argued however that it is the behaviour under assessment, not the technique that defines which concept of functional ability a measure is focused on (38). Both performance-based and self-reported measures can assess behaviours such as walking and stair climbing which are considered to be actions, akin to physical capability, or they can assess behaviours such as preparing a meal or housework which are considered to be activities, and therefore relate to disability. Performance-based measures of disability are uncommon within the literature, and can be identified with relative ease. In contrast, to distinguish self-reported measures of physical capability and disability, the wording and context of measures are particularly important. Within the literature, measures of self-reported physical capability are available, although few instruments focus exclusively on this concept (29). Appropriate measures of physical capability should therefore be chosen, with care taken to ensure that each reflects the underlying construct.

3. Dataset and descriptive analysis

Having established the theoretical framework through the conceptual definition of physical capability, the aim of this chapter is to provide context to the thesis by introducing the dataset and exploring the levels of physical capability experienced by study participants in early old age. The dataset used in this thesis is the MRC National Survey of Health and Development (NSHD).

3.1 Introduction to the dataset

The NSHD is the oldest national British birth cohort study (74). The cohort was set up in 1946 to address relevant post-war health and social policy questions regarding declining fertility rates and use of the midwifery service (75). The primary areas of research interest have changed as the cohort has aged, with a shift towards physical and mental function in midlife (75). This focus has continued because the importance of physical and mental capability has been emphasised in the ageing process and is of particular significance to the study members as they move towards older age (74).

3.1.1 Sampling framework and follow-up

A sample of 5,362 births were selected from all single, legitimate births that occurred in one week in March 1946 in England, Scotland and Wales (74). A stratified sampling technique was used to select the births, based on paternal occupational class. At the time of sample selection, the majority of the British workforce was employed in manual occupations. Consequently, all eligible births to women with husbands in non-manual and agricultural employment were selected, and one in four of all births to women with husbands in manual employment were randomly selected (75).

Since birth in 1946, the study members have been followed up 24 times: approximately every two years in childhood, and at ages 26, 36, 43, 53 and 60-64 (75). A new data collection round is currently underway in the field, when participants are aged 68-70. Of the original 5,362 participants, 2,943 (55%) were still eligible for further data collection at age 68-70. The remaining participants had either died (17.6% N=945), had previously refused (11.6% N=624), lived abroad (10.8% N=579) or were untraceable (7.5% N=402). As the current data collection round is still ongoing, the analysis in this thesis will focus on the data collected in 2006-2010, when the study members were 60-64 years old, and earlier data from the archive.

As part of the 60-64 data collection, an initial postal questionnaire was sent to all eligible study members (74). Between two months and two years after the postal questionnaire (74;76), study members were invited to visit one of six Clinical Research Facilities (CRFs). If a participant was unable or unwilling to travel to one of the CRFs, they were offered a home visit as an alternative (74). Of the 3,163 eligible study members, information was obtained during the 60-64 data collection round from 84% of participants, with 2,462 responding to the postal questionnaire, 1,690 visiting one of the CRFs and 539 receiving a home visit (77).

3.1.2 Benefits of using the NSHD

There are several advantages of using the NSHD dataset; some are generalisable to all analyses conducted using these data, whilst others are thesis specific. The data collected within the NSHD are considered to be of high quality due to the rigorous training of data collectors (generally research nurses), fact checking against hospital records and specialist coding (75). Over the life course of the cohort a wide range of high quality data have been collected prospectively, covering, amongst others, physiological, psychological and socio-economic factors. The breadth and completeness of the data enables researchers to study the influence of a large variety of factors, with appropriate account taken of potential confounding, modifying or mediating factors. The inclusion of a home visit in the 60-64 data collection round continued to enhance the dataset by reducing bias that could be introduced through social and health disadvantages of those who could not attend the CRFs (74).

Whilst these are generic advantages of the dataset, there are also several thesis-specific benefits of using the NSHD. For example, age is homogenous within a birth cohort, such as the NSHD, which overcomes the challenge of controlling for age, known to be strongly associated with physical capability. At 60-64 years the age of the cohort is relatively young compared to ageing studies in the existing literature, which means the thesis captures physical capability near the start of the typical declining trajectory. Previous research in concordance and discordance of physical capability has tended to focus on older age groups, despite the positive implications for interventions in younger age groups. The relative young age of participants within the NSHD has the additional benefit of reducing the impact of confounding by co-morbidity compared to studies in older populations, who tend to have a higher prevalence of chronic disease. For example, previous work (77) has shown that, within the NSHD study population at age 60-64, the most common clinical disorder was hypertension, with 53% of participants categorised as hypertensive. In comparison the prevalence of hypertension, using a similar clinical definition, was 83% among participants of the Newcastle 85+ study (78). Although extensive loss of function is uncommon in community-dwelling adults in early old age, the sample size is relatively large compared to other studies of physical capability, so the physical capability measures employed within the NSHD should capture an individual's ability wherever it lies on the spectrum, including those with the lowest and highest levels of capability.

In addition, the physical capability measures used within the dataset correspond to the conceptualisation of physical capability within this thesis. Several performance-based and self-reported measures of physical capability have been collected within the same data collection round for the last two waves of data collection, enabling two cross-sectional comparisons. Within each data collection round, the self-reported and performance-based measures were captured at the same time as part of the clinical assessment, removing any temporal discrepancy between the two approaches. The self-reported measures have been captured at three time points, which provides more detailed information about the declining levels of physical capability experienced across two time intervals.

It should be acknowledged that whilst there are numerous benefits to using the NSHD there are also some limitations. For example, like many cohort studies the NSHD dataset has experienced loss to follow up which could introduce bias and limit the sample size available. A more detailed discussion of the limitations and benefits of the NSHD, and how these specifically relate to this thesis can be found in Section 8.4.1.

3.2 Physical capability variables

As outlined in the previous section, the physical capability variables collected in the NSHD are a benefit of using this dataset in this thesis. The following section will describe these variables in more detail, providing information about how the data have been collected from midlife to early old age at ages 43, 53 and 60-64.

3.2.1 Performance-based measures

Data ascertainment

Performance-based measures of physical capability were collected when study members were 53 and 60-64 years old. Under the supervision of a trained nurse, and following standardised protocols (79;80), three performance-based measures were collected at both time points: grip strength, chair rise time and standing balance. Timed get up and go (TUG) was included as an additional performance-based measure in the 60-64 data collection round. For each performance-based measure a score was recorded if the participant was willing and able to perform that particular test. If the study member was unable or unwilling to complete the task the reason was noted and the nurse progressed onto the next task. In this thesis, the subgroup of participants who were unable to complete the performance-based tasks due to health reasons were incorporated in analyses where appropriate. Such individuals were considered to be an important group who could still provide viable information about their levels of physical capability despite being unable to complete the test, and their exclusion may have introduced bias to the dataset. Details of how these individuals were incorporated into the analysis will be outlined in the appropriate chapters.

Study protocol

The study protocols used for the four performance-based measures are outlined in the following section.

To measure grip strength at age 60-64, participants were asked to squeeze the handle of a Nottingham electronic dynamometer: three tests in their dominant hand and three in their non-dominant hand. The dynamometers are accurate, linear and stable to ±0.5kg, with each machine calibrated at the start of the assessment using a back-loading rig (79). The nurse provided strong vocal encouragement throughout the test to elicit maximal performance. The protocol differed slightly between data collection rounds, with only two repetitions in each hand for the data collection round at age 53. In this thesis, the maximum value achieved by each participant was used in the data analysis.

In both data collection rounds, study members were asked to perform 10 chair rises with their arms folded, as quickly as possible, and the time taken to complete the task, was recorded in seconds. The test was conducted using an armless, straight backed hard chair, with the seat approximately 46cm above the floor (79). To ensure that all performance-based variables used within this thesis followed a similar scale, with low values representing poor functional performance, the time taken to complete the task was converted into a measure of speed (number of rises per second).

Two tests of standing balance were performed by the study members in both data collection rounds. Initially participants were asked to fold their arms, stand on their preferred leg and raise their other leg a few inches above the ground, holding this position for a maximum of 30 seconds. The task was then repeated with the participants closing their eyes. For each test, the maximum time, up to 30 seconds, that the participant maintained the balance position was recorded. In line with previous work on these measures (79;81), a decision was made to use the performance time from the eyes-closed balance test within the analysis for this thesis, as a substantial ceiling effect was noted for the eyes-open test.

In the 60-64 data collection round, participants were asked to get up from a chair, walk three metres, turn around, walk three metres back and sit back in the chair to measure TUG, and the time taken to complete the task was recorded. The task was performed at normal pace, and participants were allowed the use of personal aids if required (N=24), although the use and type of aid were noted. Help from another individual was prohibited. In a similar manner to the chair rise times, the time taken to complete the task was converted to speed (m/s) to invert the scale.

3.2.2 Self-reported measures

Data ascertainment

Self-reported measures of physical capability were captured when participants were aged 43, 53 and 60-64 years old, with more detailed assessments at ages 43 and 60-64 (see Table 3.1). The questions used within the NSHD are based on the Office of Population Censuses and Surveys (OPCS) survey of disability (82). Participants were asked to report difficulties with a range of tasks including walking, bending, cooking and bathing. An initial screening question was asked to detect those who reported difficulty with each specific task of interest. This screening question was followed by one or more detailed questions asking about severity for those who indicated they experienced difficulty. Depending on the number and detail of the severity questions, the number of response categories varied across the self-reported variables. The wording and format of the questions used varied slightly between the three data collection rounds; for a complete list of all questions employed to assess self-reported physical capability, see Appendix 1.

At ages 53 and 60-64, the self-reported measures of physical capability were ascertained before the performance-based measures were undertaken to ensure that participants' responses were not influenced by their performance under standardised conditions.

Table 3.1: Tasks included in the self-reported assessments at ages 43, 53 and 60-64

	Age (yrs)		
Physical capability tasks	**43**	**53**	**60-64**
Walk 400m	✓	✓	✓
Climb flight of 12 stairs	✓	✓	✓
Grip (holding, gripping or turning things)	✓	✓	✓
Bend down and straighten up	✓		✓
Maintain balance (need to hold onto something)	✓		✓
Shop (carry heavy load in each hand)	✓		✓
Bath and shower	✓		✓
Get in and out of chair	✓		✓
Get in and out of bed	✓		✓
Dress and undress	✓		✓
Get to the toilet	✓		✓
Use toilet	✓		✓
Wash hands and face	✓		✓
Feed self (including cutting food)	✓		✓
Prepare hot meal			✓
Heavy housework			✓

Data management

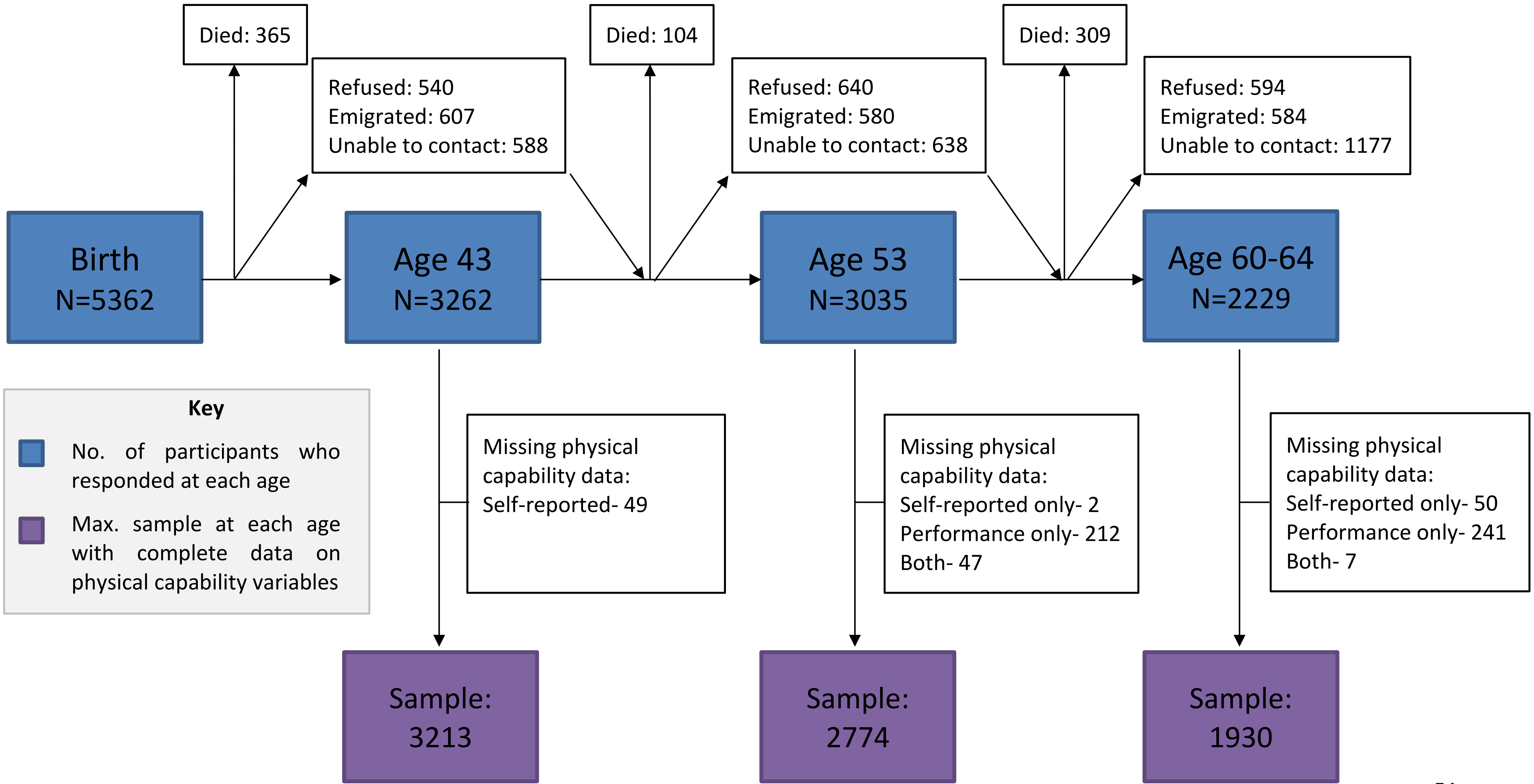
Within this thesis, the self-reported physical capability variables of primary interest come from the 60-64 data collection round; however, the data required cleaning before any analysis could be undertaken. During the 60-64 data collection round, paper-based questionnaires were used by the nurses to collect information on the self-reported variables. This approach meant that the correct gating procedure between questions could not be enforced. For each self-reported variable, the initial screening question and corresponding severity questions were cross-tabulated to check for inconsistencies. If any discrepancies were identified, these were corrected using consistent logic across all of the variables. For example, those who reported no difficulty in the screening question but went on to answer the severity question with any level of difficulty above the minimum level had the initial question recoded to indicate that they experienced difficulty. For cases where there was any uncertainty, the original transcript was consulted. One variable (self-reported bending) followed a different format to the other questions (see Appendix 1). A different approach was required to clean this variable as a result of the question format, with participants asked up to four gated questions regarding their bending capability. Instead of using cross-tabulation to check for inconsistencies between two parts, the cleaning process for this variable considered the question as a whole, starting with the last part and working backwards to ensure that each participant took a valid pathway through the entire question.

There were 16 self-reported physical capability variables within the 60-64 data collection round, assessing functional ability across a range of tasks from walking to bathing. However, a decision was made to continue the main analysis on discordance only with variables that closely matched the conceptualisation of physical capability within this thesis. Consequently, only the self-reported measures that represent "actions" rather than "activities" have been included. Using these criteria, the five self-reported measures selected for this thesis were: walking, stair climbing, gripping, balance and bending. For two of these variables there was a direct corresponding performance-based measure (grip strength and balance), and walking was the main component of TUG.

3.2.3 Study sample for analysis

The size of the study sample varies throughout this thesis depending on the specific analysis in question, and will be outlined in each of the following chapters. In general, the number of participants available for any given analysis initially depends on the response rate for the particular data collection round of interest (see Figure 3.1). To obtain the maximum sample size for each age, it is important to take into account missing data for the self-reported and performance-based physical capability measures. The maximum sample available for analysis in this thesis using complete data was 3,213 participants at age 43, 2,774 participants at age 53 and 1,930 participants at age 60-64 (see Figure 3.1). It is important to note that participants were able to leave and re-enter the study between waves of data collection. For example, 110 participants present in the maximum sample at age 60-64 had not participated in the two previous data collection rounds.

Figure 3.1: A flow chart to summarise how the sample for this thesis was obtained



3.3 Descriptive analysis

Throughout this thesis the primary variables of interest are the performance-based and self-reported physical capability variables collected at age 60-64. The aim in this section is to explore the levels of physical capability experienced by study members at this age through descriptive analysis of these variables. The data presented in this chapter uses the maximum number of observations available. Data in subsequent analyses (presented in the following chapters) have lower numbers of observations due to missing data on covariates. A detailed discussion on missing data and the implications of the chosen methods of handling this can be found in Section 8.4.1.

3.3.1 Methods

Performance-based measures

Before analysing the values obtained for each of the performance-based tests, an initial analysis compared the number of participants unable to complete each task for health reasons. A chi-squared test was used to ascertain whether there was a significant difference in the proportion recorded as unable to complete the tasks according to the sex of the participant.

Where a valid performance value was recorded for each of the performance-based tests, sex-specific histograms and summary statistics were calculated. For variables that were considered to be normally distributed, a t-test was used to compare the mean values produced for each sex. If the performance-based variable was not normally distributed, then a non-parametric equivalent to the t-test (a Mann-Whitney test) was used instead.

Self-reported measures

An initial analysis was conducted to calculate the proportion of participants who reported at least one difficulty across all five self-reported measures and the proportion who reported difficulty in all five measures. The individual measures were then analysed by calculating the sex-specific proportion of the study population who reported each level of difficulty. Mann-Whitney tests were used to compare the sex differences in these proportions, as the data were ordinal and not normally distributed.

3.3.2 Descriptive results of performance-based measures at age 60-64

Between 1.7% and 6.2% of participants were unable to complete each of the four performance-based tasks at age 60-64, with the highest proportion unable to complete the chair rise task (see Table 3.2). For each performance-based task, more women than men were recorded as unable to complete tasks for health reasons, but this sex difference was only significant for grip strength.

Table 3.2: Proportion of participants unable to complete each performance-based task for health reasons at age 60-64

Performance-based task	Unable to complete task N (%)#		
	All	Male	Female
Grip strength	49 (2.31)	13 (1.28)	36 (3.27)*
Chair rise	136 (6.19)	62 (5.90)	74 (6.45)
Standing balance	89 (4.04)	40 (3.79)	49 (4.27)
TUG	34 (1.65)	19 (1.93)	15 (1.39)

base population = all those with value obtained and those unable
Sex difference significant at * P=0.01 level

The sex-specific summary statistics and histograms produced for the four performance-based measures can be seen in Figures 3.2a-3.2d. Of the four measures, grip strength presents the most clearly defined normal distribution, and the other three histograms appear to have positively skewed distributions. The histograms for both chair rise and TUG speed have been slightly distorted by outlying values captured at the top performance level. The outlying values were within an acceptable range for both variables, and thus a decision was made not to exclude these values. As there were less than 10 outlying values for each measure, it was decided that in further analyses both of these variables would be treated as a normal distribution. When the potentially outlying values are excluded, the sex specific skewness and kurtosis values (see Table 3.3) suggest that the distributions of TUG and chair speed are comparable to that of grip strength.

Table 3.3: Sex specific skewness and kurtosis values for performance-based tasks, excluding potentially outlying values

Performance-based tasks	Men		Women	
	Skewness	Kurtosis	Skewness	Kurtosis
Grip strength	0.11	3.20	0.26	2.78
TUG	0.49	4.17	0.09	4.35
Chair speed	0.73	3.77	0.57	3.81
Balance (eyes closed)	3.06	14.1	3.26	16.5

The histogram for standing balance clearly displays a positively skewed distribution, with over half of the study sample unable to hold their balance position for more than four seconds, suggesting evidence of a floor effect. It is also worth noting that there was some evidence of a small ceiling effect, with a grouping of individuals visible at 30 seconds (the maximum time). The ceiling effect was, however, far smaller than that observed in the eyes-open standing balance test, where almost all (72%) of the participants reached the maximum time limit (not shown). In subsequent descriptive analyses (presented in the remainder of Chapter 3 and Chapter 5) the skewed nature of the distribution of standing balance time was taken into account, by ensuring that appropriate summary statistics, such as IQR and medians, and non-parametric statistical tests were used.

From the distribution and calculated summary statistics, the histograms appear broadly comparable across both sexes, with the exception of grip strength, where there is a clear distinction between men and women. For example, the mean grip strength value for men is 1.5 times greater than the corresponding value for women, and there is a wider distribution of values for men than women. When appropriate statistical tests were applied however, the average values for all four performance-based measures were found to have significant sex differences (see Figures 3.2a-3.2d).

Figure 3.2a: Mean and distribution of grip strength at age 60-64 by sex

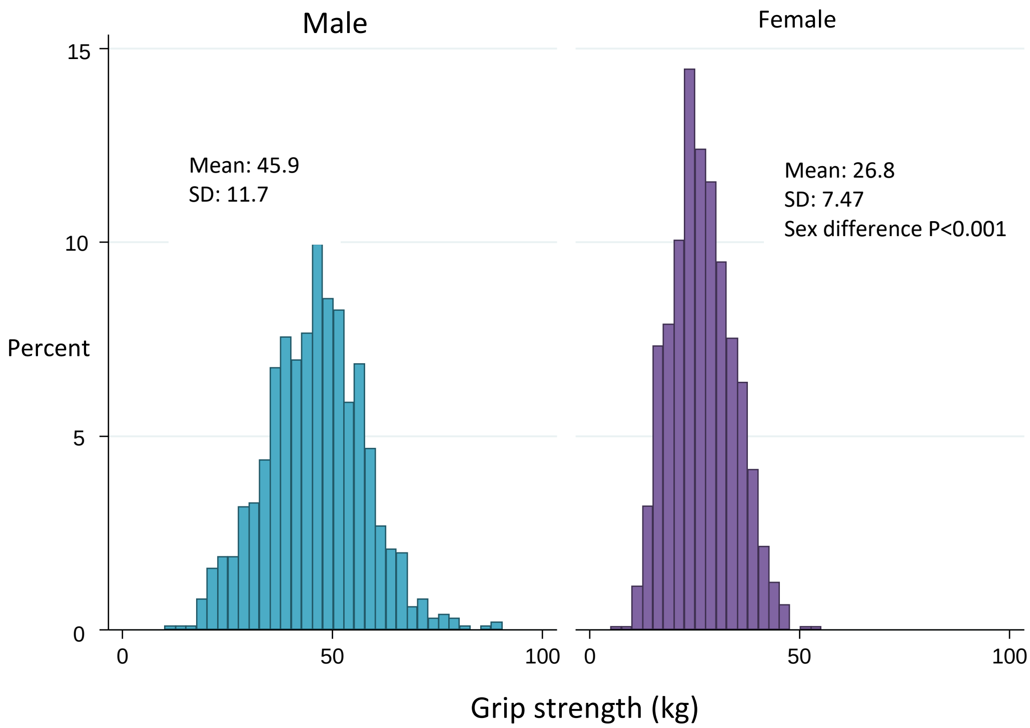


Figure 3.2b: Mean and distribution of chair rise speed at age 60-64 by sex

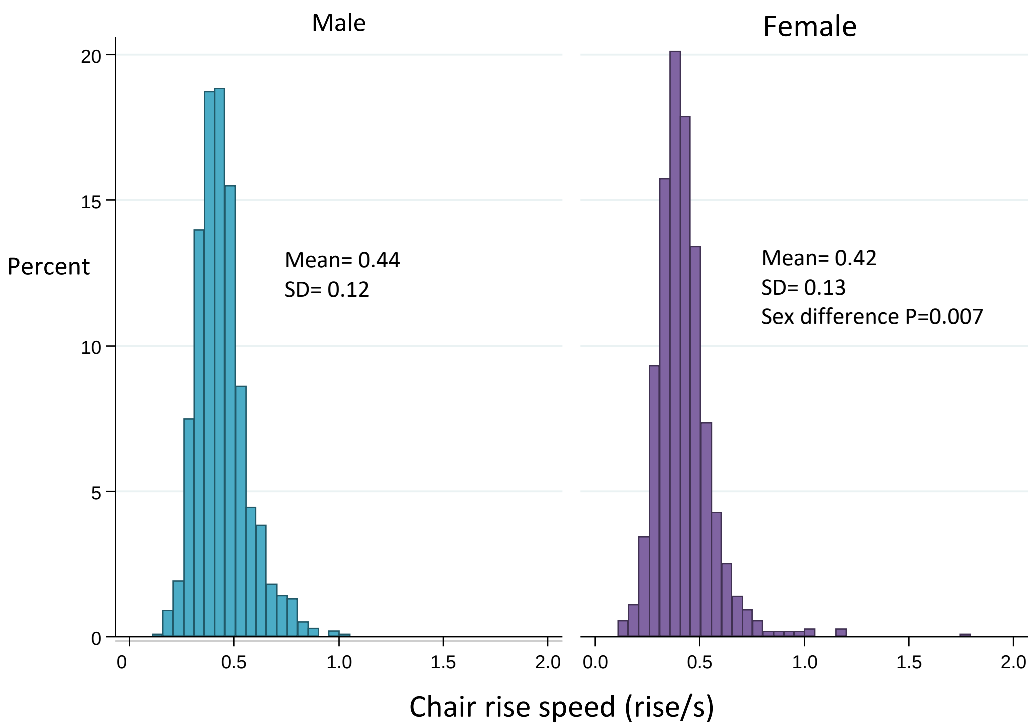
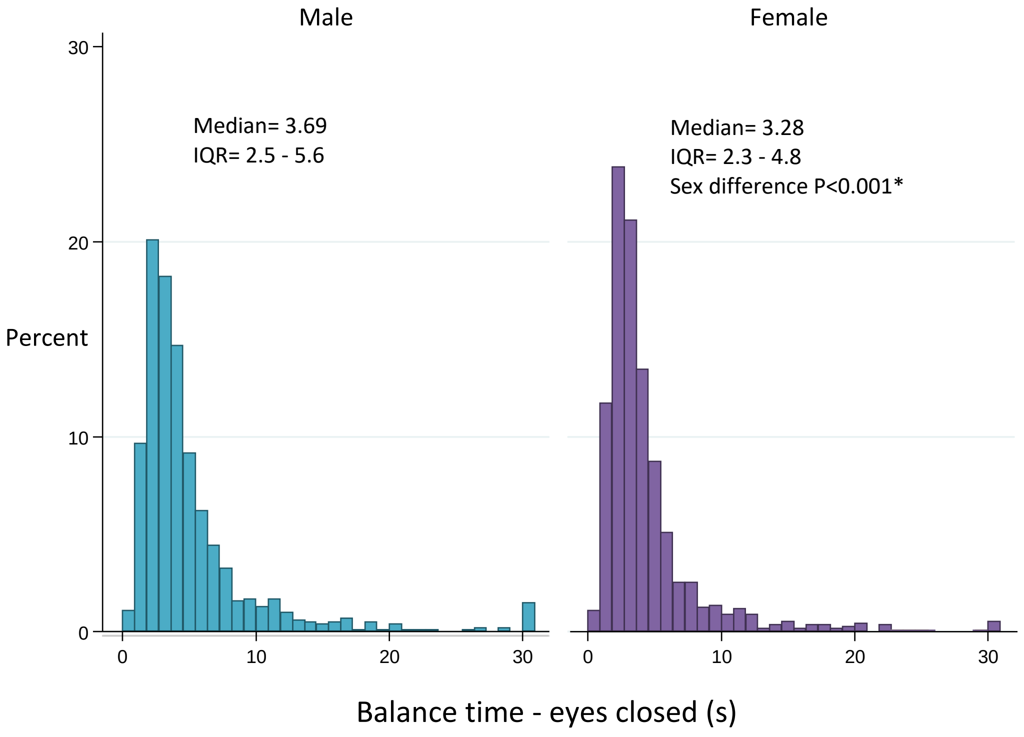
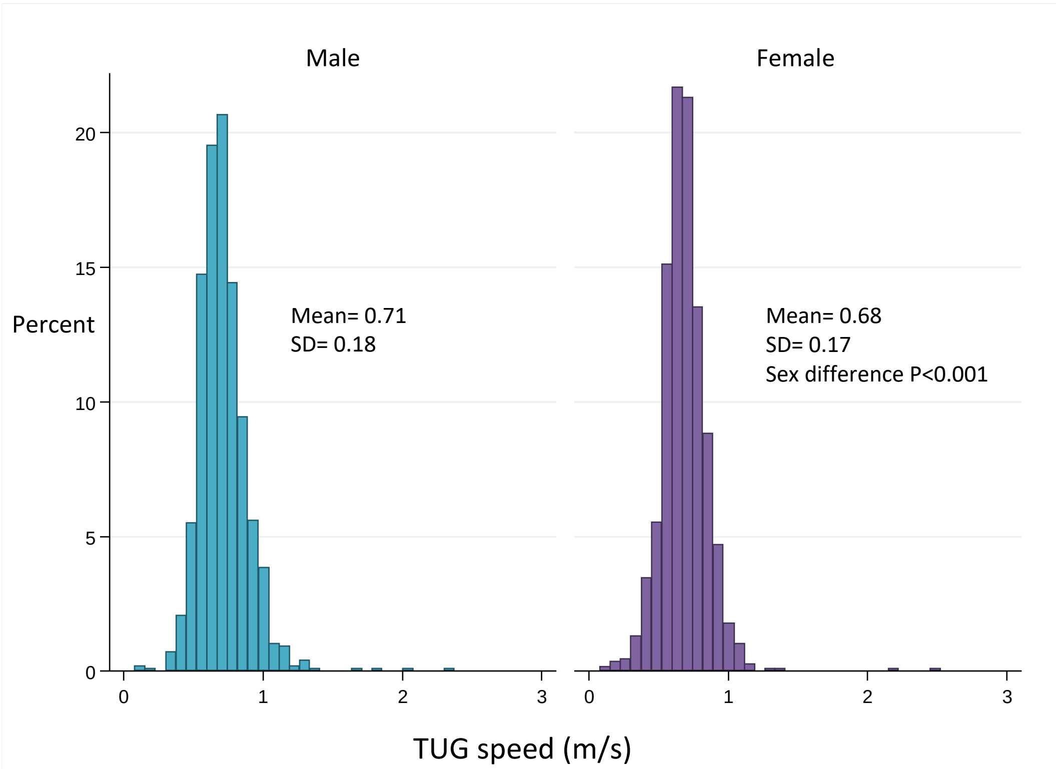


Figure 3.2c: Mean and distribution of standing balance (eyes closed) at age 60-64 by sex



*Mann-Whitney significance test applied (non-parametric)

Figure 3.2d: Mean and distribution of TUG speed at age 60-64 by sex



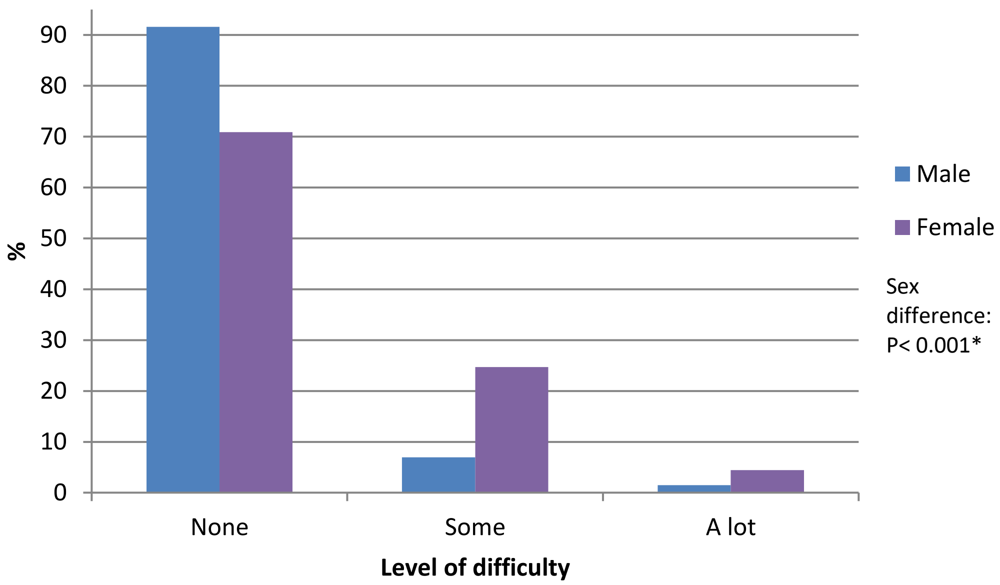
3.3.3 Descriptive results of self-reported measures at age 60-64

Across the five self-reported measures selected for the main analysis on discordance in this thesis, 40% (N=899) of the study population at 60-64 reported difficulty in at least one of the tasks, and 3.3% (N=81) of participants reported difficulty in all five. These results suggest that there may be some evidence of a ceiling effect in the self-reported measures, with the majority of participants reporting no difficulty with the five tasks assessed. However, given the relatively young age of the sample, it was expected that many participants would still maintain high levels of physical capability (as defined by no difficulty). The headline statistics mask a substantial sex difference, as only 28% of men reported difficulty in at least one task compared to 52% of women. This sex difference was considered when analysing the self-reported measures individually by stratifying the data (see Figures 3.3a-3.3e).

Amongst those who reported difficulty for each task, a graded pattern was observed, with fewer people reporting the more severe levels of functional limitations. The pattern was most clear when only three response categories were provided. Expanding the number of response categories provides more detailed information about the level of physical capability experienced by participants, and this information can be harnessed in the concordance analysis (see Chapter 5) when comparing the self-reported measures of physical capability with the performance-based measures. However, due to the negative focus of self-reported measures, which capture the loss of function, additional response categories only distinguish those who report difficulty with the tasks of daily living and thus have lower levels of physical capability. The majority of individuals in this study did not report any difficulty in early old age, and it is hard to distinguish their level of physical capability, as these individuals may have a broad range of physical capability levels but do not feel that their experience of physical capability passes the threshold to be classified as “difficulty”.

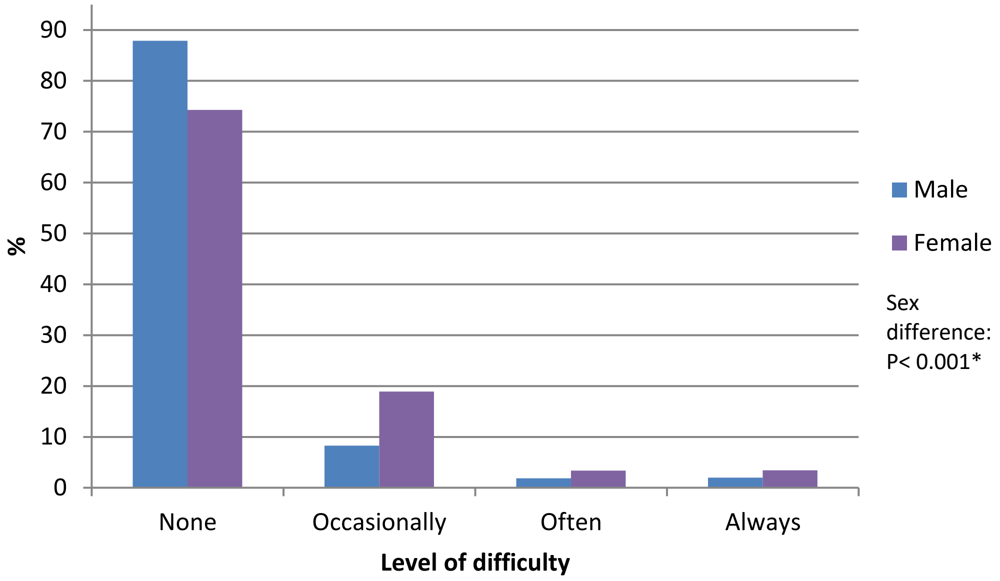
With the exception of self-reported walking, there was a significant sex difference in each of the self-reported physical capability measures within the NSHD. This sex difference can be highlighted by comparing the proportion of men and women who report no difficulties across the five self-reported measures. The number of response categories differs in each self-reported variable, so the “no difficulty” category is the only response consistent across all five measures. More than 70% of the population reported no difficulty for any given physical capability task; however, this ranged from 86-92% for men compared to only 70-88% for women. This translates into a prevalence of reported difficulty between 12-30% for women and 8-14% for men. It is also interesting to note which tasks had the highest and lowest prevalence for each sex. For women the lowest prevalence of difficulty was observed in self-reported walking and the highest in self-reported gripping. In contrast, self-reported gripping had the lowest observed difficulty prevalence in men and self-reported stair climbing had the highest prevalence.

Figure 3.3a: Proportion of participants reporting difficulty with gripping at age 60-64 by sex



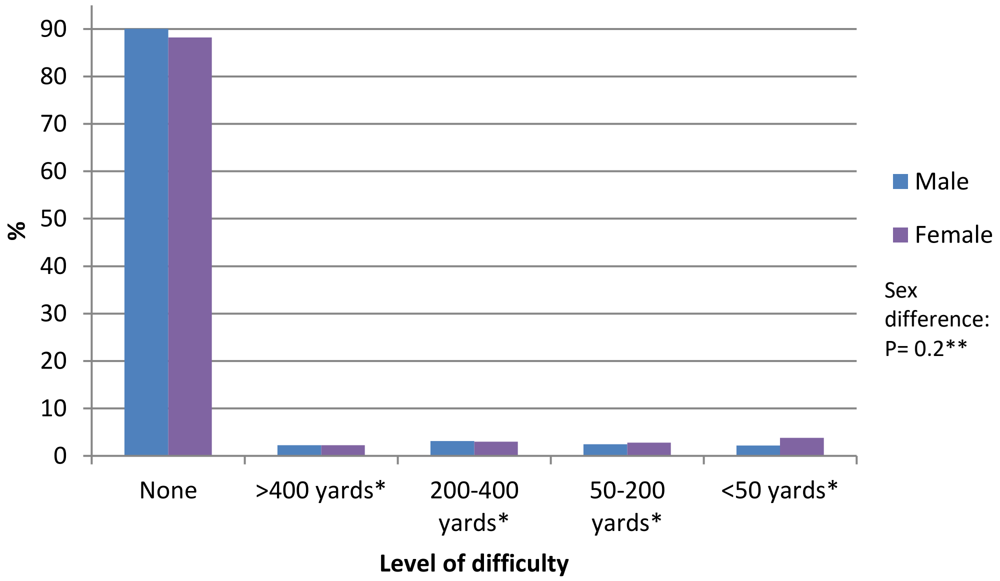
*Mann-Whitney significance test applied (non-parametric)

Figure 3.3b: Proportion of participants reporting difficulty with balance at age 60-64 by sex



*Mann-Whitney significance test applied (non-parametric)

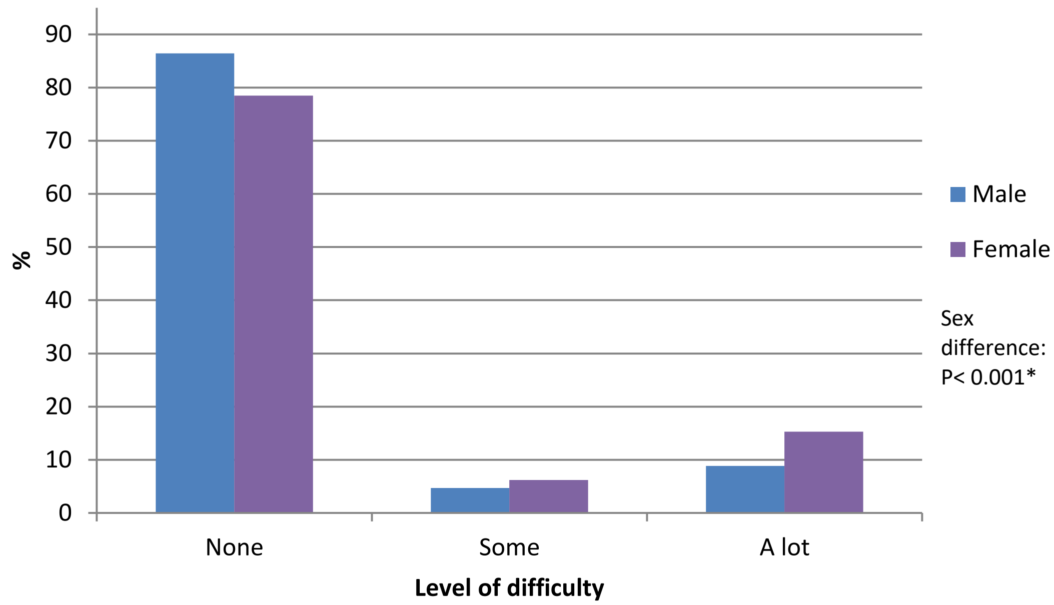
Figure 3.3c: Proportion of participants reporting difficulty with walking at age 60-64 by sex



*Distance participants can walk without stopping or severe experiencing discomfort

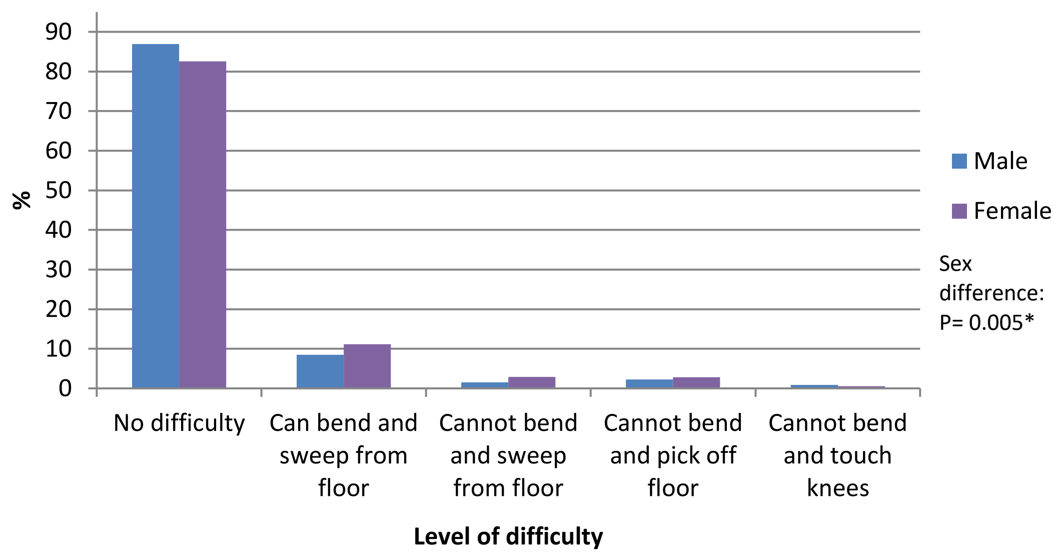
**Mann-Whitney significance test applied (non-parametric)

Figure 3.3d: Proportion of participants reporting difficulty climbing stairs at age 60-64 by sex



*Mann-Whitney significance test applied (non-parametric)

Figure 3.3e: Proportion of participants reporting difficulty with bending at age 60-64 by sex



*Mann-Whitney significance test applied (non-parametric)

When looking across the whole spectrum of self-reported capability there is a clear trend observed in self-reported gripping, with the proportion of the population reporting difficulty decreasing as the level of difficulty increased (see Figure 3.3a). The trend is however less clear in the other self-reported measures, particularly in self-reported stair climbing, the only other three-category variable (Figure 3.3d), in which a higher proportion of the population report "a lot" of difficulty rather than "some" difficulty. A downward trend is less clearly defined in the other self-reported variables, which have more than three categories, due to the small number of participants in each of the categories.

3.4 Summary of results

The results presented in this chapter provide descriptive data on physical capability across the whole spectrum of ability at age 60-64 in the NSHD, using self-reported and performance-based measures. Within the sample, some individuals have retained high levels of capability in early old age, for example those who report no difficulty with all five of the assessed self-reported tasks or those who reached the maximum 30 second time limit for the standing balance test. Equally a number of individuals with low levels of physical capability can be identified at age 60-64, for example those who report severe limitations with the tasks assessed or are unable to complete the performance-based tasks for health reasons. The wide variability in observed physical capability reflects the findings in the literature (11;21).

Both performance-based and self-reported measures demonstrated significant sex differences, with the exception of self-reported walking. In general women were found to perform less well than men and were more likely to report difficulty. The sex difference was particularly noticeable for gripping, both grip strength and self-reported gripping, and this was taken into account in future analyses (as outlined in subsequent chapters).

4. Hierarchy of Loss

In addition to the cross-sectional descriptive analysis shown in Section 3.2, the self-reported physical capability measures were used to explore the trajectories of loss in functional ability within the NSHD population. The main aim of this analysis is to identify which physical capability tasks individuals first report difficulty with. This knowledge can be used to help confirm the selection of physical capability measures made in Chapter 3, based on the conceptual definition of physical capability, which will be used in the rest of this thesis.

In order to address the aim of this chapter, the first half of the chapter focuses on a review of the existing literature that examines the order in which individuals report difficulty with the tasks of daily living. Based on the findings of this review, the second half of the chapter explores the hierarchy of loss within the NSHD population from midlife to early old age. The analysis in this chapter incorporates all 16 of the self-reported measures assessed at age 60-64, in addition to relevant self-reported physical capability measures collected at earlier ages. The work presented in this chapter has been published (83).

4.1 Literature Review

4.1.1 Introduction

As noted in Chapter 1, previous research has established that, from midlife onwards, population average levels of physical capability decline steadily with increasing age (1;80). This process is often gradual, progressing over many years, and is in contrast to the sudden loss of capability associated with a specific incident such as a stroke or accident (84). During this time, difficulties develop when performing a range of common physical tasks of daily living in a typically observed sequence (85;86), often described as the hierarchy of loss.

Previous studies of the hierarchy of loss have included tasks covering mobility, basic activities of daily living (BADLs) and instrumental activities of daily living (IADLs), to incorporate the full range of potentially relevant aspects of physical capability associated with maintenance of independence. Self-reported measures have been used predominantly, as they are readily available in a number of established studies and surveys, and because they have clear clinical significance as they specifically capture loss of capability. Consequently, such measures are easily translatable to policy makers, who use them to aid decisions regarding the delivery of health services and the maintenance of independent living (87). However, as noted in Chapter 1, self-reported measures of physical capability can be problematic, as there are large ceiling effects in community-dwelling populations (88). The impact of this ceiling effect can be attenuated through the construction of a hierarchy, which aims to distinguish between the levels of physical capability experienced across the population, in a way that provides more information than a simple summary measure, as it accounts for the relative difficulty of each task (88;89).

By establishing a hierarchy of loss, and focusing on individuals when they first report difficulty with tasks towards the start of the hierarchy, it may be possible to identify individuals who are at high risk of subsequent functional decline and consequent loss of independence. These individuals are likely to be suitable candidates for interventions to prevent further decline and promote recovery. As they have yet to develop the more severe limitations associated with tasks towards the end of the hierarchy, preventative interventions are likely to be most effective. By the time an individual reports difficulty performing tasks at the bottom of the hierarchy the opportunity to successfully intervene is limited, whereas there is evidence within the literature to suggest that individuals who report difficulty with tasks towards the top of the hierarchy can recover (90). Although the proportion of the general population who report severe limitations in early old age is relatively low, such individuals often require formal or informal care (91), which can place a burden on health and social care systems. The hierarchy of loss can be used to help prioritise resources and service allocation to assist with the maintenance of independent living (86) by implementing interventions to prevent further decline in physical capability.

4.1.2 Typically observed sequence of loss

A comprehensive literature review was undertaken to examine the existing literature on the hierarchy of loss. Papers were included in this review if they were published within the past 20 years, incorporated individual physical capability tasks and assessed a hierarchical order of the tasks included. In total 10 papers were included, and these have been summarised in Appendix 2.

The hierarchy of loss typically starts with a decline in tasks associated with strength, balance and co-ordination (90), such as climbing stairs. These are often complex tasks involving the use of lower extremities. The hierarchy of loss then progresses through to tasks which are considered less complex, many of which are linked to upper limb movement. For example, tasks associated with manual dexterity, such as feeding yourself, are the final tasks with which individuals report the experience of difficulty (90). There are individual differences between published studies in terms of the tasks included in the hierarchy (see Appendix 2), but generally the same sequence is observed at the population level across the literature. It has been suggested that this hierarchical ordering of tasks may reflect the development of biological and psychological skills in infancy, with items lost hierarchically in the reverse order to which they were gained during childhood (92). A distinction is made between the basic physiological tasks required to survive and tasks with a social or cultural element. For example, a child will first develop the skills required in order to survive, including essential tasks such as feeding. Only once these skills have developed do the more social elements evolve, such as bathing and dressing. Conversely, an adult with declining physical capability will first report difficulty with these more complex social tasks, before reporting difficulty in feeding. This ordering is inherently logical from an evolutionary perspective, as an individual maintains the key skills required to survive for the maximum length of time.

The pattern of physical capability loss described has been consistently observed across different study populations and specific sub-populations, using a variety of different methods. Several national populations, with samples from across Europe (including France, UK, and Italy) (85-87;93) and the US (91;94;95), but also from non-Western cultures such as Taiwan (90;96), have been used to successfully demonstrate the hierarchy of loss. Both cross-sectional and longitudinal study designs have yielded similar results. Those studies using cross-sectional study designs have focused on Item Response Theory-based methods (85;86;94;97;98), for example Mokken scaling, or have used descriptive statistics to analyse the prevalence of reported difficulty (95;96). The hierarchy of loss is often implied in cross-sectional studies on the basis of the prevalence of individuals reporting difficulty with specific tasks, with checks in place to ensure that people do not report difficulty with tasks lower down the hierarchical order without also reporting difficulty in the more complex tasks. In these studies it is presumed that the tasks with the highest prevalence of reported difficulty are the first tasks with which individuals report difficulty.

In comparison, longitudinal studies focus on incidence, through the use of survival analyses to establish the age of onset for difficulty with a specific task, in a given study population (87;90;91;93). These methods ideally require the continuous collection of data, but practical limitations mean data can usually only be collected at discrete time points. The analytical methods can be adapted to account for the design of data collection, for example using interval-censored data to produce discrete survival analysis curves rather than the continuous Kaplan Meier curves traditionally used for survival analysis (91). However, some authors argue that it is inappropriate to use discrete methods, as they assume an equal time interval between data collections points for all individuals which is often not feasible in reality (87). Instead the maintenance of continuous methods is recommended, particularly if the time between observations varies considerably. Although the different methods and approaches used to analyse the hierarchy of loss, both longitudinally and cross-sectionally, have their strengths and limitations, the hierarchical ordering of tasks is consistently observed across these studies, demonstrating the robust nature of the findings.

Whilst the majority of previous studies have focused on community-dwelling populations, there have also been studies of specific sub-populations. These sub-populations have included condition-specific groups such as patients with cancer, multiple sclerosis or rheumatoid arthritis (97), as well as individuals residing in institutional care homes (94). Several studies conducted sensitivity analyses to explore the impact of including certain subgroups on the hierarchical ordering of tasks, and found no difference in their results when excluding those who live in institutions (86;87).

With regards to sex differences, as expected, women generally reported higher levels of difficulty across all of the tasks assessed, and at younger ages than men. Despite these established differences in the experience of functional decline, the hierarchical ordering of the tasks was fairly consistent between men and women in studies that assessed the sex difference (86;87;91). Some differences were observed however, with women initially reporting difficulty with tasks involving strength, whilst men reported difficulty with endurance-related tasks earlier, such as walking (86;96). Where other differences were noted, it was often due to inconsistent patterns in the ordering of tasks, especially between toileting, transferring from a chair and dressing (87;91).

The stereotypical pattern associated with the hierarchy of loss appears to be independent of pathological cause (85). This may be partially explained by the complex nature of physical capability decline, which is often multi-causal in old age. However, there is some evidence that whilst the hierarchical ordering of tasks does not change with the underlying pathological cause, the number of comorbidities experienced may affect the speed of the transition through the hierarchy. For example, in one study of over 10,000 primary care patients aged 65+, who were followed up for 10 years, those with no comorbidities reported difficulty with physical capability tasks 1-5 years later than those with one comorbidity, who in turn reported difficulty 1-6 years later than individuals with two or more comorbidities (93).

4.1.3 Critical appraisal of previous literature

Although previous studies have consistently established a hierarchy of loss across several different populations, using various different study designs, there are a number of points for discussion that need to be raised. The majority of studies have included a limited number of tasks within the hierarchy (usually fewer than 10), with the most comprehensive study including 17 tasks (86). Comparisons between studies are therefore limited, as few papers include the same tasks, which makes it challenging to establish a consistent hierarchy across the whole range of tasks assessed. The tasks most commonly included within the papers reviewed were bathing, transferring from a chair and dressing. When focusing on these three tasks, the hierarchical order reported starts with bathing, considered the most complex task, and ends with dressing, which was the last of the three tasks with which individuals reported difficulty.

However, there is substantial variation around the ordering of tasks included in the hierarchy of loss, which could be due to the differences in the wording of questions used to assess any one specific task. It has been suggested that there are more variations in the phraseology of the tasks assessed than there are in the specific tasks included within studies which focus on the hierarchy of loss (87). For example, there is a choice when assessing physical capability: questions can be phrased either as “can you” or “do you”. By asking “can you”, the researcher is attempting to capture the extent of an individual’s capacity, rather than “do you”, which may incorporate the social context of an individual (42). The second option would be akin to the concept of disability, and would reflect the lived experience of an individual rather than their capability (see Chapter 2).

It is also important to consider the response options available to each question and how these might influence the results. Despite most surveys offering a range of response options, from “no difficulty”, through various levels of difficulty to “unable to perform”, nearly all of the studies reviewed chose to dichotomise these options for analysis into “no difficulty” versus “difficulty” (regardless of the level of severity). At least one paper (97) reported that this dichotomisation strengthened the hierarchical scale produced, and it would have been unsuitable to use polytomous response options for some of the analytical methods. However, the definition of “difficulty” was not consistent across the studies, with some electing to include the use of aids or help. It is possible that by including the use of aids or help that the results may be biased, as access to these will not be universally available. Therefore, the reported levels of difficulty for certain tasks within the hierarchy may actually reflect the role and influence of society, closer to the concept of disability, rather than an individual’s capability. However, there seems to be limited evidence of this bias, as demonstrated by one paper where the authors changed the dichotomisation from difficulty to those requiring help or not, and noted that the hierarchical ordering of tasks remained consistent (86).

A further limitation of the existing literature is that nearly all of the published studies in this area have focused on populations over the age of 75 years. Although this is an appropriate age for clinicians to provide support for individuals already experiencing the effects of poor physical capability, a life course approach suggests it may be beneficial to also consider individuals at a younger age. By exploring the hierarchy of loss in early old age, individuals can be identified whilst they are in the early stages of decline, when intervention may have the greatest impact; to prevent further decline and to promote recovery (90;95). However, when focused on populations over the age of 75 years, there are fewer individuals at the earlier stages of decline, and interventions aimed at prevention and recovery would be expected to be less effective (96).

When studying the hierarchy of loss, it is important to consider the dynamic nature of functional decline. As previously discussed, the progressive decline of physical capability is often multi-causal, and consequently lower levels of capability may be temporary or reversible as the underlying pathological causes interact (90). As a result of this dynamic trajectory, not everyone will follow the hierarchy of loss, with studies reporting that between 4-24% of participants did not follow the proposed hierarchy (90;95). It has been suggested that, although there may be one hierarchy that the majority of the population will follow, other pathways are also prevalent within the general population, accounting for the dynamic nature of the process. Sometimes this will only involve a small change in the ordering of tasks, particularly for those tasks that do not demonstrate a strong ordering. For example, one study of 5,000 community-dwelling individuals aged 70 years or older identified over 100 different hierarchies that met the minimum criteria for scalability (28.6% of all possible combinations) when exploring the order of loss for six tasks of daily living (99). The hierarchical ordering of tasks was fairly consistent across the 103 hierarchies identified, with individuals first reporting difficulty with bathing, going on to report difficulty with feeding and continence last; only the ordering of the three remaining tasks (dressing, going to the toilet and transferring from bed) changed (99).

In part, the presence of multiple hierarchies in a population may be explained by the differential rates of recovery reported for the various tasks incorporated within the hierarchy of loss. Although rates of recovery are reasonably low, ranging from 0.5-3.3% in one of the studies reviewed (87), they appear to be higher in tasks towards the top of the hierarchy with which individuals first experience difficulty. This may explain why it can be hard to distinguish between items towards the start of the hierarchy. As well as considering the role of recovery, an alternative reason why some individuals may deviate from the expected hierarchy is the presence of certain medical conditions (85).

For example, individuals suffering from arthritis in their hands will report difficulty with tasks associated with manual dexterity, which are usually found towards the bottom of the hierarchy, before they report difficulty with tasks involving lower limb movement which are more commonly found towards the top of the hierarchy (90). Equally those experiencing pain as a result of a medical condition may report difficulty with a specific task whilst they are still physically capable of performing that task or more complex tasks with only minor or no limitations (98). From a methodological perspective it is important that researchers do not assume that the hierarchical order applies at the individual level, unless specific assumptions have been met and appropriate models applied, otherwise the hierarchy is only applicable at the population level (88).

