2017

Measuring work disability in the U.S.: conceptual, methodological, and diagnostic considerations

BOSTON UNIVERSITY

SCHOOL OF PUBLIC HEALTH

Dissertation

MEASURING WORK DISABILITY IN THE U.S.: CONCEPTUAL, METHODOLOGICAL, AND DIAGNOSTIC CONSIDERATIONS

by

MOLLY ELIZABETH MARINO

B.S., Cornell University, 2009
M.P.H., Boston University, 2013

Submitted in the partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader

Alan M. Jette, Ph.D.
Professor of Health Law, Policy & Management

Second Reader

Lewis E. Kazis, Sc.D.
Professor of Health Law, Policy & Management
Boston University, School of Public Health

Research Assistant Professor of Medicine
Boston University, School of Medicine

Third Reader

Pengsheng Ni, MD
Research Associate Professor of Health Law, Policy & Management

Fourth Reader

Diane Brandt, Ph.D.
Protocol Manager, Rehabilitation Medicine Department
National Institutes of Health

Fifth Reader

William Frey, Ph.D.
Vice President and Study Area Director in the Health Studies Sector
Westat®, Inc.

ACKNOWLEDGEMENTS

I would first like to thank my committee chair and advisor, Dr. Alan Jette for his many years of mentorship, and his support and guidance in this work and other projects. Dr. Jette has set a tremendous example as an excellent researcher, mentor, and leader. I would also like to thank my other committee members Dr. Lewis Kazis and Dr. Pengsheng Ni for providing immeasurable support as I developed as a researcher. I learned so much from their suggestions, advice, and feedback. I also thank Dr. Diane Brant and Dr. William Frye, for being so kind to spend their time serving as external readers. I value their contribution and support.

This project uses data from a collaborative project involving researchers at Boston University, the NIH Department of Rehabilitation Medicine, Westat, Inc., and YouGov, Inc. Without all of the members of this team, I would have not been able to complete this endeavor. I would particularly like to thank my colleagues at the Health & Disability Research Institute, whom I have been working with since 2011. They have provided me with an ideal environment to learn and practice my developing skills as a health services researcher.

The faculty in the Health Services Research program in the Health Law Policy & Management have taught and supported me through two degrees from the School of Public Health. I have immense appreciation for the education they provided me, especially in challenging me to think critically about my own work. And to my compatriots in the HSR program, I value the friendships we have developed and the memories we have created over the past few years. They have been vital to my completing the program.

To my friends and family: thank you so much for all of your love throughout this process and throughout my life. I would particularly like to thank my parents for instilling in me the value of education, hard work, and life beyond school and work, and my brother and sister for being the best friends and support team I could ask for. And finally to my partner, Jonathan—I give my deepest thanks and love for years of encouragement, understanding, and love.

MEASURING WORK DISABILITY IN THE U.S.:

CONCEPTUAL, METHODOLOGICAL, AND

DIAGNOSTIC CONSIDERATIONS

MOLLY ELIZABETH MARINO

Boston University School of Public Health, 2017

Major Professor: Alan M. Jette, Ph.D., Professor of Health Law, Policy & Management

ABSTRACT

The Work-Disability Functional Assessment Battery (WD-FAB) is a self-reported measure, developed to allow the SSA to collect systematic and comprehensive information about claimants' functioning. It consists of eight scales: Basic Mobility, Upper Body Function, Fine Motor Function, Community Mobility, Cognition & Communication, Resilience/Sociability, Social Interactions, and Mood & Emotions.

Three studies were conducted to evaluate the WD-FAB and apply it as an outcome measure to examine questions relevant to work disability measurement.

Examining Activity Domain Structure of the International Classification of Functioning, Disability, and Health (ICF) empirically tests the structure of the WHO's ICF Activity subdomains by comparing it to the empirical structure of the WD-FAB. The comparison found good alignment between the physical function WD-FAB scales and ICF Mobility; several Activity subdomains converge into Cognition & Communication in the WD-FAB. Mental Functions and certain Interpersonal Interactions converge. A re-organization of the subdomains into distinct, measurable constructs is presented for future ICF revisions.

Who Applies to Social Security Disability Programs? Demographic and Functional Differences among Claimants examines how Social Security disability claimants compare sociodemographically to the working age US population, assesses differences in claimants' functional status by demographic characteristics, and showcases a method to detect Differential Item Functioning (DIF), which, once controlled for, minimizes measurement error. 17 items displayed DIF, primarily based upon gender. Claimants were sociodemographically different from the general sample and reported lower functioning. Within claimants, there were very few differences of consequence in function between different sociodemographic groups.

Determining Functional Profiles of Common Conditions explores the relationship between diagnoses and function. Common patterns of diagnoses among claimants were identified: Musculoskeletal, Cancer, Multisystem, Neurological & Sensory, and Mental conditions. Many of the diagnosis groups showed unique functional features. The identification of functional profiles for different condition groups suggests that WD-FAB scores may add value to the disability determination process.

There is no single litmus test for work disability, but incorporating self-reported experiences is becoming an increasingly common focus in the field. This work demonstrates how a conceptually grounded self-reported measure of functioning can be used to understand the condition of individuals whose health limits their ability to work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (cont.)