Some authors suggest that it may not be appropriate to assume that the hierarchy of loss is linear, where each task has a distinct loci within the hierarchy and is conditional on the individual reporting difficulty with the previous task (89). Indeed, when the relative difficulty of tasks within the hierarchy are plotted, the tasks cluster into groups or domains (86); it has been proposed that the hierarchy of loss should be based on these domains rather than single items. The domains can group tasks together based on the skills involved, for example complex tasks involving long distance mobility and balance would be in a different domain to those involving upper limb movement in a standing position, which in turn would be in a different domain to tasks involving manual dexterity in a seated position (85;86). Alternatively, some authors choose to separate tasks into domains based on whether they are mobility tasks, BADLs or IADLs. The idea is that although individuals may not report difficulty with tasks in the same order within each domain, the domains themselves will be lost in the same hierarchical order. This approach may help to account for small differences between different individuals' experience of the hierarchy of loss. The evidence seems to suggest that the hierarchy is more stable when based on domains rather than single items, as only 0.4-1.7% of participants do not fit the hierarchy when it is structured in this way (85;89). However, within the literature there are no fixed domains used consistently across the different studies because, as noted above, the tasks included vary considerably between studies. Furthermore, as many tasks will be a complex amalgamation of skills, it can be difficult to decide which domain a specific task belongs to if the choice of domain is based on the skills involved.

4.1.4 Summary of the literature

The studies reviewed have demonstrated that a consistent hierarchy of loss can be found using a variety of different study methods and populations. However, further research is still required to enhance our understanding of the hierarchy of loss. For example, the issue of whether to investigate domain or single item based hierarchies. Although a domain based hierarchy appears to capture a wider proportion of the population, and may be more appropriate to implement from a practical perspective for identifying individuals at high risk, there is still value in exploring single item hierarchies. Single item hierarchies enable researchers to gain a clearer picture of the hierarchy of loss by providing more detailed information on individual tasks to facilitate the decision about which tasks it is appropriate to group together.

When designing a study to explore the hierarchy of loss, or selecting data from a secondary source, it is important to consider the wording of the self-reported questions used to assess an individual's physical capability. A decision must be made whether to use the expression "can you" or "do you". Either of these may be appropriate from a policy perspective as the lived experience of an individual (captured by "do you") may be as informative to policy makers as an individual's physical capability. The most appropriate choice of wording depends on the research question to be addressed. The research question should also influence the decision of whether to include the use of aids and help as part of the assessment. It is important to be consistent in the wording wherever possible to limit the potential bias that may be introduced.

The articles reviewed have focused on adults predominantly over 75 years of age, but the question arises as to whether the hierarchy of loss can be established at earlier ages when there may be greater potential for intervention to prevent decline and promote recovery. Therefore, the aim of the analysis in this chapter was to explore the hierarchy of loss in the NSHD between midlife and early old age. Both cross-sectional and longitudinal approaches were used to determine whether consistency could be found between these two approaches within the same data set. The first objective was to establish the hierarchy of loss at age 60-64 years. The second objective was to test the hypothesis that the hierarchical order observed in the cross-sectional analysis at age 60-

64 would be reflected in longitudinal prevalence estimates, with tasks found to be at the top of the hierarchy at age 60-64 expected to have the greatest increase in levels of reported difficulty from age 43. The final objective was to test for evidence of progression through the hierarchy over time.

4.2 Methods

4.2.1 Cross-sectional analysis

Initial analyses were used to produce sex-specific descriptive statistics to examine the hierarchy of loss cross-sectionally when participants were aged 60-64. The analysis was restricted to those who responded to all 16 self-reported physical capability tasks assessed at age 60-64 (N=2,063) (see Section 3.2). The prevalence of reported difficulty was calculated for each of the 16 physical capability tasks assessed using the self-reported measures at age 60-64 described in Chapter 3. It was assumed that tasks with the highest prevalence of reported difficulty were the tasks with which participants first reported difficulty. To explore this idea in more detail, a variable was created to count the number of tasks with which individuals reported difficulty. It was expected that the median number of tasks individuals reported difficulty with should increase as the prevalence of reported difficulty declined.

Sex-specific Mokken scales were used to confirm the cross-sectional hierarchical order of tasks for the 16 physical capability tasks. It was important that the model assumptions of unidimensionality, single and double monotonicity were met when producing the Mokken scales (97;100). If the assumptions were not met, a sensitivity analysis was conducted, whereby alternative scales were produced by removing each conflicting item in turn. The strength of the sex-specific Mokken scales produced was assessed using the Loevinger's Scalability Coefficient (H), with values above 0.5 considered strong, and an H value between 0.4 and 0.5 demonstrating an acceptable or average strength (97;101). Two approaches were used when creating the sex-specific Mokken scales; firstly the scales were produced based on the polychotomous response categories of the self-reported measures, then the analysis was repeated using a dichotomised version of the variables (no difficulty versus any difficulty).

4.2.2 Longitudinal analysis

Having established the hierarchical order of the physical capability tasks cross-sectionally, the second stage of the analysis incorporated additional data collected when participants were aged both 43 and 53 to explore the hierarchy of loss longitudinally. The sex-specific prevalence of reported difficulty was calculated for the three tasks which were repeated in all three data collection rounds: gripping, walking 400 metres and climbing 12 stairs. The analysis was restricted to those who responded to all three data collection rounds (N=2046). It was hypothesised that the hierarchical order observed in the cross-sectional analysis would be reflected longitudinally, with tasks towards the top of the hierarchy predicted to have the greatest increase in prevalence of reported difficulty from midlife to early old age.

The final stage of the analysis explored participants' progression through the hierarchy using logistic regression to calculate sex-specific odds ratios of reported difficulty performing tasks at the bottom of the hierarchy (i.e. feeding, washing and/or toileting) at age 60-64 by reported difficulty performing tasks at the top of the hierarchy (i.e. gripping, walking and stair climbing) at age 43 (N=2,106). Participants with severe disability at age 43 were excluded (those who reported difficulty with tasks at the bottom of the hierarchy at this age (N=6)), as they were assumed to have already transitioned through the hierarchy.

An initial sex-adjusted model was run, with sex interaction formally tested. Adjustments were then made for selected covariates. Each group of covariates were added to the model separately before all covariates were incorporated into a final fully-adjusted model. This analysis was then repeated stratified by sex.

<u>Covariates</u>

Covariates were identified *a priori* and included in the analysis if they were associated with baseline levels of physical capability and the progression of physical capability decline.

Three socio-demographic factors were selected: sex, education and occupation. The highest educational level attained was recorded at age 26 and categorised into five groups: degree or higher; A levels or their equivalent (usually attained at age 18); O levels or their equivalent (usually attained at age 16); CSE, clerical course or equivalent; and none. Occupational class was recorded at age 53 (or the most recent measure available in adulthood (N=83)) and categorised using the Registrar General's Social

Classification into three groups: high (I or II); medium (III manual or non-manual); and low (IV or V).

Two lifestyle factors were selected: smoking history and participation in leisure time physical activity (LTPA) from midlife onwards. Smoking status was recorded throughout adulthood, and smoking history up to age 64 years was categorised as current, former or never smoker. Study members were asked to report how frequently they had participated in any sports, vigorous leisure activities or exercise in their spare time, per month at age 43, or in the past four weeks at ages 53 and 60-64. At each age individuals were categorised as inactive (reported no participation), moderately active (participated in relevant activities 1-4 times) or most active (participated in relevant activities five or more times) and scored 0-2 accordingly (102). These values were then summed to produce a cumulative score of participation in exercise from age 43 to 60-64 (with a range of 0 (inactive at all three ages) to 6 (most active at all three ages)).

Three health status factors were selected: obesity, depression and respiratory disease. For each of these variables, three response categories were identified that took account of baseline and incident health status: never experienced health condition of interest, present from age 43 onwards and present by age 60-64. Body mass index was calculated at ages 43 and 60-64 using measured height and weight, then dichotomised using a standard cutpoint for obesity (≥ 30 kg/m^2). Symptoms of anxiety and depression were identified using the Psychiatry Symptom Frequency scale at age 43 (103) and the General Health Questionnaire-28 at age 60-64 (104). For both scales, appropriate “caseness” thresholds (≥ 23 and ≥ 5 respectively), were used. The UK Medical Research Council’s standardised questions (105) were used to identify those with severe respiratory symptoms at ages 53 and 60-64, reflecting their experience over the past three years. Participants were considered to have severe respiratory symptoms if they reported one or more of the following: wheezy or whistling chest most days or nights, usually bringing up phlegm or coughing in the morning or during the day or night in winter for at least three months each year, or a chest infection that kept them off work or indoors for more than a week.

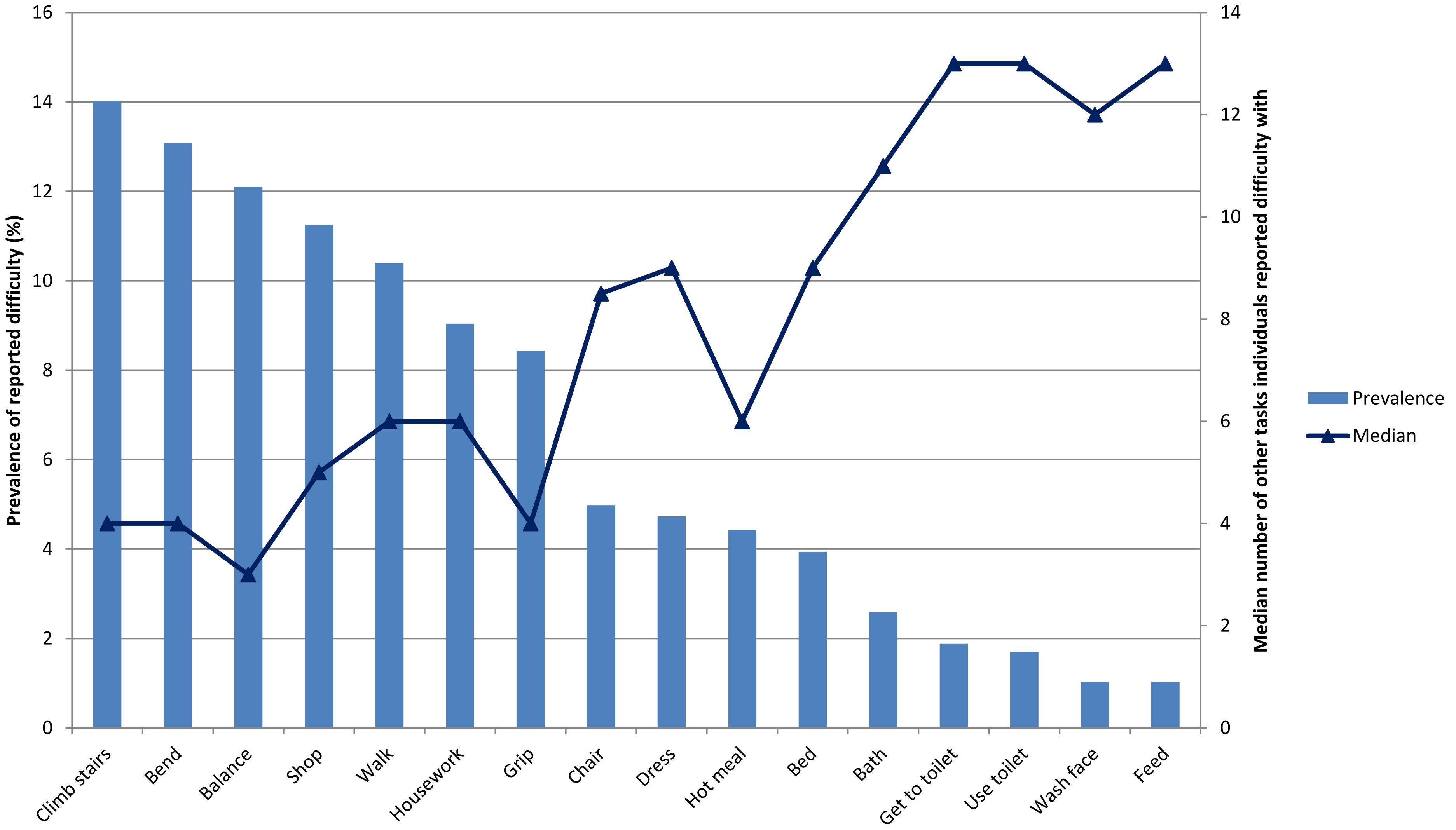
In order to minimise the potential risk of bias introduced by missing data, and to maintain statistical power, multiple imputation chained equations were used to impute the missing values of the covariates (educational level (N=99), occupational class (N=1), smoking history (N=20), participation in LTPA (N=139), obesity (N=33), depression (N=57) and respiratory disease (N=204)). Logistic regression analyses were run across 15 imputed datasets and combined using Rubin's rules (106).

4.3 Results

4.3.1 Hierarchical ordering of tasks at age 60-64

At age 60-64 nearly a third of men (30%) and over half of women (55%) reported difficulty with at least one of the 16 physical capability tasks assessed. Figure 4.1 shows the prevalence of reported difficulty for each of the physical capability tasks at age 60-64, separated for men (Figure 4.1a) and women (Figure 4.1b). Tasks associated with balance, strength and co-ordination, such as climbing stairs and carrying heavy shopping, were found to have the highest prevalence of reported difficulty. For these more complex tasks, women reported more difficulty than men, with over 20% of women and only 10% of men reporting difficulty with each of the four most prevalent tasks. Tasks associated with upper limb mobility, such as feeding yourself and washing hands and face, were found to have the lowest reported prevalence of difficulty, with less than 1% of men and women reporting difficulty with these tasks.

Figure 4.1a: Prevalence of reported difficulty with 16 physical tasks of daily living at age 60-64 for men



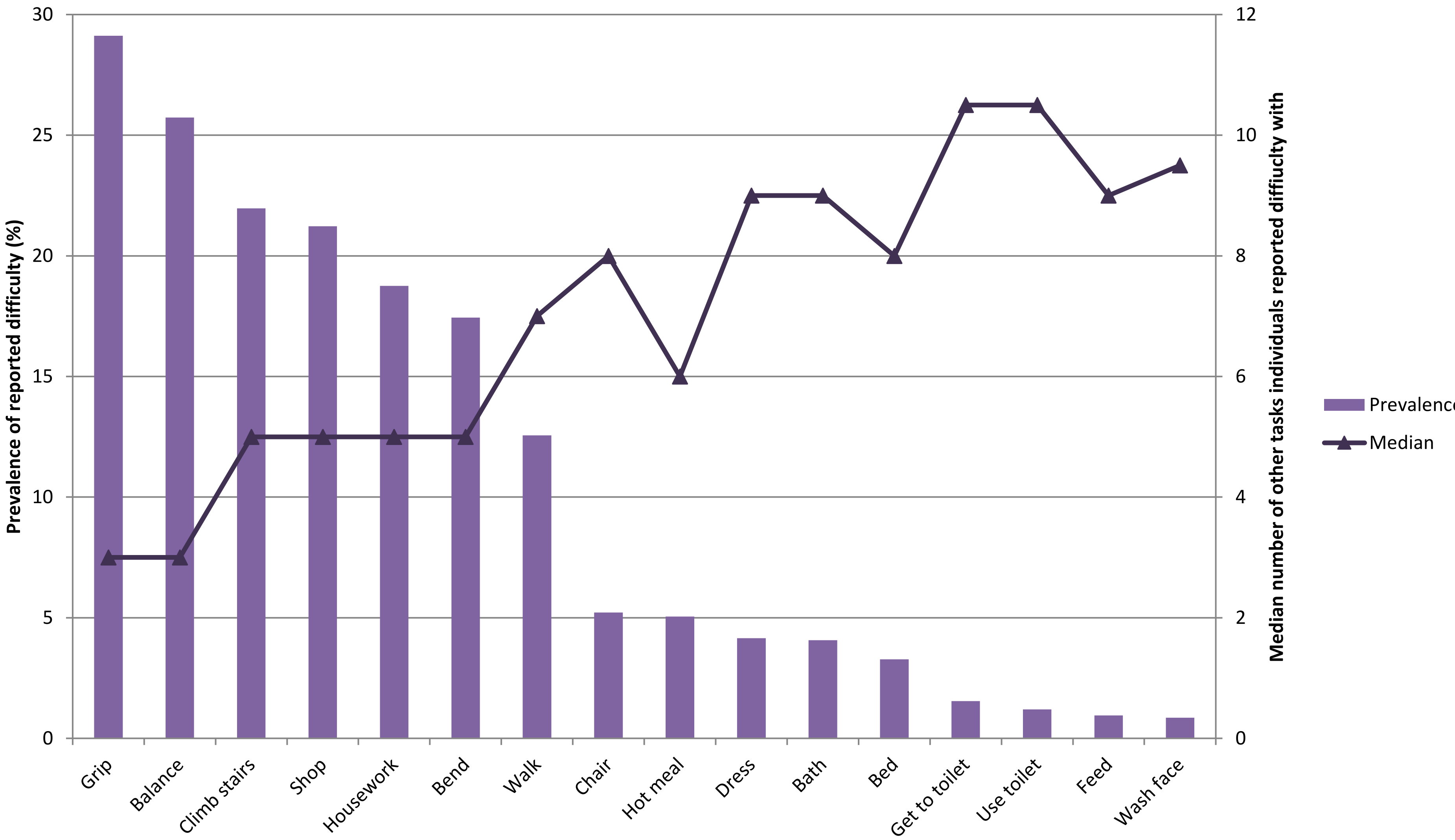


Figure 4.1b: Prevalence of reported difficulty with 16 physical tasks of daily living at age 60-64 for women

A general trend was observed among men and women, whereby the lower the prevalence of reported difficulty for a specific task the greater the median number of other tasks participants reported difficulty with. Men reported difficulty with an average of four other tasks when they reported difficulty with tasks towards the top of the hierarchy (tasks associated with balance, strength and co-ordination), through to an average of 13 other tasks when reporting difficulty with tasks at the bottom of the hierarchy (tasks associated with upper limb mobility). Women followed a similar pattern, ranging from an average of three tasks at the top of the hierarchy to 10 at the bottom.

Only the results of the binary Mokken scales produced are presented, because limitations of the model prevented the use of polychotomous variables; namely that the different self-reported physical capability variables used in this study had an unequal number of response items (as described in Chapter 3), for which the model was unable to adjust. The highest equal number of response categories across the variables was two, and the results produced using this binary approach are shown below. The Mokken scale created for women retained all 16 physical capability items, and produced a Loevinger Scalability Coefficient (H) of 0.54, indicating a strong scale. When the hierarchical order produced from the Mokken scale was compared to the order obtained from the descriptive results (Table 4.1), only two pairs of tasks (hot meal/chair and bathing/dress) were observed to be inconsistent; in each case the two conflicting items were only one position out of order.

When producing the Mokken scale for men, it was observed that two pairs of items were not forming distinct loci on the scale (dress/hot meal and get to toilet/use toilet), and were thus invalidating the assumption of monotonicity. A sensitivity analysis was conducted, removing each conflicting item in turn (Table 4.2). The four scales produced the same hierarchical ordering of tasks, and an average H value of 0.59.

The results of the sensitivity analysis were then compared to the hierarchical order obtained from the descriptive analysis (Table 4.3). The two methods produced the same hierarchical ordering of tasks, with minor variation for the two pairs of tasks that did not meet the assumption of the model.

Table 4.1: Comparison between results of descriptive analysis and Mokken scaling for women

Physical capability task	Prevalence of reported difficulty[†]		Difficulty score attributed from Mokken scaling	
	%	Rank	Rank	P(Xj=1)
Grip	29.12	1	1	0.279
Balance	25.73	2	2	0.242
Climb stairs	21.97	3	3	0.198
Shop	21.23	4	4	0.192
Housework	18.75	5	5	0.171
Bend	17.44	6	6	0.155
Walk	12.56	7	7	0.112
Hot meal	5.05	**9**	**8**	0.046
Chair	5.22	**8**	**9**	0.045
Bath	4.07	**11**	**10**	0.040
Dress	4.15	**10**	**11**	0.038
Bed	3.27	12	12	0.027
Get to toilet	1.54	13	13	0.011
Use toilet	1.2	14	14	0.009
Feed	0.95	15	15	0.009
Wash face	0.85	16	16	0.008

[†] Base population= Participants who responded to all 16 tasks
Bold values indicate where rank differs between the two approaches

Table 4.2: Mokken scaling sensitivity analysis for men

Physical capability task	Difficulty score attributed from Mokken scaling (P(Xj=1))			
	Scale 1	Scale 2	Scale 3	Scale 4
Climb stairs	0.123	0.123	0.123	0.123
Bend	0.119	0.119	0.119	0.119
Balance	0.107	0.107	0.107	0.107
Shop	0.098	0.098	0.098	0.098
Walk	0.090	0.090	0.090	0.090
Housework	0.078	0.078	0.078	0.078
Grip	0.071	0.071	0.071	0.071
Chair	0.040	0.040	0.040	0.040
Dress	0.037	0.037		
Hot meal			0.037	0.037
Bed	0.032	0.032	0.032	0.032
Bath	0.024	0.024	0.024	0.024
Get to toilet	0.016			0.016
Use toilet		0.016	0.016	
Wash face	0.010	0.010	0.010	0.010
Feed	0.007	0.007	0.007	0.007
Loevinger's scalability coefficient	0.602	0.601	0.585	0.586

*The four scales were generated by removing each conflicting item in turn

Table 4.3: Comparison between results of descriptive analysis and Mokken scaling for men

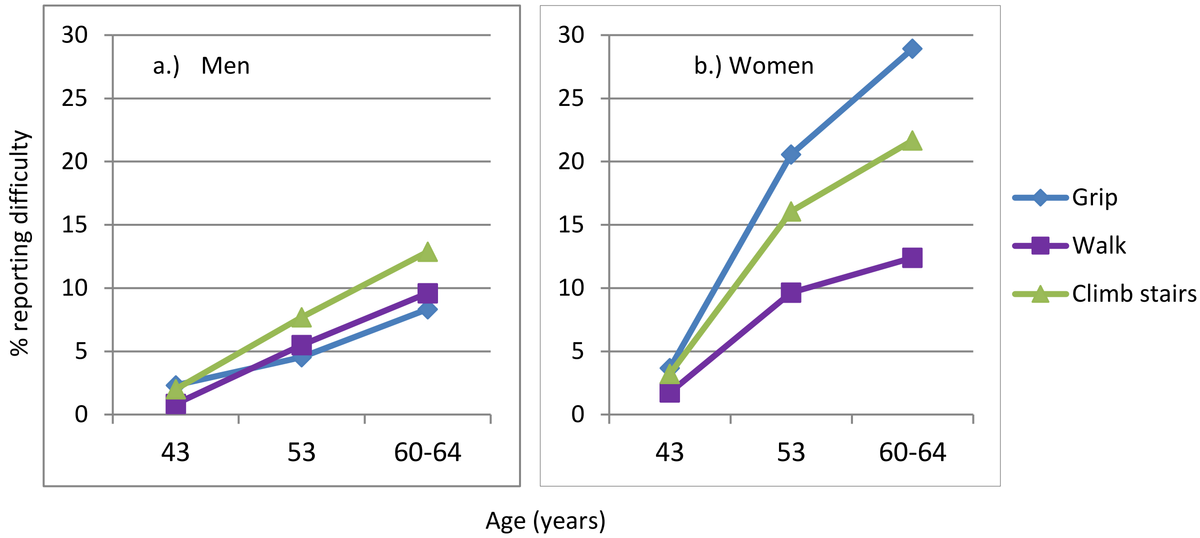
Physical capability task	Prevalence of reported difficulty[†]		Difficulty score attributed from Mokken scaling	
	%	Rank	Rank	P(Xj=1)
Climb stairs	14.03	1	1	0.1229
Bend	13.08	2	2	0.1189
Balance	12.11	3	3	0.1069
Shop	11.25	4	4	0.0979
Walk	10.4	5	5	0.0899
Housework	9.04	6	6	0.0779
Grip	8.43	7	7	0.0709
Chair	4.98	8	8	0.04
Dress	4.73	9	9=	0.037*
Hot meal	4.43	10	9=	0.037*
Bed	3.94	11	11	0.032
Bath	2.59	12	12	0.024
Get to toilet	1.88	13	13=	0.016*
Use toilet	1.7	14	13=	0.016*
Wash face	1.03	15	15	0.01
Feed	1.03	15	16	0.007

[†] Base population= Participants who responded to all 16 tasks
*Using average figure taken from sensitivity analysis

4.3.2 Longitudinal variation in levels of reported difficulty

From midlife onwards a diverging trend in the prevalence of reported difficulty was observed for the three tasks repeated in all three data collection rounds (see Figure 4.2). For women, the prevalence of difficulty increased from a range of 2-4% at age 43 to a range of 12-29% by age 60-64, with the greatest increase seen in gripping and the smallest increase in difficulty walking. The trend was less obvious in men, partially because fewer men reported difficulty with any of the tasks at all three ages. Also the prevalence of reported difficulty for gripping at age 43 does not appear to fit the trend. However, excluding gripping at age 43, men follow the same diverging trend, with prevalence of reported difficulty increasing from 1-2% at age 43 to a range of 8-13% at age 60-64, with the greatest increase in difficulty seen in climbing stairs and the smallest in gripping.

Figure 4.2: Prevalence of reported difficulty for gripping, walking and climbing stairs from age 43 to 60-64 years



4.3.3 Progression through the hierarchy

In the sex-adjusted regression model, those individuals who reported difficulty with at least one of the more complex tasks at the top of the hierarchy (gripping, walking and/or climbing stairs) at age 43 were 4.96 (95% CI: 2.70-9.11) times more likely to progress through the hierarchy of loss and to report difficulty with at least one of the complex tasks at the bottom of the hierarchy (feeding, washing and/or toileting) at age 60-64, compared to individuals who reported no difficulty at age 43 (see Table 4.4). Sex was not a significant factor in this model (P=0.9), and when formally tested there was no evidence of a sex interaction (P=0.14). Each group of covariates were added independently, and all were found to attenuate the association. It was observed that the incident health factors (depression, obesity and respiratory disease) had the most substantial effect on the sex-adjusted result.

On closer examination of the individual factors (see Appendix 3), participants with depressive symptoms were found to be significantly more likely to progress through the hierarchy than those without depressive symptoms. Participants who developed depressive symptoms between age 43 to 60-64 were 3.22 (95% CI: 1.63-6.35) times more likely to progress through the hierarchy, whilst participants who experienced depressive symptoms throughout the study period were 5.25 (95% CI: 2.70-10.23) times more likely to progress than those without depressive symptoms. The development of respiratory disease or obesity between age 43 and 60-64 did not appear to be significant; however, participants with either condition at baseline were more than twice as likely (2.57 (95% CI: 1.41-4.69) for respiratory disease and 2.72 (95% CI: 1.43-5.14) for obesity) to progress through the hierarchy than those who had not experienced these conditions. Physical activity (0.82 (95% CI: 0.70-0.96)) and education (0.75 (95% CI: 0.61-0.93)) were the only factors found to have a significant protective role on progression through the hierarchy; however the effect of both of these factors was attenuated in the fully adjusted model. Occupation and smoking status were not individually found to be significantly associated with the loss of function.

When all of the covariates were added to the fully-adjusted model, participants who reported difficulty with tasks at the top of the hierarchy at age 43 were 2.85 (95% CI: 1.45-5.60) times more likely to report difficulty with tasks at the bottom by age 60-64 compared to those who reported no difficulty at age 43 (see Table 4.4). While a qualitative comparison of sex-specific estimates suggested there may be a sex difference in this association in the fully-adjusted model, with men 1.54 (95% CI: 0.45-5.29) times more likely to progress through the hierarchy, whilst women were 5.28 (95% CI: 2.19-12.73) times more likely to progress if they reported difficulty at age 43, these differences were not found to be statistically significant when formally tested (p= 0.14).

Table 4.4: Odds of progressing through the hierarchy

	Reported difficulty at the top of the hierarchy at age 43*	**Model†**	**Odds Ratio (95% CI) of reported difficulty at the bottom of hierarchy at age 60-64‡**
Men and Women	No difficulty (ref.)		1.00
(N= 2106)	Difficulty	1	4.96 (2.70 – 9.11)
		2	4.33 (2.34 – 8.02)
		3	3.16 (1.63 – 6.14)
		4	4.04 (2.16 – 7.56)
		5	2.85 (1.45 – 5.60)
Men (N= 1000)	No difficulty (ref.)		1.00
	Difficulty	1	2.72 (0.91 – 8.09)
		2	2.29 (0.76 – 6.91)
		3	1.60 (0.50 – 5.18)
		4	2.50 (0.82 – 7.62)
		5	1.54 (0.45 – 5.29)
Women (N= 1106)	No difficulty (ref.)		1.00
	Difficulty	1	7.17 (3.36 – 15.34)
		2	6.51 (3.01 – 14.07)
		3	5.58 (2.35 – 13.26)
		4	5.49 (2.51 – 12.00)
		5	5.28 (2.19 – 12.73)

*117 men and 145 women reported difficulty gripping, walking and/or stair climbing at age 43

† Model 1: Sex-adjusted (if men and women are combined) or unadjusted (if analyses sex-specific)
Model 2: Model 1 plus socioeconomic factors (education and occupational class)
Model 3: Model 1 plus incident health variables (depression, obesity and respiratory disease)
Model 4: Model 1 plus health behaviours (smoking and exercise)
Model 5: Fully-adjusted model

‡ OR are combined estimates from models run across 15 imputed datasets
31 men and 34 women reported difficulty feeding, washing and/or toileting at age 60-64

4.4 Discussion

4.4.1 Summary of main findings

This chapter has presented evidence that the hierarchy of loss exists at younger ages than previously reported in the existing literature. This evidence has been demonstrated using both cross-sectional and longitudinal approaches. The descriptive analysis, conducted when participants were aged 60-64, suggests that individuals first report difficulty with tasks associated with strength, balance and coordination, before progressing through the hierarchy to more severe limitations and reported difficulty with tasks associated with manual dexterity. This hierarchical ordering of tasks was confirmed using Mokken scaling and reflected longitudinally when the analysis focused on the three tasks assessed across the three ages (age 43, 53 and 60-64). Further evidence that the hierarchy of loss exists at younger ages was supplied when exploring the progression of individuals through the hierarchy. If individuals were already reporting difficulty with tasks at the top of the hierarchy at age 43, they were nearly three times more likely to report difficulties at the bottom of the hierarchy by age 60-64, compared to those with no difficulty at age 43.

4.4.2 Comparison with other studies

As noted in the literature review (see Section 4.1), comparisons between studies are limited because most reported hierarchies are based on slightly different sets of tasks from each other. However, this study has more physical capability measures than most previous work, which facilitates comparisons. When the tasks of shopping, walking 400 metres and stair climbing were included in studies, they were consistently identified within the literature as among the first tasks with which individuals reported difficulty (86;95;97). The results of this study support this observation, but they also demonstrate the importance of additional tasks that are found towards the top of the hierarchy, such as balancing and bending. In line with previous findings from the literature, including some other British samples (86;87), difficulty washing and feeding were identified as the most severe limitations, and the last tasks with which individuals reported difficulty (85-87;91;94). It is interesting to note that this hierarchical order is consistent even at younger ages, when prevalence of difficulty undertaking some tasks is very low.

For all of the physical capability tasks it was observed that the prevalence of reported difficulty was lower for men, both longitudinally and cross-sectionally. This corresponds to previous research, which has shown that women tend to report more difficulty than men at all ages (22). Despite these differences, the hierarchical order of tasks presented was generally similar between men and women, corresponding with findings from two cross-sectional (86;96) and two longitudinal studies (87;91) that explicitly looked at the hierarchy of loss stratified by sex. Slight differences were observed in the ordering of tasks between men and women, particularly towards the top of the hierarchy. These tasks were associated with strength, and the known biological difference between men and women (107) may explain the differences observed, especially the large difference in the placement of gripping within the hierarchy. There is greater consistency in the hierarchical order of tasks towards the bottom of the hierarchy, despite only small differences in the prevalence of reported difficulty. These tasks represent severe limitations, as they are some of the most basic physical capability tasks required to maintain independent living. Once individuals have progressed to this stage of the hierarchy, there is limited opportunity for recovery.

4.4.3 Consideration of results

The hierarchy of loss presented in this study was primarily obtained from the cross-sectional analysis conducted when participants were aged 60-64. The hierarchical order can be implied from the observed prevalence of reported difficulty at age 60-64, with the highest reported prevalence of difficulty assumed to be the task with which individuals first experienced difficulty. As individuals progress through the hierarchy, the number of tasks with which they experience difficulty increases; combined with the knowledge that a relatively small number of people have progressed through the hierarchy to severe limitations by age 60-64, the tasks with the lowest prevalence are presumed to be the final tasks with which individuals report difficulty.

The two sex-specific Mokken scales confirmed the hierarchical ordering of tasks obtained from descriptive analysis. There were a few minor differences in the ordering of tasks between the two methods; however, it was noted that this lack of consistency only occurred when the tasks in question had very similar levels of reported difficulty, and so were hard to distinguish. In addition to the sex specific mokken scales presented in this thesis, two scales were produced using the data available at age 43 (H=0.55 for women and H=0.67 for men). These scales demonstrated a similar hierarchical order, with tasks associated with strength, balance and co-ordination, such as stair climbing, found at the top of the hierarchy and tasks associated with manual dexterity, such as feeding, observed at the bottom of the hierarchy. This highlights the consistency of the hierarchical pattern across different ages.

The hierarchical order obtained from the cross-sectional analysis was reflected longitudinally amongst the three tasks repeated at all three time points. The diverging trend in the prevalence of reported difficulty demonstrated that tasks towards the top of the hierarchy, which evidence suggests participants first reported difficulty with, had the greatest increase in the prevalence of reported difficulty from midlife to early old age, whilst tasks towards the bottom of the hierarchy had the smallest increase. The calculated prevalence of difficulty with gripping for men at age 43 did not appear to follow the general trend, but this may be explained by the slight change in wording of the grip question used between age 43 and age 53 (see Appendix 2).

The results obtained from the logistic regression analyses provide evidence that some individuals have already started the transition by age 43, and have completed it by age 60-64. Depression seems to be a significant factor in the loss of function, with individuals experiencing depressive symptoms more likely to progress through the hierarchy. This reflects previous longitudinal research where depression has been found to be significantly associated with physical capability decline (108;109). However, the relationship may be bi-directional, as the loss of capability may also increase depressive symptoms. The protective effect of physical activity may be a result of the prolonged use of certain muscle groups, with more active individuals able to maintain physical capability for longer. The most active individuals may benefit from the cumulative effects of physical activity across midlife, which has previously been shown in the literature to prevent functional decline (102;110). Baseline health, in terms of obesity and respiratory disease, was also identified as one of the significant factors influencing participants' progress through transition. It is plausible that those with poorer health at baseline in midlife are comparable to those at older ages, where progression through the hierarchy is more common. The association between reported difficulty at age 43 and progression through the hierarchy was more substantiated for women than men, but this may be an artefact of the data. By early old age, taken to be 60-64 in this study, few individuals have progressed through the hierarchy, and those who have are predominantly women. The small number of men in this group may explain why the results were only significant for women.

4.4.4 Methodological considerations

There are several methodological considerations that should be taken into account when interpreting the results of this chapter. Self-reported measures of physical capability were specifically chosen for these analyses, rather than their objective performance-based counterparts, due to the inherent nature of self-reported measures which focus on the loss of capability. This characteristic trait of self-reported physical capability measures makes them highly applicable to work on the hierarchy of loss, although there is an element of subjectivity associated. Depression for example has previously been cited as one of the factors known to affect individuals' perceptions, and was found to be a significant factor in this analysis. However, one of the studies included in the literature review (Section 4.1) observed that the same hierarchical order of tasks was obtained whether self-reported or performance-based measures of physical capability were used (95), which suggests the role of subjectivity is limited in this context. Another concern is that self-reported measures often have ceiling effects, and this is particularly evident when conducting analyses on a relatively young, and therefore capable, study population. However, these analyses have demonstrated that it is still possible to identify the hierarchy of loss in a relatively young population, despite few individuals reporting difficulty with the tasks specified.

Although the majority of the study population are believed to have followed the hierarchical order presented, as demonstrated by the strong Mokken scale produced and the diverging longitudinal trends, this cannot be assumed for all individuals. For example, when calculated it was observed that 24% of men and 40% of women did not strictly follow the hierarchy observed in this thesis. Due to the dynamic nature of physical capability, some individuals will take alternative pathways through the hierarchy of loss. The descriptive analysis provides some evidence of this variation, as the average number of tasks which participants reported difficulty with ranged from 4-

13 tasks for men, and 3-10 tasks for women, rather than the 1-16 range that would be expected if everyone followed a strictly linear hierarchical order.

Although a decision was made *a priori* to base the hierarchy of loss on single items, the results presented in this chapter do not support a truly linear hierarchical order, where difficulty with each task is conditional on the individual reporting difficulty with the previous task. However, the evidence provided contributes to the growing literature establishing the tasks at the top of the hierarchy with which individuals first report difficulty. The study has also highlighted the importance of these factors in predicting progression through the hierarchy to severe limitations. The challenge remains to identify a specific group of tasks for which reported difficulty can be used to formulate practical policy suggestions. If the domains suggested in the literature (85) are applied to the data, there is considerable overlap in the ordering of items between items in the first domain (tasks associated with complex manual dexterity and balance) and items in the second domain (tasks associated with long distance mobility and balance). Also some of the items included in this study such as gripping and bending do not fit within the domain definitions provided. Consequently, it does not seem sensible to recommend policy interventions based solely on this domain system. For both men and women, the first seven tasks in the hierarchy of loss are noticeably higher in terms of prevalence of reported difficulty, and the difficulty score attributed from the Mokken scale, compared to the remaining nine items. These items appear in different orders for the two sex-specific scales, but the seven items are consistently at the top of the hierarchical scale for both. From the evidence obtained throughout this study, it is proposed that this set of seven items could be used to identify individuals who are at high risk of progressing through the hierarchy.

4.4.5 Implications and future work

The evidence presented in this study suggests that the hierarchy of loss may operate from at least as early as midlife onwards. Some individuals in the NSHD, although a small group, have already progressed through the hierarchy to severe limitations by early old age. With their physical capability severely restricted, such individuals would probably experience a poor quality of life for their remaining years. These individuals will often require formal or informal care (91), placing a burden on health and social care systems. However, it may be possible to use the hierarchy of loss to help prioritise resources and service allocation to assist with the maintenance of independent living. If focus is placed on individuals when they first report difficulty with tasks towards the start of the hierarchy, it may be possible to identify those at high risk of subsequent functional decline, and consequently loss of independence. These individuals are likely to be suitable candidates for interventions to prevent further decline and possibly promote recovery.

The results of this chapter suggest that the first seven items in the hierarchy (gripping, balancing, stair climbing, housework, shopping, bending and walking 400 metres) may be suitable tasks to use when identifying these high risk individuals. However, further work is required to confirm that these tasks are effective predictors of future functional decline and to standardise the physical capability elements involved in each task. To facilitate this work, additional data may be required for the intervening years between age 43 and age 60-64. Such data could provide more detail about the hierarchical order of tasks, especially tasks towards the top of the hierarchy where some questions remain. Additional data would also enable a second group of high-risk individuals who could be identified in their fifties and followed up to early old age. However, this analysis is beyond the scope of the data available in the NSHD, and an alternative data source would have to be found.

Future work may also wish to extend the hierarchy of loss beyond reported difficulty to incorporate modification of tasks, as this may capture physical capability decline at an earlier stage, providing even greater opportunities for intervention.

Set within the wider context of this thesis, the results presented in this chapter add further support to the selection of the five self-reported measures used throughout the thesis (gripping, bending, balancing, walking and climbing stairs). Initially these tasks were chosen from the 16 self-reported physical capability measures available at age 60-64 on a theoretical basis, because they were closest conceptually to physical capability. However, the hierarchy of loss analyses have demonstrated that these tasks can be used to identify a group of individuals who are at high risk of functional decline.

5. Concordance between self-reported and performance-based measures of physical capability

Having established the descriptive characteristics of the five self-reported and the four performance-based measures of physical capability selected for this thesis (see Chapter 3), the aim of this chapter is to explore the relationship between the two sets of measures. To provide context, the first part of this chapter (Section 5.1) will extend Coman and Richardson's literature review (36) and explore the published literature on concordance between self-reported and performance-based measures of physical capability. Findings from this literature review will then been used to identify appropriate methods to examine the evidence of concordance and discordance within the NSHD at age 60-64. If the results establish the presence of concordance and discordance within the dataset, the second stage of the analysis will define five concordant and discordant groups to reflect the conceptual diagram outlined in Chapter 1 (Figure 1.1, p.24).

5.1 Concordance literature review

As reported in Chapter 1, when exploring the associations between self-reported and performance-based measures of physical capability there has been wide variation in the level of concordance observed within the literature. For example, the review by Coman and Richardson (36) suggested that correlation coefficients between self-reported and performance-based measures range from -0.72 to 0.6. However, this range should be interpreted with caution (see Section 1.3). The variation observed in levels of concordance may in part be explained by the different sample populations used within the studies, the different methods of analysis and the measures of physical capability selected. A comprehensive review of the literature was undertaken in order to address these three potential explanations and to inform decisions regarding the method of analysis within this thesis.

As a starting point, the review undertaken for this thesis considered all 17 papers included in Coman and Richardson's 2006 review on concordance in self-reported and performance-based measures of functional ability (36). Coman and Richardson used three selection criteria when refining the list of papers for inclusion in their review: 1) both self-reported and performance-based measures of functional ability must have been administered, 2) the study sample included people over 55, and 3) subjects were community-dwelling. However, the review in this chapter focused on concordance between measures of physical capability, so two papers from the original review were excluded as they focused on tasks akin to disability (see Section 2.2). Using similar search terms to those of Coman and Richardson, the search was updated to include papers published after 2006 (up to December 2015) (N=11). A further five papers were identified by searching the reference lists of eligible papers published before 2006. The current review also extended beyond community-dwelling populations to reflect the full range of physical capability observed within adults of early old age in the wider population, by including papers that focused on condition-specific samples (N=11). As the age range of participants in condition-specific samples was generally younger than those in community-dwelling samples, papers were only included in the review if a proportion of the sample represented the age group of interest: early old age, loosely taken to be 60-75 years of age. In total 42 papers have been included in this review, and the level of concordance between self-reported and performance-based measures of physical capability reported within these papers varied considerably (see Appendix 4).

The first potential explanation for the wide variability in levels of concordance reported in the literature, to be explored in this review is the characteristics of the study sample.

5.1.1 Study Sample

The characteristics of a study sample provide context to levels of physical capability, as certain subgroups of older adults are known to have poorer capability than others. In 11 of the studies reviewed the sample was sex-specific (eight focused on women and three on men) which meant that the effect of sex could not be examined in these studies. Of the 25 mixed-sex studies who reported the sex ratio of their sample, more than half (15 studies) had over 60% female participants. This sex ratio may imply a selection bias against males within the studies reviewed. However, the sex ratio found in the general population is age dependent, with an increasing proportion of women in older populations, which may explain the observed sample characteristics.

Setting

The majority of the papers reviewed (67%) examined community-based samples (see Table 5.2). Whilst the whole spectrum of physical capability can be found within community-based samples, such samples tend to have higher average levels of physical capability than condition-specific studies, because they do not specifically concentrate on those more likely to have limitations. However, it should be noted that five of the community-based papers did elect to focus on community sub-samples with capability limitations (111-115) such as frail adults or the most disabled third of the population. Many community dwellers remain independent, and 13 of the studies focused exclusively on these non-institutionalised populations. Individuals who maintain their independence have higher levels of physical capability. For example in one study (116) only 24% of community dwellers aged 75-84 reported that they were unable to walk half a kilometre compared to 85% of participants in institutions. The proportion of adults in institutions at this age is relatively small (7%), but when the institutionalised group were incorporated into the prevalence estimate, the proportion unable to walk half a kilometre, rose from 24% to 27%. Only three of the papers reviewed (40;116;117) included both community dwellers and those in nursing homes, perhaps making them more generalisable to the overall population of older adults. However, it should be noted that the proportion of the population residing within institutions increases substantially with age, with one study based in Finland suggesting that the proportion tripled from 7% to 24% between 75 to 99 years of age (116). Although there were no corresponding figures available for the UK in the studies reviewed, these results suggest the impact of institutionalisation on population levels of physical capability may be minimal amongst adults of early old age.

Eleven condition-specific papers (118-128) (see Table 5.1) were included in this review. People within these condition-specific subgroups have lower levels of physical capability because their diagnosed medical conditions or treatments are associated with reduced mobility, and an increased risk of pain and fatigue. It is feasible that the degree of concordance between the two sets of measures may be affected by the overall level of physical capability previously experienced by the individual, with adults experiencing low levels of capability more likely to have discordant results if they are in the process of adjusting to their declining capability. For example, those recovering from a hip fracture may reflect on previous physical capability, either before the fracture or to an earlier stage of recovery, rather than current capability (118). However, it is difficult to draw conclusions regarding the comparison of overall physical capability levels because of substantial differences in analyses between studies.

Most of the research is also cross-sectional, with only four of the papers reviewed (34;43;120;124) assessing longitudinal data. These studies reported similar levels of concordance between self-reported and performance-based measures of physical capability at each time point, but the follow-up period was limited (less than three years for all studies, except one which covered a five-year period). Consequently the impact of life course functional trajectories on concordance and discordance remains relatively unexplored.

Table 5.1: Main selection criteria for study populations of papers included in concordance literature review

Sample population	**N† (%)**
Community-dwelling	28 (66.7)
Non-institutionalised sample ‡	13
Targeted limitations	5
Population-based (community + institutionalised)	3 (7.14)
Fibromyalgia (FM) patients	2 (4.17)
Patients with hip fracture	2 (4.17)
Rheumatoid Arthritis (RA) patients	2 (4.17)
Ankylosing Spondylitis (AS) patients	1 (2.38)
Chemotherapy patients	1 (2.38)
Parkinson’s Disease (PD) patients	1 (2.38)
Knee Osteoarthritis patients	1 (2.38)
Multiple Sclerosis patients	1 (2.38)
Cerebrovascular patients (stroke or transient ischaemic attack)	1 (2.38)

† Total= 43 as one paper includes two possible selection criteria for their sample population
‡ Explicitly stated in exclusion criteria or implicit as part of detailed sample selection

Age

Another factor to consider is the mean age of the participants within the papers reviewed. The majority of studies included in this review have a reported mean age over 70 (N=28), and levels of physical capability are known to decrease with age (1;2). Even where the mean is not reported, the age range of participants is such that it is likely to be similar to those that do report a mean (see Table 5.2). In general, the studies with a younger mean age are those with condition-specific populations. As a result, a higher proportion of the participants in these condition-specific papers will have low levels of physical capability compared to the general population of early old age adults. Consequently a higher proportion of the samples under review were unable to complete the performance-based tasks, and many studies chose to assign these participants values corresponding to the lowest quartile or equivalent score. By incorporating those unable to perform the tests into the analysis, the concordance between self-reported and performance-based measures may have been strengthened. This would be particularly evident if the protocol of the study had the performance-based test before the self-reported measures, as the inability to perform the assessment would enhance participants' perception of their current capability.

Sample size and location

Sample size also varied between studies, although the majority of papers reviewed had a sample size of less than 500 (N=29) (see Table 5.2). Even if the papers with condition-specific samples (which understandably have smaller sample sizes due to explicit inclusion criteria) are excluded, 61% of the remaining papers have fewer than 500 participants, and 74% have less than 1,000. These small sizes offer enough statistical power to detect concordance between self-reported and performance-based measures, as the relationship is relatively strong; however the occurrence of discordance is rarer, and smaller sample sizes may not provide sufficient power to detect the phenomenon. A further limitation of the sample size is the limited generalisability of the results produced. Of the four studies with sample sizes over 3,000, two papers use samples from the same study population, the Established Populations for Epidemiological Study of the Elderly (EPESE). The EPESE population is also used by a third paper with a sample size of over 1,000. Although these three papers use subsamples of the same study population, the different methods of analysis mean that the concordance observed varies between the papers. Half of all the studies reviewed were based in the United States, with all but three of the remaining papers with known country of origin based in Europe. None of the papers reviewed focused specifically on a UK based population, with only one paper (119) including any participants from the UK.

Table 5.2: Range of mean ages and sample sizes of the concordance studies reviewed

	Grouped Categories	All N (%)	Community-dwelling N (%)	Condition-specific N (%)
N		42	31	11
Mean age (years)	< 60	8 (19.0)	2 (6.45)	6 (54.5)
	65-70	1 (2.38)		1 (9.09)
	70-75	10 (23.8)	10 (32.3)	
	75-80	14 (33.3)	13 (41.9)	1 (9.09)
	80+	4 (9.52)	2 (6.45)	2 (18.2)
	Unknown (<60)	3 (7.14)	2 (6.45)	1 (9.09)
	Unknown (<70)	2 (4.76)	2 (6.45)	
Sample size	<100	10 (23.8)	4 (12.9)	6 (54.5)
	100-500	19 (45.2)	15 (48.4)	4 (36.4)
	500-1000	5 (11.9)	4 (12.9)	1 (9.09)
	1000-1500	3 (7.14)	3 (9.68)	
	1500-3000	1 (2.38)	1 (3.23)	
	3000-5000	1 (2.38)	1 (3.23)	
	5000-10000	3 (7.14)	3 (9.68)	

5.1.2 Method of analysis

The second potential explanation for the variation in concordance observed within the literature is the method of analysis. In the 42 papers reviewed, three main techniques of analysis were identified: regression, correlation and cross-tabulation. Several different approaches were noted amongst studies using these techniques, with the selection of technique partially dependent on the measures available. Some studies used more than one method of analysis to confirm trends or to assess different aspects of their study objectives. When regression analysis was employed there was an even split between linear and logistic methods, with four papers employing each method. Amongst studies with correlation as the analysis of choice, there was an even divide between the use of Spearman's correlation coefficient (nine studies) and Pearson's correlation coefficient (nine studies). When studies employed the cross-tabulation technique there was more variation. The first method identified, used by five studies, involved comparing mean performance-based results across levels of self-reported physical capability and commenting on any observed discordance. The second approach, used by three studies, involved tabulating the proportion of participants who reported difficulty against performance-based level (based on quartiles) and commenting on any discordance. The remaining 12 studies used methods which involved categorising both the self-reported and performance-based measures. The most common version of this approach (used by eight of these papers) simply calculated the percentage concordance based on the agreement between the two categorised variables and calculated discordance accordingly. The second variation (used by three papers) chose one measure (usually the performance-based measure) as the "true value" and calculated the corresponding sensitivity and specificity of the other measure. One paper subtracted the self-reported categorical score from the performance-based categorical score and implemented cut-off points for discordance.

The selection of analytical method within each paper depended on the nature of the variables used and the purpose of the study. If both sets of measures were continuous and the data were normally distributed, then either linear regression or correlation methods were suitable for the analysis, assuming a linear relationship between the variables. However, if both measures were categorical then it was more appropriate to calculate the percentage of concordance or the sensitivity and specificity of the measures, with one measure taken as the "gold standard". If the self-reported measure was categorical and the performance-based measure was continuous, then logistic regression or cross-tabulation methods based on quartiles of performance or mean performance values were more suitable. Regression was often used in studies that developed their analysis to explore factors associated with discordance. It should be noted that some studies used more than one method of analysis to confirm trends or to assess different aspects of their study objectives.

Within the 18 papers that used correlation as the method of analysis, the Spearman or Pearson's correlation coefficients ranged from 0.12 to 0.79, with little difference in the range between the two approaches. Negative correlations were found where the scoring scale was reversed for one of the measures, but the range was within that already reported (-0.74 to -0.35). One criticism of the correlation (and linear regression) technique is that it requires the self-reported measure to be a continuous variable. When participants complete self-reported questionnaires they are usually asked to depict their experience within discrete response categories. The responses are often hierarchical in nature, from 'no reported difficulty' to 'unable', and are coded with successive integers, as exemplified in Chapter 3 where response options for the measures in the NSHD were shown. It is presumed that these papers treated the ordinal categorical coding as continuous, because no specific details are provided. Four papers provided the exception to this by transforming the ordinal scores (111;113;128;129), with one paper also accounting for the relative difficulty of each self-reported item when transforming the data (128). This is advantageous when using summary measures, because self-reported tasks require different levels of exertion, and consequently some tasks are more challenging to complete than others. A second criticism of the correlation approach is that, whilst the method implies discordance, it is hard to quantify the extent of true discordance. For example, it is possible that a weak correlation, as a result of measurement error, may be masking a stronger correlation rather than indicating discordance. This was demonstrated in one paper where the correlation coefficients adjusted for measurement error (disattenuated coefficients) showed stronger correlations across all measures compared to the coefficients calculated originally (130) (see Appendix 3). If the original coefficients had been used, a greater proportion of discordance may have been presumed.

Regardless of the method used for analyses, the studies which focused on categorical self-reported and continuous performance-based measures all found trends of increasingly poor mean levels of performance with stepwise decreases in self-reported capability (or vice versa). However, considerable discordance was also observed. For example, in one paper (39), across all four performance-based measures (walking speed, chair stand, standing balance and grip strength), 24% to 61% of those in the poorest performing quartile reported the highest level of capability. In another paper (34) it was noted that, whilst 93% of those in the top performance level reported the highest level of capability, only 41% of those who reported the highest level of capability were in the top performance level. Although these examples highlight some of the discordance observed between the two sets of measures, when using the methods associated with categorical self-reported and continuous performance-based measures it is difficult to calculate the magnitude of discordance.

When both self-reported and performance-based measures are categorical the methods used enable the explicit calculation of discordance with greater ease. A more detailed discussion focusing on discordance will follow in Chapter 6.

5.1.3 Measures of physical capability

The third factor potentially explaining the wide variability in concordance observed within the reviewed papers is the selection of self-reported and performance-based measures used within the analyses. A diverse mix of measurement scales were used within the papers reviewed, with some measures unique to specific studies (see Tables 5.3 and 5.4). Each scale consists of several items relating to specific tasks of physical capability; these items were either analysed separately or combined to create summary variables. More variability was noted amongst the self-reported measures, with 19 specific scales of self-report utilised compared to 13 performance-based assessments. The total number of papers in Tables 5.3 and 5.4 is greater than 42 (the number of papers reviewed) because some papers used more than one scale or assessment to represent each set of physical capability measures.

Table 5.3: Performance-based measures used in the concordance papers reviewed

Performance-based measure	Number of studies using measure
EPESE performance tests (all or subset)*	9
Tasks matched specifically to questionnaire	7
Assessment of Motor and Process Skills	3
Physical Performance Test	3
Performance Assessment of Self-Care Skills	2
Physical Capacity Evaluation	2
Short Physical Performance Battery	2
Groningen Fitness Test	1
Longitudinal Ageing Study Amsterdam tests	1
Multiple Sclerosis Functional Composite Index	1
National Institute on Ageing Battery	1
Physical Performance and Mobility Examination	1
TEMPA**	1
Assortment of tasks (no specified assessment)	9

*EPESE: Established Populations for Epidemiological Study of the Elderly
**TEMPA: Upper Extremity Performance Test for Elderly

The most commonly used performance-based assessment was the EPESE performance test included in the analysis of nine papers. The EPESE assessment incorporates four main items of performance: balance, gait speed, chair rise and shoulder rotation

(73;131). Many of the other performance-based measures incorporate these items or similar elements, particularly the Physical Performance Test or the Short Physical Performance Battery. However, these items are also featured individually in papers with assorted and matched performance-based measures. Some authors have adapted the items used within the EPESE performance test, as the protocol may not be practical for their particular study environment or resources, which limits comparability. However, nearly half of the papers reviewed include some form of the measures incorporated in the EPESE test. The dominance of the EPESE performance-based measures within the literature can be attributed to the influential work of Dr Jack Guralnik, who is also a key author of several of the reviewed papers. Where performance-based tasks have been specifically chosen to match the self-reported measures from the questionnaire used within the study, most papers selected measures of gait speed and objective measures of climbing stairs.

Table 5.4: Self-report measures used in the concordance papers reviewed

Self-reported measure	**Number of studies using measure**
SF-36 (subscales)*	5
EPESE questionnaire (all or subset)	4
Functional Status Questionnaire	3
Health Assessment Questionnaire	3
WHO questionnaire	3
Older American Resources and Services (OARS)	2
Modified Katz ADL	2
Performance Assessment of Self-Care Skills	2
Activity measures for post-acute care	1
Bath Ankylosing Spondylitis Function Index	1
Fibromyalgia Impact Questionnaire	1
Functional Assessment of Chronic Illness Therapy	1
Groningen Activity Restriction Scale	1
Incapacity Status Scale	1
Late Life Function and Disability Instrument	1
Manual ability measure	1
OECD indicator**	1
Sickness Impact Profile	1
Virtual Short Performance Physical Battery	1
Unclear	1
Assorted questions (no specified scale)	7

*SF-36: Short Form Health Survey
**OECD: Organisation for Economic Cooperation and Development

The most commonly used self-reported measure was the SF-36, a health survey comprised of eight subscales of domains. One of these domains focuses on physical functioning, and was used by five papers in their analysis. In addition to the physical functioning domain, one paper (114) used the physical component summary score, which also accounts for the bodily pain and physical role domains. Whilst the SF-36 is the most commonly used self-reported measure, only five studies reported its use compared to the nine studies that use the most common performance-based measure. This emphasises the variability in the use of self-reported measures.

Most self-reported measures ask participants to report the level of difficulty experienced when attempting the task in question (for example the measures used within the NSHD, see Section 3.2.2), but there are many differences in wording and structure which may contribute to the variability across studies. For example, some measures have the response option “unable”, whilst others do not provide this distinction from “difficulty” completing a task. In other cases, some measures allow the use of aids or personal help when assessing the difficulty of a task, whilst other measures explicitly ask about independent ability. The wording of questions may explain some of the weak concordance observed between self-reported and performance-based measures. For example, if the protocol of the performance-based assessment prohibits the use of personal help and aids, weaker correlations would be expected in studies where self-reported measures do not explicitly ask about independent ability. Concordance could also be limited if the self-reports are based on what an individual “can” hypothetically do rather than making the distinction with what they report actually doing. It has been suggested in the literature (42) that the use of hypothetical capacity questions may underestimate the prevalence of difficulty experienced within older populations. This idea has been supported in one paper (122), which asked participants about both habits (do) and abilities (can). The study found that participants were more likely to overestimate their performance when asked to assess their abilities rather than habits. The difference between ability and habit widened as the task in question became more complex.

When considering the influence of the hypothetical nature of self-reported measures on concordance, it is also important to note the order in which different physical capability measures are ascertained. For example, if performance-based measures are completed before the self-reported measures, it is possible that an individuals' perception of their physical capability may be affected; their reports are less likely to be based on their longer term perception of capacity and are more likely to be based on their known level of current performance, potentially strengthening the level of concordance observed. However, very few of the studies included in this literature review recorded the order in which physical capability measures were ascertained so it is not possible to establish the potential impact that order has on concordance within this review.

One of the main points highlighted in the Coman and Richardson review was that higher correlation coefficients were observed when the selected measures focused on the same theoretical concept. Some of the papers reviewed (120;130;132) do make reference to the theoretical framework underpinning the selection of measures, but many do not. Although the inclusion criteria for this review meant papers must have measures of physical capability as conceptualised within this thesis, some of the studies have also included measures of functional ability that are conceptualised at the societal level and can therefore be considered activities rather than tasks, for example grocery shopping. It is difficult to isolate the effect of different hierarchical levels of the theoretical construct on concordance, as most papers combine self-reported items or do not quantify concordance if the self-reported items are separated. In the three papers (114;120;130) that compare performance-based measures with either self-reported physical capability items or individual self-reported disability items, all found higher correlations when measures represented the same construct. For example, one paper (120) observed a correlation coefficient of 0.71 between their self-reported physical capability (physical function domain of SF-36) and performance-based (AMPS Motor) measures, but a coefficient of only 0.35 for the correlation between the same performance-based measure and the self-reported disability (functional wellbeing) measure.

Ten of the papers reviewed specifically matched the performance-based to the self-reported measures selected in their analysis, for example matching walking speed to self-reported walking, or chair rise time to self-reported difficulty getting out of a chair. The intention of this approach was that a more valid estimate of concordance between the two sets of physical capability measures would be produced, because both measures are focused on the same task. However, it is not always possible to match the two measures exactly; for example, Sainio and colleagues (116) elected to analyse stair climbing although the self-reported task required greater endurance (one flight of stairs) than the performance-based version (two steps). Myers et al (34) observed that there was greater concordance when the wording of the self-reported items specifically reflected the performance-based task participants were asked to conduct. Two papers (130;133) compared both matched and non-matched pairings in their analysis, and the reported results suggest that matched measures have greater concordance than non-matched measures. This relationship would be expected, given the matched pairs are influenced less by variation between tasks. However, the association between non-matched measures was significant and not dissimilar in value to the matched analysis, therefore the evidence is not particularly strong. For example, when picking up a penny from the floor (performance-based task) was correlated with the bending or stooping item of the PF-10 (self-report), the coefficient was 0.33; when the same performance-based task was correlated with other items of the PF-10, such as climbing a flight of stairs or walking, the correlation coefficients ranged from 0.17 to 0.32 (130).

The concordance between self-reported and performance-based measures varies depending upon the specific task in question. The clearest distinction is between items that have a high cognitive component and those that do not, with higher cognitive elements producing lower concordance. For example, a concordance of over 50% was observed for the mobility item of walking, compared to less than 20% for money and medication management (123). Whilst none of the papers specifically explore the relationship between severity of task and level of concordance, several papers suggest that mobility items such as walking or climbing stairs have higher discordance than less complex self-care tasks such as eating or bathing (34;132). It could be argued that self-care tasks do not conceptually fit the construct of physical capability, and if these tasks are excluded it becomes more challenging to discern a trend between the discordance of different physical capability items.

5.1.4 Summary

Within the 42 papers reviewed, the reported concordance in self-reported and performance-based measures varied significantly between studies, and it is likely that most of this variation is due to the wide range of measures, methods of analysis and study populations used.

Ideally it is important to set the research within a theoretical framework, and from this select the relevant scales and assessments of self-reported and performance-based physical capability measures. This framework will then enable researchers to ensure that the items selected represent the same theoretical construct, and will help direct appropriate methods of analysis. For example, in this thesis physical capability has been conceptualised as a continuum, and the performance-based measures reflect the underlying construct as they collect continuous data. Caution should therefore, be taken when using any method involving the categorisation of performance-based measures, as this would disrupt the continuum of the continuous data collected. However, it may not always be possible to do this given the practical limitations of the data available for analysis. When selecting self-reported and performance-based measures to investigate concordance and discordance it is important to be explicit in the protocol used, to maximise comparability between the two sets of measures. For example, it should be noted if the use of aids or personal help is incorporated in both measures in order to detect "true" discordance rather than disagreement between the two sets of measures due to the different subsamples they are comparing. However, research is often constrained by the availability of data, especially when using secondary resources, so it may not be possible to follow these recommendations.

The NSHD dataset has a number of beneficial characteristics (see Chapter 3) which will be used within this thesis to address some of the issues in the literature which have been highlighted by this review. Firstly, some of the issues related to sample selection raised within the review are not applicable. For example, study members within the dataset represent the age group of interest, early old age adults, whereas most of the previous research has focused on older ages where interventions may be less effective. The sample size of the NSHD dataset is also larger than most of the studies within this review, which will enable more statistical power in the main analyses. A potential limitation of using the NSHD dataset is that the measures of physical capability are set within the pre-existing protocol of the study. However, this problem is not unique to this dataset, and care was taken when selecting measures for this thesis to ensure that they reflect the theoretical conceptualisation of physical capability.

Using the information collated during this review, the first stage of analysis will be to identify the level of concordance between the self-reported and performance-based measures within the NSHD dataset, and explore if there is any evidence of discordance. Following this analysis, the intention is to identify two discordant groups and investigate the association with other factors of interest. The selection of these factors will be based on a second review of the papers identified in this chapter's review that extended their analyses to investigate associations with discordance (see Chapter 6).

5.2 Methods

The first stage of the analysis aimed to assess the concordance between the self-reported and performance-based measures of physical capability within the NSHD dataset when participants were 60-64 years of age. The analysis of concordance was conducted using a sex-specific approach as a result of the significant sex differences observed in the previous descriptive analysis (see Section 3.3).

5.2.1 Assessment of concordance between performance-based and self-reported measures

Two approaches were used to examine concordance based on cross-tabulation. Firstly, mean values of each of the four performance-based measures were calculated for each response category of the five self-reported measures. For the standing balance performance-based test, median and interquartile ranges were calculated instead of means to account for the skewed distribution. A non-parametric test for trend was used to ascertain whether there was a trend across the mean (or median) performance scores. The second approach used to assess concordance was to compare the proportion of the population who reported that they were unable to complete the performance-based task for health reasons across each of the self-reported response categories. A chi-squared test for trend was used to test this association. Across both approaches any observed discordance was noted.

5.2.2 Creation of summary measures

The second stage of the analysis aimed to identify five groups of individuals based on the concordance and discordance of self-reported and performance-based measures of physical capability (see Figure 1.1, p.24). In order to meet this objective, two summary variables were created. From a methodological perspective, the creation of a summary variable for the self-reported measures enabled the categorical self-reported data to be combined into a continuous variable.

<u>Summary self-reported measure</u>

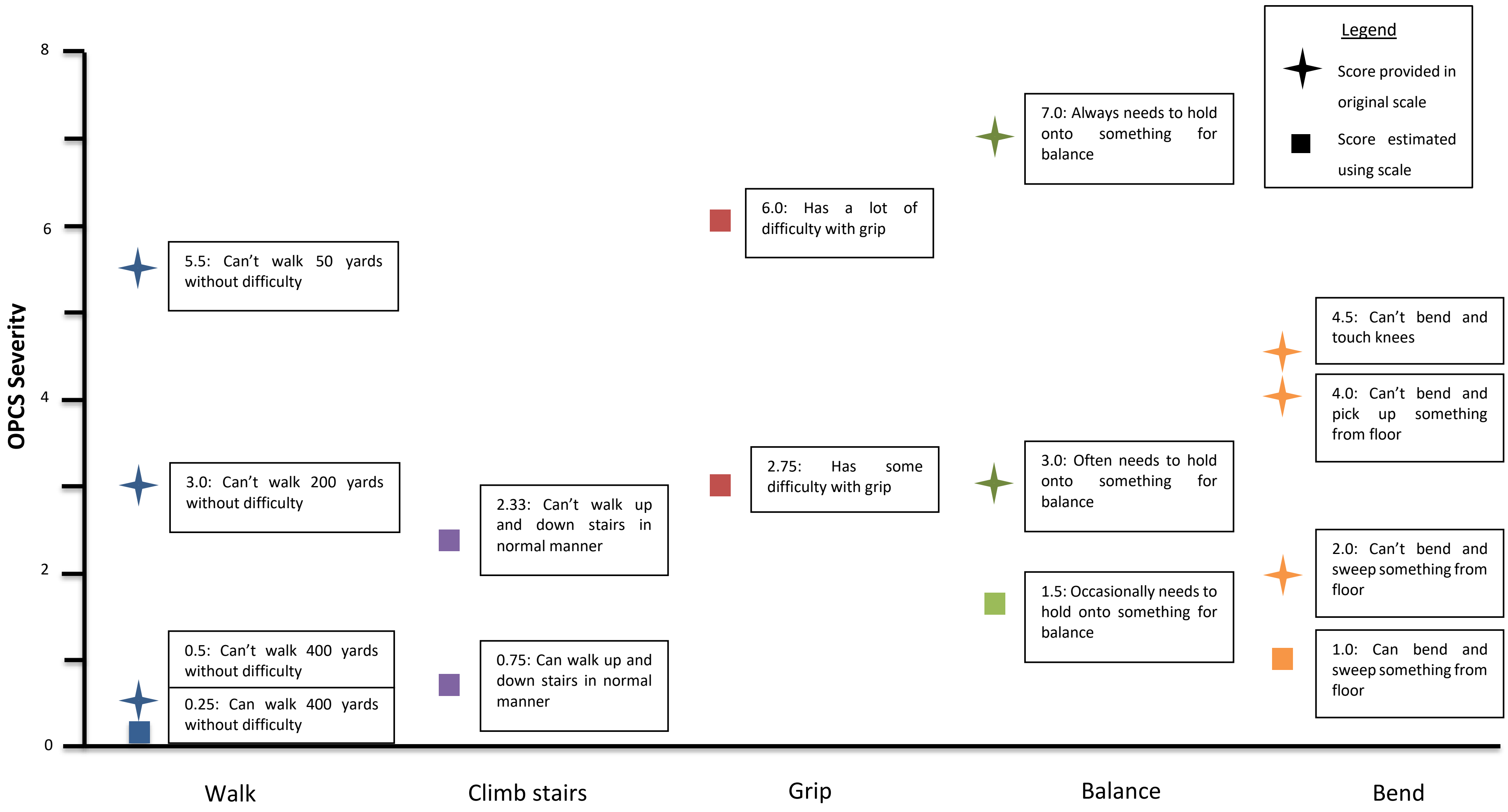
A weighted summary self-reported measure was produced, which accounted for the relative difficulty of the five self-reported physical capability tasks selected for use in this thesis (gripping, walking, bending, balancing and climbing stairs). As it was considered beyond the scope of this thesis to assess the relative difficulty of each physical capability task, a decision was made to use an existing scale to provide this information where possible. As previously mentioned (see Chapter 3), the self-reported measures used within the NSHD dataset originate from the OPCS Survey of Disability (82). As part of the published report on this survey, the OPCS produced a severity scale, and the component scores from this scale were used in this thesis.

Details of the procedure used by the OPCS to create the severity score can be found in the report (82). To summarise, the relative impact of difficulty in tasks included in the survey was considered within domains of similar tasks, across domains and in combination, based on the consensus of a large number of "judges". Judges consisted of both professionals (doctors, physiotherapists, occupational therapists and researchers working in the field) and lay people (people with disabilities and their carers). Although the scores produced for the scale were subjective, there was a high level of agreement between the judges.

In this thesis the component OPCS severity scores have been directly assigned to the corresponding self-reported response category where possible. In circumstances where it was not possible to apply a score directly, due to differences in the wording of the self-reported response category, the OPCS scale was used as a guide to estimate appropriate scores (see Figure 5.1). For the self-reported difficulty gripping variable, the highest severity score (a lot of difficulty) was estimated as an average of two scores from the OPCS scale: difficulty picking up and pouring from a full kettle (score 6.5) and difficulty unscrewing the lid of a coffee jar (score 5.5). The two scores were selected from the OPCS scale as they specifically reflected the wording of the 60-64 variable. A similar approach was taken for the highest severity stair climbing score, although three scores from the OPCS scale were used: can only walk up and down a flight of 12 stairs if... holds on and takes rest (score 3.0), holds on but does not take rest (score 2.5) and if goes sideways or one step at a time (score 1.5). The lowest severity score for each variable was estimated as halfway between 0 (score for no difficulty) and the lowest severity score obtained from the OPCS scale, either directly or as part of an estimated average. The final values assigned to each self-reported response category can be seen in Figure 5.1.

Once scores had been assigned to each task, an aggregated score was calculated for each participant to create the summary variable. Descriptive results were then produced. A Mann-Whitney non-parametric test was conducted to investigate the presence of a sex difference within the summary variable.

Figure 5.1: Severity scores for self-reported tasks of physical capability



Summary performance-based measure

The summary performance-based measure was an expansion of the method devised by Guralnik and colleagues (81) for use in the NSHD. Guralnik *et al* produced a composite score of performance using grip strength, standing balance (eyes closed) and chair rise time at age 53, with each of the three tests rescaled to a 0 (low) - 1 (high) scale before aggregation.

The principles used by Guralnik *et al* have been replicated for grip strength, standing balance and chair rise time. However, in this thesis the timed-get-up-and-go performance test was added to the composite score following similar principles.

The procedure used to rescale each of the performance tests is outlined below:

- For standing balance the time which participants maintained the balance position with their eyes closed was divided by 30 seconds (the maximum time allowed).
- For grip strength valid values were adjusted for body size by dividing strength (kg) by height (cm). The height adjusted grip strength was then divided by the sex-specific 99th percentile of adjusted grip strength (0.4261 kg/cm for men and 0.2718 kg/cm for women). Participants with adjusted grip strength above the 99th percentile were assigned the appropriate 99th percentile value before the division, thus producing the maximum score of 1.
- For chair rise the score was calculated using the equation 1-(time/x) where time is the time taken to complete 10 chair rises and x is the 99th percentile of time (51.22s). Participants who took longer than the 99th percentile were reassigned this value.
- Timed-get-up-and-go was also calculated using the equation 1-(time/x), although time in this instance referred to the time taken to get up, complete the 3 metre walk, turn and return. Participants who took longer than the 99th percentile (18.19s) were assigned these values.

Participants who were recorded as unable to complete tasks for health reasons were allocated a score of 0 for the specific task they were unable to complete, in line with the approach used by Guralnik *et al*.

Once all four of the performance variables had been rescaled, the scores were aggregated to produce the summary variable (range 0-4). Descriptive results were then produced. Although the rescaling of grip strength accounted for sex differences, a t-test was conducted on the summary measure to compare the aggregated scores between men and women.

<u>Cross-tabulating the two summary measures</u>

When the two summary measures were collated, there was an inevitable reduction in the sample size, as some individuals were missing either the self-reported (N=50) or the performance summary measure (N=255). In most cases only one component of the missing summary measure was unavailable, so a decision was made to recode the data where possible.

Data were recoded for the following cases:

- Where one component of the self-reported summary variable was missing and individuals reported no difficulty (score of 0) across the other four components (N=12), a decision was made to recode the missing value to 0 in line with the other measures.
- Where one component of the self-reported summary variable was missing and individuals reported difficulty but had not specified the severity of this difficulty, a decision was made to allocate the lowest difficulty score for the specific task in question (N=30).
- Where the participant in question was wheelchair bound and the variable had initially been coded as missing (not applicable) (N=1), a decision was made to recode self-reported balance to the most severe level of difficulty (score of 7).
- Where participants were missing the grip strength component of the performance-based measure because they did not have an appropriate height measurement at age 60-64, a decision was made to use their recorded height at age 53 (N=2).
- Where participants were missing the TUG component of the performance-based measure and used an inappropriate aid (for example a mobility scooter), a decision was made to recode the missing value to 0 in line with those who were unable to complete for health reasons (N=1).

- Where participants were missing the balance component of the performance-based measure as they only completed the eyes-open standing balance test, a decision was made to recode the missing value to 0 in line with those who were unable to complete for health reasons (N=9).

A sensitivity analysis was conducted to investigate whether bias was introduced as a result of this recoding. Descriptive statistics were produced to test the assumptions of the recoded data, by comparing the two original summary variables with the two modified variables that included the recoded data.

5.2.3 Defining the discordant and concordant groups

When both summary variables had been produced a scatter graph was plotted to facilitate the identification of discordant groups. To ensure that high values for both summary measures corresponded to high levels of physical capability, the summary self-reported measure was reversed by subtracting the aggregate score from the maximum theoretical value attained.

To identify the discordant and concordant groups, each summary measure was then split into deciles. For the performance summary measure, sex-specific deciles were used. The five concordant and discordant groups were defined on the basis of these deciles, using theoretical knowledge of the measures to provide justification for the grouping of individuals. Individuals at either end of the distribution were considered concordant if they were observed to have high levels of physical capability in both measures or low levels of physical capability in both measures. The two discordant groups were identified as those at either end of the distribution who were at opposite ends of the physical capability spectrum for each of the summary measures. Approximately equal numbers were selected into each of the four groups. All remaining participants not included in one of the two concordant or two discordant groups were categorised as a larger middle concordant group.

Descriptive analysis

Once the five concordant and discordant groups were defined, descriptive analyses were used to characterise each group in terms of their self-reported and performance-based levels of physical capability.

Power calculations

Power calculations were performed based on the expected patterns of the distributions of binary and continuous risk factors in the concordant and discordant groups. As the sample sizes of the concordant and discordant groups were fixed these calculations were used to estimate the statistical power of the main analyses. The power was calculated for a range of possible differences in the proportions or means when comparing the concordant group with one of the discordant groups, assuming a significance level (2-sided) of 0.05.

5.3 Results

5.3.1 Cross-tabulation of performance and self-reported physical capability

The first stage of the analysis compared the mean (or median) performance values and the proportion unable to complete across each self-reported response category, for men and women (see Tables 5.5 and 5.6 respectively).

For all cross-tabulation comparisons a general trend was observed, with decreasing mean (or median) performance values as the severity of difficulty increased in the self-reported measures. With the exception of self-reported gripping and standing balance in men (P=0.6), the test for trend demonstrated that this was statistically unlikely to be due to chance (P<0.02 across all tests). A general trend was also observed in the cross-tabulation of those unable to complete specific performance tasks. The proportion of participants recorded as unable to complete the performance tasks increased as the degree of difficulty in the self-reported measures increased. The results of the chi-squared test for trend suggest that it was extremely unlikely that the observed trends were due to chance.

For all cross-tabulations it was noted that, despite no self-reported difficulty, there were always some participants who were unable to complete the performance tests, which suggests discordance. The proportion who reported no difficulty despite being unable to complete the performance tests varied between 0.1-0.8% for specifically matched tasks (i.e. self-reported gripping with grip strength, walking with TUG and self-reported balance with standing balance), compared to a range of 0.1-3.3% for non-matched tasks.

Table 5.5a: Mean performance (grip strength and TUG speed) values and number of participants unable to perform test for health reasons by response category of self-reported tasks in men aged 60-64

Self-reported measures[†]	Performance-based measure							
	Grip Strength (kg)				TUG [‡] speed (m/s)			
	N[a]	Mean	SD [b]	Unable N(%)[c]	N[a]	Mean	SD [b]	Unable N(%)[c]
Grip	**1004**	**45.90**	**11.69**	**13 (1.3)**	**962**	**0.71**	**0.18**	**19 (1.9)**
No difficulty	930	46.25	11.60	4 (0.4)	894	0.72	0.18	6 (0.7)
Some difficulty	64	41.73	12.42	4 (5.9)	61	0.64	0.20	7 (10.3)
A lot	10	40.28	9.74	5 (33.3)	7	0.52	0.19	6 (46.2)
			P=0.001	*P<0.001*			*P<0.001*	*P<0.001*
Balance	**1001**	**45.90**	**11.70**	**13 (1.3)**	**960**	**0.71**	**0.18**	**19 (1.9)**
No difficulty	884	46.43	11.55	7 (0.8)	864	0.73	0.17	3 (0.3)
Occasionally	80	42.80	11.99	3 (3.6)	71	0.60	0.15	5 (6.6)
Often	19	41.30	8.64	0 (0.0)	14	0.53	0.17	3 (17.6)
Always	18	38.54	15.35	3 (14.3)	11	0.46	0.23	8 (42.1)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Walk	**1001**	**45.89**	**11.69**	**12 (1.2)**	**959**	**0.71**	**0.17**	**18 (1.8)**
No difficulty	906	46.27	11.43	6 (0.7)	882	0.73	0.17	3 (0.3)
Walk >400yds*	21	46.29	10.09	1 (4.5)	21	0.57	0.11	0 (0.0)
200-400yds*	30	42.20	15.33	2 (6.3)	27	0.54	0.12	3 (10.0)
50-200yds*	21	36.11	9.54	2 (8.7)	17	0.51	0.12	4 (19.0)
Walk <50yds*	21	44.14	15.62	1 (4.5)	11	0.41	0.14	8 (42.1)
			P=0.002	*P<0.001*			*P<0.001*	*P<0.001*
Climb stairs	**1000**	**45.89**	**11.73**	**13 (1.3)**	**958**	**0.71**	**0.18**	**19 (1.9)**
No difficulty	869	46.37	11.68	5 (0.6)	846	0.73	0.17	4 (0.5)
Normal manner	48	41.31	12.80	0 (0.0)	47	0.64	0.15	0 (0.0)
Hold/take rest	83	43.41	10.90	8 (8.8)	65	0.53	0.15	15 (18.8)
			P=0.005	*P<0.001*			*P<0.001*	*P<0.001*
Bend	**999**	**45.91**	**11.69**	**13 (1.3)**	**957**	**0.71**	**0.18**	**19 (1.9)**
No difficulty	874	46.19	11.45	6 (0.7)	853	0.73	0.18	4 (0.5)
Can sweep	84	45.16	12.72	2 (2.3)	75	0.61	0.15	5 (6.3)
Cannot sweep	15	42.03	14.14	1 (6.3)	15	0.55	0.16	1 (6.3)
Cannot pick	20	42.55	13.98	1 (4.8)	12	0.53	0.12	3 (20.0)
Cannot touch knees	6	35.53	13.83 *P=0.015*	3 (33.3) *P<0.001*	2	0.34	0.21 *P<0.001*	6 (75.0) *P<0.001*

† Top line (bold) provides values for total sample for that variable
‡ Timed Get Up and Go
[a] Number of participants who completed task (does not include those unable for health reasons)
[b] P values from non-parametric test for trend
[c] N (unable) as % of total number of participants (N + N (unable)). P values from χ^2 test for trend
*Distance participants can walk without stopping or experiencing severe discomfort

Table 5.5b: Mean performance (chair rise speed and balance time) values and number of participants unable to perform test for health reasons by response category of self-reported tasks in men aged 60-64

Self-reported measures†	Performance-based measures							
	Chair rise speed (rise/s)				Balance time (eyes closed) (s)			
	N[a]	Mean	SD [b]	Unable N(%)[c]	N[a]	Median	IQR [b]	Unable N(%)[c]
Grip	**987**	**0.44**	**0.12**	**62 (5.9)**	**1014**	**3.69**	**2.48-5.63**	**40 (3.8)**
No difficulty	933	0.44	0.12	29 (3.0)	945	3.68	2.50-5.62	20 (2.1)
Some difficulty	50	0.40	0.16	23 (31.5)	61	4.12	2.47-6.15	13 (17.6)
A lot	4	0.33	0.06	10 (71.4)	8	2.36	1.0-4.47	7 (46.7)
			P<0.001	*P<0.001*			*P=0.607*	*P<0.001*
Balance	**984**	**0.44**	**0.12**	**62 (5.9)**	**1011**	**3.69**	**2.48-5.62**	**40 (3.8)**
No difficulty	900	0.45	0.12	23 (2.5)	921	3.80	2.53-5.75	5 (0.5)
Occasionally	72	0.37	0.12	12 (14.3)	77	2.94	1.91-4.37	9 (10.5)
Often	7	0.32	0.09	12 (63.1)	7	3.32	1.57-3.53	12 (63.1)
Always	5	0.26	0.09	15 (75.0)	6	3.97	2.00-5.91	14 (70.0)
			P<0.001	*P<0.001*			*P=0.001*	*P<0.001*
Walk	**984**	**0.44**	**0.12**	**61 (5.8)**	**1011**	**3.70**	**2.49-5.63**	**39 (3.7)**
No difficulty	933	0.45	0.12	11 (1.2)	942	3.78	2.53-5.72	6 (0.6)
Walk >400yds*	13	0.39	0.09	10 (43.5)	20	3.92	2.66-5.58	4 (16.7)
200-400yds*	16	0.36	0.10	14 (46.7)	22	2.47	1.89-3.22	9 (29.0)
50-200yds*	16	0.27	0.07	9 (36.0)	17	3.07	2.03-4.10	7 (29.2)
Walk <50yds*	5	0.28	0.05	17 (77.3)	9	2.97	1.59-3.37	13 (59.1)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Climb stairs	**983**	**0.44**	**0.12**	**62 (5.9)**	**1009**	**3.72**	**2.50-5.63**	**40 (3.8)**
No difficulty	893	0.45	0.12	12 (1.3)	904	3.82	2.53-5.77	6 (0.7)
Normal manner	43	0.36	0.08	6 (12.2)	46	3.39	2.66-5.63	3 (6.1)
Hold/take rest	47	0.35	0.12	44 (48.4)	59	2.87	1.97-4.10	31 (34.4)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Bend	**984**	**0.44**	**0.12**	**60 (5.7)**	**1009**	**3.69**	**2.47-5.62**	**40 (3.8)**
No difficulty	894	0.45	0.12	17 (1.9)	901	3.80	2.50-5.75	12 (1.3)
Can sweep	68	0.37	0.12	20 (22.7)	79	3.28	2.21-4.75	11 (12.2)
Cannot sweep	10	0.35	0.12	5 (33.3)	12	3.28	1.90-4.00	3 (20.0)
Cannot pick	12	0.31	0.08	10 (45.5)	15	2.75	2.42-3.30	8 (34.8)
Cannot touch knees	0	/	/	8 (100.0)	2	6.27	5.91-6.63	6 (75.0)
			P<0.001	*P<0.001*			*P=0.013*	*P<0.001*

† Top line (bold) provides values for total sample for that variable
[a] Number of participants who completed task (does not include those unable for health reasons)
[b] P values from non-parametric test for trend
[c] N (unable) as % of total number of participants (N + N (unable)). P values from χ^2 test for trend
*Distance participants can walk without stopping or experiencing severe discomfort

Table 5.6a: Mean performance (grip strength and TUG speed) values and number of participants unable to perform test for health reasons by response category of self-reported tasks in women aged 60-64

Self-reported measures†	Performance-based measure							
	Grip Strength (kg)				TUG ‡ speed (m/s)			
	N[a]	Mean	SD [b]	Unable N(%)[c]	N[a]	Mean	SD [b]	Unable N(%)[c]
Grip	**1063**	**26.80**	**7.48**	**36 (3.3)**	**1064**	**0.68**	**0.17**	**15 (1.4)**
No difficulty	777	27.50	7.44	2 (0.3)	770	0.70	0.17	4 (0.5)
Some difficulty	257	25.17	7.21	15 (5.5)	258	0.66	0.15	4 (1.5)
A lot	29	22.42	7.27	19 (39.6)	36	0.53	0.18	7 (16.3)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Balance	**1058**	**26.81**	**7.48**	**34 (3.1)**	**1058**	**0.68**	**0.17**	**14 (1.3)**
No difficulty	806	27.08	7.35	10 (1.2)	802	0.71	0.16	1 (0.1)
Occasionally	196	26.29	7.87	12 (5.8)	201	0.64	0.15	2 (1.0)
Often	28	25.74	7.01	4 (12.5)	29	0.55	0.18	3 (9.4)
Always	28	23.95	8.35	8 (22.2)	26	0.45	0.17	8 (23.5)
			P=0.005	*P<0.001*			*P<0.001*	*P<0.001*
Walk	**1055**	**26.82**	**7.47**	**36 (3.3)**	**1055**	**0.69**	**0.17**	**15 (1.4)**
No difficulty	944	27.18	7.37	19 (2.0)	944	0.71	0.16	1 (0.1)
Walk >400yds*	22	23.22	6.79	3 (12.0)	25	0.55	0.13	0 (0.0)
200-400yds*	30	23.41	7.73	3 (9.1)	31	0.53	0.12	1 (3.1)
50-200yds*	25	24.86	8.22	4 (13.8)	25	0.53	0.23	3 (10.7)
Walk <50yds*	32	23.53	7.73	7 (17.9)	28	0.43	0.15	10 (26.3)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Climb stairs	**1056**	**26.82**	**7.47**	**35 (3.2)**	**1056**	**0.69**	**0.17**	**15 (1.4)**
No difficulty	853	27.30	7.45	12 (1.4)	849	0.71	0.16	1 (0.1)
Normal manner	63	25.40	7.54	2 (3.1)	62	0.61	0.16	2 (3.1)
Hold/take rest	140	24.52	7.13	21 (13.0)	145	0.56	0.17	12 (7.6)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Bend	**1056**	**26.81**	**7.49**	**36 (3.3)**	**1058**	**0.68**	**0.17**	**14 (1.3)**
No difficulty	891	27.21	7.50	16 (1.8)	891	0.70	0.16	2 (0.2)
Can sweep	112	24.92	6.95	7 (5.9)	113	0.61	0.19	1 (0.9)
Cannot sweep	29	24.89	7.87	4 (12.1)	33	0.52	0.16	1 (2.9)
Cannot pick	21	22.95	6.30	7 (25.0)	21	0.53	0.16	7 (25.0)
Cannot touch knees	3	21.50	9.20	2 (40.0)	0	/	/	3 (100.0)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*

† Top line (bold) provides values for total sample for that variable
‡ Timed Get Up and Go
[a] Number of participants who completed task (does not include those unable for health reasons)
[b] P values from non-parametric test for trend
[c] N (unable) as % of total number of participants (N + N (unable)). P values from χ^2 test for trend
*Distance participants can walk without stopping or experiencing severe discomfort

Table 5.6b: Mean performance (chair rise speed and balance time) values and number of participants unable to perform test for health reasons by response category of self-reported tasks in women aged 60-64

Self-reported measures[†]	Performance-based measure							
	Chair rise speed (rise/s)				Balance time (eyes closed) (s)			
	N[a]	Mean	SD [b]	Unable N(%)[c]	N[a]	Median	IQR [b]	Unable N(%)[c]
Grip	**1073**	**0.43**	**0.13**	**74 (6.5)**	**1098**	**3.28**	**2.25-4.83**	**49 (4.3)**
No difficulty	787	0.43	0.13	27 (3.3)	800	3.32	2.28-4.85	17 (2.1)
Some difficulty	257	0.42	0.15	26 (9.2)	268	3.12	2.22-4.82	14 (5.0)
A lot	29	0.32	0.12	21 (42.0)	30	2.66	1.81-3.41	18 (37.5)
			P<0.001	*P<0.001*			*P=0.021*	*P<0.001*
Balance	**1068**	**0.42**	**0.13**	**71 (6.2)**	**1092**	**3.28**	**2.25-4.84**	**47 (4.1)**
No difficulty	836	0.43	0.13	14 (1.6)	846	3.31	2.35-4.99	7 (0.8)
Occasionally	192	0.39	0.12	22 (10.3)	200	2.97	2.05-4.72	11 (5.2)
Often	23	0.38	0.10	13 (36.1)	25	2.97	1.75-3.56	11 (30.6)
Always	17	0.37	0.12	22 (56.4)	21	2.49	1.28-3.68	18 (46.2)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Walk	**1066**	**0.43**	**0.13**	**72 (6.3)**	**1090**	**3.28**	**2.25-4.84**	**48 (4.2)**
No difficulty	984	0.43	0.13	17 (1.7)	993	3.30	2.28-4.97	10 (1.0)
Walk >400yds*	22	0.37	0.10	4 (15.4)	20	2.52	1.77-3.13	5 (20.0)
200-400yds*	23	0.32	0.09	12 (34.3)	27	3.15	2.13-4.59	7 (20.6)
50-200yds*	20	0.32	0.12	12 (37.5)	24	2.53	1.95-4.10	7 (22.6)
Walk <50yds*	16	0.30	0.07	26 (61.9)	24	2.57	1.65-4.36	19 (44.2)
			P<0.001	*P<0.001*			*P=0.001*	*P<0.001*
Climb stairs	**1067**	**0.43**	**0.13**	**72 (6.3)**	**1091**	**3.25**	**2.25-4.84**	**48 (4.2)**
No difficulty	889	0.44	0.13	7 (0.8)	890	3.32	2.34-5.19	9 (1.0)
Normal manner	58	0.38	0.13	12 (17.1)	68	3.22	2.22-4.39	2 (2.9)
Hold/take rest	120	0.34	0.12	53 (30.6)	133	2.60	1.85-3.85	37 (21.8)
			P<0.001	*P<0.001*			*P<0.001*	*P<0.001*
Bend	**1067**	**0.42**	**0.13**	**73 (6.4)**	**1092**	**3.25**	**2.25-4.83**	**48 (4.2)**
No difficulty	924	0.43	0.13	16 (1.7)	930	3.28	2.27-4.91	13 (1.4)
Can sweep	106	0.39	0.15	23 (17.8)	115	3.00	2.15-4.56	11 (8.7)
Cannot sweep	23	0.35	0.11	11 (32.4)	25	3.16	1.62-4.03	9 (26.5)
Cannot pick	14	0.32	0.13	18 (56.3)	21	2.86	1.63-3.46	11 (34.4)
Cannot touch knees	0	/	/	5 (100.0)	1	3.93	3.93-3.93	4 (80.0)
			P<0.001	*P<0.001*			*P=0.007*	*P<0.001*

† Top line (bold) provides values for total sample for that variable
[a] Number of participants who completed task (does not include those unable for health reasons)
[b] P values from non-parametric test for trend
[c] N (unable) as % of total number of participants (N + N (unable)). P values from χ^2 test for trend
*Distance participants can walk without stopping or experiencing severe discomfort

5.3.2 Summary self-reported measure

The following section will detail the results of the descriptive analyses performed on the summary self-reported measure used to facilitate the identification of concordant and discordant cases within the NSHD dataset at age 60-64. Section 5.3.3 will then focus on the summary performance-based measure.

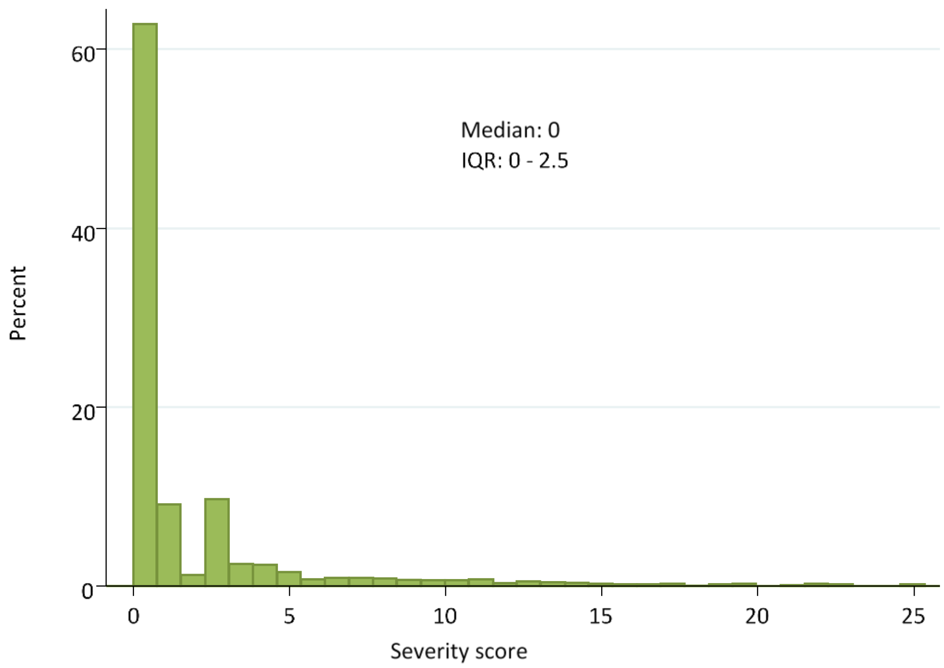
Figure 5.2 shows the distribution of the summary self-reported measure (aggregated severity score) within the NSHD at age 60-64. The summary variable incorporated data from 2,185 participants, with 55 participants excluded due to missing data. The values ranged from 0 (no reported difficulty across all five tasks) to the maximum theoretical value of 25.33 (difficulty in all five tasks).

There was a noticeable skew to the summary variable, with a large proportion (60.8%) of the population reporting no difficulty with any of the tasks included in the measure (score of 0). The Mann-Whitney test revealed a significant difference (P<0.01) between the median severity scores for men (0) and women (0.75).

Figure 5.3 presents the severity scores of participants within the NSHD excluding those who reported no difficulty. The distribution remains skewed, with an interquartile range of 1.50-6.08. Despite this skewed distribution, a decision was made not to transform the variable, as it would not be possible to correct for the large number of participants with a value of 0, and the distribution reflects the underlying theoretical construct.

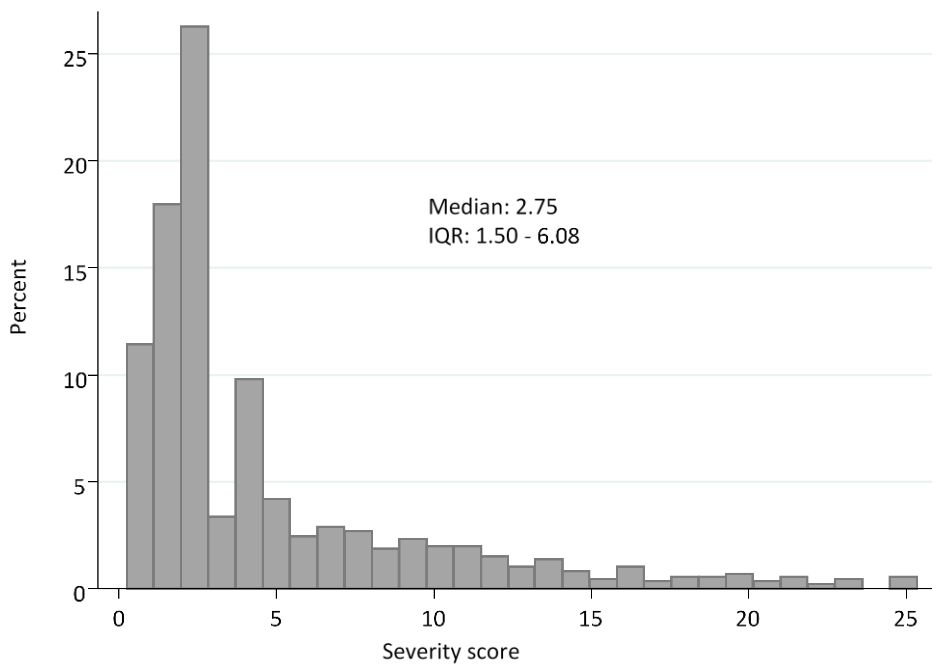
The theoretical maximum value on this scale was 25.33, whereby participants would have had to report the most severe category of difficulty across all five self-reported measures. In the NSHD sample, four participants achieved the maximum value (see Figure 5.3). The spread of participants across the full range of potential values reiterates the variability captured within the summary self-reported variable and facilitates comparability with the summary performance variable.

Figure 5.2: Histogram of summary self-reported measure at age 60-64



N=2185

Figure 5.3: Histogram of summary self-reported measure at age 60-64, excluding participants who reported no difficulty



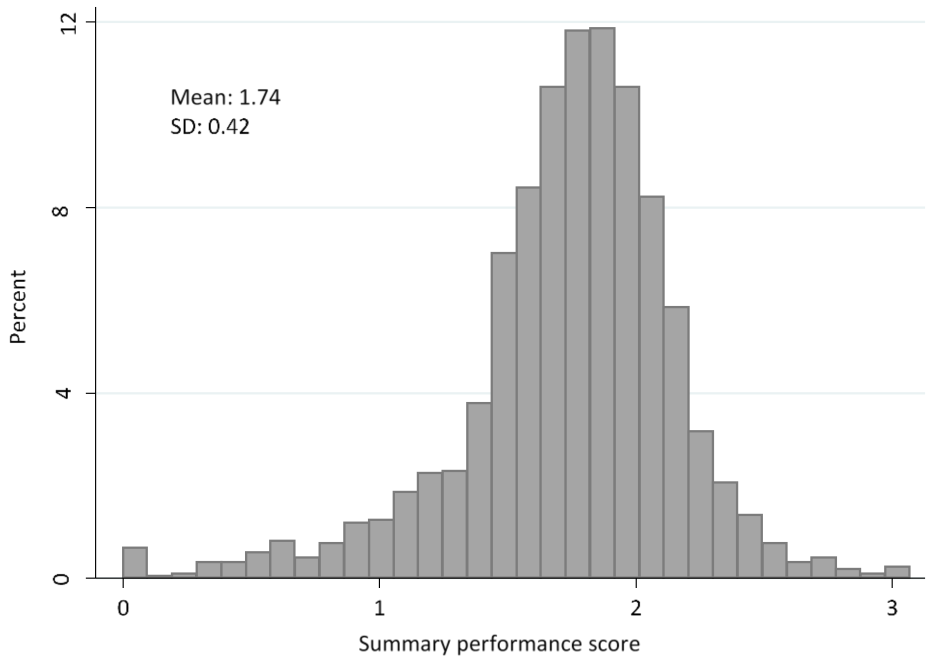
N=857

5.3.3 Summary performance-based measure

Figure 5.4 shows the values produced for the performance summary variable and the normal distribution observed within the NSHD sample. The potential range of the summary performance variable is 0 to 4, with 0 representing low physical capability and 4 representing high levels of physical capability across all four component performance tests. The maximum score achieved within the NSHD population was 3.07, and the mean score of 1.74 was less than half the potential maximum of 4. Thirteen participants were observed to have a performance summary score of 0, which indicated that they were unable to complete any of the performance tasks or were above the 99th percentile once rescaled.

The results of the t-test showed a significant sex difference (P<0.001), with men achieving higher summary performance scores than women (a mean of 1.78 compared to 1.70).

Figure 5.4: Histogram of summary performance-based measure at age 60-64



N=1,980

5.3.4 Sensitivity analysis to test assumptions of recoded data

When the two summary measures were correlated, there was a noticeable reduction in the sample size, with over 300 participants missing values for one of the summary measures (see Table 5.7).

There appears to be no significant sex difference between those with complete data and those with missing values (P=0.5). However, those with missing data have lower performance scores and higher levels of reported difficulty (P<0.001) compared to participants with complete data.

Following the protocol laid out in Section 5.2.2, data were recoded for 51 participants, increasing the sample size for the collated summary measures to 1,981 (see Table 5.8). With the recoded data incorporated, the characteristics of the extended sample are similar to the original sample. There remains no significant sex difference between those with complete data and those with missing values (P=0.5), and those with missing data continue to have lower performance scores and higher levels of reported difficulty (P<0.05) compared to participants with complete data.

When comparing those with recoded data (N=51) and those in the original sample (N=1,930) who have data available for both measures, there was no significant sex difference (P=0.3) (see Table 5.9). However, those with recoded data have lower performance scores and higher levels of reported difficulty (P<0.001) compared to those from the original sample, reflecting the decisions made when imputing data.

Descriptive statistics comparing the original and recoded summary measures (Table 5.10) show that the recoded data has had minimal impact on the two summary measures. Similar results were observed when comparisons were made sex-specifically.

Table 5.7: Characteristics of participants with missing data for the two summary measures

		Participants with data available for			P Value
		Both measures	Self-report only	Performance only	
N		1930	259	50	
Sex	% female	52.1	51.8	60.0	0.5[a]
Performance summary measure	Mean (SD)	1.74 (0.42)	-	1.51 (0.57)	0.001[b]
Self-reported summary measure	Median (IQR)	0 (0-2.33)	0 (0-3.75)	-	0.001[c]

[a] chi-squared test, [b] t-test, [c] Mann-Whitney non-parametric test

Table 5.8: Characteristics of participants with missing data for the two summary measures, after recoding

		Participants with data available for			P Value
		Both measures	Self-report only	Performance only	
N		1981	248	10	
Sex	% female	52.3	51.21	70	0.5 [a]
Performance summary measure	Mean (SD)	1.74 (0.43)	-	1.43 (0.63)	0.03[b]
Self-reported summary measure	Median (IQR)	0 (0-2.33)	0 (0-3.17)	-	0.002[c]

[a] chi-squared test, [b] t-test, [c] Mann-Whitney non-parametric test

Table 5.9: Characteristics of those with data available for both summary measures

		Original	Recoded	P Values
N		1930	51	
Sex	% female	52.12	58.82	0.3 [a]
Performance summary measure	Mean (SD)	1.74 (0.42)	1.41 (0.62)	<0.001 [b]
Self-reported summary measure	Median (IQR)	0 (0-2.33)	2.25 (0.25-6.58)	<0.001 [c]

[a] chi-squared test, [b] t-test, [c] Mann-Whitney non-parametric test

Table 5.10: Descriptive statistics of the sample comparing the original and recoded summary measures

		N	Mean/ Median	SD/IQR	Min	Max
Performance summary measure	Original	1980	1.74	0.42	0	3.07
	Recoded	1991	1.73	0.43	0	3.07
Self-reported summary measure	Original	2185	0.0	0-2.50	0	25.3
	Recoded	2229	0.0	0-2.58	0	25.3

The correlation coefficients between the two summary measures (see Table 5.11) demonstrate that the inclusion of the recoded data has had little impact on the relationship between the two measures, although the two summary measures appear slightly more concordant with the inclusion of the recoded data.

Table 5.11: Correlation coefficients between the two summary measures

Summary measures	**Correlation coefficients**					
	Male		Female		All	
Original variables	N=924	-0.5761	N=1006	-0.6025	N=1930	-0.5923
Recoded variables	N=945	-0.5867	N=1036	-0.6121	N=1981	-0.6027

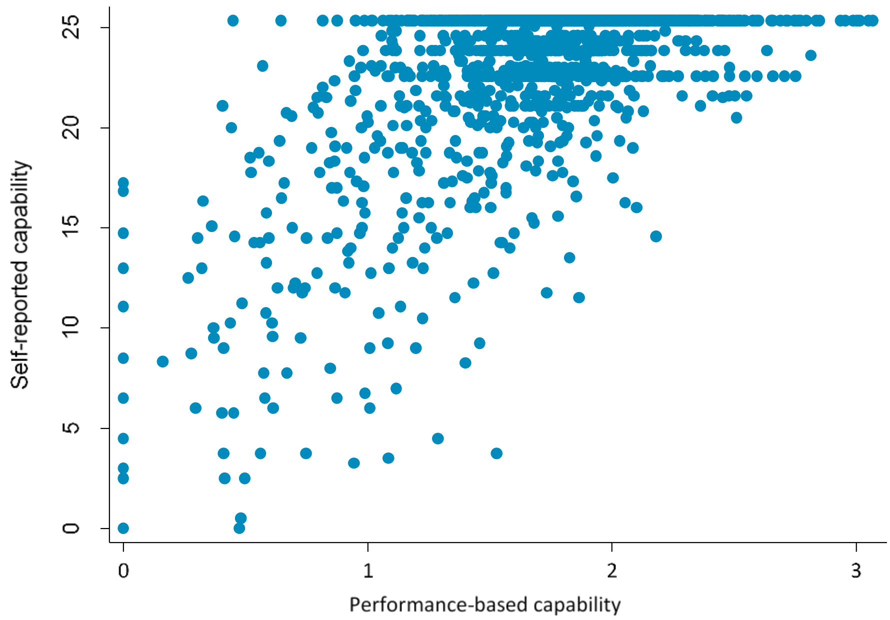
5.3.5 Identification of concordant and discordant groups

The two summary variables were plotted against each other to produce a scatter graph as shown in Figure 5.5.

Although the positively skewed nature of the self-reported summary variable is noticeable, the distribution of participants across the scatterplot highlights the variability in levels of concordance and discordance experienced by participants in early old age within the NSHD.

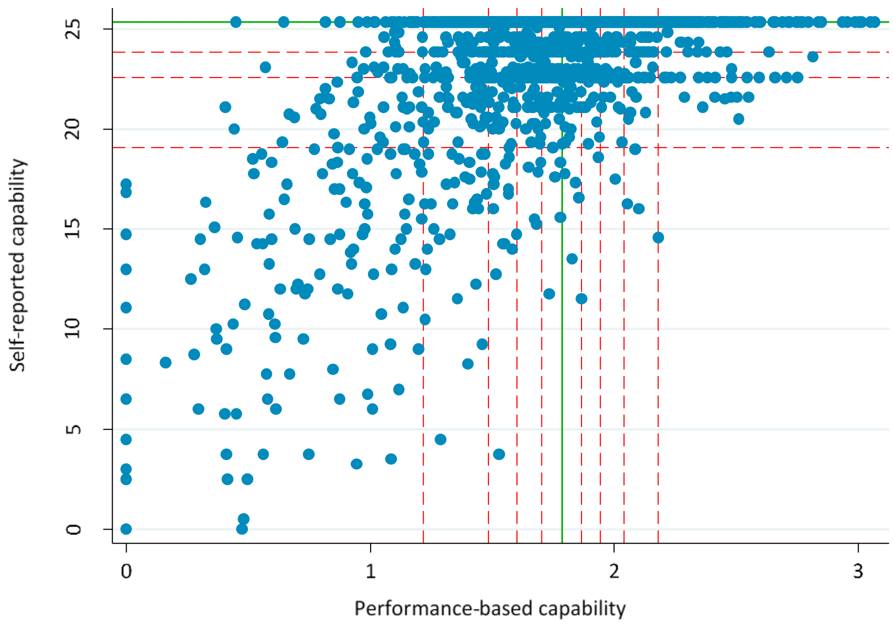
In order to facilitate the identification of three concordant and two discordant groups, the two summary measures were divided into decile groups. The skewed distribution of the self-reported variable meant that the 5^{th} -10^{th} deciles were combined into one group, as they were indistinguishable (see Figure 5.6).

Figure 5.5: Scatterplot of performance and self-reported physical capability at 60-64



N=1981

Figure 5.6: Deciles of performance and self-reported physical capability at 60-64



Red dashed lines=decile
Green line= median

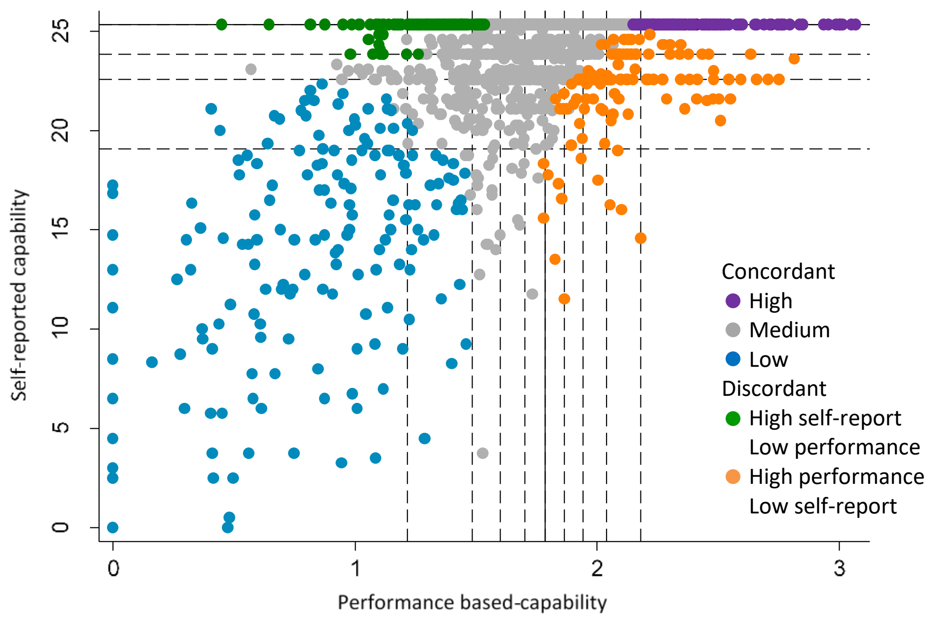
Decisions were made *a priori* about how to define each of the concordant and discordant groups based on the decile classification. It was important to account for the skewed nature of the self-reported variable when making these decisions. For example, only those in the top 10% of performance values were considered to be truly concordant with the self-reported variable (see Figure 5.7, purple) due to the large proportion of participants who reported no difficulty (the highest value on the self-reported scale).

At the high end of the performance spectrum (top 20%), participants who reported any level of difficulty were viewed as discordant, as the vast majority of participants report no difficulty across the performance spectrum. In addition to these individuals, anyone with a performance value above the median who reported severe difficulty (lowest 10% of self-reported values) was considered discordant. A stepped approach was used between the median and top 20% to identify the rest of this discordant group (see Figure 5.7, orange).

The opposing discordant group was identified by incorporating all those who reported no difficulty yet were observed to be in the bottom 20% of performance values. It was also decided that anyone who reported only minor limitations but was in the lowest 10% of performance values should also be considered discordant (see Figure 5.7, green). The concordant group at the lower end of the spectrum were identified as participants in the lowest 20% of performance values who reported severe limitations (lowest 10% of self-reported values). Participants with values in the lowest 10% of performance scores and in the second lowest decile of the self-reported variable were also included in this concordant group (see Figure 5.7, blue).

The distribution of the discordant and concordant groupings can be seen in Figure 5.7. The slight overlap visible between the discordant and concordant groups is due to the sex-specific deciles used to define the cut points in the performance-based measure.

Figure 5.7: Identification of concordant and discordant groups



On the basis of these definitions, 172 participants were allocated to the low concordant group, 157 to the high concordant group, 138 to the low performance high self-reported capability group and 135 to the high performance, low self-reported capability group (see Table 5.12). The remaining participants (N=1,379) were assigned to the middle concordant group for future analyses.

Table 5.12: Breakdown of study population (N) across deciles of performance and self-reported physical capability at age 60-64

	Performance deciles									
Self-reported deciles	1st (Low)	2nd	3rd	4th	5th	6th	7th	8th	9th	10th (High)
5th-10th (High)	29	97	119	117	126	127	141	147	151	157
4th	12	26	29	27	23	24	24	22	19	15
3rd	15	18	20	34	24	28	21	20	16	18
2nd	29	26	18	15	19	14	10	8	8	8
1st (Low)	112	31	13	6	5	4	3	2	4	0

Concordant groups: Blue= low capability (N=172), Purple= high capability (N=157), White= medium capability (N=1379)

Discordant groups: Green (overestimators) = low performance, high self-reported capability (N=138)

Orange (underestimators) = high performance, low self-reported capability (N=135)

Once the five concordant and discordant groups were defined, the self-reported physical capability levels (see Table 5.13) and the performance-based levels (see Table 5.14) of each group were explored. For both summary measures of physical capability a gradient was observed across the three concordant groups, with low levels of physical capability amongst participants in the low concordant group and high levels of physical capability for participants in the high concordant group. In terms of their self-reported physical capability, overestimators were similar to those in the high concordant group, reporting no difficulty, whilst underestimators had physical capability levels between those in the middle and low concordant groups. With regards to their performance-based characteristics, overestimators had physical capability levels between those in the low and middle concordant groups, whereas underestimators had performance-based physical capability levels between those in the middle and high concordant groups.