LIST OF FIGURES

LIST OF ABBREVIATIONS

CFA	Confirmatory Factor Analysis
DIF	Differential Item Functioning
ES	Effect Size
ICF	WHO International Classification of Functioning, Disability, and Health
IRT	Item Response Theory
LCA	Latent Class Analysis
SSA	Social Security Administration
SSDI	Social Security Disability Insurance
US	United States
WD-FAB	Work Disability Functional Assessment Battery

INTRODUCTION

The United States Social Security Administration's (SSA) disability programs provide financial support to over 10.3 million Americans through the SSDI program and 8.3 million additional individuals through the SSI program, making the programs the largest federal provider of financial assistance to disabled workers and their families.[1] In 2014 alone, over 750,000 recipients were awarded Social Security Disability Insurance (SSDI) benefits because of a work limiting health condition.[1] The Social Security Act defines work disability as an inability to engage in "substantial gainful activity due to any medically determinable physical or mental impairment that can be expected to result in death or to last for a continuous period of not less than 12 months."[1] The past two decades have seen a steady and at times precipitous rise in the numbers of SSDI applicants and beneficiaries.

Unlike many other trends in health care, the aging population is not a large driver of the volume increase, as almost all beneficiaries are under the age of 66. Once individuals reach the age of 66 years old, they are converted to the Social Security retirement program, which has a different funding source. There are, however, some demographic changes driving the growth. The influx of women in the workforce in the late 20th century increased the number of people who could become eligible if a health condition limited their ability to work. Younger people have also been applying for and receiving benefits in greater numbers, which means beneficiaries may be in the program longer. There have been shifts in the US economy, including fewer jobs for manufacturing and other traditional 'blue collar' workers, which was exacerbated by the recession that started in 2008. These shifts displaced people from the workforce, making them possibly more likely to apply for SSDI benefits.[2–4] On average, more than 15% of the newly awarded disabled workers receive their benefits due to a mental health disorder, and more than one third due to either musculoskeletal, nervous system, or sense organ conditions.[1] The proportion of individuals receiving benefits because of mental disorders has been increasing over the past two decades, as have individuals with chronic conditions such as low back pain and arthritis.[1]

SSDI application rates vary by state and sociodemographic factors, and the demographic makeup of the SSDI beneficiary population is quite different from that of the general working age population of the United States.[3–5] Compared to the U.S. population as a whole, SSDI beneficiaries are slightly more likely to be male and (although the SSA does not collect information on race and ethnicity) other studies indicate that beneficiaries are more likely to be African American/Black (around 25% compared to approximately 13% in the U.S.).[6] Although beneficiaries account for 4.8% of the US population, there are states where as much as 7% or more of the population receives SSDI benefits including Alabama, Arkansas, Kentucky, Maine, Mississippi, and West Virginia. There are other states where beneficiaries are less than 3% of the population.[1,3,4] What is not known is how much of this variation is due to differences in work disability prevalence, and difference in functional status, and how much is due to other factors, such as gender, race, ethnicity, age, education, job availability, and trends in the economy at large. Part of this dissertation focuses on better understanding variation through the better measurement of function as it related to Social Security work disability.

Challenges to Measuring and Determining Work Disability

The measurement and characterization of work disability is difficult for individuals with chronic conditions who often present with a heterogeneous array of symptoms and functional limitations. Chronic conditions can manifest themselves in a unique constellation of symptoms and functional limitations in different environments, some of which fluctuate in intensity, which can contribute to work disability. This suggests that work disability be defined by more than just the presence and severity of the condition or impairment itself. There are those with mental health or chronic musculoskeletal conditions who have minimal or no problem working; individuals who can work either part time or with accommodations; and, others who cannot work at all, and it is not always determined by the severity of a single biomedical condition.

The SSA evaluation process for determining work disability includes 5 sequential steps: First, the SSA verifies that a person is not currently employed in what is referred to as a substantially gainful activity (SGA). Second, the person applying for SSDI and SSI (referred to as a claimant) must demonstrate that a condition is present that interferes with what would be basic work activities. These first two steps are to verify the basic eligibility criteria for meeting the statutory definition of work disability. Then, the applicant's claims are compared against the SSA's 'Listing of Impairments,' which is divided into fourteen major body systems.[i] If the claimant's impairment falls short of

---

[i] SSA's 14 "Listing of Impairment" groups are: Musculoskeletal System, Special Senses and

specified severity criteria in the listings, then he/she advances to the next steps: previous (Step 4) and potential (Step 5) work capacity. Before these steps the SSA uses medical evidence and other relevant documents to complete what is called a Residual Functional Capacity (RFC) Assessment. There is both a Physical and Mental RFC. Using the RFC assessment and all of the other information about a claimant, the SSA will examine whether a claimant can perform relevant past work (Step 4) and if not any other substantially gainful work available in the national economy (Step 5). In deciding on whether or not a person can do other work, the SSA considers the person's medical conditions, age, education, work experience, and any transferable skills.

Conceptual Issues for Disability Measurement

In determining eligibility for work disability benefits, the SSA uses a definition of disability primarily rooted in the medical model, focusing on symptoms and diagnoses.[7] The relationship between symptoms and work performance is not always clear, and the weak relationship between them has been increasingly recognized as one of the fundamental challenges in work disability assessment.[8–10] The determination process does not have a systematic and standard assessment of a person's ability to perform certain tasks or activities that may affect an individual's potential ability to work; rather it relies primarily on diagnosis, and the RFC forms focus predominantly on signs, symptoms, and laboratory findings related to impairment. Although medical and diagnostic information is an integral part of understanding a persons' ability to work, the current

---

Speech, Respiratory System, Cardiovascular System, Digestive System, Genitourinary Disorders, Hematological Disorders, Skin Disorders, Endocrine Disorders, Congenital Disorders that Affect Multiple Body Systems, Neurological, Mental Disorders, Cancer, Immune System Disorders.

conceptualization of disability in the literature calls for a multidimensional approach. For example, in 2008 the IOM Committee on *The Future of Disability in America* explicitly identified conceptual confusion about disability and related concepts as a major barrier to future progress in preventing, measuring, and designing interventions focusing on disability.[11] The report recommended that a framework for disability should incorporate socioeconomic and psychosocial factors.