Table 5.13: Median self-reported physical capability summary score of the concordant and discordant groups at age 60-64

Concordant and discordant groups	N (%)	**Self-reported physical capability** Median (IQR†)
Total	1981	25.33 (23.00 – 25.33)
Concordant		
Low	172 (8.68)	14.88 (10.25 – 18.33)
Middle	1379 (69.6)	25.33 (23.83 – 25.33)
High	157 (7.93)	25.33 (25.33 – 25.33)
Discordant ‡		
Underestimators	135 (6.81)	22.58 (21.58 – 23.83)
Overestimators	138 (6.97)	25.33 (25.33 – 25.33)

† IQR = Interquartile Range
‡ Discordant groups: Underestimator = high performance, low self-reported capability
Overestimator = low performance, high self-reported capability

Table 5.14: Mean performance-based physical capability summary score of the concordant and discordant groups at age 60-64

Concordant and discordant groups	N (%)	Performance-based physical capability Mean (SD†)
Total	1981	1.74 (0.42)
Concordant		
Low	172 (8.68)	0.83 (0.39)
Middle	1379 (69.6)	1.78 (0.22)
High	157 (7.93)	2.38 (0.21)
Discordant ‡		
Underestimators	135 (6.81)	2.12 (0.22)
Overestimators	138 (6.97)	1.30 (0.18)

† SD= Standard Deviation
‡ Discordant groups: Underestimator = high performance, low self-reported capability
Overestimator = low performance, high self-reported capability

5.3.6 Power Calculations

The following power calculations have been produced based on the concordant and discordant groupings defined in Section 5.3.5. For ease of calculation the number of participants in each group has been rounded: 1700 in the concordant group and 140 in each of the discordant groups. As the two discordant groups include approximately equal numbers of participants, the power calculations shown in this section have been based on comparisons between the concordant group and one of the discordant groups, on the understanding that the results presented would be identical for the other discordant group.

The first set of power calculations explored the statistical power within the study when comparing the proportions of a binary risk factor between the concordant and one of the discordant groups (see Table 5.15). The calculations show that power within this study ranges from 20.5 – 100% given the expected distributions of the binary risk factors to be included in future analyses.

Table 5.15: Estimates of statistical power within the study when comparing the proportions of a binary risk factor in the concordant group with one of the discordant groups

	Power (%)*										
Prop. in concordant group (N=1700)	Proportion in discordant group (N=140)										
	0.25	0.3	0.35	0.45	0.5	0.55	0.65	0.7	0.75	0.8	0.85
0.3	22.7	-	24.3	94.7	99.7	100	100	100	100	100	100
0.5	100	99.8	93.5	20.5	-	20.5	93.5	99.8	100	100	100
0.7	100	100	100	100	99.7	94.7	24.3	-	22.7	72.8	98.7
0.8	100	100	100	100	100	100	97.3	77.5	30.5	-	28.1

*Significance level of 0.05 assumed and size of concordant/discordant groups accounted for

The second set of power calculations explored the statistical power within the study when comparing means of continuous risk factors (See Tables 5.16-5.18). As the concordant group contains 85% of the sample population, the mean and standard deviation values of the sample population have been used as a guide for expected factors of interest; BMI for Table 5.16, wellbeing for Table 5.17 and verbal memory for Table 5.18.

These calculations show that statistical power within the study ranges from 19.4-99.7% for BMI, 10.7-83.0% for wellbeing and 15.0-97.5% for verbal memory, given the expected mean values for each of these risk factors and a difference in means of between 0.5 and 2.0 units.

Table 5.16: Power estimates when comparing the difference in mean BMI between the concordant and one of the discordant groups

	Power (%)*				
Difference in mean†	Standard deviation				
	4.8	4.9	5.0	5.1	5.2
0.5	22.0	21.3	20.6	20.0	19.4
1.0	65.8	64.1	62.3	60.6	58.9
1.5	94.4	93.6	92.7	91.7	90.6
2.0	99.7	99.6	99.4	99.4	99.2

* Significance level of 0.05 assumed and size of concordant/discordant groups accounted for
† Concordant group assigned mean value of 28 kg/m^2

Table 5.17: Power estimates when comparing the difference in mean wellbeing between the concordant and one of the discordant groups)

	Power (%)*				
Difference	Standard deviation				
in mean†	7.8	7.9	8.0	8.1	8.2
0.5	11.3	11.1	11.0	10.8	10.7
1.0	30.8	30.2	29.5	28.9	28.4
1.5	58.9	57.9	56.8	55.8	54.8
2.0	83.0	82.1	81.1	80.1	79.2

* Significance level of 0.05 assumed and size of concordant/discordant groups accounted for
† Concordant group assigned mean value of 52 WEMWBS score

Table 5.18: Power estimates when comparing the difference in mean verbal memory between the concordant and one of the discordant groups

	Power (%)*				
Difference	Standard deviation				
in mean†	5.8	5.9	6.0	6.1	6.2
0.5	16.5	16.1	15.7	15.4	15.0
1.0	50.0	48.7	47.4	46.2	45.0
1.5	83.6	82.4	81.1	79.8	78.5
2.0	97.5	97.1	96.6	96.1	95.6

* Significance level of 0.05 assumed and size of concordant/discordant groups accounted for
† Concordant group assigned mean value of 24 words recalled

5.4 Discussion

Selected methods identified in the literature review were used to establish evidence of concordance and discordance within the NSHD. To further explore the relationship between the self-reported and performance-based measures of physical capability at age 60-64, two summary measures were produced. These measures were used to define five concordant and discordant groups.

5.4.1 Evidence of concordance and discordance in the NSHD at age 60-64

The initial analysis, based on the cross-tabulation of mean (or median) performance values against each response category of the self-reported variables, was used to explore whether there was any evidence of concordance and discordance in the NSHD at age 60-64. As expected, the mean (or median) performance values decreased as the severity of difficulty increased across the self-reported response categories. These results reflect the stepwise relationship between the performance and self-reported measures reported in the literature (39;115;134;135), suggesting a high level of concordance.

There has been less research incorporating those who were unable to complete the performance tasks for health reasons. Of the studies included in the literature review at the start of this chapter, the closest any came to analysing those unable to complete the tasks was Simonsick et al (135) who reported a stepwise relationship between the proportion able to complete a 400 metre walk and self-reported capability. The results from this thesis demonstrated a similar trend, with an increasing proportion of study members who were unable to complete the task for health reasons, as the severity of difficulty in the self-reported variable also increased.

Combined, these two sets of results provide strong evidence that the self-reported and performance-based measures of physical capability within the NSHD are concordant. However, there was also evidence of discordance, with a small proportion of the sample reporting no difficulty despite being unable to perform the individual tests for health reasons. In line with previous research comparing matched and non-matched pairings of tasks (130;133), it was observed that this level of discordance was generally lower for matched than non-matched tasks, although there was considerable overlap between the two groups, as noted in Section 5.3.1. It seems logical that specifically matched tasks should have lower levels of discordance as they focus on the same task; however the self-reported and performance-based measures in this thesis do not align exactly, which may explain why the level of discordance is similar to tasks that have not been matched.

It is also worth noting that a trend in mean performance score does not reflect the experience of individuals. For example, despite recording a mean grip strength value of 46.25kg, men who reported no difficulty gripping had grip strength values ranging from 11.8kg to 87.8kg. In each of the self-reported categories the range of performance values achieved by participants varied across the whole spectrum of physical capability, and therefore discordant cases may be found within each self-reported response category.

The cross-tabulation approach was selected for this analysis because it utilised the categorical nature of the self-reported variables and the continuous nature of the performance-based measures. Although the results presented in Section 5.3.1 address the objective of this preliminary analysis, the limitations of this method (discussed in the literature review) mean it has only been possible to ascertain the magnitude of discordance amongst those who are unable to complete each performance-based measure due to health reasons. In order to explore the level of concordance and discordance across the whole study population in more detail, an alternative approach was required.

5.4.2 Production of two summary measures

A decision was made to produce two summary measures to facilitate the identification of concordant and discordant cases. By producing summary variables, the analysis encapsulated the overall experience of physical capability for individuals rather than restricting the analysis to individual physical capability tasks.

<u>Self-reported summary measure</u>

The advantage of the self-reported summary measure produced was the creation of a continuous variable, which overcame the limitations experienced in previous research using categorical data. For example, the continuous nature of the summary self-report variable facilitated comparison with the continuous performance-based measures. Although summary self-reported measures, such as the performance domain of the SF-36, already exist, a decision was made to create a new summary measure. This decision was made because the existing summary variables do not exclusively incorporate tasks focused on physical capability, as conceptualised in this thesis. In addition, existing summary measures tend to simply aggregate ordinal categorical values, which does not enable them to account for the relative difficulty of the physical capability tasks.

The summary self-reported measure created was based on the severity scale derived by the OPCS Survey of Disability (82). The component scores of this scale were subjective, however the high level of agreement between the judges and the clear assessment procedure provide confidence that the scores assigned were not arbitrary. Unfortunately, not all of the self-reported measures available in this thesis mapped onto the OPCS scale. Where it was not possible to directly assign a value from the severity scale, an estimate was produced following a systematic procedure to ensure that values were assigned in a logical manner.

The summary variable produced has a noticeably skewed distribution, which is unsurprising in this relatively young population, due to the inherent nature of self-reported measures to focus on the loss of function. However, the proportion reporting no difficulty was only 60.8% in the summary variable, compared to the individual self-reported variables where the lowest comparable value was 81% for self-reported difficulty gripping (see Section 3.3.2). These values suggest that the summary measure captured greater variability across the spectrum of physical capability than the discrete self-reported response categories of any one variable by itself.

<u>Performance-based summary measure</u>

As previously outlined in Section 5.2.2, the performance-based summary measure was created by adapting a published method reported by Guralnik and colleagues (81). This approach facilitated the inclusion of people who reported that they were unable for health reasons to complete a specific task. This additional data may be informative when making comparisons with the self-reported measures, as it enables the summary measure to incorporate the full spectrum of physical capability, particularly those with lower levels of physical capability. The inclusion of individuals who are unable to perform specific tests for health reasons may also reduce bias in the summary measure, as the proportion who are unable to complete each task will differ depending on the task in question and the relative difficulty of that task (136).

The method used to produce the performance-based summary measure gave equal weighting to each of the performance tasks. This may be considered as a limitation of the method, however there is no clear way of weighting the component tasks. Each of the component measures have been rescaled relative to the 99th percentile of the study population, which means the relative difficulty of tasks within this sample may not be consistent with the weighting that would be applied at the general population level.

The summary measure produced has a normal distribution, but a maximum value of only 3.08 despite the theoretical maximum of 4.0. This suggests that the four measures of performance used to produce the summary measure capture different elements of physical capability, as none of the study members were able to achieve relative maximum performance across all four tasks. This reiterates the point that the summary measure is able to capture the whole spectrum of physical capability, in a way that the individual measures cannot.

<u>Cross-tabulating the two summary measures</u>

When the two summary measures were correlated, there was a reduction in the sample size of 300 people. Where possible, data were recoded to increase the sample size from 1,930 to 1,981 study members. It was not possible to impute using multiple imputation techniques, as the data were not missing at random. The procedure outlined in Section 5.2.2 restricted imputation to cases where a clear argument could be made for the manipulation of data. For example, coding those with one component of the self-reported summary measure missing to 0, when all other values are 0, corresponds to the observed results for 60% of the study population compared to only 20% who reported difficulty with one task. The most contentious decision was to recode data for those who reported difficulty but did not specify the severity of this difficulty (N=30). The decision was made to include these individuals because they had provided informative details, which could relate to their performance scores, but they did not answer the second part of the gated question. Within this relatively young and capable population, individuals who reported that they experienced difficulty were up to five times more likely to report the lowest level of difficulty than any other level of severity. Consequently, a decision was made to assume that those who had an unknown level of difficulty were most likely to report only minor limitations.

As expected, those with recoded data had lower levels of performance-based capability than the original sample. This result is understandable given that almost all of the recoded values for the summary performance-based measures were coded as 0 (the minimum value) for participants who were unable to complete a specific test for health reasons. Individuals with recoded data also had higher levels of reported difficulty in the summary self-reported measure compared to the original sample. This can be explained, given only a third of the recoded data set reported no difficulties across all five tasks, compared to 60% in the original sample.

Nevertheless, the inclusion of the recoded data has had little overall impact on the two summary measures, with the summary variables incorporating recoded data producing similar mean, median, standard deviations, interquartile ranges and correlation coefficients to the original variables. This suggests that analyses throughout the rest of this thesis would benefit from the inclusion of the recoded data, as it increases the sample size without introducing a substantial amount of bias to the results.

5.4.3 Identification of concordant and discordant groups

Both of the summary measures produced showed good variability across the spectrum of physical capability that was reflected in the spread of concordant and discordant cases when the two measures were plotted on a scatter graph. The graph suggested that it would be feasible to identify five concordant and discordant groups within the NSHD dataset at age 60-64.

Deciles were viewed as a clear way to divide participants and define the relevant groups, as they provided logical cut-points across the spectrum for the two continuous measures. Sex-specific deciles were used for the performance summary measure, because values are relative to the sample population, and adjustments have been made to account for the known sex difference in grip strength in the production of the summary measure.

As expected, given the parameters used to define the concordant and discordant groups, overestimators aligned with the self-reported physical capability levels of the high concordant group and the performance-based physical capability levels of the low concordant group. In contrast, underestimators were aligned with the self-reported physical capability levels of the low concordant groups and the performance-based physical capability levels of the high concordant group. However, both discordant groups also overlapped with the middle concordant in terms of their physical capability levels. This observation is more apparent for the underestimator group, and may be a consequence of the selection criteria used to define each group.

The power calculations presented in this chapter show that a difference of 15% or more in the proportion of a given risk factor, between the concordant and one of the discordant groups, can be detected with more than 90% power. The power calculations also demonstrate that 90% power is reached with a mean difference between the concordant and one of the discordant groups of more than 1.5 or 2.0 units, depending on the standard deviation.

Further discussion on the implications of statistical power within this thesis can be found in Section 8.4, together with a more detailed discussion about the implications of how the concordant and discordant groups have been defined, including a sensitivity analysis to explore alternative groupings.

6. Discordance between self-reported and performance-based measures of physical capability

Having established evidence of discordance between self-reported and performance-based measures of physical capability within the NSHD dataset at age 60-64, the aim of this chapter is to identify factors associated with discordance. This chapter starts with a review of the existing evidence base by focusing on published studies which explore the association between factors of interest and discordance. The findings of this review will then been used to identify suitable factors for analysis in the second part of this chapter and the following chapter.

6.1 Discordance literature review

To summarise the existing evidence base on discordance, the review in this chapter included a subset of 16 papers from the concordance review in Chapter 5, where studies aimed to investigate factors associated with discordance (see Appendix 5). A discussion of how discordance was defined and assessed in the studies reviewed will follow in Section 6.1.1.

As previously discussed (see Chapter 5), the characteristics of a study population can influence the level of concordance and discordance observed between self-reported and performance-based measures of physical capability. As the papers included in this review are a subset of the review in Chapter 5, many of the characteristics are similar. Of the 16 papers included in this discordance review, 12 focused on community-based samples and four on disease-specific samples. The majority of papers (N=12) have study populations with an average age over 70, but as many of these samples were heterogeneous in age, discordance could be explored across old age. Of the studies reviewed, the largest sample size was approximately 2,800 participants, however 63% of studies had between 100-800 participants (N=10); only the disease-specific papers had study populations of less than 100 people. Although relatively small compared to some studies, study populations of this size were still large enough to quantify the level of discordance within the samples, and they produced statistically significant results when investigating factors associated with discordance. Three of the papers were sex-specific, focusing on female populations, which prevented any analysis of sex differences in these studies. The majority of the studies were based in the United States (N=10), with the remaining papers focused on European populations, which provides a Western cultural perspective on discordance. However, none of the studies included samples from the United Kingdom.

It is important to note that all of the papers identified for this chapter's review were cross-sectional in design. It appears that there is a gap in the literature, with further research required to investigate the longitudinal factors associated with discordance. The use of longitudinal factors would enable researchers to consider discordance from a life course perspective, and potentially identify lifetime determinants of discordance.

6.1.1 Magnitude and direction of discordance

To explore the level and direction of discordance observed within their study populations, most of the papers categorised the self-reported and performance-based measures of physical capability. Cross-tabulation was then used to either calculate the percentage disagreement between the two measures or calculate the proportion of false positives (underestimation of physical capability) and false negatives (overestimation of physical capability) produced when the performance-based measures were taken as the "true value". One study did not take either of these approaches; instead the authors elected to subtract a derived continuous self-reported score from the performance score, using a threshold of a 0.25 unit difference in functional score to represent discordance (125).

Each of these methods produced a similarly diverse range of discordance estimates, with the level of discordance observed in the 16 studies ranging from 3.1-74.3%. Most studies reported a level of discordance between 15-50% (see Appendix 5). One paper developed the analysis and explored the magnitude rather than the direction of discordance by quantifying the severity of the observed discordance (121). Participants were considered to have "slight" discordance if their self-reported and performance-based measures differed by one categorical level of function (55%), or "substantial" discordance if the difference was across two or more levels of function (19.3%). When investigating factors associated with discordance, the authors split the analysis according to the magnitude of discordance and observed a dose-response relationship.

When researching discordance, it is important to distinguish between the two directions in which discordance can operate (see Figure 1.1). Firstly, participants may report that they have no difficulties, suggesting a high level of physical capability, in contrast to their performance scores which suggest a lower level of capability (coloured green in Figure 1.1). To help facilitate the discussion, these individuals shall be referred to as "overestimators". The second type of discordance operates in the opposite direction, with participants reporting lower levels of physical capability than would be expected given their performance values (coloured orange in Figure 1.1). This group shall be referred to as "underestimators". Not all of the papers reviewed examined the direction of discordance, but the eleven that did found conflicting results.

Four of the papers reported that their study populations had a tendency to overestimate their capability (71;121;123;125), whilst four studies reported a tendency to underestimate (122;131;132;137); three studies reported no distinct direction of discordance (45;72;138). When examining the proportion of individuals categorised into the two discordant groups, the difference in the direction of discordance was not always marginal. Two of the studies presented data to suggest that over 90% of discordant cases were overestimators (71;121), and one of the studies reported that at least 89% of discordant cases were underestimators (132). However, few studies formally tested the relative size of the two discordant groups, and many were less clear cut than the examples provided (see Appendix 5). The two papers that did explicitly test the direction of discordant cases calculated the percentage bias and used McNemar's test to estimate the significance of the difference in direction. One study found no difference in the proportion of discordant cases in either direction (72), and the other found discordant cases were more likely to be underestimators than overestimators (132). Where it was possible to extract the functional mobility domain (tasks conceptually closest to physical capability), there was a tendency for study participants to underestimate capability (122;132;137). The other papers covered a mixture of tasks across all domains of functional ability, and included some more socially orientated tasks akin to disability, which may have influenced the results.

When considering the direction of discordance reported, there is a need to take into account study design limitations. Ceiling and floor effects of the self-reported and performance-based measures used in a study can influence the level and direction of discordance observed (45;121;139). For example, in many studies the majority of individuals will report “no difficulty” with the self-reported measure, and these individuals can therefore only be classified as concordant or overestimators of their physical performance.

6.1.2 Factors associated with discordance

The aim of this section of the review is to summarise the key findings from the 16 studies, focusing on the factors that potentially influence discordance which were included in more than one paper, or those of particular relevance to this thesis. A wide variety of potential factors were examined within the studies under review, which have been broadly categorised as: demographic factors, health-related factors, factors associated with cognitive and physical functioning, psychological factors and socioeconomic factors (see Table 6.1). A note of caution is required, as the factors are often interdependent (40), however they have been separated to facilitate discussion.

Table 6.1: Key factors highlighted in discordance literature review

Potential risk factors	**Number of papers**	
	Investigating factor	where finding was statistically significant
Demographic factors		
Age	12	6
Sex	10	6
Health-related factors		
Self-reported/perceived health	4	4
Chronic conditions/comorbidities	3	3
Pain	2	1
Cognitive and physical functioning		
Cognitive function	7	4
Physical capability decline	1	1
Psychological factors		
Depression and anxiety	6	3
Personality	3	2
Socioeconomic factors		
Education	6	3
Employment status	1	0
Social support	4	1

To explore the association of specific factors with discordance, ten of the studies used multiple regression analyses. There were some noted differences between the studies using regression analyses, with some electing to use discordance as the dependent variable (N=6), whilst other studies chose the self-reported measure as the dependent variable and added the performance-based measure as one of the independent variables (N=4). The studies with smaller sample sizes (N=6) chose either to separate the analysis by the factor of interest, or to characterise the two discordant groups and test the significance of the difference between each group, instead of using regression analyses. The results produced by each of these methods did not differ substantially.

Demographic factors

Age was the most commonly investigated factor, with 12 of the studies electing to explore the association of age with discordance. The general trend observed in half of these papers (N=6) was that discordance increased with age (40;45;72;116;125;132). For example, in one study of community-dwelling participants in the United States, the level of discordance across five physical capability tasks was reported to be between 3.2-4.7% for participants under 75 years, compared to 4.3-10.7% for those over 75, with the highest level of discordance observed for the mobility tasks (132). However, five of the 12 papers found no significant association between age and discordance (112;121;123;129;138). It has been suggested that older individuals may underestimate their capability, particularly for mobility tasks, to reflect the cautious approach they take when completing the tasks of daily living (132). On the other hand, it is also plausible that older people may overestimate their ability, given the lower expectations of physical capability associated with age (129). However, only two studies found significant results when exploring the association between age and direction of discordance, and they found conflicting results (72;116).

Sex was the second most commonly studied factor examined (10 papers). It was observed that women were more likely to be discordant than men (40;45). When exploring the direction of this discordance the findings in the literature appear to be consistent, with women tending to be more likely to underestimate their capability and men more likely to overestimate (116;125;129;131). However, it is important to note that both men and women have been observed to overestimate as well as underestimate their physical capability (131), which may explain why four of the studies (72;112;123;138) that explored the impact of sex found no significant results.

Health-related factors

Self-perception of health was one of the main health factors examined within the reviewed papers. Consistent results were produced across the papers (N=4), regardless of whether self-rated health questions were used or questions regarding satisfaction with health and physical function. In general, those with poor health (or dissatisfied) were more likely to underestimate their capability (72;129;132). Perceived health was one of the strongest factors associated with discordance for two of the studies reviewed (72;129); for example in one study those with low self-perceived health had 23 times higher odds of being discordant for a walking task than those with high self-perceived health (72). It should be acknowledged that an individual’s self-reported assessment of their health will be influenced by the potential negative impact of various chronic conditions. However, the association between self-reported health and discordance remained even after adjustments were made to account for the number of chronic conditions an individual experienced (72).

Independent of self-reported health, the presence of chronic conditions is a potentially important factor for discordance, and was explored by three of the papers reviewed. It was reported in two of the studies that individuals with two or more chronic conditions were more likely to be discordant than those with fewer than two conditions (71;121). The direction of the discordance is not calculated in these studies, but those with two or more chronic conditions were found to have an increased risk (RR = 1.52) of being substantially discordant (121). Specific chronic conditions experienced by an individual may make certain tasks more challenging to accurately assess than others. For example, women with knee osteoarthritis were found to be most discordant on tasks within the functional mobility domain, which were most likely to be affected by their condition, as they struggled to accurately report their own capability (122). Of the specific chronic conditions explored by the studies, stroke appears to be an important factor to consider, with one study (132) reporting that prior stroke was one of the factors most strongly associated with discordance. A paper that focused specifically on cerebrovascular events (121) found that those who had experienced a stroke were 20% more likely to be slightly discordant and 130% more likely to be substantially discordant compared to those who had not experienced a stroke.

The experience of pain is one of the physical symptoms of chronic conditions that may be associated with discordance, and this factor was included in two of the papers reviewed. One found no significant association between the experience of pain and discordance (125); however a second paper found that when pain was included in a model with only four other factors (including the performance-based measure), 85% of the variance in the self-reported measure was explained by these five factors (129).

Physical and cognitive functioning

Perhaps one of the most interesting factors associated with discordance, given the dynamic nature of an individual's experience of physical capability, is the influence of recent functional decline. Although only explored by one study (129), this factor may provide useful information. This study of people aged 65 to 94 provided evidence of a strong association between retrospectively reported recent functional decline and discordance, with participants reporting lower levels of capability than expected given their performance if they reported recently experiencing a decline. This observation aligns with adaptation theory, which suggests an individual's perception of their capability is based on a comparison with previous levels of experienced capability. For an individual who has recently declined in physical capability, this could result in an exaggerated report of their limitations (129).

Within the literature it has been proposed that the order in which participants conduct the performance and self-reported assessments of physical capability may alter the level of observed discordance (129;138). It is postulated that those who conduct the performance test first would have a more accurate assessment of their capability, as they have just demonstrated the specific task in question (138). One interesting finding linked to functional decline and the association with discordance is that the effect of test order was negligible on those who had experienced a recent decline, whereas those who had not experienced a decline were considerably less likely to be discordant when the performance test was conducted first (129).

Cognitive function is also an important factor to consider when exploring the discordance between self-reported and performance-based measures of physical capability. It is plausible that individuals who experience cognitive impairment are less able to accurately assess their own capability. The results on cognitive function obtained from the papers reviewed were mixed. In general, lower levels of cognitive function appear to be associated with discordance, but there were conflicting results regarding the direction of this discordance (45;112;121;123;132). One study (45) found cognitive impairment was significantly associated with both overestimation and underestimation of physical capability. Several of the studies that investigate the association between cognitive function and discordance excluded those with severe cognitive impairments (40;45;112;121), which may weaken the association and explain some of the null findings.

<u>Psychological factors</u>

Depression was investigated as a factor of interest by six of the papers reviewed (40;112;121;125;129;137) due to the acknowledged influence it has on self-reported measures, with individuals suffering from depression more likely to have a negative perception of their physical capability. However, it is important to note that depression may also influence performance-based measures of physical capability, with depressed individuals less motivated to achieve maximal performance during their assessment (2;140). Most of the papers found no significant association between depression and discordance. One study noted that, despite being non-significant as a factor by itself, it was important to include depression in the final model, as it substantially altered the association of other variables with discordance (121). The few papers that did report significant associations observed that those who were depressed were more likely to underestimate their capability (40;112;137). However, this observation was not reported in the functional mobility domain, only in domains focused on Instrumental Activities of Daily Living (IADLs) or mixed studies. Three of the papers also looked at symptoms of anxiety as a factor of interest (40;112;125). Only one paper found significant results, with more anxious individuals most likely to underestimate their capability (112).

Personality was explored as a potential explanation of discordance in three of the papers reviewed, with personality accounting for nearly 6% of the variance in the self-reported measure of physical capability, once the performance-based measure had been accounted for (40;112). When examining the individual personality traits, there was no evidence that extroversion was associated with discordance, but there was some evidence of an association between neuroticism and discordance (40;112). Although one study found no association, a second study exclusively focusing on the association between neuroticism and discordance found that neuroticism directly accounted for 2% of the variance in the self-reported measure of physical capability once the performance-based measure had been accounted for, and a further 2% of the variance could be explained as an interaction term (139). The authors suggest that those with higher levels of neuroticism are more susceptible to functional limitations, so a small decline in physical capability will be magnified in their self-assessment.

<u>Socioeconomic factors</u>

Socioeconomic position was explored as a factor of interest in six of the studies reviewed. In the six studies that used education as their indicator of socioeconomic position, half found no association with discordance (45;112;123). The three studies which reported significant results provide conflicting evidence, with one study suggesting that more educated people are less likely to be discordant (40), whilst the two other studies presented evidence to indicate that more educated people are more likely to be discordant (72;121). In the one study where employment status was used as an indicator of socioeconomic position, no significant association was found with discordance (123).

Several of the papers have attempted to explore the influence of social support on discordance, with measures of social activity (125), marital status, social networks (121) and living arrangements (123;132). However, none of these studies found any significant associations between the factors of interest and the observed discordance, with one exception. In a study of patients with Parkinson's disease, when exploring the direction of discordance, it was observed that of those who lived with their family or spouse, 92% overestimated their physical capability, whilst only 8% underestimated their physical capability ($P<0.05$) (123). In comparison, amongst those who lived alone, 57% overestimated their physical capability and 43% underestimated their physical capability.

6.1.3 Critical review

Within the 16 papers reviewed in this chapter, the level of discordance varied between the studies and between the tasks of physical capability assessed. It was noted that the greatest level of discordance was observed for tasks relating to mobility. Personal care tasks and other Activities of Daily Living (ADLs) are routinely practised by individuals as part of their everyday lives (137), so it is understandable that people are more aware of their capability in these tasks, and are consequently more able to accurately assess their physical capability in relation to these tasks. In contrast, some people may not perform the mobility tasks as often, if at all. For example, people who live in bungalows will have fewer opportunities to climb stairs. In these situations the self-reported assessment of physical capability becomes hypothetical, whilst the performance assessment closely reflects reality, and this may explain the greater level of discordance observed for mobility tasks.

The attributes of the self-reported and performance-based measures of physical capability (discussed in Section 1.2) may also contribute to the level of discordance observed. For example, it may be necessary to consider the temporal restraints of each measure; performance-based measures record the level of physical capability of an individual at a specific point in time, whereas individuals are known to evaluate their physical capability over a period of time when self-reporting. Short-term fluctuations in physical capability, caused by illness, fatigue or pain, may contribute to discordance as a result of these attributes (45). The temporary variation in physical capability level may make the accurate assessment of self-reported physical capability more challenging when reflecting on experience over a period of time, but equally the level recorded during the performance test may not reflect the "usual" level of physical capability experienced by that individual.

An individual may be classified as experiencing difficulty with a specific task during a performance test if they take too long to complete the task. However, the individual themselves may not perceive their speed as an indication of difficulty (71). It is plausible that the level of discordance observed may be due to the subjective nature of the self-reported measure (45;131). People have different interpretations of what the word "difficulty" means to them, based on their own experiences and expectations (72). It has been shown within the wider literature that different subgroups of the population have different thresholds for reporting difficulty and inability. For example, as level of income increases, individuals report the experience of difficulty, or inability to complete a task, at higher levels of performance-based physical capability compared to those with lower income levels (141). A similar gradient is also observed with age, although, when adjusted for baseline physical capability, there was no difference in the threshold for reported difficulty between men and women. These subgroup differences may explain why some of the factors reviewed were associated with discordance.

When focusing on the direction of discordance, it was observed that amongst studies which assessed direction, many of the factors identified within the review were associated with underestimation rather than overestimation of physical capability. However, it is important to identify the factors associated with overestimation, as these individuals may have developed more resilience. Self-reported assessments of physical capability may incorporate coping strategies, which enable individuals to report relatively high levels of physical capability despite their poor performance scores (40;45). It has therefore been suggested that one improvement to consider for future research is the inclusion of questions asking about how individuals have changed the way they carry out certain tasks and whether they have changed the frequency with which they complete the tasks (71). These questions would enable researchers to explore the extent to which adaptations contribute to the overestimation of physical capability. As a note of caution, if individuals are aware that they have altered the way they complete tasks, it is possible they will report lower levels of physical capability (137). Modifications made in everyday life will not impact on an individual's performance test, so it is possible that they could underestimate their physical capability if the adaptations they employ are incorporated into their assessment of physical capability. In one of the studies, which investigated the difference between an individual's ability (can do) and their habits (does do), it was noted that participants were less likely to overestimate their physical capability if the questions were based on habits (71). The authors suggest that the lower level of discordance observed for habits may be due to the inclusion of adaptive strategies in an individual's assessment of physical capability.

Only a few of the papers included in this review explicitly justified the selection of factors for their analysis *a priori*. Where possible factors should be selected in line with the underlying theoretical construct used to frame the research question of interest.

6.1.4 Conclusion of review

This review has highlighted a wide variety of factors which have been investigated in relation to discordance. However, the results have been inconclusive for many of the factors investigated within the literature, both in terms of their associations with discordance and the direction of discordance. Only sex and self-reported health demonstrated a consistent association; women and those reporting poor health were more likely to underestimate their levels of physical capability. Although the evidence base developed from the 16 studies reviewed may be limited, it provides a foundation from which to select the factors of interest for this thesis. These factors can be grouped into two categories: firstly, socio-demographic and behavioural risk factors, which will be the focus of the remainder of this chapter, and secondly markers of health status, which will be explored in Chapter 7.

Therefore, the aim for the remainder of this chapter is to explore the magnitude and direction of discordance observed when investigating the association of socio-demographic and behavioural risk factors with discordance. Factors included in the analysis for this chapter were specifically chosen based on the results of the literature review, with four socio-demographic factors and three behavioural risk factors selected. By characterising individuals in each of the discordant groups in terms of their socio-demographic and behavioural risk factors, it may be possible to identify discordant individuals when only one type of physical capability measure is used, which could have a practical application in a community setting.

The four socio-demographic factors selected were sex, education, occupational class and marital status. Sex was chosen as one of the key factors of interest due to the strong and consistent associations reported in the literature. It seems plausible that sex may also confound the associations of several other factors with discordance. Education and occupational class were both selected as indicators of socioeconomic position. As noted in Section 6.1.2, previous studies found conflicting results for education, and comparisons between different indicators of socioeconomic position have not been made before. Marital status was included as a marker of social support, because social support was one of the few factors within the literature that appeared to be associated with overestimation of physical capability, and this may represent resilience. Although age was the most commonly explored factor within the literature review, a decision was made not to include age within the main analysis of this thesis due to the narrow age band of participants. A sensitivity analysis exploring the effect of age was conducted and a more detailed discussion of age will follow in Section 6.4.2.

The association between behavioural risk factors and discordance has not been explored in much detail in previous studies, with only two studies including such factors in their analyses (71;138). To address this gap in the literature and to help characterise discordant individuals, three key behavioural risk factors, previously shown to be associated with physical capability, were selected for analysis: smoking (142;143), physical activity (102;144) and body mass index (BMI) (145;146).

6.2 Methods

6.2.1 Variables for socio-demographic and behavioural risk factor analysis

The following section details how the socio-demographic and behavioural risk factors used in this chapter were ascertained.

Two indicators of socioeconomic position were selected for inclusion in this analysis: education and occupational class. The highest educational level attained at age 26 was recorded and categorised into four groups: degree or higher; A level or equivalent; up to O level (includes O Level, CSE, clerical course or equivalent); and none. Occupational class was recorded at age 53 (or the most recent measure available in adulthood (N=98)) and categorised using the Registrar General's Social Classification into three groups: high (I or II), medium (III manual or non-manual) and low (IV or V). The final socio-demographic factor included in the analysis was marital status, which was recorded at age 60-64 when participants reported their current marital status. Participants were categorised into one of the four categories: single (never married); married; widowed; and separated or divorced. The data were checked against the marital status information obtained in 1999, and inconsistencies were amended (N=2).

Smoking status was recorded throughout adulthood, and smoking history up to age 60-64 years (or up to age 53 if unavailable (N=142)) was categorised as: current; former; or never smoked. Physical activity was recorded at age 60-64 years, at which time participants were asked to report whether they had engaged in any sports, vigorous leisure activities or exercise in their spare time in the past four weeks. Responses were categorised as: inactive (reported no participation); moderately active (participated in relevant activities 1-4 times); or most active (participated in relevant activities five or more times). The final behavioural factor included in the analysis was BMI, which was calculated using the height and weight measurements objectively collected by a nurse at age 60-64. For this analysis, BMI was treated as both a continuous measure and also categorised using the standard WHO cut points (<18.5 kg/m^2 underweight, 18.5-24.9 kg/m^2 normal range, 25.0-29.9 kg/m^2 overweight, >30.0 kg/m^2 obese) (147;148). Underweight individuals (N=10) were grouped with those of normal weight due to the very low prevalence of this characteristic within the sample.

6.2.2 Analysis of socio-demographic and behavioural risk factors with discordance

The analytical approach used in this chapter to explore the association of socio-demographic and behavioural risk factors with discordance was separated into three stages. Firstly, a descriptive analysis was conducted to identify the direction of the association between each factor and discordance. The proportion of participants within each subcategory of the socio-demographic and behavioural risk factors was calculated across the discordant and concordant groups. The three concordant groups were combined into one group for this analysis (and all future analyses unless specifically stated), for reasons of power, to focus the interpretation of results on discordance, and so models were not driven by the difference between the three concordant groups. To ascertain whether the two discordant groups were distinct from the concordant group, chi-squared tests were used to assess the difference between the three groups. For continuous factors of interest, mean and standard deviations were calculated for the concordant and two discordant groups. In a similar process to the categorical variables, ANOVA analyses were used to test the difference in mean values across the three groups.

As an extension to the descriptive analysis, an analysis was conducted to ascertain if there was a gradient in the socio-demographic and behavioural risk factors across the three concordant groups. A non-parametric test for trend was used to formally test the gradient in both the categorical and continuous variables. The characteristics of the two discordant groups were then compared to the three concordant groups.

The second stage of the analysis used sex-adjusted multinomial logistic regression models to quantify the magnitude of the association between each factor and discordance, accounting for sex. Each of the socio-demographic and behavioural risk factors were placed into sex-adjusted models run on maximum N, with discordance as the dependent variable and the combined concordant group used as the reference category. For each of these models, sex interactions and the association with discordance were formally tested.

The final stage of the analysis used a mutually-adjusted multinomial logistic regression model to explore potential confounding and the interrelationship between factors within the analysis. A criteria of P≤0.1 from the test of association, run on the sex-adjusted models as part of the second stage of the analysis, was used to select factors for inclusion into the mutually-adjusted model. Before the mutually-adjusted model was run, each of the sex-adjusted models were repeated with a restricted sample that had complete data for all relevant covariates.

A sensitivity analysis was conducted to test whether participants excluded from the restricted sample differed from those included in the mutually-adjusted model in terms of their socio-demographic and behavioural characteristics.

6.3 Results

6.3.1 Descriptive analysis of socio-demographic and behavioural risk factors

The results of the descriptive analysis to ascertain whether the two discordant groups were characteristically distinct from the combined concordant group in terms of their socio-demographic and behavioural risk factors are shown in Table 6.2.

<u>Socio-demographic factors</u>

There was a noticeable difference in the sex distribution of the two discordant and concordant groups (see Table 6.2). Amongst participants who reported low levels of physical capability yet performed to high standards (underestimators), there was a male: female ratio of almost 20:80. In contrast, amongst participants who reported high levels of physical capability despite their poor performance (overestimators), there was a higher proportion of men, with a ratio of nearly 60:40. When the two discordant groups were compared to the concordant group, who had a roughly equal proportion of men and women, a statistically significant difference was observed (P<0.001).

In relation to education, no distinct differences were observed between the characteristics of the concordant and two discordant groups (P=0.1). However, the descriptive results observed for occupational class indicate that compared to the concordant group underestimators were more likely to have a high occupational class (54% vs. 49%), whilst overestimators were more likely to have a low occupational class (17% vs. 12%), and vice versa (see Table 6.2). Both discordant groups had around 41% of participants in a middle occupational class, compared to 39% in the combined concordant group, suggesting that this group remained consistent across the groups. When tested, the occupational class characteristics of the combined concordant and two discordant groups were statistically different (P=0.04).

No differences were observed for marital status between the two discordant and concordant groups (P=0.2).

Behavioural risk factors

In terms of their smoking history, the characteristics of the two discordant groups were different compared to the combined concordant group (P=0.001). Underestimators were more likely to be ex-smokers compared to the combined concordant group (65% vs. 56%), whilst overestimators were more likely to be current smokers (17% vs. 12%). The pattern was also observed operating in the opposite direction. Strong evidence was found that the differences observed in the smoking characteristics of the concordant and two discordant groups were not due to chance (P=0.001).

With regard to physical activity, overestimators were more likely to never participate in physical activity compared to the concordant group (73% vs. 62%) and underestimators were more likely to participate in physical activity, either sometimes or frequently (17% & 24% vs. 14% & 24%). However, the differences between the groups were only marginally significant (P=0.08).

In relation to the BMI, when used as a categorical or continuous variable, there were no characteristic differences observed between the two discordant and concordant groups.

Characteristics of the three concordant groups

When the analysis was extended to explore the characteristics of the three concordant groups (see Appendix 6), evidence of a gradient was observed across the three concordant groups for sex, education, occupational class, smoking history, physical activity and BMI (each gradient had $P \leq 0.02$). Participants in the high capability concordant group were more likely to be male, have a degree, have a high occupational class, have never smoked, frequently participate in physical activity and be normal weight. In comparison, participants in the low capability concordant group were more likely to be female, have no education, have a low occupational class, be a current smoker, never participate in physical activity and be obese. In general, when the two discordant groups were mapped onto this gradient, the characteristics of the overestimators were more closely aligned with the low concordant group and those of the underestimators were aligned with the high concordant group. The exception to this trend was sex and the ex-smoker category of the smoking history variable, which operated in the opposite direction.

Table 6.2: Socio-demographic and behavioural characteristics of the combined concordant and two discordant groups

	Proportion of sample (%)†				P Value‡
	Total Sample	Combined concordant group	Discordant groups		
			Under-estimators	Over-estimators	
Maximum N*	(N=1981)	(N=1708)	(N=135)	(N=138)	
Socio-demographic factors					
Sex	(N=1981)				
Males	47.7	48.5	21.5	63.0	
Females	52.3	51.5	78.5	37.0	<0.001
Education	(N=1873)				
None	28.9	29.1	19.5	36.0	
Up to O-Level	29.1	29.3	31.3	24.8	
A-Level or equiv	29.9	29.9	33.6	25.6	
Degree or higher	12.1	11.7	15.6	13.6	0.1
Occupational Class	(N=1970)				
Low	12.2	12.4	5.2	16.8	
Medium	39.1	38.8	41.0	41.6	
High	48.6	48.8	53.7	41.6	0.04
Marital status					
Single	3.8	3.6	2.4	7.7	
Married	79.4	79.3	79.0	81.2	
Widowed	5.6	5.7	6.5	3.4	
Separated	11.2	11.4	12.1	7.7	0.2
Behavioural risk factors					
Smoking History	(N=1956)				
Never	32.2	31.8	32.3	37.8	
Ex-smoker	56.1	56.3	64.7	45.2	
Current smoker	11.7	11.9	3.0	17.0	0.001
Physical activity	(N=1932)				
Never	62.7	62.2	59.1	72.9	
Sometimes	14.1	14.0	16.7	12.8	
Frequently	23.2	23.8	24.2	14.3	0.08
BMI §	(N=1974)				
Normal	29.3	29.2	33.3	25.7	
Overweight	41.7	41.8	38.5	44.1	
Obese	29.0	29.0	28.2	30.2	0.7
BMI (continuous)†	(N=1974)				
Mean (kg/m^2)	27.9	27.9	27.6	28.1	
SD	4.9	4.9	4.5	4.6	0.7

* Value of N varies for each factor due to missing data
† For continuous variables mean and standard deviation (SD) values provided instead of proportions and details of units provided
‡ P value from χ^2 test if variables categorical and from ANOVA if continuous variables
§ BMI categorised using standard WHO cut-points. Underweight individuals (N=10) were grouped with those of normal weight.

6.3.2 Sex-adjusted socio-demographic and behavioural risk factor models of discordance

The results from the multinomial logistic regression models used to explore the association of socio-demographic and behavioural risk factors with discordance are shown in Table 6.3. When formally tested, there was no evidence of sex interaction (P>0.07).

In general, where an association was observed between one of the factors and discordance in the descriptive analysis, it was noted that the association was stronger in the sex-adjusted models compared to the previous descriptive analysis (Section 6.3.1), suggesting that the relationship may previously have been partially confounded by sex.

<u>Socio-demographic factors</u>

When the association between sex and discordance was explored, females were 3.45 (95% CI: 2.26 – 5.25) times more likely to underestimate their physical capability compared to participants in the concordant group. In contrast, females were almost half (0.55 (95% CI: 0.39 – 0.79)) as likely to overestimate their physical capability as males, compared to the concordant group. The results obtained from the regression models provide very strong evidence for the association between sex and both directions of discordance (P<0.001).

With regard to education, participants with A-levels or a degree were more likely than those with no qualification to underestimate their physical capability levels, compared to the concordant group (RRR=1.67 (95% CI: 1.00 – 2.80) and RRR=2.74 (95% CI: 1.46 - 5.13) respectively). While the association between education and underestimation of physical capability was strong (P=0.01), there was no evidence of an association between education and overestimation of physical capability (P=0.4).

The results observed for occupational class suggest that (compared to the concordant group) participants with medium or higher occupational class were more likely to underestimate their physical capability compared to those with a lower occupational class (RRR=2.83 (95% CI: 1.26 – 6.32) and RRR=3.61 (95% CI: 1.63 – 8.01) respectively). Equally, participants in the high occupational class group were half as likely (0.51 (95%

CI: 0.31 – 0.86)) (compared to the concordant group) to overestimate their physical capability compared to those in the lower occupational class group. When tested, the association between occupational class and discordance, in either direction, was strong (P<0.03).

No association was observed between marital status and discordance (P>0.1).

Behavioural risk factors

In terms of their smoking history, participants categorised as current smokers were less likely to underestimate their physical capability (0.26 (95% CI: 0.09 – 0.75)) (compared to the concordant group) compared to those who never smoked. In contrast, participants categorised as ex-smokers were less likely to overestimate their physical capability (0.63 (95% CI: 0.42 - 0.93) (compared to the concordant group) compared to those who never smoked. Strong evidence was observed for the association between smoking history and both directions of discordance (P<0.02).

When the association between physical activity and discordance was examined, individuals who frequently partook in physical activity were less likely (compared to the concordant group) to overestimate their physical capability (0.51 (95% CI: 0.31 – 0.85)) compared to those who did not partake in physical activity. The strength of the association between physical activity and overestimation of physical capability was strong (P=0.03), however no association was observed between physical activity and underestimation of physical capability (P=0.7).

No association was observed between BMI, when used as a categorical or continuous factor, and discordance (P>0.6).

Table 6.3: Sex-adjusted models for the association between discordance and socio-demographic factors or health risk behaviours (run on max. N)

Factors of interest ‡	RRR (95% CI) †	
	Underestimators	Overestimators
Socio-demographic factors		
Sex		
Male	1.00	1.00
Female	3.45 (2.26 - 5.25)**	0.55 (0.39 - 0.79)**
	P<0.001	P=0.001
Education		
None	1.00	1.00
Up to O-Level	1.36 (0.81 – 2.29)	0.75 (0.47 – 1.21)
A-Level or equiv	1.67 (1.00 – 2.80)*	0.69 (0.43 – 1.11)
Degree or higher	2.74 (1.46 – 5.13)**	0.83 (0.46 – 1.50)
	P=0.01	P=0.4
Occupational Class		
Low	1.00	1.00
Medium	2.83 (1.26 – 6.32)*	0.73 (0.44 – 1.21)
High	3.61 (1.63 – 8.01)**	0.51 (0.31 – 0.86)*
	P=0.006	P=0.03
Marital Status		
Married	1.00	1.00
Single	0.69 (0.21 – 2.26)	2.10 (1.00 – 4.39)*
Widowed	0.86 (0.40 – 1.83)	0.69 (0.25 – 1.95)
Separated	0.97 (0.55 – 1.72)	0.69 (0.34 – 1.39)
	P=0.9	P=0.1
Health behaviours		
Smoking History		
Never	1.00	1.00
Ex-smoker	1.27 (0.87 – 1.88)	0.63 (0.42 – 0.93)*
Current smoker	0.26 (0.09-0.75)*	1.16 (0.69 – 1.95)
	P=0.007	P=0.02
Physical activity		
Never	1.00	1.00
Sometimes	1.23 (0.75 – 2.02)	0.79 (0.46 – 1.35)
Frequently	1.07 (0.69 – 1.64)	0.51 (0.31 – 0.85)**
	P=0.7	P=0.03
BMI§		
Normal	1.00	1.00
Overweight	0.91 (0.60 – 1.38)	1.17 (0.75 – 1.83)
Obese	0.87 (0.55 – 1.37)	1.21 (0.75 – 1.94)
	P=0.7	P=0.7
BMI (continuous)		
Per kg/m^2 increase	0.99 (0.96 – 1.03)	1.01 (0.97 – 1.05)
	P=0.6	P=0.6

†Relative Risk Ratio, Reference group = concordant group

‡Sex interaction formally tested but no evidence found. P values ranged from 0.07-0.56

§ BMI categorised using standard WHO cut-points. Underweight individuals were grouped with those of normal weight.

* Significant at P=0.05 level, ** significant at P=0.01 level

Blue values indicate factors with P≤0.1 and therefore will be taken forward to the mutually-adjusted model

6.3.3 Mutually-adjusted socio-demographic and behavioural risk factor model of discordance

In addition to sex, four factors (education, occupational class, smoking history and physical activity) were carried forward to the mutually-adjusted model on the basis of a selection criteria P≤0.1 for the test of association. The initial results in Table 6.4 show the sex-adjusted models for the selected factors run on a restricted sample (N=1799). All results reflected the direction of association observed in the previous analysis when models were run on maximum Ns (Section 6.3.2). In terms of occupational class, smoking and physical activity, the association with discordance weakened in the restricted sample, particularly for the association of physical activity with overestimation, which was found to be non-significant (P=0.2). However, with regards to sex and education, the association with discordance strengthened slightly in the models run on the restricted sample.

When a sensitivity analysis was conducted in terms of their socio-demographic and behavioural characteristics, no statistically significant differences were observed between participants included in the restricted sample and those excluded for missing data on covariates (see Appendix 6), with the exception of education, where those excluded were less likely to have A-Levels or a degree.

In the mutually-adjusted model (compared to the concordant group) women were 4.27 (95% CI: 2.67 – 6.83) times more likely than men to underestimate their physical capability, and current smokers were 0.34 (95% CI: 0.12 – 0.97) times less likely to underestimate their physical capability compared to participants who never smoked. In contrast (compared to the concordant group), females were 0.46 (95% CI: 0.30 – 0.69) times less likely than men to overestimate their physical capability, and ex-smokers were 0.62 (95% CI: 0.41 – 0.94) times less likely to overestimate their physical capability compared to those who never smoked.

The association of occupational class and education with discordance was strongly attenuated in the mutually-adjusted model, although (compared to the concordant group) participants with a degree were still 2.31 (95% CI: 1.12 – 4.76) times more likely to underestimate their physical capability compared to those with no education, as were participants in the middle or the higher occupational class group (2.37 (95% CI: 1.04 – 5.37) and 2.32 (95% CI: 0.99 – 5.43) respectively). The association between physical activity and discordance was fully attenuated.

Table 6.4: Comparison between sex-adjusted and fully adjusted models for socio-demographic and health risk behaviours (N=1,799)

Factors of interest	Sex-adjusted model on restricted N RRR (95% CI) †		Mutually-adjusted model RRR (95% CI) †	
Socio-demographic factors	Underestimators	Overestimators	Underestimators	Overestimators
Sex				
Male	1.00	1.00	1.00	1.00
Female	3.66 (2.33 – 5.74)**	0.49 (0.33 – 0.72)**	4.27 (2.67 – 6.83)**	0.46 (0.30 – 0.69)**
	P<0.001	P<0.001	P<0.001	P=0.002
Education				
None	1.00	1.00	1.00	1.00
Up to O-Level	1.46 (0.85 – 2.52)	0.69 (0.42 – 1.15)	1.25 (0.71 – 2.17)	0.77 (0.46 – 1.30)
A-Level or equiv	1.79 (1.05 – 3.07)*	0.72 (0.45 – 1.16)	1.45 (0.80 – 2.62)	0.87 (0.51 – 1.50)
Degree or higher	3.03 (1.59 – 5.79)**	0.80 (0.44 – 1.47)	2.31 (1.12 – 4.76)*	1.09 (0.54 – 2.22)
	P=0.008	P=0.4	P=0.1	P=0.7
Occupational Class				
Low	1.00	1.00	1.00	1.00
Medium	2.54 (1.13 – 5.70)*	0.78 (0.44 – 1.39)	2.37 (1.04 – 5.37)*	0.80 (0.45 – 1.43)
High	3.21 (1.44 – 7.16)**	0.58 (0.33 – 1.03)^{m}	2.32 (0.99 – 5.43)^{m}	0.63 (0.33 – 1.21)
	P=0.01	P=0.1	P=0.1	P=0.4
Health behaviours				
Smoking History				
Never	1.00	1.00	1.00	1.00
Ex-smoker	1.20 (0.81 – 1.79)	0.63 (0.41 – 1.00)*	1.26 (0.85 – 1.89)	0.62 (0.41 – 0.94)*
Current smoker	0.27 (0.10 – 0.78)*	1.17 (0.67 – 2.04)	0.34 (0.12 – 0.97)*	1.00 (0.56 – 1.79)
	P=0.02	P=0.03	P=0.03	P=0.05
Physical activity				
Never	1.00	1.00	1.00	1.00
Sometimes	1.29 (0.78 – 2.14)	0.86 (0.49 – 1.49)	1.07 (0.64 – 1.80)	0.92 (0.39 – 1.62)
Frequently	1.01 (0.64 – 1.60)	0.61 (0.37 – 1.02)	0.83 (0.52 – 1.33)	0.66 (0.38 – 1.12)
	P=0.6	P=0.2	P=0.7	P=0.3

† Relative Risk Ratio, Reference group = concordant group

** Significant at P=0.01 level, * significant at P=0.05 level, ^{m} marginally significant

6.4 Discussion

Based on the findings of the literature review and gaps identified in the literature, four socio-demographic and three behavioural risk factors were selected to be investigated in relation to discordance in the NSHD, using both descriptive analysis and multinomial logistic regression models.

6.4.1 Summary of findings

In the descriptive analysis, the two discordant groups were observed to have distinctly different characteristics from the concordant group in terms of their sex, occupational class and smoking history. Marginal differences were also observed in relation to their level of participation in physical activity. The results suggested that overestimators were more likely to be male, have a low occupational class, be a current smoker and never participate in physical activity, compared to participants in the concordant group. Conversely, underestimators were more likely to be female, have a high occupational class, be an ex-smoker and frequently participate in physical activity compared to those in the concordant group. In general, these characteristics aligned overestimators with participants in the high capability concordant group and underestimators with participants in the low capability concordant group.

Differences observed between the descriptive analysis and the multinomial logistic regression models were due to the important influence of sex on discordance. When sex-adjusted models were run to test the association of each factor with discordance, statistically significant associations were observed for education, occupational class, smoking history and physical activity. In line with the descriptive analysis, compared to the concordant group underestimators were more likely to be female, have a medium or high occupational class and have attained A-Levels or higher. Underestimators were also less likely to be current smokers. In contrast, overestimators were less likely to be female, have a high occupational class, be an ex-smoker or frequently participate in physical activity compared to those in the concordant group. No evidence of a sex interaction was observed, and all factors that met the selection criteria of P=0.1 (sex, education, occupational class, smoking history and physical activity) were carried forward into a mutually-adjusted model. When mutually-adjusted, the only factors that remained significantly associated with discordance were sex and smoking history, although underestimators were still more likely to have a medium or high occupational class compared to the concordant group (P<0.06) and were more likely to have a degree (P<0.05).

6.4.2 Socio-demographic factors

Sex was observed to be the factor most strongly associated with discordance in this analysis. The effect of sex grew stronger in the mutually-adjusted model, and appeared to confound the association between discordance and the other factors of interest. The findings in this chapter correspond to the results reported in the literature, with women more likely to underestimate and men more likely to overestimate their physical capability (116;125;129;131). It has been proposed that one of the reasons women are more likely to underestimate is that they may be more sensitive to symptoms of functional decline and thus report limitations more readily than their male counterparts (131). It is also possible that men and women may gauge their functional ability against different criteria when assessing their physical capability for the self-reported measures (129). The performance-based measures may also contribute to the difference in observed discordance between men and women. A recent study demonstrated marked differences in the self-reported and performance-based trajectories of men and women in old age (149). Accounting for baseline physical capability, faster rates of decline were observed in women for the self-reported trajectory, whilst men experienced faster rates of decline in their performance-based physical capability. These opposing trajectories may explain some of the observed discordance.

Two indicators of socioeconomic position were included in the analysis in this chapter: education and occupational class. The association between socioeconomic position and discordance appears to operate in the same direction when either education or occupational class were used, with participants more likely to underestimate their physical capability if they have either a degree or a higher occupational class. Previous studies in the literature may have focused on education as a marker of socioeconomic position rather than occupational class, due to the availability of the data and the universal relevance of education to all participants within a given study (150). In North

America, where many of the published studies were based, education is more commonly used as an indicator of socioeconomic position, which may also explain the high prevalence of this factor in the studies reviewed. Where the association between education and discordance has been explored, previous studies have either found the association to be non-significant (45;123) or have produced conflicting results (40;121). No information was provided about the direction of the discordance in these studies, so the findings in this chapter extend the existing evidence base. There are no studies available for comparison with regard to occupational class, but it seems intuitive that the association between occupational class and education should operate in the same direction.

Despite the high correlation observed between education and occupational class, the findings presented in this chapter suggest that the two indicators of socioeconomic position may have independent associations with discordance. There was some evidence in the mutually-adjusted model that underestimators were more likely to have a degree even after accounting for occupational class, and underestimators were also more likely to have a high occupational class accounting for education. Although education and occupational class are both indicators of socioeconomic position, each measure captures a different element of the concept (150;151), and it is unlikely that all residual socioeconomic confounding was captured by these two indicators. It is plausible that participants with different levels of socioeconomic position gauge their physical capability against different criteria in a similar manner to that previously suggested for sex. For example, those with higher education levels may be more amiable to health messages (150), and consequently have a greater sensitivity to their declining levels of physical capability. In terms of their performance-based measures, individuals with higher socioeconomic position may have access to better resources, enabling them to act on the health messages they receive, resulting in the maintenance of relatively high levels of physical capability in early old age.