The current SSDI definitions and application process does not completely capture the complexity of current understandings of disability. In a comparison of conceptual models of disability and how the SSA defines and operationalizes disability, there are several areas where the current model was identified as falling short.[7] For example, the current process relies heavily on a single diagnosis, which fails to capture the increasingly common concerns of multi-morbidity. The primary disconnect lies in the finding that the current definition focuses on impairment caused by a health condition, rather than *functional ability*, which is a broader characterization of all body functions, ability to complete various activities, and participating in various life situations. Contemporary models define disability as the result of an interaction between an individual's health and their environment; disability is the difference between an individual's capacity limitations and the environmental demands. It is also a dynamic and complex process, not a single state of being. The gap between the statutory definition of work disability and the contemporary models of disability present an opportunity for a different conceptual model that can guide measurement, inform the SSDI/SSI application process, and better understand the phenomenon of Social Security work disability on an individual, group, and population level.

WD-FAB: A New Self-Report Measure of Work-Related Functional Activities

The Work-Disability Functional Assessment battery (WD-FAB) is a newly developed self-reported measure that utilizes Item Response Theory (IRT) and Computerized Adaptive Testing (CAT) administration to measure activity limitations relevant to work and work disability.[12–16] In 2011, the work disability Functional Assessment Battery (WD-FAB) was first created to allow the SSA to collect more systematic and comprehensive information about claimants' functioning to inform the SSDI disability determination process The battery consists of eight scales: four physical factors (Basic Mobility, Upper Body Function, Fine Motor Function, and Community Mobility), and four nonphysical factors (Cognition & Communication, Resilience/Sociability, Self-Regulation, and Mood & Emotions).[17–19]

In measures that employ Item Response Theory (IRT), each item is ordered in a hierarchy along a unidimensional construct, and scores are based upon probability models that represent the likelihood that a respondent selects an answer choice, given their ability on that construct.[12,20] Computerized Adaptive Testing is a method of administering IRT based instruments. The computer algorithm selects each subsequent question based upon a respondent's answer to previous questions. This adaptive approach allows for the skipping of items irrelevant to a person's functional status, which minimizes the burden of administration without sacrificing measurement precision.[14] For work disability assessment, there are a large number of areas (or domains) that need to be measured to capture the complexity of function, and there is a wide range of function of human ability to be measured for each particular construct. Therefore, the IRT/CAT approach to measurement development is particularly suited to measuring functional status relative to work capacity.

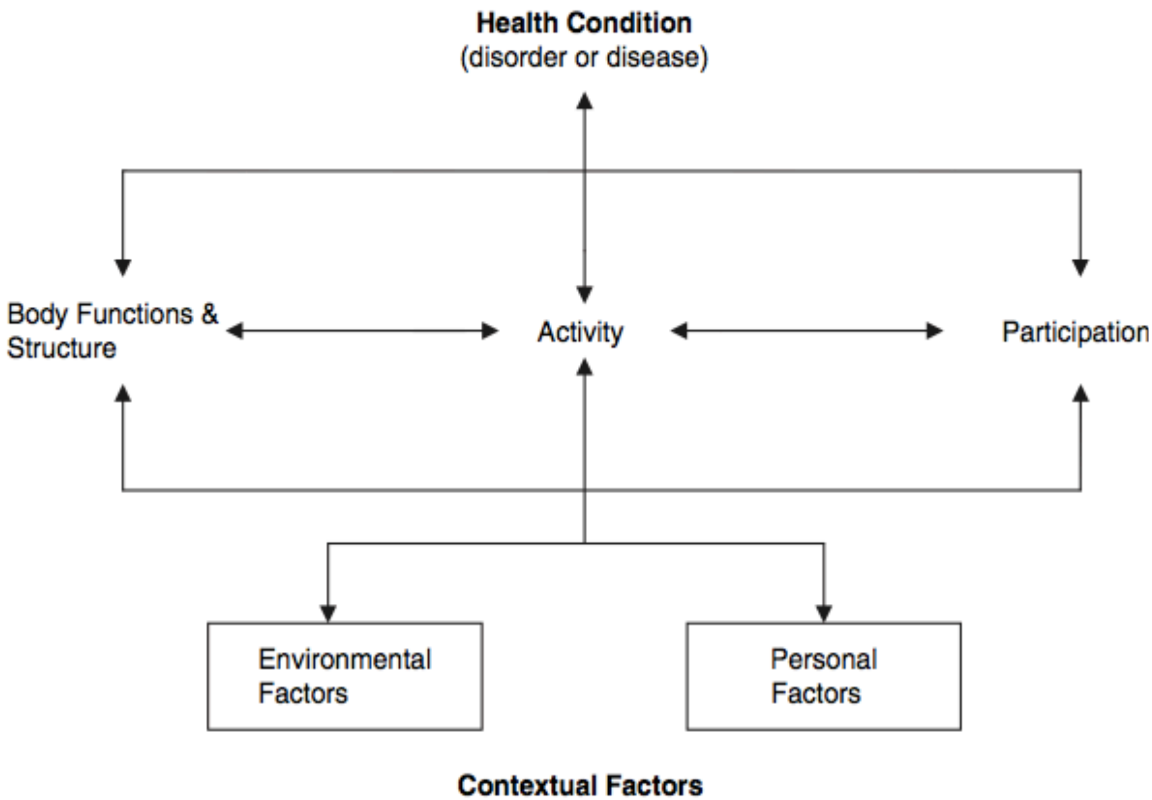
Using the International Classification of Functioning, Disability and Health (ICF) [21] as a guiding conceptual framework, items for the WD-FAB were developed using a comprehensive literature review and feedback from content experts. Cognitive interviews were conducted to assess that items were written clearly and interpreted consistently. The entire item pool was administered to two samples. One was a random sample that included a total of 3,793 SSA disability claimants who applied for benefits in the previous 2 months, stratified by urban/rural status across the ten national SSA regions, referred to as the "claimant sample." The second was a total of 2,100 working age adults, obtained using sample matching whereby samples representative of a study-appropriate target population could be constructed from large but unrepresentative pools of opt-in survey respondents, referred to as the "general sample."[22] Exploratory factor analyses were conducted to examine the structure of the items, and confirmatory factor analyses were conducted to determine that each factor was unidimensional. Item response theory analyses was used to calibrate the items using a graded response model.[23] Scores on the WD-FAB are presented on a metric based upon the score distribution of the general sample, with a mean of 50 and a standard deviation of 10, with higher scores representing higher functioning in the domain.

This dissertation analyzes data from the WD-FAB development project. The overall research goal of this dissertation is to evaluate the WD-FAB and apply it as an outcome measure to improve measurement of Social Security work disability. This study has the potential to enhance the field of work disability measurement conceptually, methodologically, and substantively.

Conceptual Framework

A primary limitation in the current SSA disability determination process is that it is symptom and deficit based, which contributes to lack of measurement prevision and likely increased variability in determinations. Current research and clinical practice indicate that health and functioning in the context of work involves a complex process influenced not only by individual disease factors as well as physical function, mental health, contextual and environmental factors. Such a multifaceted phenomenon must be studied using a comprehensive underlying theoretical framework of disability. One of the leading models for characterizing the multifaceted nature of disability is the World Health Organizations (WHO) International Classification of Functioning, Disability and Health (ICF).[21] The ICF calls for research on disability and health to be examined with a "whole person" perspective. The ICF is a biopsychosocial model of disability that characterizes disability as a result of the complex relationship between individual health characteristics, environmental factors, personal factors, and the extent to which those interactions affect a person's ability to perform activities and participate in daily roles such as work (**Figure 1**). The ICF sees disability as the result of the process between the interaction and the demands that surround the person and the person themselves. For example, a particular medical condition, or activity limitation is not enough to characterize someone has having a disability; it is a combination of these factors with the demands of an environment. This whole-person, biopsychosocial framework is well-suited as the foundational framework for examining the multidimensional issue of work disability assessment.

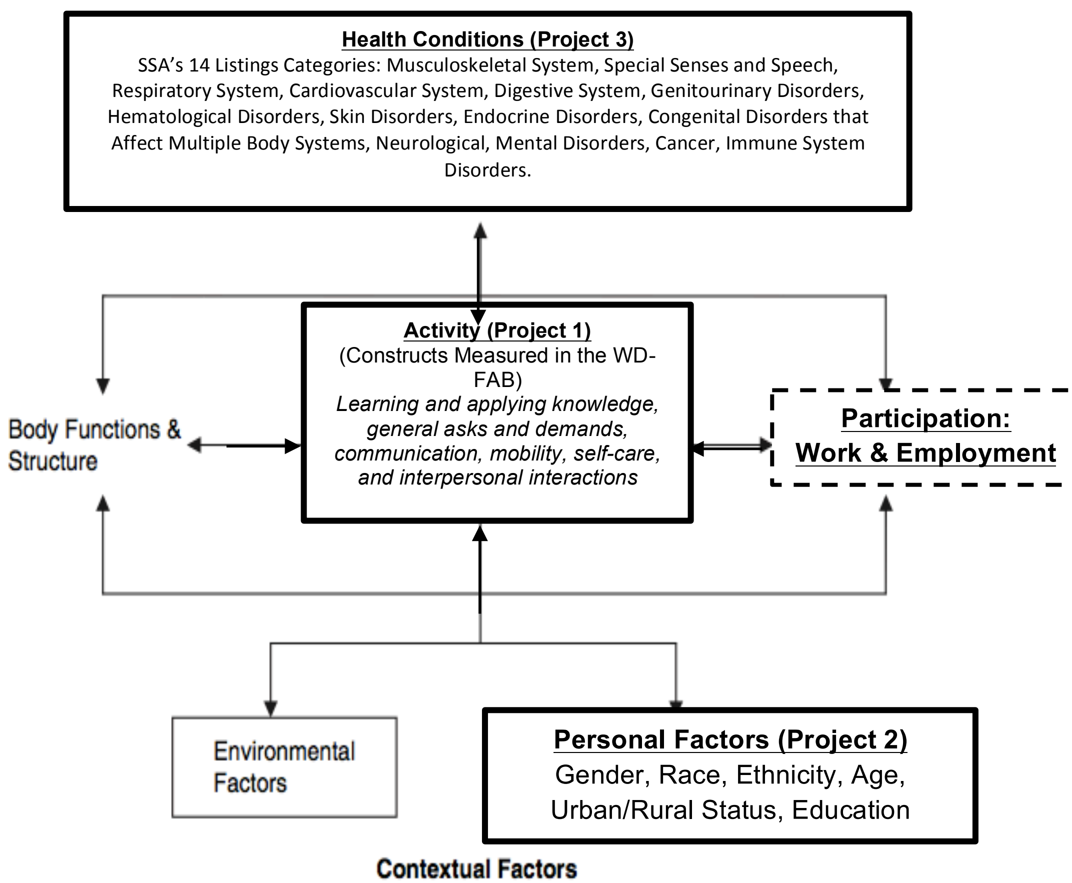
Figure 0.1. Conceptual Model: World Health Organization's International Classification of Functioning Disability and Health (ICF)