In the results presented, marital status was not found to be associated with discordance. This finding reflects the observation made in the only published study to have focused on this marker of social support (121). However, the nonsignificant result in this chapter may be partially explained by the profile of the sample. In the NSHD, nearly 80% of participants were married at age 60-64, which means that the possibility of variation in the remaining categories of marital status was fairly small. There was some suggestion that overestimators were more likely to be single, but this finding must be interpreted with caution due to the small number of participants in this category of marital status. The impact of small numbers and power in the sample will be discussed in more detail in Chapter 8.

Whilst age was highlighted within the literature review (Section 6.1.2) as a potentially important factor associated with discordance, a decision was made not to include age in the main analysis of this thesis because of the narrow age band of participants in NSHD and the confounding effect of region within the 60-64 data collection round (as initial funding was provided for a pilot study in Manchester before the clinical assessment was then rolled out across the country with the clinics in London the last to open). A sensitivity analysis was conducted using the mutually adjusted model from Section 6.3.3, to explore the relationship between discordance and age within the NSHD sample (see Appendix 6). The results indicated that age did not attenuate the association between discordance and any of the other factors included in the mutually adjusted model.

6.4.3 Behavioural risk factors

The literature contains very little evidence of the effect of behavioural risk factors on discordance. As few studies have focused on this topic, the findings in this chapter are often not comparable to previous studies and may provide a useful start to the evidence base in this area.

Of the behavioural risk factors included in the analysis, smoking history had the strongest association with discordance. The association between discordance and smoking history remained after accounting for sex, other behavioural risk factors and sociodemographic factors in the mutually-adjusted model. Underestimators were less likely to be current smokers, and overestimators were less likely to be ex-smokers. It is possible that participants who have actively made the decision to stop smoking have a greater awareness of their overall health, including a decline in their functional ability, and therefore these individuals may perceive that they have low levels of physical capability. In contrast, current smokers may have a higher threshold of perceived difficulty, as they use smoking as part of their coping strategy.

The results presented in this chapter provide some support for the association between physical activity and discordance, with overestimators less likely to frequently participate in physical activity in sex-adjusted models. Participants who are less active may have less opportunity to gauge their level of physical capability accurately, and therefore their perception of their current physical capability may be based on their previous experience prior to decline associated with age. Equally, those who are less active may have lower performance-based physical capability, as they have not gained from the potential benefits associated with participation in physical activity such as strength, power and cardiorespiratory fitness (102). Only two studies had previously explored the association between physical activity and discordance; one study reported a non-significant association (138), and the other presented descriptive results suggesting participants who overestimate their physical capability were less active than those with high levels of self-reported and performance-based physical capability (71). The findings from the latter study support those observed in this chapter, however further research may be required to replicate these findings.

No associations were found between BMI and discordance within this analysis, regardless of whether the variable was treated as a categorical or continuous factor. Only one paper included in the review investigated BMI (71), and the results presented suggested that discordant individuals had a higher BMI than concordant individuals. However, the paper presented purely descriptive analyses, so inferences about the association between BMI and discordance are limited.

6.4.4 Conclusion

The evidence presented in this chapter has extended the existing evidence base of factors associated with discordance. For some factors, such as education, this has meant providing more detail about the direction in which discordance operates; for other factors, such as occupational class and smoking history, this has involved identifying associations not previously documented.

It appears the two discordant groups do have distinct characteristics based on sociodemographic and behavioural risk factors, and these characteristics could potentially be used to help predict discordant individuals.

A discussion of the implications of these findings will be presented in Chapter 8, together with findings from Chapter 7 concerning the association between markers of health and discordance.

7. Health factors associated with discordance

7.1 Introduction

A number of markers of health status were identified as potential factors of interest in association with discordance, based on the findings of the literature review in Chapter 6. By examining the association between markers of health status and discordance, the analysis in this chapter aims to explore whether an individual's health status can be used to identify discordant individuals when only one type of physical capability measure is available. Within this thesis it is proposed that the health characteristics of the discordant groups could be used to provide further evidence in support of resilience amongst those who overestimate their physical capability.

7.1.1 Selection of health factors

In order to ascertain the health characteristics of the two discordant groups, eight markers of health status were investigated: self-reported health, cardio-metabolic disease, respiratory symptoms, pain, fatigue, depression, mental wellbeing and cognitive function. Self-reported health was selected based on the consistent results observed in the literature, and because it provides a good indicator of participants' overall health (152;153). To further explore the association between chronic disease and discordance, several different conditions were included in the analysis, some of which were discussed in the literature review (see Section 6.1.2), such as stroke (one component of cardio-metabolic disease), whilst others such as respiratory symptoms have not been previously examined. An individual's ability to accurately assess their physical capability may not only be affected by the presence of specific conditions, but also by commonly experienced symptoms such as pain and fatigue that can be associated with the experience of these chronic conditions. Consequently, both chronic conditions and symptoms were included in the analysis to clarify the association between the experience of chronic conditions and discordance.

Although the association between depression and discordance may seem intuitive, with individuals potentially more likely to underestimate their physical capability if they experience the symptoms of depression and anxiety, the majority of the studies in the literature review found no evidence of an association. This warrants further investigation, because depression is treatable, and therefore a modifiable risk factor. Wellbeing was also chosen for the analysis, as a positive association with discordance would strengthen the case for an effect of mental health on discordance. However, it is important to note that wellbeing is more than the absence of depression; it incorporates both pleasure and a sense of belonging (154). These feelings of positive mental wellbeing can exist even whilst an individual is suffering from the symptoms of depression and anxiety, suggesting the two concepts may be partially independent. The final factor to be included is cognitive function. Previous studies have found conflicting results, and have not been able to identify the direction in which discordance operates, but it may be important because it is possible that individuals with poor cognitive function are less likely to accurately assess their physical capability.

7.1.2 Longitudinal extension of analysis

As previously highlighted in Section 6.1, studies in the literature have focused on cross-sectional analysis, with little acknowledgement of how factors influencing discordance may operate across life. To address this gap in the literature, the second half of this chapter aims to explore the association of participants' psychological and functional history with discordance.

The concept of discordance focuses on the relationship between current levels of self-reported and performance-based physical capability. However, it is important to remember that in early old age (which is of particular interest in this thesis), levels of physical capability are a product of functional reserve developed in early adulthood and age-related declines in functional ability (153). Therefore, an individual's prior experience of physical capability may be highly relevant in the context of discordance. The closest any of the previous studies in the literature came to incorporating an individual's prior experience of physical capability in the analysis of discordance was a question which asked retrospectively about recent functional decline (129) (see Section 6.1.2). To develop this topic, a longitudinal approach was selected in this chapter to account for age-related declines in physical capability, using data collected prospectively on both self-reported and performance-based physical capability.

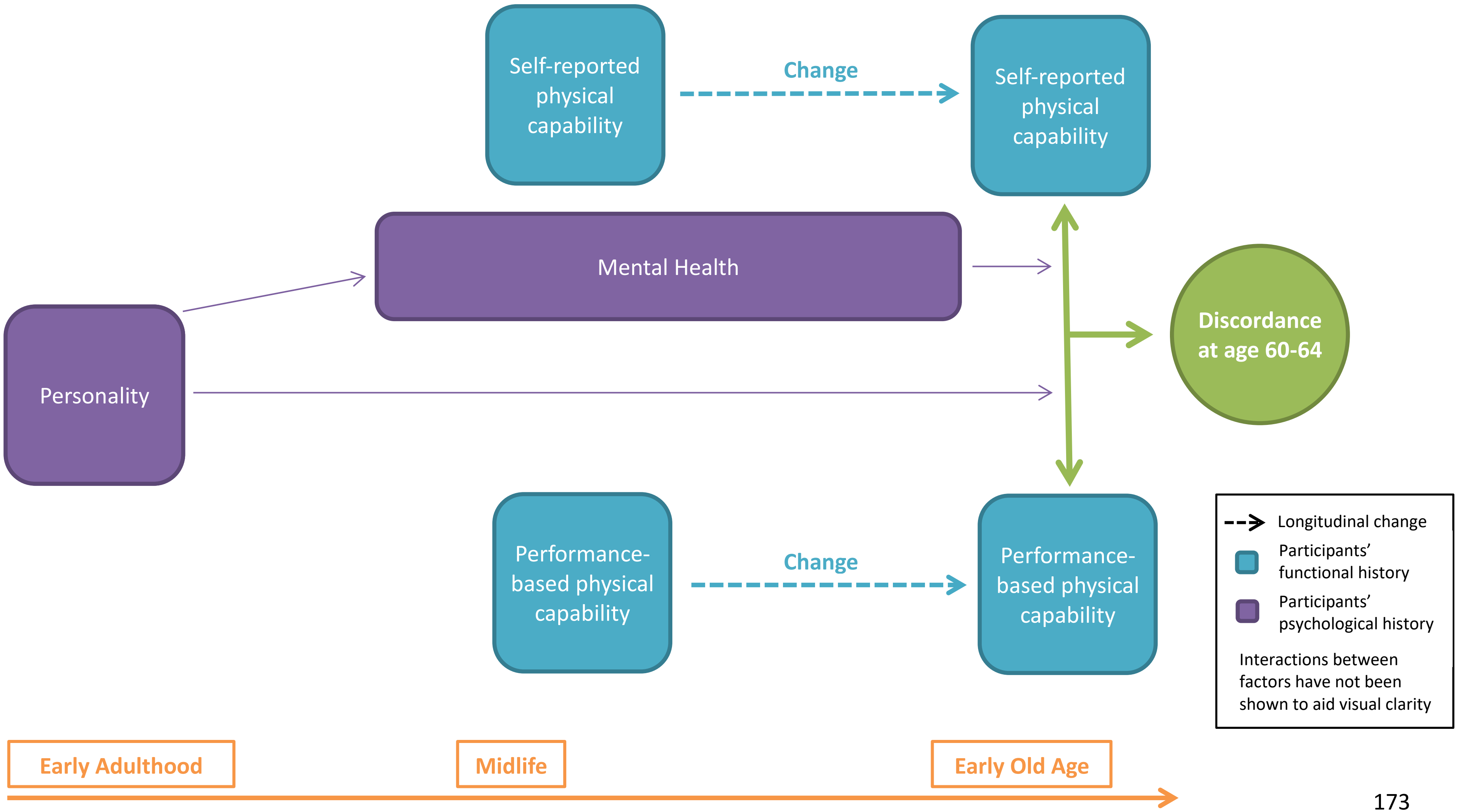
When considering an individual's physical capability history, it may also be prudent to examine their mental health, as there is evidence that poor levels of physical capability and depression have a mutually reinforcing relationship (155). Previous research focusing on the association between physical capability decline and changes in depression (155) indicates that, compared to individuals who were never depressed, individuals with chronic depression were more likely to experience a substantial decline in physical capability levels. Although this relationship holds true using either self-reported or performance-based measures of physical capability, a greater risk of decline was observed when self-reported measures of physical capability were used, suggesting that the chronicity of depression may have an effect on the discordance between measures of physical capability. As outlined in Section 7.1, the cross-sectional association between depression and discordance will be explored in the first part of this chapter, but the second part of this chapter extends the analysis to investigate whether an accumulation of risk, as indicated by the chronicity of symptoms, has a greater association with discordance than the prevalence of depression at any given point.

Personality was included as a factor of interest in the analysis of participants' psychological and functional history, because an individual's character, as defined by their personality traits, may influence the way they adapt to age-related declines in physical capability, as part of the wider ageing process (156). For example, it has been suggested in the literature that individuals with high levels of neuroticism are more sensitive to small declines in their physical capability (139;157). This sensitivity may thus result in the underestimation of their physical capability. Previous research has demonstrated that personality traits remain relatively constant throughout the life course (156), and therefore the relationship between personality and discordance can be explored using data collected at one time point earlier in adulthood in this thesis.

Drawing together participants' psychological and functional history, a conceptual diagram was created (see Figure 7.1), incorporating the longitudinal change from midlife to early old age in mental health, self-reported physical capability and performance-based physical capability. Combined with personality, these longitudinal factors may influence the relationship between self-reported and performance-based physical capability, leading to discordance in early old age.

The aim for this chapter is therefore twofold: to explore the magnitude and direction of discordance observed when investigating the association between markers of health status and discordance, and then to examine the association of participants' psychological and functional history with discordance.

Figure 7.1: Conceptual model of how participants' psychological and functional history influence discordance



7.2 Methods

7.2.1 Variables for health status analysis

In order to address the first aim of this chapter, to explore the association between discordance and health status, eight factors representing health status were chosen: self-reported general health, respiratory disease, cardio-metabolic disease, pain, fatigue, depression, wellbeing and cognitive function. The selection of these variables was discussed in Section 7.1.1, and the following section will provide details of how each variable was ascertained.

In order to assess general health, at age 60-64 participants were asked: “How is your health in general?” Responses to a 5-point Likert scale were categorised into one of three groups: excellent/very good, good or fair/poor. The first chronic condition included in the analysis was respiratory symptoms, and the UK Medical Research Council’s standardised questions (105) were used to identify those with severe respiratory symptoms at age 60-64. Participants were considered to have severe respiratory symptoms if they reported one or more of the following when reflecting on their experience over the past three years: wheezy or whistling chest most days or nights, usually bringing up phlegm or coughing in the morning or during the day or night in winter for at least three months each year, or a chest infection that kept them off work or indoors for more than a week. Cardio-metabolic disease was included as a second indicator of chronic disease, incorporating diabetes, stroke and angina or myocardial infarction (MI). These conditions were grouped due to the low prevalence of each individual condition within the NSHD at age 60-64. Consequently, participants were defined as having cardio-metabolic disease if they reported the presence of one or more these conditions. The doctor diagnosis of diabetes, stroke and angina or MI was recorded up to and including the assessment at age 60-64 to identify participants who had ever experienced these conditions.

Pain was assessed at age 60-64, with participants asked: “How much bodily pain have you had in the past four weeks?” This question was taken from the bodily pain scale of the SF-36 (158-160). Participants were dichotomised as those who did not report pain and those who reported any level of pain (mild, moderate or severe). Fatigue was also ascertained at age 60 to 64, when participants were asked: "How much of the time during the past four weeks did you feel tired?" This question was taken from the vitality scale of the SF-36 (158-160). Responses were dichotomised as those who did not report fatigue and those who experienced fatigue at least some of the time (ranging from a little to all of the time).

Two measures of mental health were incorporated into the analysis: depression and wellbeing. Symptoms of anxiety and depression were identified using the General Health Questionnaire-28 (104) at age 60-64. A threshold score of ≥5 was used to identify participants considered to be a "case". Wellbeing was assessed at age 60-64 using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS). Participants were asked to respond to 14 positively worded statements, each scored 1 (none of the time) to 5 (all of the time). An aggregate score was then produced, ranging from 14-70, with higher values indicating higher levels of wellbeing (161). Where data were missing for ≤3 items within the scale, a value was imputed based on the specific participant's mean score.

Verbal memory was used as an indicator of cognitive function. Unlike many of the studies reviewed in Section 6.1, which focused on cognitive impairment, the use of verbal memory enables this thesis to examine cognitive function across the normal range. Previous work has also shown that, in the context of ageing, verbal memory predicts clinically significant declines in global cognition (162;163). Verbal memory was assessed at age 60-64 using a 15-item word learning task, with scores representing the number of words correctly recalled over three trials (maximum score 45).

7.2.2 Variables for longitudinal analysis

The variables chosen for the longitudinal discordance analysis, used to address the second aim of this chapter, focus on the psychological and physical capability history of study members as outlined in Section 7.1. The variables fall into three categories: personality traits, depressive symptoms and categories of physical capability decline. The following section provides details about how each variable was ascertained.

Personality traits

The two measures of personality that were included in this analysis, neuroticism and extraversion, were assessed when participants were aged 26, using the shortened form of the Maudsley Personality Inventory (MPI) (164). The MPI includes six extraversion items that focus on sociability, energy and activity orientation, and six neuroticism items that focus on emotional stability, mood and distractibility. Each item had three response options, scored 0-2: no; don't know; and yes. For each personality trait, the scores from the six relevant items were aggregated to produce a score from 0-12, with high scores representing high levels of the trait of interest. Each variable was dichotomised using a cut point of ≤6 (165) to distinguish those with high levels of the specific personality trait.

Depressive typologies

To account for participants' history of symptoms of depression and anxiety in later midlife (age 53 to 60-64), a categorical variable was generated, similar to that used as a covariate in Chapter 4. At ages 53 and 60-64, the General Health Questionnaire-28 was used to assess symptoms of depression and anxiety (104), and an appropriate casesness threshold (≥5) was used to identify those with symptoms at each age. Participants were categorised as having experienced symptoms of depression and anxiety at neither age, at age 53 only, at age 60-64 only, or both (chronic depression).

Physical capability

As previously outlined in Section 3.2, information on participants' levels of physical capability has been collected from midlife onwards using self-reported measures at age 43, 53 and 60-64 and performance-based measures at ages 53 and 60-64. Two longitudinal physical capability variables were produced for inclusion in this chapter's analysis: one self-reported measure accounting for changes in levels of physical capability between age 43 and age 60-64, and one performance-based measure accounting for changes in levels of physical capability between age 53 and 60-64.

7.2.3 Construction of longitudinal physical capability variables

Longitudinal performance-based variable

In order to produce the longitudinal performance-based variable, comparable measures were required at ages 53 and 60-64. It was not possible to recreate the summary performance-based measure outlined in Section 5.2.2 at age 53, because the fourth performance test completed at age 60-64 (timed get up and go) was not conducted at the earlier assessment. Instead a decision was made to recreate the summary measure as originally described by Guralnik and colleagues (81) at both ages. This summary measure was a composite score of the remaining three performance-based measures (grip strength, standing balance (eyes closed) and chair rise time), with each test rescaled 0-1 before aggregation.

Once the two comparable measures were created, a categorical variable with four categories of change was produced, following a similar procedure to that outlined by Cooper and colleagues (166). A sex-specific standard deviation score (mean=0, SD=1) was calculated using the comparable summary measure at age 53 for those who had valid measures at both time points. The comparable summary measure created at age 60-64 was standardised using the sex-specific standard deviation and mean values from age 53. Each variable was then categorised: <-1SD, -1 to 1 SD, >1 SD before the two variables were cross-tabulated and categories of change identified as shown in Table 7.1.

Table 7.1: Categories of change for summary performance-based measure between age 53 and age 60-64

Summary performance-based SD score at:		**Age 60-64**	
Age 53	<-1 SD	-1 SD to 1 SD	>1 SD
<-1 SD	Stable Low	Maintenance†	Maintenance†
-1 SD to 1 SD	Decline	Maintenance†	Maintenance†
>1 SD	Decline	Decline	Stable high

† Maintenance group = those who maintained physical capability within "normal" range

Descriptive analyses were used to explore the sex-specific mean levels of performance-based physical capability at ages 53 and 60-64 in the total sample and in the four categories of the longitudinal variables.

Longitudinal self-reported variable

A decision was made to exclude the self-reported measures captured at age 53 from the longitudinal self-reported variable due to the limited number of tasks assessed at this age. The small number of physical capability tasks assessed at age 53 would affect comparability with the two more extensive data collection rounds at ages 43 and 60-64.

Unlike the longitudinal performance-based variable, where two new comparable measures were required, it was possible to recreate a summary self-reported measure at age 43 in the same way as that used in previous analyses at age 60-64 (outlined in Section 5.2.2). Each response category of the five self-reported measures included in the summary measure (walk, grip, bend, balance and climb stairs) were assigned a value from the OPCS severity scale. A number of modifications to the procedure outlined in Section 5.2.2 were required to ensure comparability between the two summary variables at ages 43 and 60-64. At age 43 the response categories available for the five self-reported measures of physical capability differed from those used to create the summary measure at age 60-64 (see Appendix 1). To harmonise the data, the response categories from age 43 were capped to the highest level of severity used in the summary measure at age 60-64. For the grip task, a decision was made to combine two questions from the age 43 data collection wave detailing the level of difficulty participants had unscrewing the lid of a coffee jar and pouring from a full kettle, as these questions best reflected the wording of the grip question at age 60-64.

Once the summary self-reported measure was produced at age 43, the summary scores assigned to each participant were subtracted from the score obtained from the summary measure at age 60-64. If the resulting difference between the two scores was positive, participants were considered to have experienced a decline in their level of reported physical capability between the two time points. If the resulting difference was negative then participants reported an improvement in their level of physical capability. Due to the skewed nature of the summary self-reported variable, it was predicted that many participants would report no change in their level of physical capability, remaining at the lowest level of reported difficulty throughout. Alternative approaches were considered, but the highly skewed nature of the two summary variables limited the potential methods available. A more detailed discussion of this issue will follow in Section 7.4.

Descriptive analyses were used to compare the median levels of self-reported physical capability at ages 43 and 60-64 in the total sample and in the three categories of the longitudinal variable.

7.2.4 Analysis of health and longitudinal factors

To explore the association of both health and longitudinal factors with discordance, a similar analytical approach was used to that outlined in Section 6.2.4.

Firstly, descriptive analyses were conducted to identify the direction of the association for each health and longitudinal factor with discordance. The proportion of participants within each subcategory of the health and longitudinal factors was calculated across the two discordant and combined concordant groups. Chi-squared tests were used to assess the difference between the three groups to ascertain whether the two discordant groups were distinct from the concordant group. For continuous variables, the mean and standard deviation were calculated for the concordant and two discordant groups. ANOVA analyses were then used to test the difference in mean values across the three groups.

As an extension to the descriptive analyses, an analysis was conducted to ascertain whether there was a gradient in either the health or longitudinal factors across the three concordant groups. A non-parametric test for trend was used to formally test the gradient in both categorical and continuous variables. The characteristics of the two discordant groups were then compared to the three concordant groups.

The second stage of the analysis used sex-adjusted multinomial logistic regression models to quantify the magnitude of the association between each factor and discordance, accounting for sex. Each health and longitudinal factor was placed into sex-adjusted models run on the maximum number of participants, with discordance as the dependent variable and the combined concordant group used as the reference category. For each of these models, evidence of a sex interaction and the association with discordance were formally tested. When exploring the association between markers of health and discordance, a sensitivity analysis was conducted to determine whether the individual components of the cardio-metabolic disease variable were independently associated with discordance.

When exploring the association between longitudinal factors and discordance, an extended descriptive analysis was conducted to provide more detailed information about the association between the longitudinal physical capability variables and discordance. The difference in median levels of the summary self-reported physical capability at age 43 and age 60-64, and the difference in mean levels of the summary performance-based physical capability measures at age 53 and age 60-64, were compared across the five concordant and discordant groups.

The final stage of the analysis used a mutually-adjusted multinomial logistic regression model to explore potential confounding and the interdependence of factors within the analysis. A criteria of $P \leq 0.1$ from the test of association, run on the sex-adjusted models as part of the second stage of the analysis, was used to select factors for inclusion into the mutually-adjusted model. Once factors had been selected, the sex-adjusted models were repeated using a restricted sample that had complete data for all relevant covariates before the mutually-adjusted model was run.

A sensitivity analysis was conducted to test whether participants excluded from the restricted sample differed from those included in the mutually-adjusted model in terms of their health or longitudinal characteristics.

7.3 Results

The following section can be separated into two parts. Initially, the results exploring the association between markers of health status at age 60-64 and discordance will be presented, in order to address the first aim of this chapter. Once all three stages of the analysis have been presented, the results section will move forward to the second aim of the chapter and will focus on the association of participants' psychological and functional history with discordance.

7.3.1 Descriptive analysis of markers of health status

The results of the descriptive analysis used to ascertain whether the two discordant groups were characteristically distinct from the concordant group in terms of their health status are shown in Table 7.2.

General health

With respect to self-reported health, no differences were observed between the two discordant and concordant groups (P=0.9).

Chronic conditions

When exploring the differences between the two discordant and concordant groups in terms of their respiratory symptoms and cardio-metabolic disease, no statistically significant differences were observed (P=0.2 and P=0.6 respectively). In terms of their experience of pain, underestimators were more likely to report the presence of pain compared to those in the concordant group (85% vs 66%). Conversely, overestimators were more likely to report no pain compared to those in the concordant group (36% vs 34%). When tested, the pain characteristics of the two discordant and concordant groups were distinctly different (P<0.001). The descriptive results for fatigue indicate no differences in the characteristic of the two discordant and concordant groups (P=0.2).

Mental health

With regard to depression and wellbeing at age 60-64, no differences were observed between the two discordant and concordant groups (P=0.2 and P=0.4 respectively).

Cognitive function

The descriptive results for verbal memory indicate that, compared to the concordant group, underestimators had a higher verbal memory score, suggesting higher levels of cognitive function, and overestimators had a lower verbal memory score, implying that these participants had lower levels of cognitive function. Strong evidence was found that the differences observed in the verbal memory scores of the two discordant and concordant groups were not due to chance (P=0.02).

Characteristics of the three concordant groups

The analysis was extended to explore the characteristics of the three concordant groups (see Appendix 7). Evidence of a gradient was observed across the three concordant groups for all eight factors (each gradient had P<0.01, except fatigue where P=0.1). Participants in the high concordant group were more likely to experience excellent/very good health, high levels of wellbeing and verbal memory, low levels of pain and fatigue, and fewer reports of depression, respiratory symptoms and cardio-metabolic disease. In contrast, those in the low capability concordant group were more likely to experience poor health, lower levels of wellbeing and verbal memory, high levels of pain and fatigue, and were more likely to report the presence of depression, respiratory disease and cardio-metabolic disease. When the two discordant groups were mapped onto these gradients, no clear pattern could be ascertained.

Table 7.2: Health characteristics of the combined concordant and two discordant groups

Factors of interest	Proportion of sample (%) †				P Value‡
	Total Sample	Combined Concordant group	Discordant groups		
			Under-estimators	Over-estimators	
Maximum N*	(N=1981)	(N=1708)	(N=135)	(N=138)	
General Health					
Self-reported health	(N=1820)				
Excellent/V. Good	55.1	55.3	53.6	52.9	0.9
Good	31.7	31.5	31.2	34.5	
Fair/Poor	13.2	13.1	15.2	12.6	
Chronic conditions					
Respiratory symptoms	(N=1752)				
Absent	81.5	82.2	77.1	77.8	0.2
Present	18.5	17.9	23.0	22.2	
Cardio-metabolic §	(N=1739)				
Absent	87.1	86.8	90.1	87.7	0.6
Present	12.9	13.3	9.9	12.3	
Pain	(N=1806)				
Absent	32.5	33.7	15.1	35.9	
Present	67.5	66.3	84.9	64.1	<0.001
Fatigue	(N=1812)				
Absent	7.95	8.17	3.97	9.24	
Present	92.1	91.8	96.0	90.8	0.2
Mental health					
Depression and anxiety	(N=1935)				
No	82.7	83.2	77.3	81.5	0.2
Yes	17.3	16.8	22.8	18.5	
Wellbeing (continuous)	(N=1751)				
Mean (WEMWBS score)	51.8	51.9	51.4	51.1	0.4
SD	8.0	7.9	8.0	8.4	
Cognitive function					
Verbal Memory (cont.)	(N=1936)				
Mean (words recalled)	24.4	24.5	25.1	23.1	0.02
SD	6.1	6.1	5.9	6.3	

* Value of N varies for each factor due to missing data
† For continuous variables mean and standard deviation (SD) values provided instead of proportions and details of units provided
‡ P value from χ^2 test if variables categorical and from ANOVA if continuous variables
§ Presence of diabetes, stroke or MI. When present: 86% had only 1 condition, 12% had 2 conditions and 1 % had all 3 conditions

7.3.2 Sex-adjusted markers of health status models of discordance

The results of the sex-adjusted multinomial logistic regression models used to explore the association between markers of health status and discordance are shown in Table 7.3.

When formally tested, there was no evidence of a sex interaction for any of the variables included in this analysis (P>0.06), with the exception of wellbeing (P=0.048).

General health

No association was observed between self-reported health and discordance (P=0.7).

Chronic conditions

In terms of respiratory symptoms and cardio-metabolic disease, no association was found between either of these factors and discordance. The individual components of the cardio-metabolic variable were included in the sex-adjusted models as a sensitivity analysis, and the results demonstrated that none of the individual factors were independently associated with discordance (P>0.1) (see Appendix 7). The combined cardio-metabolic disease variable was therefore appropriate to use.

With respect to their experience of pain (compared to the concordant group), participants who reported the presence of pain were 2.59 (95% CI: 1.57 – 4.29) times more likely to underestimate their physical capability compared to participants who did not report pain. The association between pain and underestimation of physical capability was very strong (P≤0.01), but no significant association was observed between pain and overestimation of physical capability (P=0.8).

No association was observed between fatigue and discordance (P>0.2).

Table 7.3: Sex-adjusted models for the association between markers of health status, ascertained at age 60-64, and discordance (run on max. N)

Factors of interest [‡]	**RRR (95% CI)** [†]	
	Underestimators	Overestimators
General Health		
Self-reported health		
Excellent/V. Good	1.00	1.00
Good	0.95 (0.63 – 1.44)	1.19 (0.79 – 1.79)
Fair/poor	1.20 (0.70 – 2.06)	1.00 (0.56 – 1.79)
	P=0.7	P=0.7
Chronic Conditions		
Respiratory disease		
Absent	1.00	1.00
Present	1.42 (0.91 – 2.22)	1.29 (0.82 – 2.04)
	P=0.1	P=0.3
Cardio-metabolic[§]		
Absent	1.00	
Present	0.86 (0.46 – 1.61)	0.85 (0.47 – 1.52)
	P=0.6	P=0.6
Pain		
Absent	1.00	1.00
Present	2.59 (1.57 – 4.29)	0.95 (0.64 – 1.41)
	P<0.001	P=0.8
Fatigue		
Absent	1.00	1.00
Present	1.91 (0.76 – 4.78)	0.93 (0.48- 1.77)
	P=0.2	P=0.8
Mental Health		
Depression and anxiety		
Not a case	1.00	1.00
Case	1.25 (0.81 – 1.92)	1.24 (0.78 – 1.95)
	P=0.3	P=0.4
Wellbeing		
Per 1 SD increase	0.93 (0.77 – 1.12)	0.89 (0.74 – 1.07)
	P=0.5	P=0.2
Cognitive Function		
Verbal Memory		
Per 1 SD increase	1.00 (0.83 – 1.19)	0.83 (0.69 – 1.00)*
	P=1.0	P=0.05

†Relative Risk Ratio, Reference group = concordant group
‡Sex interaction formally tested but no evidence found for all factors (P>0.06) except wellbeing P<0.05
§ Presence of diabetes, stroke or MI. When present: 86% had only 1 condition, 12% had 2 conditions and 1 % had all 3 conditions
* Significant at P=0.05 level, ** significant at P=0.01 level
Blue values indicate factors with P≤0.1 and therefore will be taken forward to the mutually-adjusted model

Mental health

With regards to depression and wellbeing, no association with discordance was observed. However, as previously stated, evidence of a sex interaction was found for wellbeing. In sex-specific models (see Table 7.4), males with higher WEMWBS scores were less likely (compared to the concordant group) to underestimate their physical capability (RRR per 1 SD increase = 0.62 (95% CI: 0.42 – 0.99)). The association between wellbeing and underestimation was statistically significant for men (P=0.02), but no association was observed between wellbeing and overestimation of physical capability for men (P=0.1) or between wellbeing and discordance in either direction for women (P>0.6).

Table 7.4: Sex-specific models for the association between wellbeing and discordance

	RRR (95% CI) †	
	Underestimators	Overestimators
Male		
Wellbeing (continuous)		
Per 1 SD increase	0.62 (0.42 – 0.91)* P=0.02	0.82 (0.65 – 1.05) P=0.1
Female		
Wellbeing (continuous)		
Per 1 SD increase	1.05 (0.85 – 1.30) P=0.6	0.98 (0.73 – 1.33) P=0.9

† Relative Risk Ratio, Reference group = concordant group
* Significant at P=0.05 level
Blue values indicate factors with P≤0.1 and therefore will be taken forward to the mutually-adjusted model

Cognitive function

In terms of verbal memory, those with a higher verbal memory score were less likely (compared to the concordant group) to overestimate their physical capability (RRR per 1 SD increase = 0.83 (95% CI: 0.69 – 1.00)). When tested, the association between verbal memory and overestimation was reasonably strong (P=0.05), however no association was observed between verbal memory and underestimation of physical capability (P=1.0).

7.3.3 Mutually-adjusted markers of health status model

Three factors (pain, wellbeing and verbal memory) were carried forward to the mutually-adjusted model on the basis of a selection criteria P≤0.1 for the test of association. The mutually-adjusted model also accounted for sex, and an interaction term between sex and wellbeing.

The initial results in Table 7.5 show the sex-adjusted models for the selected factors run on a restricted sample (N=1561). All results reflected those observed in the previous analysis when models were run on maximum N (Section 7.3.2). However, in terms of wellbeing and verbal memory, the association with discordance weakened in the restricted sample, particularly for the association between verbal memory and overestimation of physical capability, which was found to be non-significant (P=0.4). However, the association between pain and underestimation of physical capability strengthened slightly in the models run on the restricted sample.

When a sensitivity analysis was conducted, in terms of their health status, no statistically significant differences were observed between participants included in the restricted sample and those excluded for missing data on covariates (see Appendix 7).

In the mutually-adjusted model (compared to the concordant group), participants who experience pain were 2.88 (95% CI: 1.66 – 4.98) times more likely to underestimate their physical capability compared to those who did not report pain. The association between wellbeing and discordance in men was strongly attenuated, with little evidence (P=0.1) that males with higher WEMWBS scores were less likely (compared to the concurrent group) to underestimate their physical capability (RRR per 1 SD increase = 0.71 (95% CI: 0.45 – 1.11)). The association between verbal memory and discordance was fully attenuated.

Table 7.5: Comparison between sex-adjusted and fully adjusted models for socio-demographic and health risk behaviours (N=1561)

	Sex-adjusted model on restricted N RRR (95% CI)†		**Mutually-adjusted model** RRR (95% CI)†	
	Underestimators	Overestimators	Underestimators	Overestimators
Pain				
Absent	1.00	1.00	1.00	1.00
Present	2.85 (1.65 – 4.91) P<0.001	0.89 (0.58 – 1.34) P=0.6	2.88 (1.66 – 4.98)** P<0.001	0.85 (0.56 – 1.29) P=0.4
Wellbeing (Male)				
Per 1 SD increase	0.67 (0.44 – 1.03) m P=0.07	0.83 (0.63 – 1.08) P=0.2	0.71 (0.45 – 1.11) P=0.1	0.84 (0.64 – 1.09) P=0.2
Wellbeing (Female)				
Per 1 SD increase	1.05 (0.84 – 1.30) P=0.7	0.96 (0.69 – 1.34) P=0.8	1.10 (0.87 – 1.37) P=0.4	0.93 (0.67 – 1.30) P=0.7
Verbal Memory Per 1 SD increase	1.03 (0.85 – 1.26) P=0.8	0.92 (0.75 – 1.13) P=0.4	1.05 (0.86 – 1.28) P=0.6	0.92 (0.75 – 1.13) P=0.4

†Relative Risk Ratio, Reference group = concordant group
** Significant at P=0.01 level, * significant at P=0.05 level, m marginally significant

In order to address the second aim of this chapter (to explore the association of participants' psychological and function history with discordance), two longitudinal physical capability variables were created. The following section provides the results of the descriptive analysis, detailing the changes experienced by participants in their performance-based and self-reported physical capability from midlife to early old age.

7.3.4 Descriptive analysis of longitudinal physical capability created

Longitudinal performance-based physical capability

The two comparable summary performance-based measures used to create the longitudinal measure each had a potential range of 0 to 3, with 0 representing low levels of physical capability and 3 indicating high levels of physical capability across all three components of the performance-based measure. At age 53, the mean score was 1.42 for men and 1.31 for women. By age 60-64, mean levels of performance-based physical capability had declined for both sexes to a value of 1.28 for men and a value of 1.21 for women. When tested, the decline in mean performance-based scores between age 53 and age 60-64 was significant ($P<0.001$) for both men and women.

Using the four categories of change, approximately 20% of both men and women were classified as showing clear evidence of decline in their performance-based physical capability measures (see Table 7.6). Among this group, mean performance-based physical capability declined between age 53 and 60-64 from a value of 1.69 to 1.11 for men and 1.55 to 1.01 for women. Approximately 6% of the sample were categorised as experiencing stable low levels of physical capability between age 53 and age 60-64, and 2% were classified as experiencing stable high levels of physical capability in the same time period.

Longitudinal self-reported physical capability

The two summary self-reported measures used to create the longitudinal measure each had a potential range of 0 (no reported difficulty across all five tasks) to a maximum theoretical value of 25.33 (difficulty in all five tasks). Both summary measures had a noticeable skew, with a median score of 0 observed for men and women at both ages. However, strong evidence of an increase in the levels of reported difficulty between ages 43 and 60-64 was found for men and women (P<0.001). These results are consistent with those observed in Chapter 4 when exploring the hierarchy of loss.

When participants were categorised based on their experience of self-reported physical capability between ages 43 and 60-64, 25% of men and 49% of women were classified as having declined. The level of decline observed was relatively small, with an increase in median difficulty score from 0 to 2.58 for men and 0 to 2.75 for women. Similar levels of change were observed in the 2% of both men and women who were categorised as having improved (see Table 7.7). When the levels of self-reported physical capability at ages 43 and 60-64 were compared for each sub-category, both the decline and improved group demonstrated strong evidence of difference between the two time points (P<0.001), whereas there was no evidence of a difference for those who were categorised as having maintained their capability (P=1.0).

Table 7.6: Mean levels of the two comparable summary performance-based measures at age 53 and 60-64 in each category of change, by sex

	Men				Women			
	N (%)	Mean (SD)			N (%)	Mean (SD)		
		Age 53	Age 60-64	P Value †		Age 53	Age 60-64	P Value†
Total	877	1.42 (0.39)	1.28 (0.33)	<0.001	982	1.31 (0.35)	1.21 (0.34)	<0.001
Stable Low	51 (5.82)	0.77 (0.23)	0.69 (0.29)	0.07	62 (6.31)	0.66 (0.24)	0.60 (0.29)	0.1
Decline	208 (23.7)	1.69 (0.45)	1.11 (0.37)	<0.001	196 (20.0)	1.55 (0.38)	1.01 (0.39)	<0.001
Maintenance ‡	600 (68.4)	1.36 (0.26)	1.36 (0.22)	0.9	700 (71.3)	1.27 (0.24)	1.30 (0.22)	0.01
Stable High	18 (2.05)	2.16 (0.26)	2.04 (0.24)	0.1	24 (2.44)	1.99 (0.27)	1.87 (0.17)	0.05

† P Value from t test comparing mean values of summary performance-based measures at age 53 and age 60-64

‡ Maintenance group = those who maintained physical capability within "normal" range

Table 7.7: Median level of the two comparable summary self-reported measures at age 43 and 60-64 in each category of change, by sex

	Men				Women			
	N (%)	Median (IQR) †			N (%)	Median (IQR) †		
		Age 43	Age 60-64	P Value		Age 43	Age 60-64	P Value
Total	967	0 (0-0)	0 (0–0.5)	<0.001	1056	0 (0–0)	0 (0–2.75)	<0.001
Improved	18 (1.86)	1.75 (1–2.75)	0 (0–0)	<0.001	19 (1.80)	2.75 (1.5–6.75)	0 (0–2.75)	<0.001
No change	706 (73.0)	0 (0–0)	0 (0–0)	1.00	525 (49.7)	0 (0–0)	0 (0–0)	1.00
Decline	243 (25.1)	0 (0–0)	2.58 (1.5–6.0)	<0.001	512 (48.5)	0 (0–0)	2.75 (1.75–5.88)	<0.001

† Low values (no reported difficulty) correspond to high levels of physical capability

‡ P Value from Mann-Whitney comparing values of summary self-reported measures at age 43 and age 60-64

7.3.5 Descriptive analysis of participants' psychological and functional history

The descriptive analysis used to ascertain whether the two discordant groups were characteristically distinct from the concordant group in terms of their psychological and functional history is shown in Table 7.8.

<u>Personality</u>

When exploring the personality traits, no significant characteristic differences were observed between the two discordant and concordant groups for extraversion. With respect to neuroticism, underestimators were more likely to have a high level of neuroticism compared to the concordant group (54% vs. 45%). In contrast, overestimators were more likely to have lower levels of neuroticism (64% vs. 55%). When tested, the neurotic tendencies of the two discordant and concordant groups were distinctly different (P=0.02).

<u>Depression from age 53 onwards</u>

In terms of their experience of anxiety and depressive symptoms from age 53 onwards, participants who underestimated their physical capability were more likely to have symptoms at both ages compared to the concordant group (14% vs. 6%). Conversely, compared to the concordant group, overestimators were less likely to have symptoms at both ages (4.7% vs. 6.2%) and more likely to have symptoms only at age 60-64 (14% vs. 10%). Strong evidence was found that the differences observed in the experience of anxiety and depression symptoms in the two discordant and concordant groups were not due to chance (P=0.02).

<u>Longitudinal physical capability variables (functional history)</u>

The descriptive results for physical capability indicate underestimators were more likely to have experienced a decline in self-reported physical capability (99% vs. 34%) compared to the concordant group, whilst overestimators were more likely to have no change in their self-reported physical capability (89% vs. 65%). With respect to participants' performance-based physical capability, underestimators were more likely to have maintained stable high levels of capability compared to the concordant group (8.5% vs. 2.0%), whereas overestimators were more likely to have declined (60% vs. 19%) or have stable low capability (16% vs. 5%) compared to the concordant group.

When tested, the physical capability characteristics (measured using either the self-reported or performance-based longitudinal variables) of the two discordant and concordant groups were distinctly different (P<0.001).

Table 7.8: Characteristics of the two discordant groups using participants' functional and psychological history

Factors of interest	**Proportion of sample (%)**				**χ²**
	Total Sample	Combined concordant group	Discordant groups		**P Value**
			Under-estimators	Over-estimators	
Maximum N*	1981	1708	135	138	
Personality					
Extraversion					
Less	34.1	33.9	40.0	31.2	
More	65.9	66.1	60.0	68.9	0.3
Neuroticism					
Less	54.8	54.7	45.8	63.9	
More	45.2	45.3	54.2	36.1	0.02
Depression					
Late midlife experience					
No symptoms	70.7	71.1	63.3	72.2	
Symptoms at age 53	12.7	12.9	13.3	9.52	
Symptoms at age 60-64	10.0	9.80	9.17	13.5	
Symptoms at both ages	6.64	6.22	14.2	4.76	0.02
Physical capability					
Self-reported	(N=1799)				
Improved	1.83	1.87	0.85	2.33	
No change	62.0	64.5	0.00	89.2	
Decline	36.1	33.7	99.2	8.53	<0.001
Performance-based	(N=1759)				
Stable low	5.57	5.24	0.00	15.7	
Decline	21.7	19.2	16.1	60.0	
Maintenance†	70.4	73.5	75.4	24.4	
Stable high	2.33	2.03	8.47	0.00	<0.001

* Value of N varies for each factor due to missing data

† Maintenance group = those who maintained physical capability within "normal" range

Characteristics of the three concordant groups

When the descriptive analysis was extended to explore the psychological and functional characteristics of the three concordant groups (see Appendix 8), evidence of a gradient was observed for all factors (each gradient P<0.01, except neuroticism P=0.07) apart from extraversion (P=0.6). Participants in the low concordant group were more likely to have high levels of neuroticism, symptoms of anxiety and depression at both ages or at age 60-64 only, declining levels of physical capability in both the self-reported and performance-based longitudinal measures, and low stable levels of performance-based capability. In contrast, participants in the high concordant group were more likely to have low levels of neuroticism, have maintained their level of physical capability longitudinally (using either measure), and have high stable levels of performance-based capability. Participants in the high concordant group were also less likely to experience symptoms of anxiety and depression either at both ages or at age 60-64 only. When the two discordant groups were mapped onto these gradients, no clear pattern could be ascertained.

7.3.6 Sex-adjusted psychological history models of discordance

The results of the sex-adjusted multinomial logistic regression models used to explore the association of participants' psychological and functional history with discordance are shown in Table 7.9. When tested, there was no evidence of a sex interaction for any of the variables included in the analysis (P>0.2).

Personality

No association was found between either personality trait and discordance (P>0.2).

Depression

In terms of their midlife experience of symptoms of anxiety and depression (compared to the concordant group), participants who experienced symptoms at both ages were 2.04 (95% CI: 1.15 – 3.63) times more likely to underestimate their physical capability compared to those who have never experienced any symptoms. When tested, there was some evidence (P=0.09) that the association between the experience of symptoms and underestimation of physical capability was not due to chance. However, no association was observed between the experience of symptoms and overestimation of physical capability (P=0.4).

<u>Longitudinal physical capability (functional history)</u>

With regards to the longitudinal physical capability variables, a decision was made to dichotomise the variable into those who experienced decline and those who did not, given the absence of participants in certain categories when the longitudinal and discordant categories were cross-tabulated (see Table 7.8).

The results observed for self-reported decline suggested (compared to the concordant group) participants who experienced a decline were less likely to overestimate their physical capability than those who did not report a decline (RRR= 0.20 (95% CI: 0.11 – 0.37)). Strong evidence was observed for this association (P<0.001). No results have been presented for the association between self-reported decline and underestimation of physical capability as there were insufficient numbers in the baseline category to ensure stability in the model.

When the association between performance-based decline and discordance was examined, individuals who experienced a decline were more likely (compared to the concordant group) to overestimate their physical capability (RRR= 6.20 (95% CI: 4.17 – 9.21)) compared to those who did not experience a decline in their performance-based physical capability. The strength of the association between performance-based decline and overestimation of physical capability was strong (P<0.001), but no association was observed between performance-based decline and underestimation of physical capability (P=0.5).

Table 7.9: Sex-adjusted models for the association between discordance and participants' functional or psychological history (run on max. N)

Factors of interest ‡	RRR (95% CI) †	
	Underestimators	Overestimators
Personality		
Extraversion		
Less	1.00	1.00
More	0.85 (0.58 – 1.26)	1.07 (0.72 – 1.60)
	P=0.4	P=0.7
Neuroticism		
Less	1.00	1.00
More	1.14 (0.78 – 1.67)	0.76 (0.51 – 1.12)
	P=0.5	P=0.2
Depression		
Late midlife experience		
No symptoms	1.00	1.00
Symptoms at age 53	1.02 (0.58 – 1.80)	0.74 (0.39 – 1.41)
Symptoms at age 60-64	0.89 (0.46 – 1.72)	1.49 (0.86 – 2.60)
Symptoms at both ages	2.04 (1.15 – 3.63)*	0.87 (0.37 – 2.04)
	P=0.09	P=0.4
Physical capability		
Self-reported		
No decline	--	1.00
Decline	--	0.20 (0.11 – 0.37(=) **
		P<0.001
Performance-based		
No decline	1.00	
Decline	0.84 (0.50 – 1.39)	6.20 (4.17 – 9.21)**
	P=0.5	P<0.001

‡ No evidence of a sex interaction for any of these models: P values range from 0.2-0.7
†Relative Risk Ratio, Reference group = concordant group
*Significant at P=0.05 level, ** significant at P=0.01 level
-- Unstable model

7.3.7 Extended analysis of longitudinal physical capability and discordance

The results of the extended descriptive analysis, to investigate the mean/median levels of the component measures used to create the longitudinal physical capability variables across the five concordant and discordant groups, are shown in Tables 7.10 and 7.11.

Strong evidence of a decline in mean levels of performance-based physical capability was observed in the low concordant, middle concordant and overestimator groups for both men and women (P<0.001) (see Table 7.10). Overestimators appear to have declined to a greater extent than the middle concordant group. When the magnitude of decline was compared between the overestimator and the low concordant group, and between the overestimator and the middle concordant group in both men and women, a statistically significant difference was found (P<0.002).

In women, strong evidence of an increase in mean performance-based physical capability levels was observed in the high concordant and underestimator groups (P<0.005). A similar trend was observed in the high concordant group for men, but the increase was not statistically significant (P=0.08). There was no difference in the mean levels of performance-based physical capability for male underestimators (P=0.5), but this may be due to the small number of men in this particular group.

From the self-reported perspective of physical capability, strong evidence of an increase in levels of reported difficulty was observed, for both men and women, in the low concordant, middle concordant and underestimator groups (P<0.0001) (see Table 7.11). Underestimators appear to have declined in terms of their self-reported physical capability (reported more difficulty) to a lesser extent than those in the low concordant group, but more than the middle concordant group. When the magnitude of this decline was compared between underestimators and the low concordant group, and between underestimators and the middle concordant group, a significant difference was found in both men and women (P<0.001). No evidence of a change in the reported levels of difficulty was observed in men and women for the high concordant or the overestimator group. In both groups, the majority of individuals reported no difficulty at both time points, as demonstrated by the IQR (0-0).

Table 7.10: Median level of the two comparable summary performance-based measures at age 53 and 60-64 in each concordant and discordant group

	Men				Women			
	N (%)	Mean (SD)			N (%)	Mean (SD)		
		Age 53	Age 60-64	P value †		Age 53	Age 60-64	P value†
Total	825	1.42 (0.39)	1.28 (0.33)	<0.001	934	1.32 (0.35)	1.23 (0.34)	<0.001
Concordant								
Low	50 (6.06)	1.09 (0.38)	0.60 (0.28)	<0.001	93 (9.96)	1.00 (0.40)	0.61 (0.30)	<0.001
Middle	602 (73.0)	1.43 (0.36)	1.30 (0.18)	<0.001	637 (68.2)	1.33 (0.31)	1.23 (0.19)	<0.001
High	76 (9.21)	1.76 (0.42)	1.85 (0.24)	0.08	68 (7.28)	1.57 (0.33)	1.73 (0.19)	0.0008
Discordant								
Underestimators	24 (2.91)	1.64 (0.32)	1.58 (0.23)	0.5	94 (10.1)	1.45 (0.38)	1.56 (0.21)	0.005
Overestimators	73 (8.85)	1.19 (0.29)	0.90 (0.16)	<0.001	41 (4.50)	1.17 (0.28)	0.91 (0.16)	<0.001

† P Value from t test comparing mean values of summary performance-based measures at age 43 and age 60-64

Table 7.11: Mean level of the two comparable summary self-reported measures at age 43 and 60-64 in each concordant and discordant group

	Men				Women			
	N (%)	Median (IQR) ‡			N (%)	Median (IQR) ‡		
		Age 43	Age 60-64	P value†		Age 43	Age 60-64	P value†
Total	860	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	<0.001	939	0.0 (0.0 – 0.0)	0.0 (0.0 – 2.75)	<0.001
Concordant								
Low	55 (6.40)	0.0 (0.0 – 0.75)	9.58 (6.33 – 14.6)	<0.001	84 (8.95)	0 (0 – 1.00)	10.3 (7.0 – 16.3)	<0.001
Middle	616 (71.6)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	<0.001	649 (69.1)	0.0 (0.0 – 0.0)	0.0 (0.0 – 2.33)	<0.001
High	83 (9.65)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	0.5	66 (7.03)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	1.00
Discordant								
Underestimators	25 (2.91)	0.0 (0.0 – 0.0)	2.75 (1.5 – 3.75)	<0.001	92 (9.80)	0.0 (0.0 – 0.0)	2.75 (2.13 – 3.75)	<0.001
Overestimators	81 (9.42)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	0.3	48 (5.11)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	0.2

‡ Low values (no reported difficulty) correspond to high levels of physical capability
† P Value from Mann-Whitney comparing values of summary self-reported measures at age 43 and age 60-64

7.4 Discussion

The following discussion has been separated into two sections so that each of the aims outlined in Section 7.1 can be addressed separately.

7.4.1 Exploring the association between markers of health and discordance

Based on findings in the literature review and gaps identified in the literature, eight markers of health status were selected to address the initial aim of this chapter: to explore the association between health and discordance using both descriptive analysis and multinomial logistic regression models.

<u>Summary of findings</u>

In the descriptive analysis the two discordant groups were observed to have distinctly different characteristics to the concordant group in terms of their experience of pain and verbal memory. The results indicated that underestimators were more likely to report at least mild pain and have higher verbal memory scores than participants in the combined concordant group. Conversely overestimators were more likely to report no pain and have lower verbal memory score than those in the concordant group. No evidence of differences between the two discordant groups and the concordant group was found for the other makers of health status, and when the two discordant groups were mapped onto the health gradients produced by the three concordant groups, no clear pattern was observed.

In the sex-adjusted models, pain and verbal memory were observed to be associated with discordance. Underestimators were more likely to report the experience of pain, and overestimators were less likely to have high verbal memory score. Evidence of a sex interaction was found for wellbeing, with a small effect evident for male underestimators and no association found for women. When mutually-adjusted, only pain remained associated with discordance.

<u>Physical health</u>

In the literature, perceived health was reported to be one of the strongest factors associated with discordance (72;129), with consistent results suggesting that individuals who reported poor health were more likely to underestimate their physical capability (72;112;129;132). It has been suggested that individuals with poor health may affiliate the physical symptoms of a chronic health condition with symptoms of functional decline, leading them to underestimate their capability (72). However, the analysis in this chapter found no significant association between self-reported health and discordance. This finding may be partially explained by the high proportion of the sample who report excellent or very good health (over 50% across all subgroups).

The presence of chronic conditions is often considered to be an important marker of ageing, and the findings in the literature provide evidence to suggest that individuals are more likely to be discordant the more chronic conditions they experience (71;121), although no direction of discordance was reported. The findings presented in this chapter did not reflect the observations from the literature, as no association was found between the presence of chronic conditions and discordance. This was particularly noticeable for stroke, which some studies have suggested was a highly important factor (121;132). However, it is important to remember when interpreting the associations of these markers of physical health with discordance that, compared to the older populations used in previous studies, at age 60-64 relatively few participants within the NSHD reported the presence of chronic conditions. For example, less than 2% of the study population reported that they had experienced a stroke compared to 7% in one of the papers cited in the literature review (132). As individuals age, they experience more chronic conditions, and this may explain why no significant findings were observed in this relatively young sample for the association between chronic conditions and discordance.

The underlying belief is that individuals with chronic conditions may equate the increased effort required to complete a certain task with experiencing difficulty (71). However, the analysis presented also found no association between levels of fatigue and discordance, providing little support to this theory.

The only factor linked to physical health that was found to be associated with discordance in mutually-adjusted models was pain. The studies reviewed in the literature produced mixed findings in relation to pain, and did not report the direction of any observed discordance (125;129). The results presented from this analysis provide additional information to the existing evidence base, and indicate that those who report pain are more likely to underestimate their physical capability. Two potential explanations have been offered for the observed discordance that may work in combination to magnify the problem. Firstly, it is possible that people will be motivated to perform well during the performance assessment knowing they only have to complete the task once, and so they ignore the pain temporarily (129). This may result in relatively high performance-based scores compared to a more realistic measure of their normal capability. Secondly, individuals experiencing pain may report difficulties more readily at high levels of physical capability because they are unable to complete the tasks in a manner that meets their expectations (129). This in turn may result in lower self-reported levels of physical capability than would be expected given their lived experience. When these two explanations operate in conjunction, pain increases the possibility of discordance by temporarily allowing high performance-based levels of physical capability whilst lowering self-reported levels of physical capability.

It is important to note that certain chronic conditions, such as arthritis, could be potential causes of pain and so explain the associations found between pain and discordance. Although arthritis was recognised as an important potential risk factor for discordance, it was not possible to include it as a factor within the analysis in this chapter as no appropriate data were available within the NSHD at age 60-64 and this should be acknowledged as a limitation of the data. However, a sensitivity analysis conducted using the best available data (presence of knee osteoarthritis collected at age 53 (167)) found no evidence of an association between knee osteoarthritis and discordance.

<u>Mental health and cognitive function</u>

Although half of the studies in the literature review (N=3) found no significant association between depression and discordance, those that did observed people with symptoms of depression were more likely to underestimate their physical capability (40;112;137). This direction of discordance seems entirely plausible given the negative perceptions such individuals are likely to have of their physical capability. Despite this persuasive argument, the results of the cross-sectional analysis reflect the majority of the previous studies, with no significant association between symptoms of depression or anxiety and discordance. It is possible that the effect of depression on self-reported measures of physical capability may be countered by the influence of depression on performance tests, which could explain the non-significance of the association. For example, whilst individuals with depression may underestimate their self-reported physical capability, it is also possible that they are unlikely to perform to their maximum performance-based capability, resulting in apparently concordant results between the two measures.

Whilst wellbeing has not been explicitly explored in previous studies, it seems logical that the association with discordance should operate in the opposite direction to that for depression and anxiety. The results of this chapter offer limited support to this hypothesis, with some evidence found that males with higher wellbeing scores were less likely to underestimate their physical capability. This does operate in the expected direction, however the effect was small and fully attenuated in the mutually-adjusted models. Further research may be required to replicate these findings and explore the association between wellbeing and discordance in more detail.

The final marker of health explored in this analysis was verbal memory as an indicator of cognitive functioning. The literature suggested that people with cognitive impairments (poor cognitive functioning) were more likely to be discordant, but there were conflicting results about the direction of this discordance (45;112;121;123;132). Individuals with low cognitive function are probably less able to accurately assess their own physical capability, but the results in the literature suggest that cognitive impairment does not appear to systematically bias individuals' opinions one way or the other. The analysis presented in this chapter suggested that those with lower levels of cognitive functioning were more likely to overestimate their physical capability, but after mutually adjusting for other indicators of health status the association became non-significant.

7.4.2 Exploring the association of participants' functional and psychological history with discordance.

To address the second aim in this chapter, five factors representing participants' functional and psychological history were selected, and the association between these factors and discordance was explored using descriptive analysis and multinomial logistic regression models.

Summary of findings

In the descriptive analysis, the two discordant groups were observed to be distinctly different to the concordant group in terms of their neuroticism, experience of depression in midlife, longitudinal self-reported capability and their longitudinal performance-based capability. The results suggest that underestimators were more likely to have high levels of neuroticism, chronic depression, declining levels of self-reported physical capability and stable high levels of performance-based physical capability. In contrast, overestimators were more likely to have low levels of neuroticism, symptoms of depression at age 60-64 only, no change in their self-reported physical capability, and either a stable low or declining performance-based physical capability. No statistically significant differences between the two discordant groups and the concordant group were observed for extraversion. When the two discordant groups were mapped onto the gradients produced by the concordant group, no clear pattern was discernible.

In sex-adjusted models, experience of depression during midlife, self-reported decline and performance based-decline in physical capability were found to be associated with discordance. Those with chronic experience of symptoms of anxiety and depression were more likely to underestimate their physical capability, whilst individuals who experienced a performance-based decline were more likely to overestimate their physical capability. In contrast those who experienced self-reported decline were less likely to overestimate their physical capability. The extended descriptive analysis indicated that participants who experienced the greatest declines in their levels of physical capability were in the low concordant group, whereas those who experienced moderate declines in their self-reported physical capability were in the underestimating group, and those who experienced moderate declines in their performance-based physical capability were in the overestimating group.

Personality

Two personality factors were included in the analysis of this chapter: extraversion and neuroticism. In line with previous research (40;112), no association was found between extraversion and discordance in this thesis. Based on previous findings in the literature (139), it was expected that those with higher levels of neuroticism would be more likely to underestimate their physical capability. The descriptive analysis presented in this chapter provides some evidence in support of this hypothesis, but the association was attenuated in the sex-adjusted models. In this sample, where 55% of women have high levels of neuroticism compared to 34% of men (see Appendix 8), sex appears to be a confounding factor for the association between neuroticism and underestimation of physical capability.

Anxiety and depression from age 53 onwards

As previously discussed in Section 7.4.1, it seems plausible that individuals with depression could potentially underestimate their physical capability given the negative perceptions such individuals are likely to have about their physical capability. When the association between depression and discordance has been examined cross-sectionally, both in this chapter and in the literature, the expected association has not been observed. However, the results presented in this chapter suggest that when the association between depression and discordance is explored longitudinally, individuals with chronic depression (i.e. symptoms of depression and anxiety at more than one time point) were more likely to underestimate their physical capability. A sensitivity analysis (see Appendix 8) was conducted using an alternative depression variable, which categorised participants into one of six life course profiles of depressive and anxious symptomology (168). The results of this sensitivity analysis support the findings reported in this chapter, with some evidence that individuals with repeated symptoms throughout adult life were more likely to underestimate their physical capability. It has been proposed in the literature that depression may influence levels of physical capability, through various behavioural pathways such as poor treatment compliance or risky health behaviours, and biological pathways that inhibit the neurological and endocrine systems (155). Individuals with chronic depression are likely to activate these pathways more frequently and persistently than individuals who experience depression for a relatively short time period, resulting in a stronger association between physical capability and depression. As the effects of depression appear to influence self-reported physical capability to a greater extent, this continued exposure to the behavioural and biological mechanisms may explain why the association between depression and discordance was only evident in the longitudinal analysis in this chapter.

<u>Physical capability</u>

Only one study has previously examined the association between an individual's physical capability history and discordance. However, the reliance on retrospective measures in this previous study restricted the ability of the analysis to account for the dynamic nature of an individual's experience of physical capability. The analysis in this chapter has extended the existing evidence base by exploring the association between participants' physical capability history and discordance using longitudinal self-reported and performance-based measures of physical capability from midlife to early old age. Although it was not possible to run the sex adjusted model for the association between self-reported decline and underestimation of physical capability, the descriptive results and the corresponding sex adjusted overestimation model, indicate the results in this thesis replicate the findings of the previous study. However, given the parameters used to define the discordant groups this association was expected, as was the association between performance-based decline and overestimation of physical capability. Further discussion on this subject will follow when discussing the methodological considerations of the approach used when analysing the longitudinal physical capability variables.

Where this thesis provides new insight is the in the extended descriptive analysis. The results presented in this chapter suggest that individuals who experience a moderate decline in their performance-based physical capability may overestimate their physical capability, and those who experience a moderate decline in their self-reported physical capability may underestimate their physical capability. It is possible that those who experience severe declines in either their self-reported or performance-based physical capability may have a greater awareness of their current physical capability due to the substantial impact this decline has had on their lives. In contrast, those who only experience a moderate decline may be more prone to discordance because they are still adjusting to their new physical capability level.

There are some methodological considerations that should be taken into account when interpreting the longitudinal physical capability findings presented in this chapter. For example, there is an element of circularity when using the longitudinal physical capability measures. In the analysis presented in this chapter, the longitudinal physical capability measures are seen as a potential factor of interest that may explain some of the observed discordance within the NSHD. However, these measures look at changes in the level of physical capability between midlife and early old age, whilst the outcome variable in this analysis (discordance) is dependent on the level of physical capability experienced at the final point in early old age. As a consequence, declining levels of self-reported physical capability are only applicable to those who underestimate their physical capability, and declining levels of performance-based physical capability are only applicable to those who overestimate their physical capability. Despite this circularity, observed declines in either measure of physical capability still provide useful information in relation to discordance, as the results presented in this chapter suggest the rate of decline differs between individuals who are concordant and those who are discordant.

When creating each longitudinal measure, only two time points were included due to the availability of data within the NSHD. It is therefore important to acknowledge the limitations of analysing change using two time points. For example, it is not possible to assume a constant shape of change for all individuals, as some may have started to change immediately after the first data collection, whilst others may have experienced a delay before their change in levels of physical capability (169). Equally, measurement error at either time point or regression to the mean between the two time points may lead to inaccurate conclusions, as each of these limitations may mask the true change in levels of physical capability (169-171). Efforts were made to try and overcome some of these challenges in the performance-based measure using a method to categorise change in physical capability previously published (166). This method enabled the identification of four categories of change, providing the ability to distinguish between individuals who have declined to low levels of physical capability by age 60-64 from those who have experienced sustained low levels of physical capability throughout later midlife. Alternative methods (172) were considered, but these could only distinguish individuals with high levels of decline, leaving a group with very heterogeneous longitudinal change for comparison. To a lesser extent the method used in this chapter also faced this challenge, with some heterogeneity observed in the majority of participants within the "maintenance" group.

Given the substantial skew of the two self-reported summary measures used to create the longitudinal self-reported measure, it was not possible to account for the limitations of two time points in the same way as the longitudinal performance-based measure. However, the 20-year gap between data collection points improves the plausibility that any change observed is meaningful in the context of healthy ageing, as any reported difficulty at age 43 is unlikely to be age-related decline in physical capability. Unlike some papers in the literature that focus on the incidence of self-reported limitations, the change observed between the two time points in this thesis also accounted for the progression of mild to severe difficulty, which may be useful information when implementing policy to alleviate the impact of declining levels of physical capability (173).

Ideally, to address the methodological considerations discussed, and to explore the association between individuals' longitudinal physical capability and discordance in more detail, the analysis should be repeated using more than two time points. This may require an alternative dataset or repeating the analysis following a further round of data collection in the NSHD. It may also be prudent to analyse the association between individuals' physical capability and discordance using data collected at more frequent time periods to capture the longitudinal change in more detail.

7.4.3 Conclusion

The evidence presented in this chapter has continued to extend the existing evidence base of factors associated cross-sectionally with discordance, by providing more detail about the direction in which discordance operates for factors such as pain, and by documenting the association between discordance and factors such as wellbeing which have not been previously explored.

There were no meaningful differences between the two discordant groups and the concordant group in terms of their markers of health status at age 60-64, with the exception of pain. The findings reported in this chapter suggest that individuals who experience pain may underestimate their physical capability, and these individuals could be suitable candidates for an intervention to reduce the impact of age-related declines in physical capability.

The conceptual diagram proposed at the beginning of this chapter described how past psychological and functional history could influence discordance in early old age. The association between longitudinal factors and discordance has not been previously documented, so the results presented from this analysis go some way to addressing one of the major gaps identified in the discordance literature. Within this chapter, some evidence was found to support the conceptual diagram, with long-term mental health and the two longitudinal physical capability measures demonstrating an association with discordance.

The following chapter will provide a more detailed discussion of the results presented in this chapter, outlining the implications for these results and the key themes presented throughout the thesis.

8. Conclusions

8.1 Introduction

The focus of this thesis has been on physical capability in early old age, with a particular emphasis on discordance between self-reported and performance-based measures of physical capability. By identifying the factors associated with discordance, it is possible that interventions can be developed to minimise the impact of age related declines in physical capability for individuals.

Having explained the importance of maintaining physical capability into old age, especially in the context of global population ageing, Chapter 1 described the two sets of measures commonly used to assess physical capability, and the attributes of these measures. Chapter 2 outlined the theoretical constructs underpinning this thesis, by critically reviewing the existing conceptual frameworks within the literature and clarifying the terminology used to describe the relevant concepts, distinguishing between physical capability and disability. Using self-reported and performance-based measures of physical capability, Chapter 3 described the levels of physical capability observed in the NSHD population at age 60-64. The results presented suggest that, within the study population, levels of physical capability varied widely across the whole spectrum of ability, with many participants reporting and performing to relatively high levels of physical capability in early old age, whilst others had low levels of physical capability.

Chapter 4 investigated the hierarchical order with which participants first report the experience of difficulty with the tasks of daily living. These findings informed the selection of self-reported measures used to create a summary self-reported measure of physical capability at age 60-64. Chapter 5 used this summary self-reported measure and compared it with a summary performance-based measure to identify three concordant and two discordant groups, as outlined *a priori* in Figure 1.1. Chapters 6 and 7 then examined the associations between discordance and a variety of factors identified *a priori*, based on the findings of a comprehensive literature review, to identify potential determinants of discordance. Comparisons were drawn between the characteristics of the combined concordant group and of participants in the two discordant groups to ascertain the direction and magnitude of the association with discordance.

The aim of this chapter is to conclude this thesis by reviewing the findings and key themes, formulating potential policy implications and addressing the strengths and limitations of the methods used.

8.2 Discordance between self-reported and performance-based measures of physical capability

It is important to focus on discordance in early old age because of the conceivable implications that underestimation or overestimation of physical capability may have, for both the individual in terms of their quality of life, and broader society through the increased financial burden placed on the health and social care sector as a result of inaccurately assessing physical capability. For example, individuals who overestimate their physical capability without making suitable adaptations or developing coping strategies may be at risk of future health complications if they continue to perceive relatively high levels of physical capability despite their poor performance (137;157). In contrast, individuals who underestimate their physical capability may be relying on assistive devices and personal help more than necessary (137), when appropriate interventions could enable them to appreciate their potential capability, resulting in an improved quality of life with full engagement in the activities of daily living required to maintain independence (138).

8.2.1 Potential explanations for the observed discordance

It is important to consider whether the observed discordance is: 1) meaningful; 2) an artefact of error in the measurement of self-reported or performance-based measures; or 3) an unintended consequence of the method used to identify discordance.

One explanation for discordance that has been proposed in the literature is that self-reported and performance-based measures may reflect different concepts within the scope of functional ability, with self-reported measures akin to disability and performance-based measures conceptually closer to physical capability (112;139). However, in this thesis considerable care was taken when selecting the self-reported and performance-based measures to ensure that all measures used when defining discordance met the conceptual definition of physical capability outlined in Section 2.2.1. It is therefore unlikely that the discordance observed in this thesis is due to the self-reported and performance-based measures relating to entirely different constructs of functional ability.

Although it is unlikely that the self-reported and performance-based measures used in this thesis reflect entirely different concepts, it is plausible that the attributes of each measure (discussed in Section 1.2) could result in self-reported and performance-based measures assessing slightly different elements of the same concept. The potential impact of these different attributes on the level of discordance observed in any study was highlighted as part of the critical review of the discordance literature in Section 6.1.3. Of particular note are three attributes: the influence of time (45;112;118), the inclusion of adaptations (40;45;112;174) and the potential hypothetical nature of self-reported measures (45;137).

The first attribute to be discussed is the influence of time. An individual's experience of physical capability throughout their life is dynamic, with temporary fluctuations in levels of physical capability caused by illness or injury that act largely independently of age-related declines later in life. Performance-based measures assess physical capability at a specific point in time, and this may not be an accurate assessment of an individual's overall physical capability. For example, if an individual is experiencing temporary incapacity due to injury or illness, the performance-based results will indicate that they have low levels of physical capability. In contrast, the individual's self-reported assessment of their physical capability may reflect their usual capability in recent weeks, indicating higher levels of physical capability. In this example the individual would be classified as an overestimator. However, it is possible that if the performance-based assessment were repeated a week later, and if the individual had recovered from their temporary incapacity, they would then be concordant. Unless specific questions are asked to ascertain whether the individual feels their physical capability on the day of assessment differs to that normally experienced, the temporal restraints of self-reported and performance-based measures will remain a potential explanation for any observed discordance.

The second attribute of self-reported and performance-based measures to be considered is the inclusion of adaptations. When asked to self-report their physical capability, it has been suggested that individuals consider how they complete the task in question within their daily lives, and thus incorporate the use of assistive devices or other coping strategies in their perception of their current physical capability levels. The protocol used for many performance-based measures prohibits the use of aids, which may explain why some individuals struggle with the task when following the strict protocol of the performance-based assessment but overestimate their capability, as they do not perceive that they have any difficulty normally when completing the task. For example, when getting out of a chair, many people use the armrest for support, in an action that is almost subconscious, but when completing the chair rise test for the performance-based assessment participants in the NSHD were required to fold their arms (see Section 3.2.1), which may make the manoeuvre more challenging. However, in the NSHD the exception to this was the TUG assessment, where assistive devices were allowed (see Section 3.2.1). If an individual's physical capability was low enough that they required an aid to walk the six metres necessary to complete the task, it is probable that such an individual would have an accurate assessment of their lower physical capability and would be classified as concordant, even if the protocol prohibited assistive devices. This may explain why the initial analyses in Chapter 5 (see Section 5.3.1) found the lowest level of discordance between TUG and self-reported walking when compared to any other specifically matched tasks. The proportion who reported no difficulty despite being unable to complete the performance tests varied between 0.1-0.3% for TUG and self-reported walking, compared to a range of 0.3-0.8% for the other matched tasks.

The third attribute of the physical capability measures that could influence the level of observed discordance is the potentially hypothetical nature of self-reported measures. It is plausible that some individuals may not have had much recent opportunity, or need, to complete certain physical capability tasks, such as walking as far as a quarter of a mile or climbing stairs. In such circumstances, the self-reported assessment of physical capability for these individuals may be based on what they believe they would be capable of, rather than reflecting on any recent experience. In contrast, performance-based measures of physical capability involve the active completion of a specific task as part of the assessment, and therefore by nature cannot be hypothetical. It is also worth noting that performance-based measures focus on a particular action in a prescribed format, whereas each individual may have a different approach to completing the specified physical capability task when free from the standardised protocol, and so may assess their physical capability differently in self-reported measures. The greater the number of alternative approaches that can be taken to complete the physical capability task in everyday life, the higher the potential for discordance, as it becomes more challenging for individuals to assess the hypothetical situation. Consequently, the likelihood of a difference between the self-reported and performance-based assessment of physical capability increases.

The difference in the hypothetical nature of assessment between self-reported and performance-based measures explains why it is important to be consistent in the ordering of physical capability assessments. For example, if the performance-based test is administered first, evidence suggests the level of observed discordance tends to be lower (129;138), as individuals may incorporate their recent experience of performance-based capability into their self-reported perception of their current physical capability levels. In the NSHD, all participants completed the self-reported measures before completing the performance-based assessment (as outlined in Section 3.2.2), so the observed discordance may reflect the hypothetical nature of the self-reported tasks; however, at age 60-64, when the majority of individuals were independent and actively involved with most tasks of daily living, it is unlikely that many of the participants based their self-reported perception of physical capability on a hypothetical situation.

In addition to the attributes of the self-reported and performance-based measures of physical capability, it is also possible that the observed discordance may be the result of the process used to identify discordance. Given the relatively high level of correlation between self-reported and performance-based measures of physical capability quoted in the literature (36) and observed in this thesis (r=0.6), it may be reasonable to assume that some of the discordance observed could be due to measurement error in either set of physical capability measures (121). However, steps were taken to minimise the effect of measurement error in this thesis.

To maximise reproducibility, standardised protocols were used when assessing performance-based physical capability, as outlined in Section 3.2.1. Nurses were trained to follow this protocol in order to reduce intra and inter operator variability. Evidence in the literature suggests that all four performance-based measures of physical capability used in this thesis have high levels of reliability (175-178). The self-reported measures used in this thesis are based on a questionnaire designed by the Office of Population Censuses and Surveys (OPCS), which has proven validity (82). When using similar questions, evidence in the literature found each self-reported measure demonstrated high levels of consistency, which was not influenced by age or baseline levels of physical capability (179). At age 60-64 nurses received training regarding the administration of this questionnaire to reduce intra and inter operator variability. However, as highlighted in Section 3.2.2, it was not possible to ensure that the correct gating procedure was enforced throughout the data collection. As a consequence, the data obtained from the self-reported measures were cleaned to improve validity.

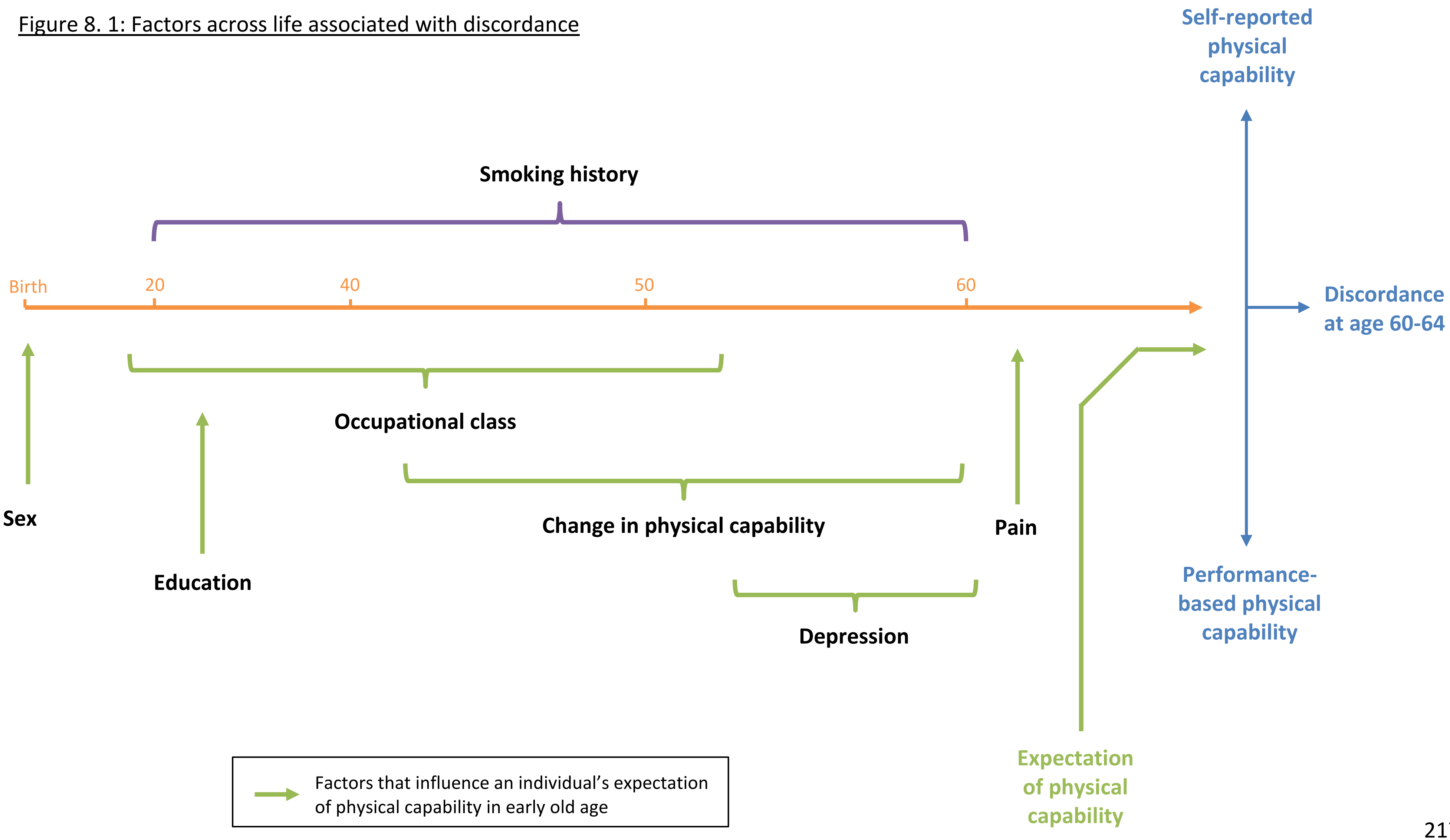
As previously discussed in Section 1.2, performance-based measures are often thought to have better validity than self-reported measures (34;36;44). However, it is plausible that both measures are unable to fully capture the concept of physical capability. For example, both self-reported and performance-based measures are known to have floor and ceiling effects (35;38), and these limitations prevent the assessment of physical capability across the whole spectrum of ability, as suggested by the conceptual definition of physical capability (see Section 2.2). Measures with floor or ceiling effects are unable to detect any variations among individuals below (or above) a certain threshold, whereas measures without these limitations would be able to differentiate these individuals, thus introducing the potential for discordance. In this thesis some evidence of a floor effect for the balance performance-based measure was observed (see Section 3.3.2), and among the self-reported measures there was an evident ceiling effect. As a consequence, information about individuals' levels of physical capability may have been lost at opposing ends of the spectrum for the two measures of physical capability, contributing to the observed discordance. Despite the steps taken in this thesis and by other researchers (139) to try and overcome the inevitable skew introduced by floor and ceiling effects, it is probable that some of the observed discordance may be due to the lack of sensitivity in one or other of the physical capability measures used when categorising discordant individuals.

8.2.2 Factors associated with discordance

Whilst the attributes of the self-reported and performance-based measures of physical capability and potential measurement error may contribute to some of the observed discordance, this thesis has identified a number of factors across the life course which could meaningfully influence levels of discordance. Of the factors identified in this thesis, sex was found to be most strongly associated with discordance, with men more likely to overestimate and women more likely to underestimate their physical capability. This finding is consistent with the literature (116;125;129;131), and emphasises the importance of allowing for sex when assessing physical capability. Several of the factors explored in this thesis, including marriage and extraversion, demonstrated no evidence of an association with discordance, reflecting the results previously shown in the literature (40;112;121). However, in this thesis it has been possible to extend the existing evidence base by clarifying the conflicting and previously unexplored direction of association for factors such as education and pain. The analysis in this thesis also included some factors whose association with discordance had not been previously examined. In some cases, for example fatigue and wellbeing, no evidence of an association was observed, but for occupational class and smoking history the results presented show strong evidence of an association with discordance and extend the existing evidence base. All of the previous studies identified in the literature review were cross-sectional in design, and thus limited in their ability to investigate how factors across life might influence discordance. An attempt was made in this thesis to address this gap in the literature by exploring the association between longitudinal factors, such as chronic depression and discordance. It should be acknowledged that some of the factors included in the analysis throughout this thesis, such as smoking, have been collected across life, so these may also contribute to addressing the gap in the literature.

Details about how each of the factors identified in this thesis could influence discordance have previously been outlined in Sections 6.4 and 7.4. From this discussion a new theory emerges, which suggests that, in addition to their direct influence on discordance, these factors could influence discordance indirectly through an individual's expectation of physical capability in early old age. A conceptual diagram to illustrate this emerging theory is shown in Figure 8.1.

Figure 8. 1: Factors across life associated with discordance



It has been suggested in the literature that an individual's perception of their physical capability may be the product of their prior physical capability levels and their expectation of future physical capability levels (129). Discordance could potentially result from the difference between reality and an individual's expectation of physical capability, based on an internal reference point. This internal reference point for physical capability may operate in a similar manner to that proposed for overall health, whereby individuals with similar levels of health may each appraise their health differently based on different internal reference points which are influenced by socio-demographic factors and their life experience (152). For example, individuals with higher levels of education are more likely to take in health messages throughout life, and therefore have aspirations to maintain relatively high levels of physical capability in early old age. Consequently, such individuals may be sensitive to more subtle declines in their physical capability, as they are no longer able to meet their aspirations, and so underestimate their capability. However, it is important to note that expectations not only influence an individual's self-reported assessment of their physical capability, but could also affect their performance-based physical capability. For example, when a group of older adults focused on their negative expectations of ageing (by comparing their physical capability with that of younger individuals), up to 50% reductions in grip strength were observed (180). Evidence in the literature suggests that expectations can be self-fulfilling, because they can alter an individual's behaviours, and as a result modify their future health outcomes (181;182). Using the example of individuals with high education levels, this could mean such individuals invest in frequent physical activity to help maintain relatively high levels of performance-based physical capability.

8.3 Implications

Evidence has been shown in this thesis for an association of several factors across life with discordance. Using these factors, it may be possible to predict discordant individuals when only one type of physical capability measure is used in assessment, based on the characteristic differences observed between the concordant and discordant groups.

Sex was observed to be the strongest factor associated with discordance in this thesis. Therefore, it is important to account for the sex of individuals when assessing their physical capability, as women are more likely to underestimate their physical capability and men are more likely to overestimate their physical capability. The results presented in this thesis for the association between other factors of interest and discordance accounted for sex. Whilst some of the associations were attenuated by this adjustment, those that remained should be considered when assessing physical capability using either self-reported or performance-based measures.

The results presented in Chapter 6 suggested that, in terms of their socio-demographic and behavioural risk factors, overestimators appear to align with those in the low capability concordant group and underestimators align with those in the high capability concordant group. This alignment can be used to aid the identification of discordant individuals. For example, amongst individuals with low self-reported levels of physical capability, a high educational level could potentially be indicative of someone who underestimates their capability, because this socio-demographic characteristic is usually associated with individuals in the high concordant group.

In Chapter 7, no discernible differences were detected between the concordant groups and the two discordant groups in terms of their markers of health status, with the exception of pain and chronic depression. For both of these factors, participants who underestimated their physical capability were aligned with the characteristics of the low concordant group, whilst participants who overestimated their capability were aligned with the high concordant group. In a similar manner to the process described for the sociodemographic factors, these characteristic differences can be used to help identify discordant individuals. However, pain and chronic depression should be used in combination with performance-based measures of physical capability to identify discordant individuals, as it is not possible to distinguish between those with low levels of self-reported physical capability and individuals who underestimate their physical capability.

With respect to the categories of physical capability decline, it is important to note that each measure of physical capability can only be used to indicate one direction of discordance, given the parameters used to identify each discordant group. For example, underestimators by definition have high levels of performance-based capability, so a decline in levels of performance-based physical capability would not be applicable to individuals in this group. Despite this limitation, individuals' category of physical capability decline could be used to predict discordant individuals when only one type of physical capability measure is available, if the direction of discordance and the appropriate physical capability measure are used in combination.

Once discordant individuals are identified, such individuals may require interventions to help minimise the impact of age-related declines in physical capability. Underestimators have maintained relatively high levels of physical capability, and suitable interventions could be used to help these individuals achieve their potential capability. For example, individuals who underestimate their physical capability due to pain may require pain relief or adaptations developed with an occupational therapist to successfully complete the tasks of daily living. By enabling individuals who underestimate their physical capability to realise their potential capability, interventions may encourage such individuals to fully engage with the physical activities of daily life, improving their quality of life and supporting their independence. When healthcare professionals assess the care needs of individuals in old age, the assessments are often based on self-reported measures of physical capability. However, following this procedure, individuals who overestimate their physical capability, and potentially have an unmet need for care, are not identified. Unfortunately, this thesis was unable to find any factors specifically associated with an increased likelihood of overestimation of discordance, so based on the findings presented it is not possible to identify individuals who overestimate their physical capability when using the self-reported measure in isolation. If performance-based measures were introduced to the assessment of care needs, overestimators could be identified and interventions applied to ensure these individuals received the appropriate care required to maintain independent living. In addition, performance-based measures would aid the identification of discordant individuals who underestimate their physical capability due to pain or chronic depression.

It is important to recognise that many of the existing interventions to minimise the impact of age-related declines in physical capability are implemented in old age, when such interventions may be less effective (96). The results presented in Chapter 4 suggest that interventions should not be delayed until old age, as high risk individuals may be identified in midlife at a point when the reversal or prevention of further decline may be possible (71;183). By early old age, some individuals have already experienced substantial decline in functional ability, and as the population ages this proportion is likely to increase unless preventative action is taken.

8.4 Methodological considerations

When interpreting the results presented in this thesis, it is important to account for several methodological considerations raised throughout the analytical procedure.

8.4.1 The NSHD dataset

One of the first features to consider is the choice of dataset within this thesis. As previously highlighted in Section 3.1.2, there are numerous benefits to using the NSHD when exploring concordance and discordance between self-reported and performance-based measures of physical capability in early old age. The socio-demographic characteristics of participants in the NSHD broadly correspond to those observed in the general population amongst individuals born in the UK of a similar age (76), suggesting that the results found in this thesis should be applicable to adults in early old age in Britain.

It should be noted that the inclusion of home visits at age 60-64 helped to reduce the bias introduced by the health and social disadvantage of those unable to attend the Clinical Research Facilities. However, despite this, the mean levels of health of participants within the NSHD is generally better than that in the general population, as demonstrated by the lower prevalence of self-reported long term limiting illness at age 60-64 in the NSHD, when compared with those of a similar age using 2001 census data (26% vs 34%) (76). Throughout this thesis steps were taken to include those who were unable to complete the performance-based physical capability tests for health reasons, to minimise any additional health bias introduced through incomplete data.

When the NSHD study was set up in 1946, the sampling framework only included single, legitimate births from England, Scotland and Wales (74), which means that children born to unmarried women and people who have moved to the UK since birth have not been incorporated into the dataset. Although this selection criterion may be appropriate for adults who are currently aged 60-64, given the small proportion of children born to unmarried women in 1946 and the small proportion of ethnic minority individuals amongst this age group (76), future cohorts may have different socio-demographic characteristics. This may limit the generalisability of the results to future cohorts, given the increasing ethnic diversity of the population and changing family structures. However, the findings presented in this thesis often reflect those previously observed in the literature, so similar associations between the factors of interest and discordance may be detected in future British cohorts.

One of the generic advantages of the NSHD emphasised in Section 3.2.1 was the inclusion of a home visit at age 60-64, to enable the retention of participants who could not attend the clinical research facilities (CRFs). Whilst this has obvious positive implications for response rates and the limitation of bias introduced through the social and health disadvantage of those who could not attend the CRFs, there are some negative implications that apply specifically to this thesis. When examining patterns of missing data for the performance-based physical capability measures, it was noted that individuals who participated in a home visit were more likely to be missing the TUG assessment compared to those who had attended a CRF. The explanation for this pattern of missing data is the limitation of space in participants' houses. The TUG assessment requires a space more than three metres in length, which could be set aside in the CRF, but may have been challenging to find within the confines of a participant's home. As the data for the TUG assessment was not missing at random, it was not possible to use multiple imputation when trying to increase the sample size for the summary performance-based measure.

As with all cohort studies, the NSHD has experienced attrition since its inception, but as previously discussed, the sample remains broadly representative of the general population. Figure 3.1 illustrated how the maximum sample was obtained for each data collection round, resulting in a sample of 1,930 participants with suitable physical capability data available at age 60-64, from a possible 2,229 participants. Where possible, data were imputed for the two summary physical capability measures produced in this thesis to increase the sample size. The sensitivity analysis conducted in Section 5.3.4 implied that no substantial bias was introduced during the imputation process, but the sample size increased by 2.6%, potentially providing more power to later analyses. The analyses in this thesis have focused on complete cases, excluding participants with data missing for at least one covariant. When the socio-demographic and health characteristics of participants included in the analysis of this thesis (complete cases) were compared to those with missing data (see Appendix 9), it was noted that complete cases were more likely to be female, have higher educational levels and a higher occupational class, have never smoked, and have symptoms of depression at age 60-64 or chronic depression. As these characteristics are associated with an increased likelihood of underestimation of physical capability, it is possible that some bias may have been introduced through the exclusion of participants with missing data for both physical capability variables. Throughout Chapters 6 and 7, further sensitivity analyses were conducted to examine whether participants excluded from the mutually-adjusted models due to missing data on covariates differed from those included in each of the sex-adjusted models. In general, no differences were observed, suggesting the attenuation of the association between certain factors and discordance prior to mutual adjustment may be a consequence of low statistical power.

Low statistical power was likely to be a limitation within this thesis, as demonstrated when the observed differences in proportions and means between the concordant and discordant groups (reported in Sections 6.3.1, 7.3.1 and 7.3.4) are compared to those used to estimate statistical power (see Section 5.3.6). For example, the proportion of participants, at age 60-64, who reported no symptoms of anxiety and depression was 83% in the concordant group and 77% in the underestimator discordant group. When similar figures were used in the power calculations the estimated power was only 30%. Given the sample size restrictions, determined by the fixed size of the total sample and the a priori definitions used to identify the concordant and discordant groups, the power calculations indicate that there was insufficient power to detect some of the small differences in the distributions of some of the factors, observed between the concordant and two discordant groups.

8.4.2 Production of summary physical capability measures

In order to identify concordance and discordance between the self-reported and performance-based measures of physical capability within the NSHD at age 60-64, two summary physical capability measures were created (see Section 5.2.2). The methodological considerations of each of these measures have previously been discussed in Section 5.4.2. Summary measures of physical capability were chosen in this thesis, rather than a focus on individual physical capability tasks, for several reasons.

Firstly, by creating summary measures it is possible to incorporate different tasks into one measure, and each of these tasks may capture a different element of physical capability. This approach provides a more detailed picture of an individual's overall physical capability level, which is required to complete the physical tasks of daily living, and increases the chance that any impact of age-related decline is successfully captured (116).

The second reason for choosing summary measures of physical capability in this thesis is from a methodological perspective. The self-reported measures of physical capability used in the NSHD are categorical, and this places certain constraints on the potential analytical approaches available when the data are used in conjunction with the continuous data produced from the performance-based measures. The creation of a summary self-reported variable enabled the categorical self-reported data to be combined into a continuous variable, which was then more comparable to the performance-based data.

It should be acknowledged that use of individual physical capability measures when exploring discordance has the benefit of a targeted approach, which may be useful in a clinical context. However, because of the limitations associated with this approach it was decided that this outweighed the potential benefit for this thesis. When individual physical capability measures are used to assess discordance in a particular task, it is crucial that the selection of the self-reported and performance-based measures is carefully considered. Otherwise any observed discordance may be an artefact of the difference between the self-reported and performance-based measures selected, rather than meaningful discordance. For example, some individuals may underestimate their capability due to the greater endurance required in a self-reported measure compared to the corresponding performance-based measure, such as a reflection on their ability to walk 400 metres in the self-reported measure versus the three to six metres usually required to complete the gait speed performance-based test.

It is also important to note that, even when the two measures are carefully aligned, performance-based measures focus on a specific aspect of function within the physical capability task, whereas self-reported measures are more complex, given they are situated within the context of everyday life. For example, the grip strength performance-based measure assesses a very specific movement, whilst the corresponding self-reported measure assesses the ability of an individual to use this movement as part of a physical task of daily living, for example holding a heavy kettle.

8.4.3 Methods for identifying concordance and discordance

The literature review in Section 5.1 identified a number of different methodological approaches used in the literature to analyse concordance and discordance. The method used depended on the research question and the type of physical capability measures available. Initially, this thesis used a cross-tabulation approach to compare the mean (or median) performance-based values and the proportion of participants unable to complete the performance-based tasks for health reasons, across the response categories of each self-reported measure. The stepwise relationships observed indicated that the self-reported and performance-based measures of physical capability at age 60-64 were largely concordant, but evidence of discordance was also observed.

Having established evidence of discordance in the NSHD at age 60-64, an alternative method was required to classify participants as concordant or discordant, in order to explore the association between factors of interest and discordance. Many of the studies reviewed in the discordance literature review (see Section 6.1) categorised both the self-reported and performance-based measures to help identify discordant individuals, and used regression analyses to investigate factors associated with discordance. A similar approach was taken in this thesis, although the categorisation of the self-reported and performance-based measures was developed to reflect the conceptual diagram shown in Figure 1.1. Conventionally, discordant groups have been identified by calculating the percentage disagreement between the two measures of physical capability or the proportion of false positives and false negatives produced when the performance-based measures were taken as the "true value". Neither of these approaches account for the variability experienced in levels of physical capability within the concordant and discordant groups. As a result, reasonably minor differences between self-reported and performance-based measures experienced within the normal range of physical capability could be classified as discordant when these differences are potentially not meaningful. Consequently, to avoid such misclassification, in this thesis all participants in the middle range of the self-reported and performance-based measures of physical capability were categorised as concordant.

Deciles were used to categorise the continuous summary self-reported and performance-based measures of physical capability, with the high concordant, low concordant and two discordant groups defined based on these deciles. The definition of each group was clearly stated *a priori,* and followed a logical division of the sample population, accounting for the skewed nature of the self-reported summary measure. By design, roughly equal numbers of participants were categorised into each of the discordant groups, which prevents any discussion about which direction of discordance was more prevalent within the NSHD.

It is important to note that the grouping of individuals outlined in this thesis is only one way of defining the five concordant and discordant groups, because there are no distinct cut points between each group (i.e. those chosen are arbitrary). A sensitivity analysis was conducted (see Appendix 9) to explore the potential impact of alternative groupings. Given the distribution of the sample population, it was not possible to assign equal numbers of participants to each discordant group in the alternative groupings, with more individuals categorised into the overestimator group compared to the underestimator group. The two alternative groupings have more power than the original grouping, but they also have less specificity, as they have a greater chance of including concordant individuals within the two discordant groups.

To investigate whether alternative groupings would have influenced the results presented in this thesis, the mutually-adjusted model from Chapter 6 was repeated. Regardless of the parameters used to define the concordant and discordant groups, the association between sex and discordance remained constant and strong. Occupational class and smoking were both consistently associated with discordance across the different groupings, although the magnitude of the association was greater in the original analysis. Interestingly the association between physical activity and discordance, which was attenuated in the original mutually-adjusted model, remained statistically significant when the two alternative groupings were used. The original parameters used in this thesis to define the concordant and discordant groups focused on individuals at the edge of the spectrum for self-reported and performance-based measures of physical capability. The level of concordance or discordance experienced at the edge of the spectrum is likely to be higher than that observed in the middle, which could explain the stronger associations observed in the original grouping, compared to the two alternative groupings.

Approximately 140 participants were categorised into each of the discordant groups based on the original parameters used in this thesis. With relatively small numbers in each of the discordant groups, the analysis presented may have been affected by limitations of low statistical power. This may have led to an inability to detect small but potentially meaningful associations; for example physical activity, which was found to be a statistically significant factor when alternative groupings were used. The limitations of statistical power meant that it was not possible to investigate the interdependence of factors in detail, but it is important to acknowledge that the factors identified within this thesis are unlikely to operate independently.

Data-driven methods, such as K-means clustering, were considered as an alternative way to categorise the two summary variables and identify discordance. However, these methods were not considered appropriate in this thesis given the highly skewed nature of the summary self-reported measure.

8.5 Future work

Whilst the results presented in this thesis have extended the existing evidence base when exploring factors associated with discordance, there are several avenues of research which require further work in either the NSHD or other datasets.

Firstly, it is important that the findings in this study are replicated using data from other British populations and at different ages, to establish whether the findings are generalisable to the wider population in this country. The use of alternative data sources may also help overcome some of the limitations experienced in this thesis, particularly with regard to the summary self-reported physical capability measure. At age 60-64, many of the participants within the NSHD reported no difficulty with the five physical capability tasks assessed using the self-reported measures, resulting in a skewed summary measure. Individuals only report difficulty with a certain task when they feel their levels of physical capability have declined below a certain threshold (184). In order to differentiate those above this threshold, and reduce the skewed distribution of self-reported physical capability, it may be necessary to ask individuals whether they have modified the way they complete the physical tasks of daily living. Evidence in the literature suggests that individuals who modify tasks without perceiving that they experience any difficulty may represent an intermediary group between those who report difficulty and those who do not (184). In datasets where questions about modification are used to differentiate varying levels of capability within the group who initially report no difficulty, it is possible that a less skewed distribution of self-reported physical capability would be observed. This would enable the development of analytical approaches that were not considered suitable in this thesis due to the skewed nature of the self-reported summary measure. In light of this consideration, questions about the modification of tasks have been included in the current round of data collection in the NSHD, when participants are aged 68-70. Once this data collection is complete, it may therefore be appropriate to repeat the analysis presented in this thesis using the newly collected data to address the replication of results at different ages, and to overcome some of the methodological limitations experienced in this thesis.

Secondly, there are a number of potentially interesting factors which may be associated with discordance that have yet to be investigated. For example, it was initially hypothesised that individuals who overestimated their physical capability may demonstrate resilience, but it was beyond the scope of this thesis to examine this factor in more detail. A recent systematic review of physical resilience (185) highlighted the lack of a conceptual definition and measures available within the literature to assess physical resilience. The authors proposed a conceptual model of physical resilience, whereby physical resilience was perceived to be a characteristic of an individual that influences their ability to maintain or recover levels of physical capability following the effects of a stressor (185). Although further work is required to establish measures of physical resilience, in the context of healthy ageing it is a key factor of interest which may help address the underlying aim of discordance research: to minimise the impact of age-related declines in physical capability.

Another potential factor of interest that has yet to be examined within the literature is the influence of prior discordance. Within the papers reviewed for this thesis, no study had explored discordance at more than one age within the same sample. It is possible that discordance at one age may increase the probability of an individual being discordant at a later age. The final factor of interest that could be important to future work is individuals' attitudes to ageing. By exploring the association between individuals' attitudes to ageing and discordance, it may be feasible to determine more precisely the extent to which individuals' expectations influence the level of observed discordance, as proposed in Figure 8.1. This analysis could be conducted using data within the NSHD, as questions pertaining to individuals' attitudes to ageing have been included in the new data collection round at age 68-70.

The final avenue for future research is to investigate the impact of discordance on future physical capability trajectories and health outcomes. In many of the studies reviewed the potential negative consequences of discordance are discussed, but there is limited evidence to demonstrate these outcomes. Such research could strengthen the relative importance of discordance and the value of identifying the two discordant groups. Such research may also provide some evidence of resilience amongst individuals who overestimate their physical capability, if these individuals experience less severe age-related declines compared to concordant individuals.

8.6 Concluding remarks

This thesis has focused on discordance between self-reported and performance-based measures of physical capability in early old age within a nationally representative British birth cohort. The findings presented suggest that a range of factors across life, including sex, occupational class and the presence of pain, are associated with discordance. These factors should be taken into account when either self-reported or performance-based measures are used in isolation to assess physical capability, to ensure all those in early old age with needs related to their physical capability are identified.

References

1. Cooper R, Hardy R, Aihie Sayer A, Ben-Shlomo Y, Birnie K, Cooper C, et al. Age and gender differences in physical capability levels from mid-life onwards: the harmonisation and meta-analysis of data from eight UK cohort studies. PLoS One 2011;6(11):e27899.

2. Cooper R, Hardy R, Aihie Sayer A. A life course approach to physical capability. In: Kuh D, Cooper R, Hardy R, Richards M, Ben-Shlomo Y, editors. A Life Course Approach to Healthy Ageing. Oxford, UK: Oxford University Press; 2014. p.16-31

3. Riley JC. Estimates of Regional and Global Life Expectancy, 1800-2001. Pop Dev Rev 2005;31(3):537-543.

4. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. Lancet 2009;374(9696):1196-1208.

5. Sander M, Oxlund B, Jespersen A, Krasnik A, Mortensen EL, Westendorp RG, et al. The challenges of human population ageing. Age Ageing 2015;44(2):185-187.

6. Rowland DT. Demographic Methods and Concepts. Oxford, UK: Oxford University Press; 2003.

7. Office for National Statistics. Population Ageing in the United Kingdom, its Constituent Countries and the European Union. 2012.

8. Lloyd-Sherlock P, McKee M, Ebrahim S, Gorman M, Greengross S, Prince M, et al. Population ageing and health. Lancet 2012;379(9823):1295-1296.

9. Rowe JW, Kahn RL. Successful Aging 2.0: Conceptual Expansions for the 21st Century. J Gerontol B Psychol Sci Soc Sci 2015;70(4):593-596.

10. Robinson S, Briggs R, O'Neill D. Cognitive aging, geriatrics textbooks, and unintentional ageism. J Am Geriatr Soc 2012;60(11):2183-2185.

11. Beard JR, Bloom DE. Towards a comprehensive public health response to population ageing. Lancet 2015;385(9968):658-661.

12. Ailshire JA, Beltran-Sanchez H, Crimmins EM. Becoming centenarians: disease and functioning trajectories of older US Adults as they survive to 100. J Gerontol A Biol Sci Med Sci 2015;70(2):193-201.

13. Lowsky DJ, Olshansky SJ, Bhattacharya J, Goldman DP. Heterogeneity in Healthy Aging. J Gerontol A Biol Sci Med Sci 2014;69(6):640-649.

14. O'Neill D. The art of the demographic dividend. Lancet 2011;377(9780):1828-1829.

15. Chatterji S, Byles J, Cutler D, Seeman T, Verdes E. Health, functioning, and disability in older adults--present status and future implications. Lancet 2015;385(9967):563-575.

16. Christensen K, Thinggaard M, Oksuzyan A, Steenstrup T, Andersen-Ranberg K, Jeune B, et al. Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. Lancet 2013;382(9903):1507-1513.

17. Falk H, Johansson L, Ostling S, Thogersen AK, Staun M, Host DL, et al. Functional disability and ability 75-year-olds: a comparison of two Swedish cohorts born 30 years apart. Age Ageing 2014;43(5):636-641.

18. Liang Y, Song A, Du S, Guralnik JM, Qiu C. Trends in disability in activities of daily living among Chinese older adults, 1997-2006: the China Health and Nutrition Survey. J Gerontol A Biol Sci Med Sci 2015;70(6):739-745.

19. Jagger C, Matthews FE, Wohland P, Fouweather T, Stephan BC, Robinson L, et al. A comparison of health expectancies over two decades in England: results of the Cognitive Function and Ageing Study I and II. Lancet 2016;387(10020):779-786.

20. Select Committee on Public Service and Demographic Change. Ready for Ageing? House of Lords; 2013. Report No.: HL Paper 140.

21. Keevil VL, Hayat S, Dalzell N, Moore S, Bhaniani A, Luben R, et al. The physical capability of community-based men and women from a British cohort: the European Prospective Investigation into Cancer (EPIC)-Norfolk study. BMC Geriatr 2013;13(1):93.

22. Murray ET, Hardy R, Strand BH, Cooper R, Guralnik J, Kuh D. Gender and life course occupational social class differences in trajectories of functional limiations in midlife: findings from the 1946 British Birth Cohort. J Gerontol A Biol Sci Med Sci 2011;66(12):1350-1359.

23. Bowling A, Dieppe P. What is successful aging and who should define it? BMJ 2005;331:1548-1551.

24. Cosco TD, Prina AM, Perales J, Stephan BC, Brayne C. Lay perspectives of successful ageing: a systematic review and meta-ethnography. BMJ Open 2013;3(6):e002710.

25. Deep CA, Jeste DV. Defintions and predictors of successful aging: a comprehensive review of larger quantitative studies. Am J Geriatr Psychiatry 2006;14(1):6-20.

26. Rowe JW, Kahn RL. Successful aging. Gerontologist 1997;37(4):433-440.

27. Cooper R, Kuh D, Cooper C, Gale CR, Lawlor DA, Matthews F, et al. Objective measures of physical capability and subsequent health: a systematic review. Age Ageing 2011;40(1):14-23.

28. Cooper R, Kuh D, Hardy R. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. BMJ 2010;341:c4467.

29. Gill TM. Assessment of function and disability in longitudinal studies. J Am Geriatr Soc 2010;58(Suppl):308-312.

30. Ouden MEM, Schuurmans MJ, Arts IE, Schouw YT. Physical performance characteristics related to disability in older persons: a systematic review. Maturitas 2011;69(3):208-219.

31. Verbrugge LM, Jette AM. The disablement process. Soc Sci Med 1994;38(1):1-14.

32. Keeler E, Guralnik JM, Tian H, Wallace RB, Reuben DB. The impact of functional status on life expectancy in older persons. J Gerontol A Biol Sci Med Sci 2010;65(7):727-733.

33. Bean JF, Olveczky DD, Kiely DK, LaRose SI, Jette AM. Performance-based versus patient-reported physical function: what are the underlying predictors? Phys Ther 2011;91(12):1804-1811.

34. Myers AM, Holliday PJ, Harvey KA, Hutchinson KS. Functional performance measures: are they superior to self-assessments? J Gerontol 1993;48(5):196-206.

35. Reuben DB, Seeman TE, Keeler E, Hayes RP, Bowman L, Sewall A, et al. Refining the categorization of physical functional status: the added value of combining self-reported and performance-based measures. J Gerontol A Biol Sci Med Sci 2004;59(10):1056-1061.

36. Coman L, Richardson J. Relationship between self-report and performance measures of function: a systematic review. Can J Aging 2006;25(3):253-270.

37. Freedman VA, Kasper JD, Cornman JC, Agree EM, Bandeen-Roche K, Mor V, et al. Validation of new measures of disability and functioning in the national health and aging trends study. J Gerontol A Biol Sci Med Sci 2011;66(9):1013-1021.

38. Guralnik JM, Ferrucci L. Assessing the building blocks of function: utilizing measures of functional limitation. Am J Prev Med 2003;25:112-121.

39. Kivinen P, Sulkava R, Halonen P, Nissinen A. Self-reported and performance based functional status and associated factors among elderly men: The Finnish cohorts of the Seven Countries Study. J Clin Epidemiol 1998;51(12):1243-1252.

40. Kempen GI, van Heuvelen MJ, van den Brink RH, Kooijman AC, Klein M, Houx PJ, et al. Factors affecting contrasting results between self-reported and performance-based levels of physical limitation. Age Ageing 1996;25(6):458-464.

41. Andrews AW. Self-report measures of mobility. Top Geriatr Rehabil 2012;28(1):2-10.

42. Glass TA. Conjugating the "tenses" of function: discordance among hypothetical, experimental, and enacted function in older adults. Gerontologist 1998;38(1):101-112.

43. Hoeymans N, Feskens EJ, van den Bos GA, Kromhout D. Measuring functional status: cross-sectional and longitudinal associations between performance and self-report (Zutphen Elderly Study 1990-1993). J Clin Epidemiol 1996;49(10):1103-1110.

44. Guralnik JM, Branch LG, Cummings SR, Curb JD. Physical performance measures in aging research. J Gerontol 1989;44(5):141-146.

45. Fors S, Thorslund M, Parker MG. Do actions speak louder than words? Self-assessed and performance-based measures of physical and visual function among old people. Eur J Ageing 2006;3(1):15-21.

46. Tanji H, Gruber-Baldini AL, Anderson KE, Pretzer-Aboff I, Reich SG, Fishman PS, et al. A comparative study of physical performance measures in Parkinson's disease. Mov Disord 2008;23(13):1897-1905.

47. Jette AM. Toward a common language for function, disability, and health. Phys Ther 2006;86(5):726-734.

48. World Health Organisation. International Classification of Functioning, Disability and Health:ICF. Geneva, Switzerland: World Health Organisation; 2001.

49. Nagi SZ. A study in the evaluation of disability and rehabilitation potential: Concepts, methods and procedures. Am J Public Health Nations Health 1964;54:1568-1579.

50. Nagi SZ. Some conceptual issues in disability and rehabilitation. In: Sussman MB, editor. Sociology and Rehabilitation.Washington DC, United States: American Sociological Association; 1965. p. 100-113.

51. Nagi SZ. An epidemiology of disability among adults in the United States. Milbank Mem Fund Q Health Soc 1976;54(4):439-467.

52. World Health Organisation. International Classification of Impairments, Disabilities and Handicaps: A Manual of Classfication Relating to the Consequences of Disease. Geneva, Switzerland: World Health Organisation; 1980.

53. Thyberg M, Arvidsson P, Thyberg I, Nordenfelt L. Simplified bipartite concepts of functioning and disability recommended for interdisciplinary use of the ICF. Disabil Rehabil 2015;37(19):1783-1792.

54. Nordenfelt L. Action theory, disability and ICF. Disabil Rehabil 2003 16;25(18):1075-1079.

55. Ravenek MJ, Skarakis-Doyle E, Spaulding SJ, Jenkins ME, Doyle PC. Enhancing the conceptual clarity and utility of the international classification of functioning, disability & health: the potential of a new graphic representation. Disabil Rehabil 2013;35(12):1015-1025.

56. Whiteneck G. Conceptual models of disability: Past, present and future. In: Jette AM, Field MJ, Martin L, editors. Workshop on Disability in America: A new look.Washington DC, United States: The National Academies Press; 2006. p. 50-66.

57. Jette AM. Toward a common language of disablement. J Gerontol A Biol Sci Med Sci 2009;64(11):1164-1168.

58. Freedman VA. Adopting the ICF language for studying late-life disability: a field of dreams? J Gerontol A Biol Sci Med Sci 2009;64(11):1172-1174.

59. Guralnik J, Ferrucci L. The challenge of understanding the disablement process in older persons: commentary responding to Jette A.M. Toward a common language of disablement. J Gerontol A Biol Sci Med Sci 2009;64(11):1169-1171.

60. World Health Organisation. Towards a Common Language for Functioning, Disability and Health: ICF. Geneva, Switzerland: World Helath Organistaion; 2002.

61. Snyder AR, Parsons JT, Valovich McLeod TC, Curtis BR, Michener LA, Sauers EL. Using disablement models and clinical outcomes assessment to enable evidence-based athletic training practice, part I: disablement models. J Athl Train 2008;43(4):428-436.

62. Masala C, Petretto DR. From disablement to enablement: conceptual models of disability in the 20th century. Disabil Rehabil 2008;30(17):1233-1244.

63. Freedman VA. Late-life disability trends: An overview of current evidence. In: Jette AM, Field MJ, Martin L, editors. Workshop on Disability in America: A new look.Washington DC, United States: The National Acadamies Press; 2006. p. 101-112.

64. Gill TM, Gahbauer EA, Han L, Allore HG. Trajectories of disability in the last year of life. N Engl J Med 2010;362(13):1173-1180.

65. Field MJ, Jette AM, Martin L. Workshop on Disability in America: A New Look - Summary and Background Papers. Washington DC, United States: The National Academies Press; 2006.

66. Cerniauskaite M, Quintas R, Boldt C, Raggi A, Cieza A, Bickenbach JE, et al. Systematic literature review on ICF from 2001 to 2009: its use, implementation and operationalisation. Disabil Rehabil 2011;33(4):281-309.

67. Madden R, Fortune N, Cheeseman D, Mpofu E, Bundy A. Fundamental questions before recording or measuring functioning and disability. Disabil Rehabil 2013;35(13):1092-1096.

68. Farin E, Fleitz A, Frey C. Psychometric properties of an International Classification of Functioning, Disability and Health (ICF)-oriented, adaptive questionnaire for the assessment of mobility, self-care and domestic life. J Rehabil Med 2007;39(7):537-546.

69. Ruaro JA, Ruaro MB, Souza DE, Frez AR, Guerra RO. An overview and profile of the ICF's use in Brazil--a decade of history. Rev Bras Fisioter 2012;16(6):454-462.

70. Institute of Medicine. The Future of Disability in America. Washinton DC, United States: The National Academies Press; 2007.

71. Brach JS, VanSwearingen JM, Newman AB, Kriska AM. Identifying early decline of physical function in community-dwelling older women: performance-based and self-report measures. Phys Ther 2002;82(4):320-328.

72. Ferrer M, Lamarca R, Orfila F, Alonso J. Comparison of performance-based and self-rated functional capacity in Spanish elderly. Am J Epidemiol 1991;149(3):228-235.

73. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994;49(2):85-94.

74. Kuh D, Pierce M, Adams J, Deanfield J, Ekelund U, Friberg P, et al. Cohort profile: updating the cohort profile for the MRC National Survey of Health and Development: a new clinic-based data collection for ageing research. Int J Epidemiol 2011;40(1):1-9.

75. Wadsworth M, Kuh D, Richards M, Hardy R. Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). Int J Epidemiol 2006;35(1):49-54.