The two main categories of factors that contribute to health and disability in the ICF are *Health Conditions* and *Contextual Factors*. Health conditions are diseases, disorders and injuries. Contextual factors include *Personal Factors* such as age, sex, and social/cultural background and *Environmental Factors* such as physical characteristics of the natural and built environments, legal, organizational, and social policies and structures, and social attitudes. The ICF characterizes functioning into three levels; *Body Functions & Structures*, *Activity*, and *Participation*. *Body Functions* represents physiological and psychological functioning; *Body Structures* refers to the anatomy of organs and systems, including their component structures; *Activities* is defined as the performance of specific tasks or actions, such as rising from a chair or walking upstairs; and *Participation* refers to "involvement in a life situation" including performing the requirements of social roles. Activities and Participation encompass a wide range of complexity of tasks and activities, all of which could be considered potentially relevant to assess a person's ability to work. For each of the main ICF components several key domains have been identified. Activities and Participation are combined to provide common domains including: learning and applying knowledge; general tasks and demands; communication; mobility; self-care; domestic life; interpersonal interactions and relationships; major life areas; and community, social and civic life. The ICF is dynamic. Individuals are constantly interacting with other people and the environment; different contexts place different demands on people.

To make a framework specific to work disability, it is important to explicate exactly what 'activities' and 'participation' means in the context of work disability. Although the ICF presents activities and participation as conceptually different, they list all of the activities and life situations together, often blurring the lines between the two. 'Work and Employment' is listed within the activities and participation chapters of the ICF; though as a dynamic model, activities and areas of participation can influence and are influenced by each other along with the other domains in the model. The ICF includes 'Personal Factors' and 'Environmental Factors,' and those domains require special consideration when addressing the issue of work disability determination. Individuals make the choice to apply for SSDI benefits, and this choice is likely influenced by a variety of personal and environmental factors: how they perceive their own health, functioning, the job availability/options to them, and the social and cultural beliefs and attitudes about their condition and about the SSDI and other federal programs.

Figure 0.2. Conceptual Model for the proposed Dissertation, modified from World Health Organization's International Classification of Functioning Disability and Health (ICF)



Study one: Examining Activity Domain Structure of the International Classification of Functioning, Disability, and Health (ICF)

- *Purpose*: The ICF has been used for a variety of purposes, including as the conceptual foundation for measure development, including the WD-FAB. The ICF was developed via a worldwide collaborative consensus process among content experts. The overall structure of the framework has not been empirically examined. The aim of this paper is to empirically test the hypothesized/purported structure of the ICF Activity sub domains by comparing it to the empirical structure of the WD-FAB.
- *Methods:* Independent coders employed an ICF linking methodology to assign each item in the WD-FAB to an ICF subdomain. WD-FAB scales were developed using a series of exploratory factor analyses, and confirmatory factor analyses. The final models were based upon both statistical fit and content coverage. A cross tabulation of all items were grouped by ICF Chapter (subdomain) and WD-FAB scale was completed.

Study two: Who Applies to Social Security Disability Programs? Demographic and Functional Differences among Claimants

- *Purpose*: Social Security Administration disability claimants are unique: they have been out of work; have specific health conditions that they feel limits their ability to work their current or other jobs; and have made the choice to file an insurance claim. It is necessary to understand who these individuals are socio-demographically and if there is any relationship between individual characteristics and function. Individuals with different demographic characteristics have different experiences that may influence responses on self-reported measures; it is also important to know if individuals with different demographic profiles respond to questions in a systematically different way. This paper has three aims: to examine the how claimants compare demographically to the working age US population, to identify Differential Item Functioning (DIF) of items in the WD-FAB by gender (men v. women), race (white v. not white), and age (age <55 and ≥55) in Claimants, and to assess differences in claimants' functional status by gender, age, race, ethnicity, Urban/Rural Status, and education.

- *Methods:* Two-Step IRT based DIF methods were conducted to assess DIF across three pairs of groups: age (<55 vs. ≥55), gender, and race (white vs. non-white). Once WD-FAB scores are adjusted for DIF, independent t-tests were conducted to assess differences in the functioning scores across key subgroups. Where statistical differences existed in mean scores between the groups, analysis of magnitude of the mean differences was performed by calculating Cohen's d effect size.

Study three: Determining Functional Profiles of Common Conditions

- *Purpose*: The SSA disability determination process, and particularly its Listing of Impairments relies on the assumption that there are groups of individuals identifiable based upon medical impairments, and these impairments are related to functional impairments. The bases for the SSA listings are anatomic, diagnostic, and functional. In the past, 90% of beneficiaries received their determination at the listing stage; that number is currently at 52%. This brings into question the utility of the listings. However, in the absence of a gold standard, the listings themselves are difficult to measure and validate. In addition, the assumption that individuals in a single listing diagnostic group will have similar levels of functioning in similar ways has not been empirically assessed. This paper has two aims, first to identify common patterns of allegations among SSI and SSDI claimants; and second to develop unique functional profiles for the different diagnostic categories.

- *Methods:* Based upon the definitions in the SSA's "Blue Book," claimant's allegations were classified into a potential listing/body system group. A Latent Class Analyses (LCA) was then conducted based upon the 14 dichotomous listing/body system group variables. Independent t-tests and ANOVA with post-hoc Tukey tests used to assess differences in the average functioning scores across the latent classes identified. Radar plots were then generated to display the means scores for each group to illustrate the unique functional phenotype of each group.

References

1. Social Security Administration. Annual statistical report on the social security disability insurance program, 2014.

2. Wu AY. How has the changing nature of work contributed to SSDI growth? 2015.

3. Coe NB, Haverstick K, Munnell AH, Webb A. What explains state variation in SSDI application rates? *Center for Retirement Research at Boston College Working Paper*. 2011(2011-23).

4. Coe NB, Haverstick K, Munnell AH, Webb A. Why do state disability application rates vary over time? *Center for Retirement Research at Boston College Working Paper*. 2012(12-2).

5. Von Wachter T, Song J, Manchester J. Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. *Am Econ Rev*. 2011:3308–3329.

6. Wilson KB. Predicting vocational rehabilitation acceptance based on race, education, work status, and source of support at application. *Rehabilitation Counseling Bulletin*. 2000;43(2):97–105.

7. Brandt DE, Houtenville AJ, Huynh MT, Chan L, Rasch EK. Connecting contemporary paradigms to the social security administration's disability evaluation process. *Journal of Disability Policy Studies*. 2011:1044207310396509.

8. D'Amato A, Zijlstra F. Toward a climate for work resumption: The nonmedical determinants of return to work. *J Occup Environ Med*. 2010;52(1):67–80. doi: 10.1097/JOM.0b013e3181c75291 [doi].

9. Ganellen RJ. Assessing normal and abnormal personality functioning: Strengths and weaknesses of self-report, observer, and performance-based methods. *J Pers Assess*. 2007;89(1):30–40.

10. Krause N, Frank JW, Dasinger LK, Sullivan TJ, Sinclair SJ. Determinants of duration of disability and return-to-work after work-related injury and illness: Challenges for future research. *Am J Ind Med*. 2001;40(4):464–484.

11. Field MJ, Jette A. *The future of disability in America.* National Academies Press; 2007.

12. Lord FM. *Applications of item response theory to practical testing problems.* Routledge; 1980.

13. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care*. 2007;45(5 Suppl 1): S22–31. doi: 10.1097/01.mlr.0000250483.85507.04 [doi].

14. Norris JM. Computer-adaptive testing: A primer. *Language Learning & Technology*. 2001;5(2):23–27.

15. Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika*. 1989;54(3):427–450.

16. Wang S, Wang T. Precision of warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*. 2001;25(4):317–331.

17. Meterko M, Marfeo EE, McDonough CM, et al. The work disability functional assessment battery (WD-FAB): Feasibility and psychometric properties. *Arch Phys Med Rehabil*. 2014.

18. Ni P, McDonough CM, Jette AM, et al. Development of a computer-adaptive physical function instrument for social security administration disability determination. *Arch Phys Med Rehabil*. 2013;94(9):1661–1669.

19. Marfeo EE, Ni P, Haley SM, et al. Development of an instrument to measure behavioral health function for work disability: Item pool construction and factor analysis. *Arch Phys Med Rehabil*. 2013;94(9):1670–1678.

20. Hambleton R, Pitoniak M. Testing and measurement: Advances in item response theory and selected testing practices. 2002.

21. WHO I, World Health Organization. International classification of functioning. *Disability and Health (ICF), endorsed by all*. 2007;191.

22. Rivers D. Sample matching: Representative sampling from internet panels. *Polimetrix White Paper Series*. 2006.

23. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*. 1969.

CHAPTER ONE: EXAMINING THE ACTIVITY DOMAIN STRUCTURE OF THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY, AND HEALTH (ICF)

Introduction

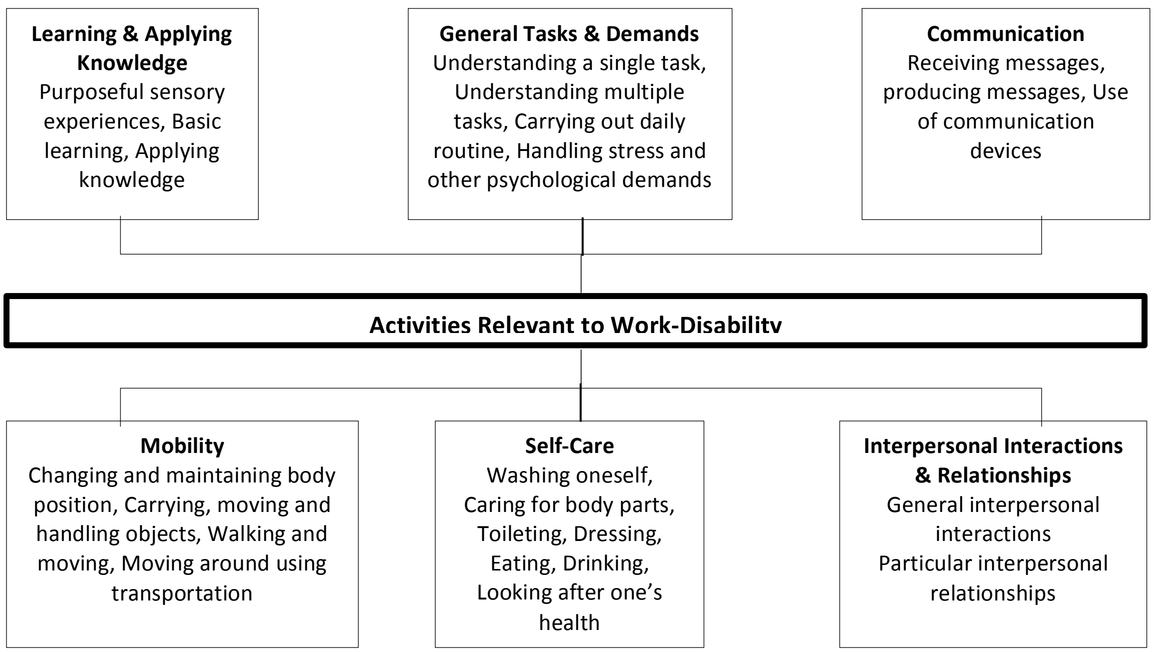
The World Health Organization's International Classification of Functioning Disability and Health (ICF) is one of the leading contemporary disablement models for characterizing the multifaceted nature of disability.[1] The ICF characterizes functioning into three levels; *Body Functions & Structures*, *Activity*, and *Participation*. *Body Functions* represents physiological and psychological functioning; *Body Structures* refers to the anatomy of organs and systems, including their component structures; *Activities* is defined as the performance of specific tasks or actions, such as rising from a chair or walking upstairs; and *Participation* refers to "involvement in a life situation" including performing the requirements of social roles. The ICF is dynamic; individuals are constantly interacting with other people and the environment; different contexts require different demands on individuals.