76. Stafford M, Black S, Shah I, Hardy R, Pierce M, Richards M, et al. Using a birth cohort to study ageing: representativeness and response rates in the National Survey of Health and Development. Eur J Ageing 2013;10:145-157.

77. Pierce MB, Silverwood RJ, Nitsch D, Adams JE, Stephen AM, Nip W, et al. Clinical disorders in a post war British cohort reaching retirement: evidence from the first national birth cohort study. PLoS One 2012;7(9):e44857.

78. Collerton J, Davies K, Jagger C, Kingston A, Bond J, Eccles MP, et al. Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. BMJ 2009;339:b4904.

79. Kuh D, Bassey EJ, Butterworth S, Hardy R, Wadsworth ME. Grip strength, postural control, and functional leg power in a representative cohort of British men and women: associations with physical activity, health status, and socioeconomic conditions. J Gerontol A Biol Sci Med Sci 2005;60(2):224-231.

80. Kuh D, Karunananthan S, Bergman H, Cooper R. A life-course approach to healthy ageing: maintaining physical capability. Proc Nut Soc 2014;73(2):237-248.

81. Guralnik JM, Butterworth S, Wadsworth ME, Kuh D. Childhood socioeconomic status predicts physical functioning a half century later. J Gerontol A Biol Sci Med Sci 2006;61(7):694-701.

82. Martin J, Meltzer H, Elliot D. OPCS Surveys of Disability in Great Britain, Report 1: The Prevalence of disability among adults. London,UK: HMSO; 1988.

83. Wloch E, Cooper R, Kuh D. Is the Hierarchy of Loss in Functional Ability Evident in Midlife? Findings from a British Birth Cohort. PLoS One 2016;11(5):e0155815.

84. Ferrucci L, Guralnik JM, Simonsick E, Salive ME, Corti C, Langlois J. Progressive versus catastrophic disability: a longitudinal view of the disablement process. J Gerontol A Biol Sci Med Sci 1996;51(3):123-130.

85. Ferrucci L, Guralnik JM, Cecchi F, Marchionni N, Salani B, Kasper J, et al. Constant hierarchic patterns of physical functioning across seven populations in five countries. Gerontologist 1998;38(3):286-294.

86. Kingston A, Collerton J, Davies K, Bond J, Robinson L, Jagger C. Losing the ability in activities of daily living in the oldest old: a hierarchic disability scale from the Newcastle 85+ study. PLoS One 2012;7(2):e31665.

87. Jagger C, Arthur AJ, Spiers NA, Clarke M. Patterns of onset of disability in activities of daily living with age. J Am Geriatr Soc 2001;49(4):404-409.

88. Fieo RA, Austin EJ, Starr JM, Deary IJ. Calibrating ADL-IADL scales to improve measurement accuracy and to extend the disability construct into the preclinical range: a systematic review. BMC Geriatr 2011;11:42.

89. Barberger-Gateau P, Rainville C, Letenneur L, Dartigues JF. A hierarchical model of domains of disablement in the elderly: a longitudinal approach. Disabil Rehabil 2000;22(7):308-317.

90. Yeh CJ, Wang CY, Tang PF, Lee MC, Lin HS, Chen HY. Hierarchy of higher-level physical functions: a longitudinal investigation on a nationally representative population

of community-dwelling middle-aged and elderly persons. Disabil Rehabil 2012;34(15):1271-1276.

91. Dunlop DD, Hughes SL, Manheim LM. Disability in activities of daily living: patterns of change and a hierarchy of disability. Am J Public Health 1997;87(3):378-383.

92. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL: A standardised measure of biological and psychosocial function. JAMA 1963;185:914-919.

93. Seidel D, Crilly N, Matthews FE, Jagger C, Brayne C, Clarkson PJ. Patterns of functional loss among older people: a prospective analysis. Hum Factors 2009;51(5):669-680.

94. Gerrard P. The hierarchy of the activities of daily living in the Katz index in residents of skilled nursing facilities. J Geriatr Phys Ther 2013;36(2):87-91.

95. Weiss CO, Fried LP, Bandeen-Roche K. Exploring the hierarchy of mobility performance in high-functioning older women. J Gerontol A Biol Sci Med Sci 2007;62(2):167-173.

96. Chen HY, Wang CY, Lee MY, Tang PF, Chu YH, Suen MW. A hierarchical categorisation of tasks in mobility disability. Disabil Rehabil 2010;32(19):1586-1593.

97. Kempen GI, Myers AM, Powell LE. Hierarchical structure in ADL and IADL: analytical assumptions and applications for clinicians and researchers. J Clin Epidemiol 1995;48(11):1299-1305.

98. Roorda LD, Roebroeck ME, van Tilburg T, Lankhorst GT, Bouter LM, Measuring Mobility Study Group. Measuring activity limitations in climbing stairs: Development of a hierarchical scale for patients with ower-extermity disorders living at home. Arch Phys Med Rehabil 2004;85(6):967-971.

99. Lazaridis EN, Rudberg MA, Furner SE, Cassel CK. Do activities of daily living have a hierarchical structure? An analysis using the longitudinal study of aging. J Gerontol 1994;49(2):47-51.

100. Diesfeldt HF. Executive functioning in psychogeriatric patients: scalability and construct validity of the Behavioral Dyscontrol Scale (BDS). Int J Geriatr Psychiatry 2004;19(11):1065-1073.

101. Olsen LR, Mortensen EL, Bech P. The SCL-90 and SCL-90R versions validated by item response models in a Danish community sample. Acta Psychiatr Scand 2004;110(3):225-229.

102. Cooper R, Mishra GD, Kuh D. Physical activity across adulthood and physical performance in midlife: findings from a British birth cohort. Am J Prev Med 2011;41(4):376-384.

103. Lindelow M, Hardy R, Rodgers B. Development of a scale to measure symptoms of anxiety and depression in the general UK population: the psychiatric symptom frequency scale. J Epidemiol Community Health 1997;51(5):549-557.

104. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med 1979;9(1):139-145.

105. Medical Research Council. Questionnaire on respiratory symptoms and instructions for interviewers. MRC; 1976.

106. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011;30(4):377-399.

107. Wells JC. Sexual dimorphism of body composition. Best Pract Res Clin Endocrinol Metab 2007;21(3):415-430.

108. Brady AO, Straight CR, Evans EM. Body composition, muscle capacity, and physical function in older adults: an integrated conceptual model. J Aging Phys Act 2014;22(3):441-452.

109. Hays JC, Saunders WB, Flint EP, Kaplan BH, Blazer DG. Social support and depression as risk factors for loss of physical function in late life. Ageing and Mental Health 1997;3(1):209-222.

110. Dodds R, Kuh D, Aihie SA, Cooper R. Physical activity levels across adult life and grip strength in early old age: updating findings from a British birth cohort. Age Ageing 2013;42(6):794-798.

111. Hergenroeder AL, Brach JS, Otto AD, Sparto PJ, Jakicic JM. The Influence of Body Mass Index on Self-report and Performance-based Measures of Physical Function in Adult Women. Cardiopulm Phys Ther J 2011;22(3):11-20.

112. Kempen GI, Steverink N, Ormel J, Deeg DJ. The assessment of ADL among frail elderly in an interview survey: self-report versus performance-based tests and determinants of discrepancies. J Gerontol B Psychol Sci Soc Sci 1996;51(5):254-260.

113. Rallon CR, Chen CC. Relationship between performance-based and self-reported assessment of hand function. Am J Occup Ther 2008;62(5):574-579.

114. Stretton CM, Latham NK, Carter KN, Lee AC, Anderson CS. Determinants of physical health in frail older people: the importance of self-efficacy. Clin Rehabil 2006;20(4):357-366.

115. Young Y, Boyd CM, Guralnik JM, Fried LP. Does self-reported function correspond to objective measures of functional impairment? J Am Med Dir Assoc 2010;11(9):645-653.

116. Sainio P, Koskinen S, Heliovaara M, Martelin T, Harkanen T, Hurri H, et al. Self-reported and test-based mobility limitations in a representative sample of Finns ages 30+. Scand J Public Health 2006;34(4):378-386.

117. Cress ME, Schechtman KB, Mulrow CD, Fiatarone MA, Gerety MB, Buchner DM. Relationship between physical performance and self-perceived physical function. J Am Geriatr Soc 1995;43(2):93-101.

118. Farag I, Sherrington C, Kamper SJ, Ferreira M, Moseley AM, Lord SR, et al. Measures of physical functioning after hip fracture: construct validity and responsiveness of performance-based and self-reported measures. Age Ageing 2012;41(5):659-664.

119. Latham NK, Mehta V, Nguyen AM, Jette AM, Olarsch S, Papanicolaou D, et al. Performance-based or self-report measures of physical function: which should be used in clinical trials of hip fracture patients? Arch Phys Med Rehabil 2008;89(11):2146-2155.

120. Mallinson T, Cella D, Cashy J, Holzner B. Giving meaning to measure: linking self-reported fatigue and function to performance of everyday activities. J Pain Symptom Manage 2006;31(3):229-241.

121. Owens PL, Bradley EH, Horwitz SM, Viscoli CM, Kernan WN, Brass LM, et al. Clinical assessment of function among women with a recent cerebrovascular event: a self-reported versus performance-based measure. Ann Intern Med 2002;136(11):802-811.

122. Rogers JC, Holm MB, Beach S, Schulz R, Cipriani J, Fox A, et al. Concordance of four methods of disability assessment using performance in the home as the criterion method. Arthritis Rheum 2003;49(5):640-647.

123. Shulman LM, Pretzer-Aboff I, Anderson KE, Stevenson R, Vaughan CG, Gruber-Baldini AL, et al. Subjective report versus objective measurement of activities of daily living in Parkinson's disease. Mov Disord 2006;21(6):794-799.

124. Stuifbergen AK, Morris M, Becker H, Chen L, Lee HY. Self-report versus performance measure in gauging level of function with multiple sclerosis. Disabil Health J 2014;7(4):413-418.

125. van den Ende CH, Hazes JM, Le CS, Breedveld FC, Dijkmans BA. Discordance between objective and subjective assessment of functional ability of patients with rheumatoid arthritis. Br J Rheumatol 1995;34(10):951-955.

126. van Weely SF, van Denderen JC, Steultjens MP, van der Leeden M, Nurmohamed MT, Dekker J, et al. Moving instead of asking? Performance-based tests and BASFI-questionnaire measure different aspects of physical function in ankylosing spondylitis. Arthritis Res Ther 2012;14(2):R52.

127. Waehrens EE, Amris K, Fisher AG. Performance-based assessment of activities of daily living (ADL) ability among women with chronic widespread pain. Pain 2010;150(3):535-541.

128. Waehrens EE, Bliddal H, Danneskiold-Samoe B, Lund H, Fisher AG. Differences between questionnaire- and interview-based measures of activities of daily living (ADL) ability and their association with observed ADL ability in women with rheumatoid arthritis, knee osteoarthritis and fibromyalgia. Scand J Rheumatol 2012;41(2):95-102.

129. Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH. Discrepancies between self-reported and observed physical function in the elderly: the influence of response shift and other factors. Soc Sci Med 1999;48(11):1549-1561.

130. Sherman SE, Reuben D. Measures of functional status in community-dwelling elders. J Gen Intern Med 1998;13(12):817-823.

131. Merrill SS, Seeman TE, Kasl SV, Berkman LF. Gender differences in the comparison of self-reported disability and performance measures. J Gerontol A Biol Sci Med Sci 1997;52(1):19-26.

132. Kelly-Hayes M, Jette AM, Wolf PA, D'Agostino RB, Odell PM. Functional limitations and disability among elders in the Framingham Study. Am J Public Health 1992;82(6):841-845.

133. Louie GH, Ward MM. Association of measured physical performance and demographic and health characteristics with self-reported physical function: implications for the interpretation of self-reported limitations. Health Qual Life Outcomes 2010;8:84.

134. Fried LP, Young Y, Rubin G, Bandeen-Roche K. Self-reported preclinical disability identifies older women with early declines in performance and early disease. J Clin Epidemiol 2001;54(9):889-901.

135. Simonsick EM, Newman AB, Nevitt MC, Kritchevsky SB, Ferrucci L, Guralnik JM, et al. Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC study. J Gerontol A Biol Sci Med Sci 2001;56(10):644-649.

136. Cooper R, Strand BH, Hardy R, Patel KV, Kuh D. Physical capability in mid-life and survival over 13 years of follow-up: British birth cohort study. BMJ 2014;348:g2219.

137. Rogers JC, Holm MB, Raina KD, Dew MA, Shih MM, Begley A, et al. Disability in late-life major depression: patterns of self-reported task abilities, task habits, and observed task performance. Psychiatry Res 2010;178(3):475-479.

138. Marsh AP, Wrights AP, Haakonssen EH, Dobrosielski MA, Chmelo EA, Barnard RT, et al. The Virtual Short Physical Performance Battery. J Gerontol A Biol Sci Med Sci 2015;70(10):1233-1241.

139. Jang Y, Mortimer JA, Haley WE, Graves AB. The role of neuroticism in the association between performance-based and self-reported measures of mobility. J Aging Health 2002;14(4):495-508.

140. Demakakos P, Cooper R, Hamer M, de OC, Hardy R, Breeze E. The bidirectional association between depressive symptoms and gait speed: evidence from the English Longitudinal Study of Ageing (ELSA). PLoS One 2013;8(7):e68632.

141. Melzer D, Lan TY, Tom BD, Deeg DJ, Guralnik JM. Variation in thresholds for reporting mobility disability between national population subgroups and studies. J Gerontol A Biol Sci Med Sci 2004;59(12):1295-1303.

142. Rapuri PB, Gallagher JC, Smith LM. Smoking is a risk factor for decreased physical performance in elderly women. J Gerontol A Biol Sci Med Sci 2007;62(1):93-100.

143. Strand BH, Mishra G, Kuh D, Guralnik JM, Patel KV. Smoking history and physical performance in midlife: results from the British 1946 birth cohort. J Gerontol A Biol Sci Med Sci 2011;66(1):142-149.

144. Brach JS, Simonsick EM, Kritchevsky S, Yaffe K, Newman AB. The association between physical function and lifestyle activity and exercise in the health, aging and body composition study. J Am Geriatr Soc 2004;52(4):502-509.

145. Hardy R, Cooper R, Aihie SA, Ben-Shlomo Y, Cooper C, Deary IJ, et al. Body mass index, muscle strength and physical performance in older adults from eight cohort studies: the HALCyon programme. PLoS One 2013;8(2):e56483.

146. Lang IA, Llewellyn DJ, Alexander K, Melzer D. Obesity, physical function, and mortality in older adults. J Am Geriatr Soc 2008;56(8):1474-1478.

147. World Health Organisation. Physical status: the use and interpretation of anthropometry. World Health Organ Tech Rep Ser 1995;854:1-452.

148. World Health Organisation. Obesity: preventing and managing the global epidemic. World Health Organ Tech Rep Ser 2000;894:1-253.

149. Botoseneanu A, Allore HG, Mendes de Leon CF, Gahbauer EA, Gill TM. Sex Differences in Concomitant Trajectories of Self-Reported Disability and Measured Physical Capacity in Older Adults. J Gerontol A Biol Sci Med Sci 2016 Apr 12 [epub ahead of print].

150. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey SG. Indicators of socioeconomic position (part 1). J Epidemiol Community Health 2006;60(1):7-12.

151. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey SG. Indicators of socioeconomic position (part 2). J Epidemiol Community Health 2006;60(2):95-101.

152. Jurges H. True health vs response styles: Exploring cross-country differences in self-reported health. Health Econ 2007;16(2):163-178.

153. Kuh D, Richards M, Hardy R, Ben-Shlomo Y. Life course epidemiology, ageing research and maturing cohort studies: a dynamic combination for understanding healthy ageing. In: Kuh D, Cooper R, Hardy R, Richards M, Ben-Shlomo Y, editors. A Life Course Approach to Healthy Ageing. Oxford, UK: Oxford University Press; 2014. p.3-15.

154. Weich S, Brugha T, King M, McManus S, Bebbington P, Jenkins R, et al. Mental well-being and mental illness: findings from the Adult Psychiatric Morbidity Survey for England 2007. Br J Psychiatry 2011;199(1):23-28.

155. Penninx BW, Deeg DJ, van Eijk JT, Beekman AT, Guralnik JM. Changes in depression and physical decline in older adults: a longitudinal perspective. J Affect Disord 2000;61(1-2):1-12.

156. Shenkin SD, Laidlaw K, Allerhand M, Mead GE, Starr JM, Deary IJ. Life course influences of physical and cognitive function and personality on attitudes to aging in the Lothian Birth Cohort 1936. Int Psychogeriatr 2014;26(9):1417-1430.

157. Suchy Y, Williams PG, Kraybill ML, Franchow E, Butner J. Instrumental activities of daily living among community-dwelling older adults: personality associations with self-report, performance, and awareness of functional difficulties. J Gerontol B Psychol Sci Soc Sci 2010;65(5):542-550.

158. McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32(1):40-66.

159. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993;31(3):247-263.

160. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30(6):473-483.

161. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007;5:63.

162. Jacobs DM, Sano M, Dooneief G, Marder K, Bell KL, Stern Y. Neuropsychological detection and characterization of preclinical Alzheimer's disease. Neurology 1995;45(5):957-962.

163. Masur DM, Sliwinski M, Lipton RB, Blau AD, Crystal HA. Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. Neurology 1994;44(8):1427-1432.

164. Eysenck HJ. A Short Questionnaire for the Measurement of Two Dimensions of Personality. J Appl Psychol 1958;42:14-17.

165. Shaw GK, Hare EH. The Maudsley Personality Inventory (Short Form): Distribution of scores and test-retest reliability in an urban population. Br J Psychiatry 1965;111:226-235.

166. Cooper R, Muniz-Terrera G, Kuh D. Associations of behavioural risk factors and health status with changes in physical capability over 10 years of follow-up: the MRC National Survey of Health and Development. BMJ Open 2016;6(4):e009962.

167. Wills AK, Black S, Cooper R, Coppack RJ, Hardy R, Martin KR, et al. Life course body mass index and risk of knee osteoarthritis at the age of 53 years: evidence from the 1946 British birth cohort study. Ann Rheum Dis 2012;71(5):655-660.

168. Colman I, Ploubidis GB, Wadsworth ME, Jones PB, Croudace TJ. A longitudinal typology of symptoms of depression and anxiety over the life course. Biol Psychiatry 2007;62(11):1265-1271.

169. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling change and Event Occurence. 1st Edition ed. New York, United States: Oxofrd Unviersity Press; 2003.

170. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005;34(1):215-220.

171. Fitzmaurice G. A conundrum in the analysis of change. Nutrition 2001;17(4):360-361.

172. Frerichs RJ, Tuokko HA. A comparison of methods for measuring cognitive change in older adults. Arch Clin Neuropsychol 2005;20(3):321-333.

173. Deeg DJ. Longitudinal characterization of course types of functional limitations. Disabil Rehabil 2005;27(5):253-261.

174. Bravell ME, Zarit SH, Johansson B. Self-reported activities of daily living and performance-based functional ability: a study of congruence among the oldest old. Eur J Ageing 2011;8(3):199-209.

175. Bohannon RW. Intertester reliability of hand-held dynamometry: a concise summary of published research. Percept Mot Skills 1999;88(3):899-902.

176. Bohannon RW. Test-retest reliability of the five-repetition sit-to-stand test: a systematic review of the literature involving adults. J Strength Cond Res 2011;25(11):3205-3207.

177. Bohannon RW, Schaubert KL. Test-retest reliability of grip-strength measures obtained over a 12-week interval from community-dwelling elders. J Hand Ther 2005;18(4):426-428.

178. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39(2):142-148.

179. Rathouz PJ, Kasper JD, Zeger SL, Ferrucci L, Bandeen-Roche K, Miglioretti DL, et al. Short-term consistency in self-reported physical functioning among elderly women: the Women's Health and Aging Study. Am J Epidemiol 1998;147(8):764-773.

180. Swift HJ, Lamont RA, Abrams D. Are they half as strong as they used to be? An experiment testing whether age-related social comparisons impair older people's hand grip strength and persistence. BMJ Open 2012;2(3):e001064

181. Coudin G, Alexopoulos T. 'Help me! I'm old!' How negative aging stereotypes create dependency among older adults. Aging Ment Health 2010;14(5):516-23.

182. Shrira A, Litwin H. The effect of lifetime cumulative adversity and depressive symptoms on functional status. J Gerontol B Psychol Sci Soc Sci 2014;69(6):953-965.

183. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. Proc Natl Acad Sci U S A 2015;112(30):4104-4110.

184. Fried LP, Bandeen-Roche K, Williamson JD, Prasada-Rao P, Chee E, Tepper S, et al. Functional decline in older adults: expanding methods of ascertainment. J Gerontol A Biol Sci Med Sci 1996;51(5):206-214.

185. Whitson HE, Duan-Porter W, Schmader KE, Morey MC, Cohen HJ, Colon-Emeric CS. Physical Resilience in Older Adults: Systematic Review and Development of an Emerging Construct. J Gerontol A Biol Sci Med Sci 2016;71(4):489-495.

186. Daltroy LH, Phillips CB, Eaton HM, Larson MG, Partridge AJ, Logigian M, et al. Objectively measuring physical ability in elderly persons: the Physical Capacity Evaulation. Am J Public Health 1995;85(4):558-560.

187. Reuben DB, Valle L, Hays RD, Siu AL. Measuring physical function in community-dwelling older persons: a comparison of self-administered, interviewer-administered, and performance-based measures. J Am Geriatr Soc 1995;43(1):17-23.

188. Alexander NB, Guire KE, Thelen DG, Ashton-Miller JA, Schultz AB, Grunawalt JC, et al. Self-reported walking ability predicts functional mobility performance in frail older adults. J Am Geriatr Soc 2000;48(11)1408-1413.

189. van den Brink CL, Tijhuis M, Kalmijn S, Nissinen A, Giampaoli S, Kivinen P, et al. Self-reported disability and its association with performance-based limitation in elderly men: a comparison of three European countries. J Am Geriatr Soc 2003;51(6):782-788.

Appendices

Appendix 1: Self-reported measures of physical capabilities in NSHD since 1989

Topic	Age (Year)		
	Age 43 (1989)	Age 53 (1999)	Age 60-64 (2006-10)
Summary of questionnaire	Initial set of questions to assess if participant experienced difficulty with specific activities, asked in self-competition booklet. If participant answered "yes" to any of the initial questions the nurse then asked the appropriate supplementary disability questions to gauge the severity of the difficulty. The participant should only have answered the questions which related to the specific task that they reported difficulties with in the initial questions. In some cases this did not occur.	Questions asked as part of the nurse interview using CAPI Version 5. For two part questions, participants only asked second part if they responded "yes" to the first part.	Set of performance questions asked as part of the nurse clinic schedule. 15 participants, who did not complete the nurse schedule, did self-complete the performance questions. Most questions where two part questions: the first part assessed if the participant had any difficulty with the specific task and "if yes" the second part of the question tried to assess the severity of difficulty experienced. In some cases the gating was not followed correctly.
Grip **	Do you have…difficulty holding, gripping or turning things…due to long term health problems or disabilities, either physical or mental? s.) Using your [right/left] hand only can you… ○ Pick up and carry a 5 lb bag of potatoes ○ Turn a tap on and off ○ Pick up a small object such as a safety pin If no to any activity above ask the following two questions • Using your [right/left] hand only can you… ○ Pick up ad carry a pint of milk ○ Pick up and hold a mug of tea or coffee ○ Turn the control knobs on a cooker • I'm going to read out some things which involve holding, gripping or turning and I'd like you to tell me whether it is difficult for you to do them (without using special gadgets)	Do you have difficulty because of long term health problems holding something heavy like a full kettle or removing a stiff lid from a jar?	Do you have difficulty because of long term health problems holding something heavy like a full kettle or removing a stiff lid from a jar? Is it difficult because of long term health problems to…. go shopping and carry a full bag of shopping in each hand?

<table>
<tr><td></td><td>Wring out light washing (eg a tea towel)
Unscrew the lid of a coffee jar
Pick up and pour from a full kettle
Serve food from a pan using a spoon or ladle
Use a pen or pencil
Use a pair of scissors
Tie a bow in laces or string</td><td></td><td></td></tr>
<tr><td>Arm **</td><td>Do you have…difficulty using arms to reach or stretch for things…due to long term health problems or disabilities, either physical or mental?
s.) using [right/left] arm how difficult is it for you to…
Hold your arm out in front of you to shake hands with someone?
Put your hand up to your head to put a hat on?
Put your hand behind your back to tuck in a blouse/shirt
Put your arm above your head to reach for something above you?</td><td>Do you have difficulty because of long term health problems using either arm to reach high above your head or to reach behind to a shirt in or do up a zip?</td><td></td></tr>
<tr><td>Walking</td><td>Do you have…difficulty walking for a quarter of a mile on the level…due to long term health problems or disabilities, either physical or mental?
s.) select appropriate level of disability:
Can walk …. without stopping or severe discomfort
More than a quarter of a mile
More than 200 yards (but less than a quarter of a mile)
More than 50 yards
Only walk a few steps
Cannot walk at all</td><td>a.) Do you find it difficult to walk for a quarter of a mile on the level because of long term health problems?
b.) How far can you walk without stopping or severe discomfort?</td><td>a.) Do you find it difficult to walk for a quarter of a mile on the level because of long term health problems?
b.) If yes, how far can you walk without stopping or severe discomfort?
More than 400 yards
200-400 yards
50-200 yards
Less than 50 yards</td></tr>
</table>

Stair climbing	Do you have…difficulty walking up and down steps or stairs…due to long term health problems or disabilities, either physical or mental? s.) select appropriate level of disability: • can walk up and down a flight of stairs in a normal manner • can only walk up and down a flight of stairs without stopping or holding on if goes sideways or one step at a time • can only walk up and down a flight of 12 stairs if holds on (doesn't need rest) • can only walk up and down a flight of 12 stairs if holds on and takes a rest • can walk up and down at least one step but cannot manage a flight of 12 steps • cannot walk up and down one step	a.) Do you find it difficult walking up and down stairs because of long term health problems? b.) Can you walk up and down a flight of 12 stairs in a normal manner without holding on or taking a rest?	c.) Do you find it difficult walking up and down stairs because of long term heal problems? d.) If yes, can you walk up and down a flight of 12 stairs in a normal manner without holding on or taking a rest?
Balance	Do you have…falling or difficulty keeping balance…due to long term health problems or disabilities, either physical or mental? s.) select appropriate level of disability • Does not need to hold onto something to keep balance • Needs to hold on occasionally • Often needs to hold on • Always need to hold on	Do you easily fall or have difficulty walking up and down stairs because of long term health problems?	Do you easily fall or have difficulty keeping your balance because of long term health problems? Do you need to hold onto something to keep your balance?
Falling	s.) select appropriate level of disability • Has not fallen in the past year • Has fallen once or twice in the past year • Has fallen between 3 and 11 times in the past year • Has fallen 12 or more time in the past year	a.) Have you fallen at all in the past 12 months that is since [LFF month and year]? b.) How many times have you fallen in the past 12 months? (answer:	a.) Have you fallen at all in the past 12 months? b.) If yes, how many times have you fallen in the past 12 months c.) On how many of these occasions have you injured yourself badly enough to see a doctor?

		categories same as those used in '89)	(answer Q b and c: continuous number)
Bending	Do you have…difficulty bending down and straightening up, even when holding onto something…due to long term health problems or disabilities, either physical or mental? s.) select appropriate level of disability • Can bend down to sweep up something from the floor and straighten up • Can bend down to pick up something from the floor and straighten up • Can bend down far enough to touch knees and straighten up • Cannot bend down far enough to touch knees and straighten up		a.) Do you have difficulty bending down and straightening up, even when holding onto something because of long term health problems? b.) Can you bend down to sweep something from the floor and straighten up? c.) Can you bend down to pick up something form the floor and straighten up? d.) Can you bend down far enough to touch your knees and straighten up? **
Heavy housework			Is it difficult because of long term health problems to… do heavy housework?
Prepare hot meal			Is it difficult because of long term health problems to… prepare a hot meal?
Washing hands and face	Is it difficult for you…washing hands and face…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems…washing hands and face? If yes, can you do it without aids or personal help?
Bathing or showering	Is it difficult for you…washing all over…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems… bathing or showering? If yes, can you do it without aids or personal help?
Dressing or undressing	Is it difficult for you…dressing and undressing…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems… dressing or undressing? If yes, can you do it without aids or personal help?

Getting in and out of a chair	Is it difficult for you…getting in and out of a chair…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems….getting in and out of a chair? If yes, can you do it without aids or personal help?
Getting in and out of bed	Is it difficult for you…getting in and out of bed…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems…getting in and out of bed? If yes, can you do it without aids or personal help?
Getting to the toilet	Is it difficult for you…getting to the toilet…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems…getting to the toilet? If yes, can you do it without aids or personal help?
Using the toilet	Is it difficult for you…using the toilet…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems… using the toilet? If yes, can you do it without aids or personal help?
Feeding yourself	Is it difficult for you…feeding yourself, including cutting up food…due to long term health problems o disabilities, either physical or mental? s.) can do without help?		Is it difficult because of long term health problems… feeding yourself, including cutting up food? If yes, can you do it without aids or personal help?

s.) indicates supplementary disability question
… indicates that there was an initial question which asked if the respondent had difficulties "with any of the following activities" (or words to that effect), then listed the topics beneath. In this table the specific activity was been inserted into the initial question in place of "any activities"
** In relation to this question also asked: How frequently at home or at work do you use your hands in strong movements such as squeezing water out of a towel, playing racquet sports, digging the garden or carrying heavy items such as a suitcase, briefcase, bucket or shopping bag. (answer: category)
** In relation to this question also asked: How frequently at home or at work do you use your arms to reach above head, such as dusting high places, replacing light bulbs, reaching objects on high shelves, or gardening activities like pruning or trimming high hedges. (answer: category)
** these questions were gated, if answered yes at any stage respondent directed to the next topic

Appendix 2: Review of studies examining the hierarchy of loss

Ordered in reverse chronological order and by first author's surname

Author	Study population *Study *Size *Age *Gender (% female) *Sample *Country *Longitudinal/ CS *Exclusion criteria	Method *No. of tasks included *Definition of difficulty *Position on aids and personal assistance *Analysis approach used *If longitudinal: time interval between collection points	Hierarchical order of tasks *From most complex and difficult (first lost) to least (last lost) Most prevalent hierarchy presented	Results
Gerrad 2013 (94)	*National Nursing Home survey *13113 *80.6 ± 13.0 *71.5% *Nursing facility residents *Unites States *Cross-sectional	*6 tasks *Dichotomised: independent vs dependent (any level of difficulty) *Not reported *Rasch analysis used to ascertain item difficulties before items graded	*Bathing, dressing, toileting, transferring, maintaining continence, eating	*For all items more difficult than eating, the majority of patients who could perform each item could also perform all easier items
Kingston et al 2012 (86)	*Newcastle 85+ *839 *85+ *62% *10% institute remaining community based *UK *Cross-sectional *end stage terminal illness	*17 tasks *Response dichotomised as no difficulty / some difficulty (some, only with an aid, unable to do this) *PCA to assess number of dimensions, followed by Mokken scaling.	*cutting toenails, shopping, use steps, walk 400 yds, heavy housework, full wash, cooking, move around house, transfer from chair, light housework, transfer from toilet, dressing, transfer from bed, wash hands and face, feeding	*Women experienced significantly more difficulty with all items except dressing and light housework *When scores of relative difficulty plotted areas of clustering of items noted, corresponding to 4 domains (mixed upper/lower body combined with balance) *Removal of cognitive items strengthen scale

Yeh et al 2012 (90)	*Health and Living Status *2729 *62 (median at baseline) *37.9% *Community-dwelling *Taiwan *Longitudinal *loss to follow up	*8 tasks *Response dichotomised: no difficulty/ any difficulty (some, a lot, unable to perform) *Without use of aids or personal help *Survival analysis (interval method used) Median ages of incident disability ranked to identify hierarchy. *'96 as baseline, '99, '03 and '07 as follow up	*Running 20-30m, carrying 11kg, squatting, climbing 2-3 floors, walking 200-33m, standing for 15mins, grasping with fingers, raising arms up	*Eight curves show good separation between hierarchical levels *Progression was same in men and women except men developed disability in squatting earlier than carrying 11kg *median age at onset earlier in women *5.6% did not follow hierarchical progression from mild, moderate to severe
Chen et al 2010 (96)	*Health and Living Status *2626 *75.36 ± 6.3 *48.1% *Community-dwelling *Taiwan *Cross-sectional *< 65, hospitalised, institutionalised	*9 tasks *dichotomised: no difficulty/ any difficulty (some, a lot, unable to perform) *Not reported *analysed descriptive data from individuals with summed difficulty tasks from 1 to 9 to determine order of appearance of difficulty. Chi squared tests used to test categorical trends	* Standing 2h, running 20-30m, lifting 11kg, squatting, climbing 2-3 floors, walking 200-300m, standing 15min, grasping with fingers, raising arms. *Categorised into three groups: Mild (4 tasks), moderate (3 tasks) and severe (two tasks)	*as difficulty with mobility tasks increased stepwise increase in % female, mean age, number of chronic illness and number of falls (all trends significant) *Males and females showed same trend in terms of mobility hierarchy and relationship with ADL tasks
Seidel et al 2009 (93)	*MRC cognitive Function and Ageing Study *12186 *76 ± 7 *59% *Primary care patients *UK *Longitudinal	*7 tasks: 4 IADL and 3 capability *dichotomised: no difficulty / any difficulty (some difficulty and not able to perform) *Not reported *prevalence (baseline) and incidence (2 year follow up) calculated, stratified by age and gender. Survival analysis to age of onset, stratified by gender and no of comorbidities *follow up 1, 2, 3, 6, 8 and 10 years	*Shopping, housework, locomotion, transport, reaching, cooking, dexterity	*Higher prevalence for women *Most of those reporting full capability at initial assessment lost their locomotion ability within 2 years, followed by reaching and thinking. *Incidence lower for men than women *Similar order for those with comorbidities although transition starts earlier

Weiss et al 2007 (95)	*WHAS II *436 *74 *100% *Community-dwelling *US *Longitudinal? *difficulty in >1 domain, MMSE <24	*4 tasks *reported difficulty or modification of task (the way or how often) *Not reported *Prevalence and development of onset cases calculated and patterns tabulated. *36 month follow up data used	*Walking ½ mile, climbing 10 steps, transferring from bed to chair, dressing	*76.6% of individuals followed hierarchy *At 36 month 88.4% fit hierarchy but climbing occurred alone more commonly than in combination with walking ½ mile
Jagger et al 2001 (87)	*Unknown *1344 *78 (median) *59% *Registered at GP *UK *Longitudinal *hospitalisation, missing data	*7 tasks *Dichotomised: independent (without difficult) vs difficulty or does not perform *Uses help or aids classed as difficulty *Cox proportional hazard regression models used to compare age and sex for time of onset. *intervals vary between rounds and individuals: 22.3±8, 21.6±5.6, 9.1±4.3, 17.3±2.2	*Bathing, mobility around home, toileting, dressing, transfer from chair, transfer from bed and feeding	*Same order for men and women *Median age at onset of disability was younger in women than men *Order of loss identical for all three age groups, but risk rose with increased age *73.7% with all 7 activities restricted met hierarchy. Order compatible with hierarchy for at least 60% across all stages
Ferrucci et al 1998 (85)	*Several *	*15 tasks *dichotomised: nondisabled (able to perform with/without difficulty but without help) and disabled (unable to perform without help) *Guttman scalogram analysis used	*doing heavy housework, cutting toenails, daily shopping, light housework, cooking, bathing or showering, walking at least 400m, using stairs, moving around outdoors, dressing, using the toilet, transfer from bed, walking between rooms, washing hands and face and feeding	*Data from ELSA: pattern suggests scalability based on domains is more stable than scalability performed on single items. Only 0.4 – 1.7% did not fit hierarchy
Dunlop et al 1997 (91)	*Longitudinal Study of Ageing *5092 *78 *63.9%	*6 tasks *Dichotomised no difficultly vs difficulty *Not reported	*Walking, bathing, transferring (bed or chair), dressing, toileting, feeding	*median age to onset older for men *order same for both genders except for women dressing and toileting are reversed. However median ages very close

	*Community-dwelling *US *Longitudinal	*Restricted to those with no disability at baseline for survival analysis for interval-censored data *1984, 1986, 1988, 1990		indicating may not be strong ordering for these two tasks in women *Hierarchy with second best compatibility rate reverses order of dressing and toileting
Kempen et al 1995 (97)	*Unknown *182 *74.5±7.7 *53.8% *Community-dwelling *Canada *Cross-sectional	*12 tasks *3 combinations: 1) no difficulty, some/a lot, don't do 2) no difficulty vs everything else 3) don't do vs everything else *PCA to investigate presence of underlying dimension, followed by Mokken scaling with suitable H and rho coefficents	*Going up and down stairs, cutting toenails, vacuuming, getting up from chair, preparing full meal, laundry, regular shopping, making beds, bathing or showering washing dishes, dressing, prepare light meal	*Mokken scaling: H coefficients indicated strong hierarchical scale for analysis 2 and average for analysis 1. Rho above minimum requirement of 0.7 and assumption of double monotony verified *do not report items in hierarchical order, order presumed from results

Appendix 3: Logistic regression results from the mutually-adjusted model in the hierarchy of loss analysis, detailing individual covariates

Table A3.1: Results of sex-adjusted logistic regression analysis with covariates

Covariates		Odds ratio (95% CI) of reported difficulty at the bottom of the hierarchy at age 60-64				
		Model 1	Model 2	Model 3	Model 4	Model 5
Difficulty at 60-64		**4.96 (2.70-9.11)***	**4.33 (2.34-8.02)***	**3.16 (1.63-6.14)***	**4.04 (2.16-7.56)***	**2.85 (1.45-5.60)***
Sex		1.02 (0.62-1.70)	0.92 (0.55-1.55)	0.79 (0.47-1.35)	1.03 (0.62-1.72)	0.75 (0.43-1.29)
Occupation		-	1.29 (0.87-1.92)	-	-	1.16 (0.76-1.76)
Education		-	**0.75 (0.61-0.93)**	-	-	0.81 (0.64-1.01)
Obesity	Developed	-	-	1.17 (0.59-2.33)	-	1.11 (0.55-2.22)
	Baseline	-	-	**2.72 (1.43-5.14)***	-	**2.25 (1.17-4.34)**
Respiratory	Developed	-	-	0.79 (0.27-2.33)	-	0.74 (0.25-2.21)
	Baseline	-	-	**2.57 (1.41-4.69)***	-	**2.34 (1.25-4.39)***
Depression	Developed	-	-	**3.22 (1.63-6.35)***	-	**3.19 (1.59-6.40)***
	Baseline	-	-	**5.25 (2.70-10.23)***	-	**5.01 (2.59-9.69)***
LTPA		-	-	-	**0.82 (0.70-0.96)**	0.95 (0.80-1.12)
Smoking	Ex-smoker	-	-	-	1.43 (0.70-2.91)	1.24 (0.60-2.58)
	Current	-	-	-	1.27 (0.44-3.63)	0.76 (0.24-2.34)

Model 1: Sex-adjusted (if men and women are combined) or unadjusted (if analyses sex-specific)
Model 2: Model 1 plus socioeconomic factors (education and occupational class)
Model 3: Model 1 plus incident health variables (depression, obesity and respiratory disease)
Model 4: Model 1 plus health behaviours (smoking and exercise)
Model 5: Fully-adjusted model
P<0.05 in bold, P<0.01 with asterisk

Table A3.2: Results of logistic regression analysis with covariates for men

Covariates		Odds ratio (95% CI) of reported difficulty at the bottom of the hierarchy at age 60-64				
		Unadjusted	Model 1	Model 2	Model 3	Fully adjusted
Difficulty at 60-64		2.72 (0.91-8.09)	2.29 (0.76-6.91)	1.60 (0.50-5.18)	2.50 (0.82-7.62)	1.54 (0.45-5.29)
Sex		-	-	-	-	-
Occupation		-	1.00 (0.53-1.90)	-	-	0.86 (0.43-1.71)
Education		-	**0.74 (0.55-0.99)**	-	-	**0.70 (0.50-0.97)**
Obesity	Developed	-	-	1.42 (0.55-3.69)	-	1.33 (0.50-3.55)
	Baseline	-	-	2.10 (0.75-5.88)	-	1.84 (0.64-5.22)
Respiratory	Developed	-	-	0.39 (0.05-3.10)	-	0.39 (0.05-3.14)
	Baseline	-	-	**3.06 (1.31-7.13)***	-	**3.86 (1.55-9.63)***
Depression	Developed	-	-	**4.16 (1.47-11.77)***	-	**4.51 (1.56-13.01)***
	Baseline	-	-	**5.30 (2.09-13.44)***	-	**5.43 (2.07-14.20)***
LTPA		-	-	-	0.87 (0.70-1.07)	1.00 (0.79-1.26)
Smoking	Ex-smoker	-	-	-	1.07 (0.44-2.64)	0.75 (0.28-2.01)
	Current	-	-	-	0.36 (0.04-2.91)	0.14 (0.00-1.39)

Model 1: Sex-adjusted (if men and women are combined) or unadjusted (if analyses sex-specific)
Model 2: Model 1 plus socioeconomic factors (education and occupational class)
Model 3: Model 1 plus incident health variables (depression, obesity and respiratory disease)
Model 4: Model 1 plus health behaviours (smoking and exercise)
Model 5: Fully-adjusted model
P<0.05 in bold, P<0.01 with asterisk

Table A3.3: Results of logistic regression analysis with covariates for women

Covariates		Odds ratio (95% CI) of reported difficulty at the bottom of the hierarchy at age 60-64				
		Unadjusted	Model 1	Model 2	Model 3	Fully adjusted
Difficulty at 60-64		**7.17 (3.36-15.34)***	**6.51 (3.01-14.07)***	**5.58 (2.35-13.26)***	**5.49 (2.51-12.00)***	**5.28 (2.19-12.73)***
Sex		-	-	-	-	-
Occupation		-	1.54 (0.91-2.60)	-	-	1.43 (0.82-2.48)
Education		-	0.75 (0.55-1.03)	-	-	0.85 (0.61-1.20)
Obesity	Developed	-	-	0.93 (0.33-2.64)	-	0.84 (0.29-2.43)
	Baseline	-	-	**3.09 (1.32-7.22)***	-	**2.60 (1.07-6.32)**
Respiratory	Developed	-	-	1.34 (0.37-4.92)	-	1.28 (0.34-4.75)
	Baseline	-	-	2.30 (0.98-5.39)	-	1.64 (0.66-4.05)
Depression	Developed	-	-	2.28 (0.85-6.09)	-	2.11 (0.78-5.72)
	Baseline	-	-	**6.68 (2.74-16.29)***	-	**6.63 (2.68-16.42)***
LTPA		-	-	-	**0.76 (0.60-0.96)**	0.89 (0.70-1.15)
Smoking	Ex-smoker	-	-	-	1.89 (0.66-5.40)	1.78 (0.61-5.18)
	Current	-	-	-	2.55 (0.68-9.59)	2.01 (0.48-8.38)

Model 1: Sex-adjusted (if men and women are combined) or unadjusted (if analyses sex-specific)
Model 2: Model 1 plus socioeconomic factors (education and occupational class)
Model 3: Model 1 plus incident health variables (depression, obesity and respiratory disease)
Model 4: Model 1 plus health behaviours (smoking and exercise)
Model 5: Fully-adjusted model
P<0.05 in bold, P<0.01 with asterisk

Appendix 4: Review of studies examining concordance between self-reported and performance-based measures of physical capability

Ordered chronologically and by first author's surname

Author, Year	Study Details: *Study population *Country *Sample size (N) *Age *Gender (% female) *Type of sample *Exclusion criteria *Study design	Self-reported measure: *Measure used *Specific section/domain used *Theoretical constructs *Type of data produced *Wording of question/response categories	Performance-based measure: *Measure used *Theoretical construct *Type of data produced *Items included	How is concordance operationalised? *Method of analysis used *Specific details	Results
Kelly-Hayes et al 1992 (132)	*Framingham cohort *United States *N=1453 *x̄ =72 (63-94) yrs *68% female *Community *Nursing home and incomplete data *Cross-sectional	*Questionnaire compiled for study *N/A *Physical capability and disability *Categorical *"Uses no help to perform", "uses device", "uses assistance of another person", "does not perform activity"	*Tasks relating to questionnaire *Physical capability *Categorical (same format as self-report) *Dressing, grooming, feeding, transferring, walking and stair climbing	*Cross-tabulation *Between each self-report and performance-based measure to determine frequency and direction of discordance	*Discordance: Dressing 3.9% Grooming 4.5% Feeding 3.1% Chair transfer 3.4% Walking 6.5% Stair climbing 6.5% *Self-reported disability greater than performance limitations (at least 89% of discordant cases)
Myers et al 1993 (34)	*Study specific *Canada *N=99 *74.4 ± 7.67 yrs *61% female *Community *None stated *Longitudinal	*Unclear *N/A *Physical capability and disability *Categorical *"No difficulty", "some difficulty", "a lot of difficulty", "do not do"	*Task relating to questionnaire *Physical capability *Continuous (time) and categorical (same format as self-report) *14 items: writing, reading time, using phone (2),	*Cross-tabulation *Mean time compared across two self-report groups: "no difficulty" and "difficulty" (some or a lot). Calculate % agreement between perceived difficulty	*No significant difference in mean performance time across nearly all items, especially mobility *Greater than 80% concordance found: time, phone, cooking and reaching. Only 55%

			opening container (5), reading, cooking, sweeping, reaching and walking	(self-report) and observer ratings (performance).	concordance with walking. *Concordance lower at follow-up, although similar numbers quoted
Guralnik et al 1994 (73)	*EPESE *United States *N=5174 *Older than 71 yrs *Unknown *Community *Resident in nursing home, telephone interview or proxy informant *Cross-sectional	*EPESE questionnaire *Lower extremity function *Physical capability *Categorical *Classified as having difficulty if needed help from another person or unable to perform ADL. To assess higher level mobility, asked about stairs and walking ½ mile.	*EPESE assessment *Physical capability *Categorical (based on quartiles or stands) *Standing balance, walking speed and chair rise	*Cross-tabulation and linear regression *Percent reporting need for help across induvial and summary performance test scores	*Higher performance capability associated with stepwise decrease in proportion reporting difficulty for both individual test (Table 2) and summary score (Table 3) *Self-reported measures explain 42% of variance in summary performance
Cress et al 1995 (117)	*No access to original paper *Unknown *N=417 *$\bar{x}$ =75.4 (64-91) yrs *Unknown *Community and nursing home *Unknown *Cross-sectional	*Sickness Impact Profile *Physical domain *Physical capability *Unknown *Unknown	*EPESE assessment *Physical capability *Unknown *Gait speed, balance, chair stands and grip strength	No access to original paper	Speed/care=0.53 Speed/amb=0.4 Speed/mobil=0.36 Grip/care=0.22 Grip/amb=0.21 Grip/mobil=0.14 Chair/care=0.39 Chair/amb=0.37 Chair/mobil=0.4 Bal/care=0.28 Bal/amb=0.25 Bal/mobil=0.28
Daltroy et al 1995 (186)	*Study specific *United States *N=289 *65-97 yrs (balanced across range)	*Health Assessment Questionnaire *N/A *Physical capability *Unknown	*Physical Capacity Evaluation *Physical capability *Continuous (rescaled 0 to 100)	*Spearman correlation *No details provided	*Correlation coefficient for PCE and HAQ global scores = -0.74 *35% of subjects indicated no disability on

	*54% female *Community *None stated *Cross-sectional	*Unknown	*Grip, writing, key turning, card turning, pegboard, dressing, shoulder extension and rotation, foot tapping, balance and TUG		the HAQ, yet PCE scores ranged over half the scale for this subgroup
Reuben et al 1995 (187)	*No access to original paper, study specific *Unknown *N=83 *64-92 yrs *54% female *Community *Unknown *Cross-sectional	*Functional Status Questionnaire (FSQ), Interview modified Katz ADL, OARS and SF 36 *FSQ: 2 scales of physical function BADL (3 items) and IADL (6 items) *Physical capability and disability *Unknown *Unknown	*Physical Performance Test *Physical capability *Unknown *7 items: writing, eating, lifting, dressing, picking up a penny, turning 360 degrees and walking	*Pearson correlation *No access to original paper	PM/modified Ka 0.30 (P<0.01) PM/OARS 0.56 (P<0.001) PM/FSQBADL 0.55 (P<0.001) PM/FSQIADL 0.56 (P<0.001) PM/SF36 0.26 (P<0.05) *Correlations are attenuated as a result of the lower reliability of measurement
Van den Ende et al 1995 (125)	*Study specific *Netherlands *N=51 *55 ± 13 yrs *52% female *Rheumatoid Arthritis patients *Suffered from other diseases affecting locomotor system *Cross-sectional	*Health Assessment Questionnaire *Dutch translation *Physical capability *Categorical *"Without difficulty", "some difficulty", "much difficulty", "unable", additional option to indicate use of aid	*Tasks relating to questionnaire *Physical capability *Categorical (scoring system based on expert consultation) *16 representative tasks from HAQ covering all categories except outdoors	*Cross-tabulation *Self-report and performance-based measure rescaled 0-3 then difference calculated (P-SR). If negative, underestimation of capability; if positive, overestimation. Considered discordant if difference >0.25 units.	*Average difference between self-report and performance score = 0.09 ± 0.39 *Range: -0.88 to 1.00 *Discordance >0.25 units for 24 patients (47%) *15 overestimated (6F, 9M) and 9 underestimated (7F, 2M)
Hoeymans et al 1996 (43)	*Zutphen Elderly Study *Netherlands	*WHO questionnaire *N/A	*EPESE assessment *Physical capability	*Spearman correlation *Coefficients calculated for	*Correlation coefficients between summary scores 0.22 in '90 and 0.39 in '93

	*N=494 *75.1 ± 4.6 yrs in '90 77.7 ± 4.2 yrs in '93 *100% male *Community *None stated *Longitudinal	*Physical capability and disability *Categorical *Dichotomised (need for help versus no need) in three dimensions (BADL, mobility and IADL)	*Categorical (dichotomise low performance) *Standing balance, walking speed, chair stand and external shoulder rotation	summary scores and for individual tests separately. Longitudinal correlation calculated between change in performance test and change in self-report, adjusting for baseline self-report	*Coefficients for mobility dimension '90/'93 Mob/bal=0.07 / 0.22 Mob/shoul=0.08 / 0.16 Mob/walk=0.17 / 0.32 Mob/chair=0.12 / 0.25 *Longitudinal: r=0.2
Kempen et al 1996a (40)	*Groningen Longitudinal Ageing Study *Netherlands *N=624 *Over 57 years *Unknown *Community *Severe cognitive impairment *Cross-sectional	*OECD indicator *Freedom of movement *Physical capability *Unknown *"Are you able to…", "can you..."	*Groningen Fitness Test for the Elderly *Physical capability *Continuous *Walking endurance, reach (x2), balance and grip	*Multiple regression analyses *Examine extent of variance explained in self-reported ADL by performance-based levels	Motor functioning R^2 = 33.1 (P<0.001) (Table 1)
Kempen et al 1996 b (112)	*Groningen Longitudinal Ageing Study *Netherlands *N=753 *M=71.9 ± 8.8 yrs F=73.6 ± 7.6 yrs *72% female *Community (but frail) *Severe cognitive impairment	*Groningen Activity Restriction Scale *11 item ADL subscale *Physical capability and disability *Categorical *"Can do it fully independently without any difficulty", "…with some difficulty", "…with great difficulty", "can only do it with someone's help". Note	*Longitudinal Ageing Study, Amsterdam *Physical capability *Continuous (time) *Putting on and taking off a jacket, walking 6m with a 180° turn halfway, 5 chair rises	*Multiple regression analyses *Examine extent of variance explained in self-reported ADL by performance summary score and individual items	*38.8% of ADL sum score variance explained by performance tests (P<0.01) *12.5% dressing variance explained by jacket test (P<0.01) *15.3% getting round house variance explained by walking test (P<0.01)

	*Cross-sectional	walking question "if necessary, with a cane".			*13.5% chair variance explained by chair rise (P<0.01)
Merrill et al 1997 (131)	*EPESE *United States *N=1458 *Over 71 yrs *68% female *Community (non-institutional) *Partially completed interviews *Cross-sectional	*EPESE questionnaire *N/A *Physical capability and disability *Categorical *"Does not need help", "needs help", "unable" or "no difficulty", "little/some difficulty", "a lot of difficulty", "unable"	*EPESE assessment *Physical capability *Categorical *Standing balance, walking, chair rise, shoulder rotation	*Cross-tabulation *Sensitivity and specificity calculated for corresponding measures, with performance-based measures considered "true positive". False negative = under report disability, false positive = over report disability	*Range of values across three matched analysis Sensitivity: 0.15-0.71 (M) 0.29-0.82 (F) Specificity: 0.91-0.94 (M) 0.83-0.90 (F) False +ve: 0.06-0.09 (M) 0.08-0.17 (F) False -ve: 0.29-0.85 (M) 0.18-0.71 (F)
Kivinen et al 1998 (39)	*Seven Countries Study *Finland *N=470 *76.5 ± 4.7 yrs *100% male *Community *Institutionalised or disabled living at home in remote municipalities *Cross-sectional	*WHO questionnaire *N/A *Physical capability and disability *Categorical *"No difficulty", "difficulty but without help", "only with help"	*EPESE assessment with addition of grip strength *Physical capability *Analysed as continuous (ordinal categorical scale) *Walking speed, chair stand, standing balance and grip strength	*Spearman correlation and cross-tabulation with logistic regression *Means and OR calculated within cross-tabulation analysis were adjusted for age using direct standardisation	*Correlation coefficients for ADL capacity and: standing balance=0.6053 chair stand=0.5050 walking=0.4775 grip=0.4284 (all P<0.001) *Clear stepwise decrease in ADL capacity with worse performance *Systematic increase in risk of disability with worse performance but discrepancies noted
Sherman and Reuben 1998 (130)	*Project Safety Net (RCT) *United States *N=363	*SF 36 and Functional Status Questionnaire (FSQ)	*National Institute on Ageing battery and Physical Performance Test *Physical capability	*Pearson correlation *Correlations calculated between scales, paired items	*Correlation coefficients: BADL/PPT: 0.37 (0.57) BADL/NIA: 0.42 (0.68) IADL/PPT: 0.49 (0.67)