The ICF has been used for a variety of purposes, including as the conceptual foundation for measure development. A systematic literature review of publications related to the ICF found 54 papers corresponding to developing core sets (lists of categories that are relevant for specific health conditions and health care contexts), and 45 papers on the development of new tools (including self-reported outcome measures).[2] The ICF was developed via a worldwide collaborative consensus process among content experts.[3] The content of the framework has been assessed in several international studies.[4–8] The classification is exhaustive—meaning its many distinct classifications comprehensively cover the entire range of issues that impact health and functioning. The few studies that have examined the ICF's structure report that the ICF is somewhat limited in its ability to distinguish between different constructs. For example, the classification does not differentiate between different types of emotional functions, such as anger and happiness. Empirical analyses of the *Activities and Participation* domain have not found a distinction between the two concepts, and suggest that they both may actually blend across the multiple subdomains.[9,10] The overall structure, however, has not been systematically and empirically examined, and the need for such evaluation has been described in the literature.[3,7,11,12]

The Work-Disability Functional Assessment Battery (WD-FAB), is a new self-reported measure of activity limitations relevant to work and work disability.[13–17] It was created to allow the Social Security Administration (SSA) to collect systematic and comprehensive information about claimants' perceptions of their functioning to inform the disability determination process. The ICF's whole-person approach was well-suited as the foundation for examining the multidimensional issue of work disability assessment, and the WD-FAB uses 'Activity' as defined in the ICF as its conceptual foundation (**Figure 1.1**).[18] The framework encompasses a wide range of tasks and activities, all of which are considered potentially relevant to assess a person's ability to work. The key domains covered were: Learning and Applying Knowledge; General Tasks and Demands; Communication; Mobility; Self-Care; Domestic Life; Interpersonal Interactions and Relationships; Major Life Areas; and Community, Social and Civic life. The items for the WD-FAB were developed using a comprehensive literature review and feedback from content experts, while using many of the Activity subdomains in the ICF [1] as a guiding conceptual framework.[19] Cognitive interviews were conducted to assess that items were written clearly and interpreted consistently.

The aim of this paper is to empirically test the hypothesized/purported structure of the ICF Activity sub domains by comparing it to the empirical structure of the WD-FAB.

Figure 1.1 Activities from the ICF hypothesized within the WD-FAB Content Model



Methods

Step 1: ICF Linking

The ICF linking methodology[8,20,21] has been used to evaluate a variety of health status measures.[22] The process involves assigning each item in the measure to the ICF, using the unique classification and nomenclature of the framework. As previously described, the ICF has two parts, each containing two separate components. Part 1 covers functioning and disability and includes Body Functions and Structure and Activities and

Participation. Part 2 covers Contextual Factors and includes Environmental Factors and Personal Factors. In the ICF classification, the letters b, s, d, and e, which refer to the components of the classification, are followed by a numeric code starting with the chapter number (one digit) followed by the second level (two digits), and the third and fourth level (one digit each). For example, the *Activity* classification contains the following nested codes: d3: communication, d350: conversations, and d3501: sustaining a conversation.

Each item from the WD-FAB was organized into a single column, each in their own row of a master spreadsheet. Two independent coders (MEM and MDS) classified each item of the WD-FAB into its most precise ICF Activity category based upon the content of the item stem and the definitions of the codes at each level in the ICF, using the complete ICF book (which contains examples) as a reference. A key feature of the ICF is the taxonomy which includes both negative and positive terms for each concept, which facilitated item classification. Throughout this process, the coders were blinded to which scale of the WD-FAB the item belongs, in order to minimize bias in ICF category placement. Each coder kept notes regarding questions or rules applied during the coding process. The two then met to reconcile differences in coding. When a decision was not clear, the coders consulted with content experts who are familiar with the ICF and the WD-FAB to determine appropriate classification.

Step 2: WD-FAB Scale Development

This study uses data from the development of the WD-FAB. The entire item pool was administered to two samples: 3,720 SSA disability claimants who applied for benefits in the last 2 months, stratified by urban/rural status across the ten national SSA regions, and 2,100 working age adults, obtained using sample matching whereby samples representative of a study-appropriate target population could be constructed from large but unrepresentative pools of opt-in survey respondents.[23]

We conducted a series of exploratory factor analyses to examine the structure of the items. The factor structure was determined based upon their interpretability, and items with loadings greater that 0.3 were included. Next, confirmatory factor analyses (CFA) were conducted to determine that each factor was unidimensional. The fit criteria used were Root Mean Square Error of Approximation (RMSEA) ≤ 0.1, Confirmatory Fit Index (CFI) & Tucker-Lewis Index (TLI) ≥ 0.9.[24–26] After we determined the model with optimal fit, we examined each solution from a conceptual perspective to ensure that the items retained contained relevant content. Therefore, the final model is based upon both statistical fit and content coverage.

Step 3: Content Organization

A cross tabulation of all items were grouped by ICF Chapter (subdomain) and WD-FAB scale was completed. This allowed for an initial summary of how many ICF constructs the WD-FAB items represent, and what ICF content was within each WD-FAB scale. The content within each scale was examined individually to characterize the content of the WD-FAB in ICF terms. Finally, a revised organization of Activities was created based upon the findings.