	*75.9 ± 5.9 yrs *Unknown *Community *Not have one of: urinary incontinence, depression, history of falling or impaired functional status (cognitive impairment or institutional) *Cross-sectional	*BADL and IADL subscale of FSQ and physical function domain of SF-36 *Physical capability and disability *Continuous (score 0-100) *“Not limited at all”, “limited a little”, “limited a lot” or “usually done with no difficulty”, “some difficulty”, “much difficulty”, “usually did not do because of health”, “usually did not do for other reasons”	*Categorical *Balance, walking, writing, eating, dressing, picking up penny, lifting and turning	within scales and theoretical component of scale	IADL/NIA: 0.50 (0.73) PF 10/PPT: 0.46 (0.59) PF 10/NIA: 0.48 (0.65) *Correlation is highest when specific movement of PF 10 matched with specific movement of PPT (0.42) and NIA (0.49) *Matched item coefficients generally no better than pairs of dissimilar items eg. dress 0.24 vs 0.16-0.30 or penny 0.33 vs 0.17-0.32
Daltroy et al 1999 (129)	*Study specific *United States *N=289 *80 ± 7.4 yrs *54% female *Community *None stated *Cross-sectional	*Health Assessment Questionnaire *Pincus’ modifications *Physical capability and disability *Ordinal categorical *“No difficulty”, “some difficulty”, “much difficulty”, “unable”	*Physical Capacity Evaluation *Physical capability *Continuous *Grip, writing, key turning, card turning, pegboard, dressing, shoulder extension and rotation, foot tapping, balance and TUG	*Pearson correlation *HAQ score underwent probit transformation before correlation	HAQ probit and PCE were strongly and negatively correlated r=-0.72 P<0.001
Ferrer et al 1999 (72)	*Health Interview Survey of Barcelona *Spain *N=626 *79 ± 5.16 yrs *65% female *Community *Not live in Barcelona, institutionalised or	*Questionnaire compiled for study *2 items respond to performance-based measures *Physical capability *Categorical *“Difficulty”, “without difficulty” or “need for help”, “without need for help”	*EPESE assessment *Categorical (based on quartiles) *Physical capability *Gait speed and chair rise	*Cross-tabulation *Sensitivity and specificity were calculated for each comparison, for these analyses performance tests considered to be “true positive”; % bias calculated to indicate	*Walk: (similar for difficulty) Specificity 98% Kappa 0.55 False negatives 42% (unable to complete walk but did not report need for help) Chair: Specificity 92%

	had cognitive impairments *Cross-sectional			direction of discordance	Kappa 0.55 False negative 37% *Bias less than 25% and non-significant, around 50% over reported
Alexander et al 2000 (188)	*No access to original paper *Unknown *N=221 *x̄ =79.9 (60-102) yrs *Unknown *Community *Evidence of clinical dementia *Cross-sectional	*EPESE questionnaire *N/A *Physical capability and disability *Categorical *"Does not need help", "needs help", "unable" or "no difficulty", "little/some difficulty", "a lot of difficulty", "unable"	*Assortment of tasks *Physical capability *Unknown *Various versions of walking, balance and chair rise tests	No access to original paper	*Correlation with performance: Katz ADL walking 0.15-0.33 Total ADL score 0.21-0.35 Total Nagi score 0.11-0.24 *Total RB score most strongly related to all 3 PM (0.21-0.44), particularly with walking (0.44) *Self-reported walking was most strongly related to the 3 performance-based measures
Fried et al 2001 (134)	*Women's Health and Ageing Study II *United States *N=436 *51.9% 71-75 yrs *100% female *Community (2/3 least disabled, non-institutional) *Report difficulty in more than 1 domain, cognitive	*Questionnaire compiled for study *N/A *Physical capability *Categorical *"For health or physical reasons do you have any difficulty…?"; if no difficulty reported, "have you changed… due to underlying health problems?"	*Assortment of tasks *Physical capability *Continuous *Gait speed, chair rise, balance and strength (grip, hip flexion, knee extensor)	*Cross-tabulation with logistic regression *Mean performance compared across levels of self-reported function, odds of modification and difficulty calculated adjusting for age, race and education	*3 categories of self-report associated in step-wise or threshold relationship with walking speed, exercise tolerance, strength and balance *Task modification intermediate category in stepwise trend in mean and odds ratio

	impairment, partial data *Cross-sectional				
Simonsick et al 2001 (135)	*Health ABC Study *United States *N=3075 *x̄ =73.6 (70-79) yrs *52% female *Community *Reported ADL difficulty, life threatening cancer, intention to leave area within 3 years *Cross-sectional	*Questionnaire compiled for study *N/A *Physical capability *Categorical *"Do you have difficulty…?" If no, asked how easy task is to complete and ease of performing more demanding level. Regardless of difficulty level, asked if tired when completing and if completed less often than 12 months ago.	*EPESE assessment (modified) *Physical capability *Continuous (derived ratio score using maximal performance from data and published sources) *Balance, gait speed, chair rise and walking endurance	*Cross-tabulation and Pearson correlation *Cross tabulate mean performance against categories of self-report	*Mean performance scores, walking speed, % completing 400m walk all increased, and time to walk 400m decreased with increasing self-report capability *% able to complete 400m walk <6 mins across self-report groups: 85% "very easy" 70% "somewhat easy" 44% "not so easy" 51% "difficulty" *Correlations generally weak: 0.13-0.35 for women and 0.16-0.29 for men
Brach et al 2002 (71)	*Pre-existing RCT of walking *United States *N=170 *74.3 ± 4.3 yrs *100% female *Community *Resident in care home, were too sick to participate or unable to attend clinic	*Functional Status Questionnaire *Subscales of BADL, IADL and social activity selected *Physical capability and disability *Continuous (0-100 scale) *"No difficulty", "some difficulty", "much difficulty", "do not do because of health reasons"	*Physical Performance Test *Physical capability *Categorical *Writing, eating, lifting, dressing, picking up penny, turning 360 degrees and walking	*Cross-tabulation *2 items from self-report matched with performance tasks for analysis; both sets of measures dichotomised and sensitivity, specificity, false positive and negatives calculated	*High specificity (0.97-0.98) but low sensitivity (0.08-0.09) *More than 50% (99) concordant. Only 2 reported more difficulty than performance indicated, reverse discordance found in 69 women

	*Cross-sectional				
Jang et al 2002 (139)	*Charlotte County Healthy Ageing Study *United States *N=459 *72.4 ± 6.22 yrs *50.3% female *Community *None stated *Cross-sectional	*Modified Katz ADL *N/A *Physical capability *Categorical *“No difficulty”, “some difficulty”, “a lot of difficulty”, “unable to do”	*Assortment of tasks *Physical capability *Categorical (based on tertiles or stands) *Walking speed, standing balance, chair rise	*Pearson correlation *Correlation coefficients calculated between individual items and total scores	*Individual item correlation coefficients ranged from 0.21 to 0.45 *Correlation between total scores was 0.54 (P< 0.001)
Owens et al 2002 (121)	*Women’s Estrogen for Stroke Trial (RCT) *United States *N=620 *75% >65 (46-91) yrs *100% female *Postmenopausal women after cerebrovascular event *Cognitive impairment (MMSE<17) *Cross-sectional	*Barthel Index *N/A *Physical capability and disability *Categorical (cut points from previous research) *Not stated	*Physical Performance Test *Physical capability *Categorical (cut points from previous research) *Writing, eating, lifting, dressing, bending, turning and walking	*Cross-tabulation *Calculate % agreement, slight clinical disagreement defined as difference of one level, substantial clinical difference 2 or more levels	*Clinical agreement for 25.7%, slight disagreement for 55% and substantial disagreement for 19.3% *Agreement between task specific measures was low (22-26%), similar to overall results
Rogers et al 2003 (122)	*Project specific *United States *N=57 *81 ± 5.01 yrs *100% female *Knee OA patients *Other disabling pathology,	*PASS self-report *N/A *Physical capability and disability *Categorical *Asked about capability, “could do”	*PASS home *Physical capability and disability *Categorical *26 items: 5 functional mobility (FM), 3 personal care (PC), 14 cognitive IADLs, 4 physical IADLs	*Cross-tabulation *Percentage agreement calculated for each domain	*FM: 57.5% concordant, 5.3% overestimation of performance, 35.8% underestimation *PC: 69%, 17%, 13.5% *Cog IADL: 68.1% 17%, 8.4% *Phys IADL: 46.1%, 11.8%, 41.7%

	institutionalised, under 70 *Cross-sectional				
Van den Brink et al 2003 (189)	*FINE Study *Finland, Italy and Netherlands *N=1161 *75.4, 75.4, 77.0 yrs (in each country) *100% male *Community *None stated *Cross-sectional	*WHO Questionnaire *All items except toenail cutting *Physical capability and disability *Categorical *Responses unknown but categorised as: “Not disabled”, “disabled in IADLs only”, “disabled in IADLs and mobility”, “disabled in IADLs, mobility and ADLs”; any not categorised classified following procedure	*EPESE assessment *Physical capability *Categorical (reverse coded in line with self-report) *Balance, gait speed, chair rise and shoulder rotation	*Logistic regression *For each country association between self-report and performance assessed using polytomous logistic regression, adjusting for age, SES, household composition and prevalence of chronic disease	*Calculated odds ratio: Finland: 1.35 (1.21-1.50) Dutch: 1.23 (1.13-1.34) Italy: 1.30 (1.18-1.43) Combined: 1.28 (1.21-1.35) *Explore association further: men with IADL disability worse performance across all tests compared to those without, those with ADL disability only worse performance in balance and chair
Reuben et al 2004 (35)	*EPESE Study *United States *N=5138 *$\bar{x}$ =78.4 yrs *65% female *Community *Nursing home, proxy/telephone interview *Cross-sectional	*EPESE questionnaire *Selected items *Physical capability and disability *Categorical *Responses unknown, but categorised as: “Independent”, “dependent in mobility and independent in all ADLs”, “dependent in mobility and 1 or more ADLs”; any not categorised excluded	*EPESE assessment *Physical capability *Categorical (based on quartiles) *Balance, gait speed and chair rise	*Cross-tabulation *Cross classified 3 states of self-reported functional status and 4 states of PPS function to create 12 unique categories; only 11 categories had (N>10) enough for analysis (last category included those with low self-report but high performance)	*93% of those in top level of performance were at top level of the self-reported hierarchy *41% of those who were at top level of self-report were at top performance level *Those who self-reported ADL disability almost uniformly scored in bottom category of PPS
Fors et al 2006 (45)	*SWEOLD *Sweden	*Questionnaire compiled for study	*Tasks related to questionnaire	*Cross-tabulation	*Kappa statistic: Lower body: 0.50

	*N=492 *77 yrs or over *Unknown *Community (institutionalised persons included) *Proxy or telephone interviews *Cross-sectional	*Lower and upper body function *Physical capability *Categorical *Can you… without difficulty? "Yes", "no"	*Physical capability *Categorical (without difficulty, with difficulty, unable) *Chair rise and wrist rotation whilst holding 1kg	*Items from performance and self-reported measures matched, Kappa statistic and % discordance calculated	Upper body: 0.22 *Discordance: Lower body: 14.9% (7% underestimate, 7.9% overestimate) Upper body: 13.2% (7.1% underestimate, 6.1% overestimate)
Mallison et al 2006 (120)	*Study specific *United States (presumed) *N=64 at time 1 N=48 at time 2 *52.5 ± 13.6 yrs *75% female *Chemotherapy patients *Can't read English, motor or neurological problems *Longitudinal	*FACIT and SF 36 *Physical and functional wellbeing subscales of FACIT and physical function domain of SF 36 *Physical capability and disability *Unknown *Not stated	*Assessment of motor and process skills and assortment *Physical capability *Continuous (convert raw scores to logit) *Six-minute walk, grip strength, 2 observed activities of everyday life	*Spearman correlation *Correlation coefficients calculated at Time 1 (2 weeks after initial treatment) and Time 2 (between 3rd and 4th cycles of chemotherapy)	*Correlation coefficients (T1, T2): PF10/6mw: 0.36, 0.48 PF10/AMPs motor: 0.71, 0.56 FWB/6mw: 0.35, 0.33 *All other coefficients were not significant and <0.3
Sainio et al 2006 (116)	*Health 2000 Survey *Finland *N=2795 *70% 55-74 yrs 30% 7599 yrs *60% female *Community (include institutions) *None stated *Cross-sectional	*Questionnaire compiled for study *N/A *Physical capability *Categorical *"Without difficulties", "minor difficulties", "major difficulties", "not at all"; additionally: use of mobility aid	*Tasks relating to questionnaire *Physical capability *Categorical (able vs disabled) *Squatting and climbing stairs	*Cross-tabulation *Analysis conducted on self-report and performance test of stair climbing. Performance test defined as "standard" when direction of disagreement calculated. Age and	*Prevalence of limitation higher in performance (29%) than in self-report (26% P<0.001) at ages 75-99, 55-74 prevalence similar *Discordance observed: 17% *Kappa coefficient describing agreement

				gender analysed separately as interaction detected.	between methods was 0.58 *Over reporting 10%, under reporting 34%
Shulman et al 2006 (123)	*Study specific *United States *N=76 *65.6 ± 11.4 yrs *32% female *Diagnosed Parkinson's Disease *None stated *Cross-sectional	*OARS (modified) *N/A *Physical capability and disability *Categorical *"No difficulty", "slower or with greater difficulty", "need some help", "need moderate help", "completely unable"; asked to give best and worst score for past week	*Tasks relating to questionnaire *Physical capability *Categorical *Walking, eating, dressing, managing medication and handling money (asked if performing at best or worst to match with self-report)	*Cross-tabulation *Observed performance test compared to subjective score on each task to assess discordance; categorised as concordant, overraters or underraters regardless of magnitude of discrepancy	*Across all 5 items: 41.3% on average reported concordant 13% overrated 44% underrated *Greatest concordance occurred for dressing (53.6%) and walking (50.7%), least for money management (18.8%) *Kappas ranged 0.09-0.32, only walking and dressing were in the "fair" range
Stretton et al 2006 (114)	*FITNESS (RCT) *Australia and New Zealand *N=243 *x̄ =79 (74-84) yrs *53% female *Frail hospital admissions *Interventions indicated/ contraindicated, terminal illness or severe cognitive impairment	*SF 36 *Physical function domain and physical component summary score *Physical capability *Presumed continuous *Not stated	*Assortment of tasks *Physical capability *Continuous *Berg Balance Test, gait speed, TUG, knee extensor strength	*Spearman correlation *Coefficients calculated for each subscale of SF 36 against performance-based measures, analysis extended using linear regression and non-parametric generalised additive models to determine if there was curvilinear relationship between variables	*Correlations with (PCS, PF): Strength: 0.27, 0.47 Berg: 0.55, 0.78 TUG: 0.55, 0.75 Gait: 0.48, 0.73 *In univariate regression Berg predicted 51% of PF score, but relationship appears to be curvilinear

	*Data treated as cross-sectional				
Latham et al 2008 (119)	*Study specific (RCT treat muscle wasting) *8 European countries *N=108 *78.9 ± 8.1 yrs *73% female *Hip fracture *Fracture caused by major trauma/ bone pathology, condition affecting recovery, drug or alcohol abuse *Cross-sectional	*Activity measure for post-acute care (AM-PAC) and SF 36 *Physical mobility and personal care scale of AM-PAC and physical function domain of SF-36 *Physical capability and disability *Continuous *Not stated	*SPPB and assortment *Physical capability *Continuous *Balance, gait speed, chair rise, six minute walk, observed ADLs	*Spearman correlation *Correlations calculated using data collected at week 12 of follow up	*Correlation coefficients across all performance-based measures: AM-PAC mobility: 0.64-0.67 AM-PAC personal: 0.49-0.63 PF: 0.67-0.73 *Note higher correlations for two physical capability measures (AM-PAC mobility and PF
Rallon and Chen 2008 (113)	*Study specific *United States *N=30 *56.07 ± 11.87 yrs *67% female *Hand functional limitations *Cognitive impairments *Cross-sectional	*Manual ability measure *N/A *Physical capability *Ordinal categorical *"Easy", "a little hard", "very hard", "cannot do"	*TEMPA: upper extremity performance test for elderly *Physical capability *Numerical (score -3 to 0) *9 tasks related to ADLs, 5 bilateral and 4 unilateral	*Spearman correlation *Ordinal ratings of MAM 36 were transformed into interval measures through Rasch Rating Scale Analysis	*Correlation coefficients: Functional rating (FR) FR right hand/MAM=0.38* FR left hand/MAM=0.38* FR bilateral/MAM=0.68** FR combined/MAM=0.79** *P<0.05 ** P<0.01
Louie and Ward 2010 (133)	*NHANES 3 *United States *N=5396 *70.7 ± 0.2 yrs *57% female	*Questionnaire compiled for study *N/A *Physical capability *Categorical	*Assortment of tasks *Physical capability *Categorical (quartiles or dichotomised as complete or partial)	*Logistic regression *To maximise specificity of association, studied performance tests in	*Worse performance significantly associated with odds of reporting worse limitations across all tasks, for both

	*Community *Partial data *Cross-sectional	*When by yourself and without use of aids: “No difficulty”, “some difficulty”, “much difficulty”, “unable”	*Chair rise, eight foot walk, lock and key test, shoulder range of motion and active hip and flexion	relation to corresponding self-reported functions; adjusted for age, gender, ethnicity, education, co-morbidity, smoking and BMI	unadjusted and adjusted models (adjusted OR ranging from 1.13 to 7.84) *Associations not uniquely specific to corresponding task
Rogers et al 2010 (137)	*Pre-existing RCT *United States *N=148 *73.95 yrs *77% female *Community *<65 yrs, diagnosed with dementia *Cross-sectional	*PASS self-report *N/A *Physical capability and disability *Categorical *Abilities scale (can do) and habits scale (does do)	*PASS home *Physical capability and disability *Categorical *26 items: 5 functional mobility (FM), 3 personal care (PC), 14 cognitive IADLs, 4 physical IADLs	*Cross-tabulation *Calculate percentage agreement	*Concordance ranged from 12.9-64.1% *Overestimation of performance ranged from 12.5-87.1% *Underestimation of performance ranged from 0.0-41.5%
Waehrens et al 2010 (127)	*Study specific *Denmark (presumed) *N=50 *43.4 ± 9.6 yrs *100% female *Chronic pain or fibromyalgia *Language barrier, help with self-care, major psychiatric disorder or other medical condition *Cross-sectional	*Fibromyalgia Impact Questionnaire (FIQ) *Physical Function subscale *Physical capability and disability *Continuous (mean ordinal categorical score multiplied by 3.3 to achieve maximum of 10) *4 category scale not specified	*Assessment of Motor and Process Skills (AMPS) *Physical capability *Continuous (ordinal scale adjusted for task difficulty and rater severity, expressed as logits) *2 standardised ADL tasks appropriate to person	*Spearman correlation *Correlations calculated at baseline; if correlations were found, evaluate strength of correlation by calculating the percentage of the variation of the data that could be explained by the association between the two variables	*Correlation between FIQ and AMPS motor: r_s=-0.35, P=0.015 Amount of variation explained by association =12.25% *Found no correlation between FIQ and AMPS process ability measures: r_s=-0.02, P=0.92
Young et al 2010 (115)	*Women’s Health and Ageing Study	*Questionnaire compiled for study	*Assortment of tasks *Physical capability	*Cross tabulate	*Worse performance associated with lowest

	*United States *N=987 *x̄ =78 (65-101) yrs *100% female *Community (1/3 most disabled) *Difficulty in less than 2 functional domains, cognitive impairment *Cross-sectional	*N/A *Physical capability and disability *Categorical *If difficulty reported, asked about severity (little/some/a lot/unable); if no difficulty, asked if they had modified way they performed task or performed task less often	*Continuous *Knee extensor strength, balance (functional reach), gait speed, chair rise	*Mean values of performance-based measures compared across self-reported categories	self-reported functioning group *Statistically significant linear trend ($P<0.001$) observed across all comparisons except the association between functional balance and bed transfer
Bravell et al 2011 (174)	*OCTO Twin Study *Sweden *N=222 *83.2 ± 2.94 yrs *66% female *Community *One twin deceased, cognitive impairment, incomplete data *Cross-sectional	*Questionnaire complied for study *IADL, PADL and mobility scales *Physical capability and disability *Continuous *Each item scored 0 “unable to perform” to 4 “able to perform independently”, total calculated for each scale	*Assortment of tasks *Physical capability *Continuous (3-point scale across all items or time) *Balance: sway, tandem stands, gait and TUG Upper body dexterity: hand and arm manipulation Upper body strength and flexibility: lifting, pouring and reaching	*Pearson correlation and regression *Regression adjusted for gender, age, marital status, depression and objective health measures; semi-partial correlation coefficients (sr^2) calculated to show the unique contribution of a significant independent variable to the total R^2 for the equation	*Correlation coefficients: IADL: 0.34-0.54 PADL: 0.37-0.51 Mobility: 0.30-0.49 *sr^2 from regression: IADL/balance: 0.21 IADL/upper strength: 0.18 PADL/upper strength: 0.27 Mobility/balance: 0.19 Mobility/upper strength: 0.25
Hergenroeder et al 2011 (111)	*Study specific *United States *N=50 *51.2 ± 5.4 yrs *100% female *Sedentary adults	*Late life function and disability instrument (LLFDI) *Function and disability component *Physical capability and disability	*Assortment of tasks *Physical capability *Continuous *Gait speed, chair rise, six minute walk	*Linear regression *Advanced lower extremity score of LLFDI was used as independent variable and gait speed as the	*Beta coefficient=0.621, $P<0.001$ *Adjusted R^2=0.373 *F=30.17, $P<0.001$ *Gait=0.01 (LLFDI) +0.53

	*Impaired movement, major surgery, cancer, chest pain with activity, cardiac event < 6 months *Cross-sectional	*Continuous (raw score transformed 0-100 scale) *Not stated		dependent; model not affected by collinearity	
Farag et al 2012 (118)	*Study specific (RCT exercise intervention) *Australia *N=148 *84 ± 8 yrs *81% female *Hip fracture *None stated *Cross-sectional	*Questions compiled for study *N/A *Physical capability *Ordinal categorical *"Very poor", "poor", "fair", "good", "very good"	*Physical Performance and Mobility Examination (PPME) and assortment of tasks *Physical capability *Continuous *Knee extensor strength, gait speed, PPME: chair rise, balance range, step test, body sway and lateral stability	*Pearson correlation *Performance-based measures were matched to self-reported items for correlation at programme completion (16 weeks from start)	*Correlation coefficients: SR strength/knee: 0.17 SR strength/chair: 0.22 SR mobility/gait: 0.40 SR mobility/PPME: 0.45 SR balance/balance: 0.33 SR balance/step: 0.37 SR balance/sway: 0.31 SR balance/stability: -0.14
Van Weely et al 2012 (126)	*Study specific *Netherlands *N=126 *45.9 ± 11.5 yrs *29% female *Ankylosing spondylitis *Pulmonary, cardiovascular or neurological comorbidity affecting ADLs *Cross-sectional	*Bath Ankylosing Spondylitis Functional Index (BASFI) *N/A *Physical capability *Unknown *Not stated - No real details provided	*Tasks relating to questionnaire *Physical capability *Continuous (scores for pain and exertion also noted) *Climbing stairs, bending, reaching up, putting on socks, reclining and decline from chair, getting up off floor, looking over shoulder and physically demanding exercise	*Linear regression *BASFI as dependent variable; looking over shoulder component different domain to others and analysed separately	*Univariate for 7 performance-based measures and BASFI: R-squared=0.31 β=0.56 (P<0.001) *Multivariate: R-squared=0.54 Perform β=0.19 (P=0.018) Exertion β=0.38 (P<0.001) Pain β=0.26 (P=0.008) *Results similar for looking around variable
Waehrens et al 2012 (128)	*Study specific *Denmark (presumed)	*Questionnaire compiled for study	*Assessment of Motor and Process Skills (AMPS) *Physical capability	*Pearson correlation *Coefficients calculated across	*Low to moderate correlations (0.37-0.72) seen across and within

	*N=146 *50.9 ± 12.08 yrs *100% female *Rheumatoid arthritis or fibromyalgia *Medical condition preventing participation, did not meet diagnostic criteria *Cross-sectional	*ADL questionnaire and ADL interview *Physical capability and disability *Continuous (raw ordinal score converted into linear measure) *7 response categories (several may apply) recoded to ordinal scale: competent, ineffective, markedly deficient and unable	*Continuous (convert raw score adjusting for rater severity, skill item difficulty and ADL task challenge) *2 culturally relevant and familiar ADL tasks of appropriate challenge	whole sample and within each diagnostic group. 14 excluded from analysis because achieved maximal ADL-Q/-I scores (occurs when instrument lacks sufficient challenge to person). If score converted to linear, would result in high SE.	diagnostic groups between AMPS motor scale and ADL-Q/-I *Correlations between AMPs process and ADL-Q/-I were generally low across and within diagnostic groups (0.1-0.52)
Stuifbergen et al 2014 (124)	*Study specific *United States *N=60 *54.3 ± 7.9 yrs *76.7% female *Multiple sclerosis patients *Incomplete data across 5 time points, distance from clinic *Longitudinal	*Incapacity Status Scale *Gross and fine motor *Physical capability *Continuous (composite score of ordinal response options) *5 point scale: 0 "no difficulty" to 4 "functioning only with great difficulty or with assistance"	*Multiple Sclerosis Functional Composite Index *Physical capability *Continuous (standardised z-score) *Timed 25-Foot Walk (T25FW), 9-Hole Peg Test (9HPT)	*Pearson correlation *Correlations calculated for each of the 5 time points	*Pattern of correlations among measures is similar at each time point: *9HPT & ISS = 0.4-0.5 *T25FW & ISS = 0.5-0.7 *9HPT & Fine = 0.5-0.7 *T25FW & Gross = 0.8-0.9
Marsh et al 2015 (138)	*Study specific *United States *N=110 *80.6 ± 5.19 yrs *72.7% female *Community *Major medical or psychiatric	*vSPPB *N/A *Physical capability *Continuous (aggregated ordinal categorical scores) *"Can you…?" then chose video animation that best reflected their capability	*SPPB *Physical capability *Continuous (aggregated ordinal categorical scores based on quartiles) *Balance tasks, chair rise, timed walk	*Spearman correlation *Correlations calculated for total and component parts of each measure	*SPPB vs vSPPB total: 0.6 *Task-specific: 0.39-0.55 *Mixed: 0.25-0.59

	condition, cognitive impairment *Cross-sectional				

SPPB = Short Physical Performance Battery

Appendix 5: Review of studies examining discordance between self-reported and performance-based measures of physical capability

Ordered chronologically and by first authors surname

Author, Year	Study Details: *Study population *Country *Sample size (N) *Age *Gender (% female) *Type of sample *Exclusion criteria *Study design	Methods: *Treatment of variables *Identification of discordance *Analysis of associated factors	Factors investigated	Evidence of discordance: *Magnitude *Direction: A= low low B= low P high SR C= high P low SR D= high high	Factors associated with discordance
Kelly-Hayes et al 1992 (132)	*Framingham cohort *United States *N=1453 *x̄ =72 (63-94) yrs *68% female *Community *Nursing home and incomplete data *Cross-sectional	*SR and P measures categorised in same way (0-3) *Percentage agreement with McNemar's test used to determine significance of discrepancies *Multiple logistic regression models: discordance used as dependent variable	Age, cognitive impairment, perceived health, whether a person lives alone or with others, documentation of stroke or coronary heart disease	*Discordance: 3.1-6.6% *Direction highly systematic, at least 89% of time reported disability was greater than observed	*Age, cognitive impairment, prior stroke and perceived health strongly associated with discordance *Mobility more than self-care tasks associated with discordance
Van den Ende et al 1995 (125)	*Study specific *Netherlands *N=51 *55 ± 13 yrs *52% female	*SR and P measures categorised in same way (0-3) *Subtraction of functional scores with cut-off point of 0.25 to signify discordance	Age, gender, disease duration, number of prescribed second line drugs, number of swollen joints, duration of morning stiffness, pain,	*Difference between small scores was 0.09 but ranged from -0.88 to 1.00	*Discordance correlated moderately well with gender, age, disease duration and number of prescribed drugs

	*Rheumatoid Arthritis patients *Suffered from other diseases affecting locomotor system *Cross-sectional	*Forward multiple regression analysis with discordance as dependent variable	depression, anxiety, social activity	*Discordance >0.25: 15 over estimators (6f 9m) and 9 under estimators (7f 2m)	*Discordance= -0.26 +0.21 x male gender + 0.01 x years of disease duration + 0.06 x number of drugs
Kempen et al 1996a (40)	*Groningen Longitudinal Ageing Study *Netherlands *N=624 *Over 57 years *Unknown *Community *Severe cognitive impairment *Cross-sectional	*SR transformed and P continuous *N/A *Multiple regression analyses: Self-reported measure used as dependent variable	Age, gender, education, cognitive functioning (memory, verbal intelligence), depressive symptoms, feelings of anxiety, neuroticism, extroversion, mastery	N/A	*Discrepancies substantially affected by level of affective functioning (4.4% variance) and personality (5.9%) *Discrepancies higher for older subjects, females, those who are depressed or anxious, those with lower cognitive functioning, mastery and level of education
Kempen et al 1996b (112)	*Groningen Longitudinal Ageing Study *Netherlands *N=753 *M= 72.9 ± 8.8 yrs F=73.6 ± 7.6 yrs *72% female	* SR transformed and P continuous *N/A *Multiple regression analyses: Self-reported measure used as dependent variable	Age, gender, education, cognitive functioning, depressive symptoms, feelings of anxiety, neuroticism, extroversion, physical competence, mastery	N/A	*Variance explained by affective functioning (0.6-1.7%) and personality (2.7-8.2%) *Physical competence, and to a lesser extent mastery, influence the accuracy of the SR

	*Community (but frail) *Severe cognitive impairment *Cross-sectional				*Null findings for age, sex and education
Merrill et al 1997 (131)	*EPESE *United States *N=1458 *Over 71 yrs *68% female *Community (non-institutional) *Partially completed interviews *Cross-sectional	*SR and P measures dichotomised *Sensitivity and specificity *Log linear analyses provide significance testing for gender differences in false negatives and positives	Gender	*For each comparison, over reporting of disability is more common than underreporting	*Both men and women discordant either way, but more women under report ability and men over report ability *False negatives= Male: 0.29-0.47, female: 0.18-0.42 *False positives= Male: 0.08-0.13, female: 0.08-0.29
Daltroy et al 1999 (129)	*Study specific *United States *N=289 *80 ± 7.4 yrs *54% female *Community *None stated *Cross-sectional	*SR (transformed) and P measures rescaled 0-100 *N/A *Stepwise forward and backward regression models used to determine variables that predicted self-reported disability controlling for observed function	Joint pain or stiffness, prescription medication, urban/rural, arthritis, gender, depression, memory problems, exercise or walk 3+ times a week, dissatisfaction with function, functional decline in past 6 months	N/A	*33% of variance explained by increase in disability over last six months, level of dissatisfaction with ability in last week, gender and current pain or stiffness
Ferrer et al 1999 (72)	*Health Interview Survey of Barcelona	*SR and P measures dichotomised *Sensitivity and specificity	Proxy respondent, gender, age, level of education, perceived health	*For each comparison, although false	*Perceived health was most strongly and consistently associated

	*Spain *N=626 *79 ± 5.16 yrs *65% female *Community *Not live in Barcelona, institutionalised or had cognitive impairments *Cross-sectional	*Percent bias calculated to detect direction of disagreement *Logistic regression used to identify possible factors associated with under and over estimation of limitation		positives rates were higher than false negatives, a lack of systematic direction of disagreement *A: 2.5-14.6, B: 1.8-9.6, C: 2.0-12.9, D: 63-93.8	with discordance, influence remained after adjusting for number of chronic conditions *Proxy respondents, old age, education tend to underreport significantly less difficulty *Disagreement not influenced by sex
Brach et al 2002 (71)	*Pre-existing RCT of walking *United States *N=170 *74.3 ± 4.3 yrs *100% female *Community *Resident in care home, too sick to participate or unable to attend clinic *Cross-sectional	*SR and P measure dichotomised *Sensitivity and specificity *Characterise each concordant/discordant group, only descriptive stats due to uneven number in each group	Age, BMI, satisfied with health, comorbidities, used assistive device, median number of steps taken per day	*Eating, dressing and bathing: false positives 0.03, false negatives 0.92 *Walking: false positives 0.02, false negatives 0.91 *If discordant, more likely to over report ability *A=7, B=2, C=69, D=92 (count)	*Concordant groups lower BMI and more chronic conditions than discordant *Those with high P were younger *Those with low SR more likely to use assistive device *Those with high SR more active (more steps taken per day)
Jang et al 2002 (139)	*Charlotte County Healthy Ageing Study *United States *N=459	*SR and P measure categorised into four levels *N/A	Neuroticism	N/A	*Independent of actual performance, neuroticism 2% additional variance

	*72.4 ± 6.22 yrs *50.3% female *Community *None stated *Cross-sectional	*Hierarchical regression model for self-reported mobility			*Interaction term between actual performance and neuroticism further 2% variance
Owens et al 2002 (121)	*Women's Estrogen for Stroke Trial (RCT) *United States *N=620 *75% >65 (46-91) yrs *100% female *Postmenopausal women after cerebrovascular event *Cognitive impairment (MMSE<17) *Cross-sectional	*SR and P measure categorised into four levels *Percentage agreement (slight disagreement if difference of one level, substantial if 2 or more levels) *Association between baseline characteristics and disagreement assessed in bivariate (chi-squared) and multivariate (logistic regression with discordance as dependent variable)	Age, ethnicity, education, comorbidities, stroke as index event, severe index event, cognitive impairment, current smoker, active treatment assignment, depressive symptoms, no health locus of control, religion not a source of strength, unmarried, few friends and relatives, inadequate social support	*Concordance for 25.7%, slight disagreement for 55% and substantial disagreement for 19.3% *93.8% of slight and all of substantial over reported their function	*Several sociodemographic and clinical characteristics but none of the psychosocial associated *Substantial disagreement more likely to be non-white, less educated, have 2+ comorbid conditions, stroke as index event, severe index event and cognitive impairment * Non-significant depression retained in model as it substantially changed association between other variables
Rogers et al 2003 (122)	*Project specific *United States *N=57 *81 ± 5.01 yrs *100% female	*SR and P measures categorised in same way (0-3) *Percentage agreement and ANOVA	Domains of physical capability	*Discordance ranged from 30.5-53.5% *Most discordant for functional mobility and	*Significant variation across domains

	*Knee OA patients *Other disabling pathology, institutionalised, under 70 *Cross-sectional	* Analysis is separated into domains: personal care, functional mobility, physical IADL and cognitive IADL		physical IADL, least for personal care and cognitive IADL *Overestimated functional independence for personal care and cognitive IADL and underestimated it for functional mobility and physical IADL	
Fors et al 2006 (45)	*SWEOLD *Sweden *N=492 *77 yrs or over *Unknown *Community (institutionalised persons included) *Proxy or telephone interviews *Cross-sectional	*SR and P measures dichotomised *Percentage agreement *Multivariate models were analysed using logistic regressions with discordance as dependent variable	Age, gender, education, cognitive impairment	*No clear tendency toward under or over estimations *Lower body function: B=7.0%, C=7.9% Upper body function B=7.1%, C=6.1%	*Women were more likely to both over and under estimate limitations for upper and lower body function, but statistically significant only for upper body *Oldest group showed an increased non-significant risk of discrepancies *Cognitive status showed strong associations with under and over estimation in upper body limitations

Sainio et al 2006 (116)	*Health 2000 Survey *Finland *N=2795 *70% 55-74 yrs 30% 75-99 yrs *60% female *Community (include institutions) *None stated *Cross-sectional	*SR and P measures dichotomised *Percentage agreement	Age, gender	*N/A *A=19, B=10, C=7, D=64 (%)	*Discordance higher among those 75-99 than those 55-74 years among both sexes, and higher among women than men in both age groups *Over reporting more common among women and increased with age, while the opposite was true for under reporting
Shulman et al 2006 (123)	*Study specific *United States *N=76 *65.6 ± 11.4 yrs *32% female *Diagnosed Parkinson's Disease *None stated *Cross-sectional	*SR and P measures categorised in same way (0-5) *Percentage agreement *Differences between under and over reporters examined using t and χ^2 tests	Cognitive impairment, living situation, duration of illness, age, gender, education, employment status	*N/A *A+D: 40.8-53.6, B: 26.0-43.4, C: 11.5-23.3 (dressing, walking and eating) % If money and medication added, A+D: 18.8-53.6, C: 26.0-79.7, D: 1.4-23.3 %	*Under reporters of disability more likely to live with family, higher MMSE score, shorted duration of illness *Non-significant results: age, sex, education, employment status, mental/ physical quality of life
Rogers et al 2010 (137)	*Pre-existing RCT *United States *N=148 *73.95 yrs *77% female *Community	*SR and P measures categorised in same way (0-3) *Percentage agreement and Fisher's exact tests	Depression, cognitive impairment	*Pattern of discordance differed depending on domain and group	*Cognitive and physical IADL: control group more likely to overestimate than either depressed group

	*<65 yrs, diagnosed with dementia *Cross-sectional	*Analysis separated by group: Depressed + no mild cognitive impairment (MCI), depressed + MCI, control (neither)		*Pattern comparable across groups for functional mobility and personal care: concordance (56.6-79.7), underestimation (15.6-37.7), overestimation (0.0-7.6)	
Marsh et al 2015 (138)	*Study specific *United States *N=110 *80.6 ± 5.19 yrs *72.7% female *Community *Major medical or psychiatric condition, cognitive impairment *Cross-sectional	*SR and P assessed in same way (continuous scores from aggregated ordinal categories) *N/A *Multiple linear regression analyses	Age, sex, moderate to vigorous physical activity, order of assessment	N/A	*Null findings for: age, sex, physical activity *Order of tasks important

Appendix 6: Extended results for the association of socio-demographic and behavioural risk factors with discordance

Table A6.1: Socio-demographic and behavioural characteristics of the study population by sex

	Proportion of the sample (%) †			P Value‡
	Total Sample	Male	Female	
Maximum N*	(N=1981)	47.7	52.3	
Socio-demographic factors				
Education	(N=1873)			
None	28.9	30.4	27.7	
Up to O-Level	29.1	21.2	36.3	
A-Level or equiv	29.9	30.6	29.3	
Degree or higher	12.1	17.9	6.73	<0.001
Occupational Class	(N=1970)			
Low	12.2	8.00	16.1	
Medium	39.1	33.7	44.1	
High	48.6	58.3	39.8	<0.001
Marital status	(N=1784)			
Single	3.76	4.14	3.41	
Married	79.4	83.4	75.8	
Widowed	5.61	2.60	8.31	
Separated	11.21	9.82	12.5	<0.001
Behavioural risk factors				
Smoking History	(N=1956)			
Never	32.2	27.9	36.1	
Ex-smoker	56.1	60.3	52.4	
Current smoker	11.7	11.8	11.5	<0.001
Physical activity	(N=1932)			
Never	62.7	63.6	61.9	
Sometimes	14.6	13.5	14.6	
Frequently	23.4	22.9	23.4	0.7
BMI §	(N=1974)			
Normal	29.2	25.2	32.98	
Overweight	41.8	47.02	36.85	
Obese	29.0	27.77	30.17	<0.001
BMI (continuous) †	(N=1974)			
Mean (kg/m^2)	27.9	27.9	27.9	
SD	4.88	4.10	5.51	0.5

* Value of N varies for each factor due to missing data
† For continuous variables mean and standard deviation (SD) values provided instead of proportions and details of units provided
‡ P value from χ^2 test if variables categorical and from ANOVA if continuous variables
§ BMI categorised using standard WHO cut-points. Underweight individuals were grouped with those of normal weight.

Table A6.2: Socio-demographic and behavioural characteristics of the three concordant groups

	Total Sample	Proportion of sample (%) † Concordant groups Combined	Low	Medium	High	P value ‡
N*	1981	1708	172	1379	157	
Socio-demographic factors						
Sex	(N=1981)					
Males	47.70	48.54	37.21	49.31	54.14	
Females	52.30	51.46	62.79	50.69	45.86	0.002
Education	(N=1873)					
None	28.94	29.14	46.11	28.00	19.86	
Up to O-Level	29.10	29.26	34.13	29.92	17.81	
A-Level or equiv	29.90	29.94	16.77	30.22	42.47	
Degree or higher	12.07	11.67	2.99	11.86	19.86	0.001
Occupational Class	(N=1970)					
Low	12.23	12.42	23.98	11.52	7.69	
Medium	39.14	38.79	44.44	38.92	31.41	
High	48.63	48.79	31.58	49.56	60.90	0.001
Marital status	(N=1784)					
Single	3.76	3.56	6.90	3.09	3.42	
Married	79.43	79.33	71.03	80.19	80.14	
Widowed	5.61	5.70	9.66	5.59	2.74	
Separated	11.21	11.41	12.41	11.02	13.70	0.7
Behavioural risk factors						
Smoking History	(N=1956)					
Never	32.21	31.75	21.76	32.48	36.31	
Ex-smoker	56.13	56.34	58.82	56.14	55.41	
Current smoker	11.66	11.91	19.41	11.39	8.27	0.001
Physical activity	(N=1932)					
Never	62.73	62.21	84.24	61.63	43.42	
Sometimes	14.08	13.98	4.85	14.89	15.79	
Frequently	23.19	23.82	10.91	23.48	40.79	0.001
BMI §	(N=1974)					
Normal	29.2	29.2	18.0	29.0	43.3	
Overweight	41.8	41.8	28.1	43.2	43.3	
Obese	29.0	29.0	53.9	27.8	13.4	0.001
BMI (continuous)	(N=1974)					
Mean	27.90	27.90	30.91	27.76	25.89	
SD	4.86	4.91	7.28	4.49	3.73	0.001

* Value of N varies for each factor due to missing data
† For continuous variables mean and standard deviation (SD) values provided instead of proportions and details of units provided
‡ P value from test of trend
§ BMI categorised using standard WHO cut-points. Underweight individuals were grouped with those of normal weight.

Table A6.3: Socio-demographic and behavioural characteristics of participants with complete data and those missing data on covariates for the mutually-adjusted model

	Proportion of sample (%)			P Value*
	Total Sample	Complete cases	Missing data on covariates	
Socio-demographic factors				
Education	(N=1873)	(N=1799)	(N=74)	
None	28.9	28.3	44.6	
Up to O-Level	29.1	29.1	28.4	
A-Level or equiv	29.9	30.4	18.9	
Degree or higher	12.1	12.2	8.11	P=0.01
Occupational Class	(N=1970)	(N=1799)	(N=171)	
Low	12.2	12.1	13.5	
Medium	39.1	39.0	40.9	
High	48.6	48.9	45.6	P=0.7
Behavioural risk factors				
Smoking History	(N=1956)	(N=1799)	(N=157)	
Never	32.2	32.2	31.9	
Ex-smoker	56.1	56.1	56.7	
Current smoker	11.7	11.7	11.5	P=1.0
Physical activity	(N=1932)	(N=1799)	(N=133)	
Never	62.7	62.5	66.2	
Sometimes	14.6	14.5	8.27	
Frequently	23.4	23.0	25.6	P=0.1

* P value from χ^2 test for categorical variables

Table A6.4: Mutually adjusted model for socio-demographic and health risk behaviours, accounting for age

	RRR (95% CI) †	
	Underestimators	Overestimators
Sex		
Male	1.00	1.00
Female	4.27 (2.67 – 6.83)**	0.45 (0.30 – 0.68)**
	P<0.001	P=0.002
Education		
None	1.00	1.00
Up to O-Level	1.25 (0.71 – 2.17)	0.76 (0.45 – 1.28)
A-Level or equiv	1.45 (0.80 – 2.62)	0.85 (0.50 – 1.46)
Degree or higher	2.31 (1.12 – 4.76)*	1.09 (0.54 – 2.22)
	P=0.1	P=0.6
Occupational Class		
Low	1.00	1.00
Medium	2.37 (1.04 – 5.37)*	0.81 (0.45 – 1.45)
High	2.32 (0.99 – 5.44)^{m}	0.67 (0.35 – 1.28)
	P=0.1	P=0.5
Health behaviours		
Smoking History		
Never	1.00	1.00
Ex-smoker	1.27 (0.85 – 1.89)	0.62 (0.41 – 0.94)*
Current smoker	0.34 (0.12 – 0.97)*	1.00 (0.56 – 1.79)
	P=0.03	P=0.05
Physical activity		
Never	1.00	1.00
Sometimes	1.07 (0.64 – 1.80)	0.93 (0.53 – 1.62)
Frequently	0.83 (0.52 – 1.33)	0.66 (0.38 – 1.12)
	P=0.7	P=0.3
Age (continuous)		
Per 1 month increase in age	1.00 (0.99 – 1.01)	1.02 (1.00 – 1.04)*
	P=1.0	P=0.01

†Relative Risk Ratio, Reference group = concordant group

** Significant at P=0.01 level, * significant at P=0.05 level, ^{m} marginally significant

Appendix 7: Extended results for the association between markers of health status and discordance

Table A7.1: Health characteristics of the study population by sex

Factors of interest	Proportion of the sample (%) †			P Value ‡
	Total Sample	Male	Female	
Maximum N*	(N=1981)	47.7	52.3	
General Health				
Self-reported health	(N=1820)			
Excellent/V. Good	55.1	57.3	53.0	
Good	31.7	28.8	34.3	
Fair/Poor	13.2	13.8	12.7	0.05
Chronic conditions				
Respiratory	(N=1752)			
Never had	81.5	80.8	82.2	
Diagnosed	18.5	19.2	17.8	0.5
Cardio-metabolic §	(N=1739)			
No	87.1	83.6	90.3	
Yes	12.9	16.4	9.7	<0.001
Pain	(N=1806)			
Absent	32.5	37.2	28.3	
Present	67.5	62.8	71.7	<0.001
Fatigue	(N=1812)			
Absent	7.95	9.70	6.38	
Present	92.1	90.3	93.6	0.01
Mental health				
Depression	(N=1935)			
Not a case	82.7	87.3	78.5	
Case	17.3	12.7	21.5	<0.001
Wellbeing (cont.)	(N=1751)			
Mean (WEMWBS score)	51.8	51.9	51.8	
SD	7.96	7.75	8.16	0.8
Cognitive function				
Verbal Memory (cont.)	(N=1936)			
Mean (Words recalled)	24.4	23.2	25.5	
SD	6.11	5.95	6.04	<0.001

* Value of N varies for each factor due to missing data
† For continuous variables mean and standard deviation (SD) values provided instead of proportions and details of units provided
‡ P value from χ^2 test if variables categorical and from ANOVA if continuous variables
§ Presence of diabetes, stroke or MI. When present: 86% had only 1 condition, 12% had 2 conditions and 1 % had all 3 conditions

Table A7.2: Health characteristics of the three concordant groups

Factors of interest	Proportion of sample (%)†					P value ‡
	Total Sample	Concordant groups				
		Combined	Low	Medium	High	
N	1981	1708	172	1379	157	
General Health						
Self-reported health	(N=1820)					
Excellent/V. Good	55.05	55.33	10.81	58.56	71.81	<0.001
Good	31.70	31.54	27.70	32.84	24.16	
Fair/Poor	13.24	13.13	61.49	8.60	4.03	
Chronic conditions						
Respiratory disease	(N=1752)					
Never had	81.51	82.15	61.43	84.15	85.31	<0.001
Diagnosed	18.49	17.85	38.57	15.85	14.69	
Cardio-metabolic §	(N=1739)					
No	87.06	86.77	68.61	88.14	92.36	<0.001
Yes	12.94	13.23	31.39	11.86	7.64	
Pain	(N=1806)					
Absent	32.5	33.7	5.63	35.1	48.3	<0.001
Present	67.5	66.3	94.4	64.9	51.7	
Fatigue	(N=1812)					
Absent	7.95	8.17	4.86	8.32	10.1	0.1
Present	92.1	91.8	95.1	91.7	89.9	
Mental health						
Depression	(N=1935)					
Not a case	82.69	83.21	60.49	85.30	88.82	<0.001
Case	17.31	16.79	39.51	14.70	11.18	
Wellbeing (cont.)	(N=1751)					
Mean (WEMWBS score)	51.84	51.94	47.74	52.37	52.92	<0.001
SD	7.96	7.93	8.43	7.75	7.57	
Cognitive function						
Verbal Memory (cont.)	(N=1936)					
Mean (Words recalled)	24.42	24.48	22.15	24.52	26.44	<0.001
SD	6.11	6.11	6.43	5.94	6.51	

* Value of N varies for each factor due to missing data
† For continuous variables mean and standard deviation (SD) values provided instead of proportions and details of units provided
‡ P value from test of trend
§ Presence of diabetes, stroke or MI. When present: 86% had only 1 condition, 12% had 2 conditions and 1 % had all 3 conditions

Table A7.3: Sex-adjusted models for the association between the three component factors of the cardio-metabolic variable and discordance

Components of cardio-metabolic disease ‡	RRR (95% CI) †	
	Underestimators	Overestimators
Diabetes		
Absent	1.00	1.00
Present	1.06 (0.50 – 2.25)	0.84 (0.38 – 1.86)
	P=0.9	P=0.7
Angina or MI		
Absent	1.00	1.00
Present	0.83 (0.33 – 2.10)	0.44 (0.16 – 1.23)
	P=0.7	P=0.1
Stroke		
Absent	1.00	1.00
Present	0.00 (0.00-0.00)	1.73 (0.59 – 5.05)
	P=1.0	P=0.3

†Relative Risk Ratio, Reference group = concordant group
‡Sex interaction formally tested but no evidence found for all factors (P>0.07)

Table A7.4: Health characteristics of participants with complete data and those missing data on covariates for the mutually-adjusted model

	Proportion of sample (%)			P Value*
	Total Sample	Complete cases	Missing data on covariates	
Chronic conditions				
Pain	(N=1806)	(N=1561)	(N=245)	
Absent	32.5	32.7	31.4	
Present	67.5	67.3	68.6	0.7
Mental health				
Male Wellbeing (cont.)	(N=822)	(N=722)	(N=100)	
Mean (WEMWBS score)	51.9	52.0	51.1	
SD	7.75	7.65	8.38	0.3
Female Wellbeing (cont.)	(N=929)	(N=839)	(N=90)	
Mean (WEMWBS score)	51.8	52.0	50.4	
SD	8.16	8.10	8.62	0.08
Cognitive function				
Verbal Memory (cont.)	(N=1936)	(N=1561)	(N=375)	
Mean (Words recalled)	24.4	24.5	24.1	
SD	6.11	6.11	6.13	0.2

* P value from χ^2 test for categorical variables

Appendix 8: Extended results for the association of participants psychological and functional history with discordance

Table A8.1: Psychological and functional history characteristics of study population by sex

	Proportion of the sample (%)			P Value †
	Total Sample	Male	Female	
Maximum N*	(N=1981)	47.7	52.3	
Personality				
Extraversion	(N=1785)			
Less	34.1	29.2	38.5	
More	65.9	70.8	61.5	<0.001
Neuroticism	(N=1784)			
Less	54.8	66.3	44.6	
More	45.2	33.7	55.4	<0.001
Depression				
Late midlife experience	(N=1838)			
No symptoms	70.7	77.7	64.5	
Symptoms at age 53	12.7	10.6	14.5	
Symptoms at age 60-64	10.0	7.79	12.0	
Symptoms at both ages	6.64	3.95	9.00	<0.001
Physical capability				
Self-reported trajectory	(N=1799)			
Improved	1.83	1.86	1.81	
Reference	62.0	73.8	51.22	
Decline	36.1	24.3	47.0	<0.001
Performance trajectory	(N=1759)			
Stable low	5.57	5.45	5.67	
Decline	21.7	23.8	19.8	
Reference‡	70.4	68.7	72.0	
Stable high	2.33	2.06	2.57	0.2

* Value of N varies for each factor due to missing data
† P value from χ^2 test for categorical variables
‡Reference group = those who maintained physical capability within “normal” range

Table A8.2: Psychological and functional history characteristics of three concordant groups

	Total Sample	Concordant groups				P value†
		Combined	Low	Medium	High	
Maximum N*	(N=1981)	1708	172	1379	157	
Personality						
Extraversion	(N=1785)					
Less	34.1	33.9	31.0	34.3	33.6	
More	65.9	66.1	69.0	65.7	66.4	0.6
Neuroticism	(N=1784)					
Less	54.8	54.7	46.8	55.5	57.1	
More	45.2	45.3	53.2	44.5	42.9	0.07
Depression						
Late midlife experience	(N=1838)					
No symptoms	70.7	71.1	48.7	73.6	72.7	
Symptoms at age 53	12.7	12.9	14.3	12.3	16.0	
Symptoms at age 60-64	10.0	9.80	21.4	8.31	10.7	
Symptoms at both ages	6.64	6.22	15.6	5.75	0.67	<0.001
Physical capability						
Self-reported trajectory	(N=1799)					
Improved	1.83	1.87	2.16	1.90	1.34	
Reference	62.0	64.5	0.00	67.5	98.7	
Decline	36.1	33.7	97.8	30.6	0.00	<0.001
Performance trajectory	(N=1759)					
Stable low	5.57	5.24	39.9	1.86	0.00	
Decline	21.7	19.2	49.7	15.9	17.4	
Reference ‡	70.4	73.5	10.5	82.2	61.8	
Stable high	2.33	2.03	0.00	0.08	20.8	0.002

† P value from χ^2 test for categorical variables

‡ Reference group = those who maintained physical capability within "normal" range

Table A8.4: Sex-adjusted models for the association between discordance and depression (assessed using Colman's trajectories (168)

Factors of interest ‡	RRR (95% CI) †	
	Underestimators	Overestimators
Colman's trajectories		
Absence	1.00	1.00
AOMS	1.47 (0.89 – 2.41)	0.61 (0.33 – 1.13)
RMS	1.20 (0.78 – 1.83)	1.14 (0.75 – 1.71)
ASGAO	0.33 (0.10 – 1.07)	1.32 (0.70 – 2.48)
AOSS	2.08 (0.93 – 4.67)	1.25 (0.48 – 3.26)
RSS	1.64 (0.36 – 7.54)	--
	P=0.09	P=0.5

*A: Absence of symptoms, AOMS: Adult-onset moderate symptoms, RMS: Repeated moderate symptoms, ASGAO: Adolescent symptoms with good adult outcome, AOSS: Adult-onset severe symptoms, RSS: Repeated severe symptoms over the life course

Appendix 9: Sensitivity analysis for missing data and alternative discordant groupings

Table A9.1: Comparison of the socio-demographic and health characteristics of participants included in the analytical sample for this thesis and those excluded due to missing data

Factors of interest	Sample (%)		P value
	Complete cases†	Missing data ‡	
Sex	(N=1981)	(N=2697)	
Males	47.7	55.4	
Females	52.3	44.6	<0.001
Education	(N=1873)	(N=2559)	
None	28.9	47.8	
Up to O-Level	29.1	26.2	
A-Level or equiv	29.9	18.8	
Degree or higher	12.1	7.23	<0.001
Occupational Class	(N=1970)	(N=2230)	
Low	12.2	21.4	
Medium	39.1	45.9	
High	48.6	32.8	<0.001
Smoking History	(N=1956)	(N=1197)	
Never	32.2	26.1	
Ex-smoker	56.1	46.6	
Current smoker	11.7	27.3	<0.001
Depression			
Late midlife experience	(N=1838)	(N=1064)	
No symptoms	70.7	77.4	
Symptoms at age 53	12.7	18.1	
Symptoms at age 60-64	10.0	3.20	
Symptoms at both ages	6.64	1.14	<0.001

† Complete cases = participants with data for both measures of physical capability and individual covariates

‡ Missing data = participants missing data for both measure of physical capability but have data for individual covariates

Table A9.2: Breakdown of study population across deciles of summary performance-based and summary self-reported physical capability measures, showing alternative cut points for concordant and discordant groups*

	Performance deciles									
Self-reported deciles	1st (Low)	2nd	3rd	4th	5th	6th	7th	8th	9th	10th (High)
5th-10th (High)	29	97	119	117	126	127	141	147	151	157
4th	12	26	29	27	23	24	24	22	19	15
3rd	15	18	20	34	24	28	21	20	16	18
2nd	29	26	18	15	19	14	10	8	8	8
1st (Low)	112	31	13	6	5	4	3	2	4	0

* Original grouping shown in darkest shade. Progressively lighter shades indicate each alternative grouping.

Concordant groups: Blue= low capability, Purple= high capability, white= medium capability)

Discordant groups: Green= low performance, high self-reported capability (Overestimators)

Orange= high performance, low self-reported capability (Underestimators)

Table A9.3: Proportion of participants in the concordant and discordant group using different cut points to define each set of groups.

	Concordant and discordant groupings N (%)*		
	Original	Alternative 1	Alternative 2
Concordant			
Low	172 (8.68)	211 (10.7)	229 (11.6)
Middle	1379 (69.6)	987 (49.8)	888 (44.8)
High	157 (7.93)	308 (15.5)	308 (15.5)
Discordant			
Underestimators	135 (6.81)	192 (9.69)	244 (12.3)
Overestimators	138 (6.97)	283 (14.3)	312 (15.7)

*Cut points used to define each grouping shown in Table A9.2

Table A9.4: Comparison of results for alternative discordant groupings from the fully adjusted models for socio-demographic and health risk behaviours (N=1799)

Factors of interest	Alternative Grouping 1 RRR (95% CI) †		Alternative grouping 2 RRR (95% CI) †	
	Underestimators	Overestimators	Underestimators	Overestimators
Sex				
Males	1.00		1.00	
Females	3.87 (2.61 - 5.73)**	0.59 (0.43 – 0.78)**	3.19 (2.28 – 4.48)**	0.61 (0.47 – 0.81)**
	P<0.001	P<0.001	P<0.001	P=0.001
Education				
None	1.00	1.00	1.00	1.00
Up to O-Level	1.16 (0.74 – 1.80)	0.95 (0.66 – 1.36)	1.32 (0.88 – 1.97)	1.00 (0.71 – 1.43)
A-Level or equiv	1.01 (0.62 – 1.66)	0.87 (0.59 – 1.28)	1.18 (0.76 – 1.83)	0.83 (0.57 – 1.22)
Degree or higher	1.55 (0.83 – 2.90)	0.91 (0.53 – 1.54)	1.36 (0.76 – 2.44)	1.02 (0.62 – 1.69)
	P=0.4	P=0.9	P=0.5	P=0.7
Occupational Class				
Low	1.00	1.00	1.00	1.00
Medium	1.69 (0.93 – 3.07)[m]	1.00 (0.64 – 1.54)	1.85 (1.07 – 3.20)*	1.02 (0.67 – 1.56)
High	1.67 (0.89 – 3.14)	0.88 (0.55 – 1.41)	1.82 (1.02 – 3.22)*	0.91 (0.58 – 1.45)
	P=0.2	P=0.7	P=0.08	P=0.8
Smoking History				
Never	1.00	1.00	1.00	1.00
Ex-smoker	1.19 (0.84 – 1.69)	0.75 (0.55 – 1.02)[m]	1.20 (0.87 – 1.64)	0.83 (0.61 – 1.11)
Current smoker	0.55 (0.27 – 1.12)	1.12 (0.73 – 1.72)	0.59 (0.32 – 1.09)[m]	1.11 (0.73 – 1.69)
	P=0.07	P=0.06	P=0.05	P=0.2
Physical activity				
Never	1.00	1.00	1.00	1.00
Sometimes	1.28 (0.83 – 1.98)	0.90 (0.60 – 1.35)	1.05 (0.69 – 1.58)	0.91 (0.62 – 1.34)
Frequently	0.92 (0.62 – 1.37)	0.62 (0.42 – 0.90)*	0.85 (0.60 – 1.22)	0.56 (0.38 – 0.81)**
	P=0.4	P=0.04	P=0.6	P=0.009

†Relative Risk Ratio, Reference group = concordant group ** Significant at P=0.01 level, * significant at P=0.05 level, [m] marginally significant