Results

ICF Linking

The WD-FAB contains a total of 309 unique items, each of which were classified independently by two investigators. At the chapter/subdomain level, there was disagreement on 34 items (11%) after the initial independent coding. At the three-digit code level, there was 73.46% total agreement between the two coders. Agreement was better on the physical function items (86.98%) than the mental health function items (61.35%). After meeting to reconcile differences, there were only eleven physical function items and nine mental function items that were brought to a third content expert for consultation. Because some ICF subdomains only have classification up to the third level and others the fourth, all data presented here will focus on the three-digit classification of each item. **Table 1.1** provides an example of WD-FAB item content and final ICF classification.

The WD-FAB contains content from nine chapters of the ICF. Eight of those chapters are within Activities, and one chapter—Mental Functions—are within Body Functions. Although Body Functions and Mental Functions were not initially planned to be part of the coding, it was clear to both coders that many items did belong in this category. This was particularly true for items related to temperament, energy/drive, attention, memory, thought, cognitive and emotional functions. The only Activity chapter not included in the WD-FAB was Major Life Areas, which involves education, work & employment, and economic life.

Table 1.1 Examples of WD-FAB Items by Scale and Final ICF Classification

WD-FAB Scale	Item Content	ICF Chapter	Final ICF Code
Basic Mobility	Are you able to bend to look under a car?	Mobility	d4105 Bending
Upper Body Function	Are you able to push open a heavy door?	Mobility	d4451 Pushing
Upper Extremity Fine Motor Function	Are you able to remove a gas cap from a car?	Mobility	d4453 Turning or twisting the hands or arms
Community Mobility	Are you able to get on to a bus or train?	Mobility	d4702 Using public motorized transportation
Cognition & Communication	I have trouble putting my thoughts together.	Learning & Applying Knowledge	d163 Thinking
Self-Regulation	I have difficulty following the rules.	Interpersonal Interactions and Relationships	d7203 Interacting according to social rules
Self-Regulation	People know that I get angry very easily.	Mental Functions	b1521 Regulation of emotion
Resilience/Sociability	I am able to adjust in to other people's ways.	Interpersonal Interactions and Relationships	d7202 Regulating behaviors within interactions
Resilience/Sociability	If I make a mistake I know I can deal with it.	Mental Functions	b1266 Confidence
Mood & Emotions	In the past 7 days, it was hard to keep up enthusiasm to get things done.	Mental Functions	b1301 Motivation

WD-FAB Scale Development

The WD-FAB development project led to the creation of eight scales: four physical function scales (Basic Mobility, Upper Body Function, Fine Motor Function, and Community Mobility), and four mental health function scales (Cognition & Communication, Resilience/Sociability, Self-Regulation, and Mood & Emotions).[27–29] Results from the CFA indicated acceptable fit statistics across all mental health subdomains in both samples and all scales demonstrated acceptable fit statistics meeting the RMSEA<=0.1, CFI & TLI >=0.9 criteria. **(Table 1.2)**

	Claimant Sample			**Working-Age Adult Sample**		
Subdomain	**CFI**	**TLI**	**RMSEA**	**CFI**	**TLI**	**RMSEA**
Basic Mobility	0.95	0.95	0.08	0.99	0.99	0.05
Upper Body Function	0.98	0.98	0.07	0.99	0.98	0.05
Fine Motor Function	0.96	0.95	0.07	0.99	0.99	0.03
Community Mobility	0.95	0.95	0.05	0.94	0.94	0.08
Cognition & Communication	0.98	.097	0.04	0.90	0.90	0.06
Self-Regulation	0.91	0.91	0.06	0.91	0.09	0.02
Resilience/Sociability	0.91	0.91	0.07	0.92	0.92	0.07
Mood & Emotions	0.95	0.94	0.08	0.95	0.94	0.080

Table 1.2 Confirmatory Factor Analysis Results for the WD-FAB

Basic Mobility involves moving or changing body position or location by transferring from one place to another by crawling, walking, running, etc. Upper Body Function contains activities such as reaching, lifting, pulling, pushing, and carrying. Fine Motor Function is primarily the manipulation of objects requiring dexterity. Community Mobility is defined as using various forms of transportation, public and driving. Cognition & Communication involves conveying information and thoughts, receiving and producing messages, and carrying on conversations through language, signs, and symbols; applying knowledge that is learned, thinking, solving problems, and making decisions. The Resilience/Sociability scale relates to adaptability, coping with stress, confidence, and conscientiousness. Self-regulation involves carrying out the actions and tasks required for basic and complex interactions with people in a contextually and socially appropriate manner. Finally, the Mood & Emotions scale measures feelings, temperament, energy, and vitality.

Content Organization

Table 1.3 shows the count of items cross tabulated by WD-FAB scale and ICF Chapter. The Cognition & Communication scale draws primarily from the Learning & Applying Knowledge, General Tasks & Demands, and Communication Chapters, but also contains items related to Self-Care, Interpersonal Interactions and Relationships, and Mental Functions. Most items from the Basic Mobility, Upper Body Function, Fine Motor Function, and Community Mobility scales map into the ICF Mobility Chapter. However, those scales also contained items related to General Tasks & Demands, Communication, Self-Care, and Domestic Life.

The Resilience/Sociability scale is primarily based on Interpersonal Interactions & Relationships, but also include Learning & Applying Knowledge, General Tasks & Demands, Community, Social & Civic Life, and Mental Functions. Most of the items in the Mood & Emotions and Self-Regulation scales are not directly found in any chapter in the Activities sections of the ICF, but rather are better classified within the Body Functions subdomain, Mental Functions. However, these scales also drew from the Interpersonal Interactions & Relationships, Learning & Applying Knowledge, and General Tasks & Demands.

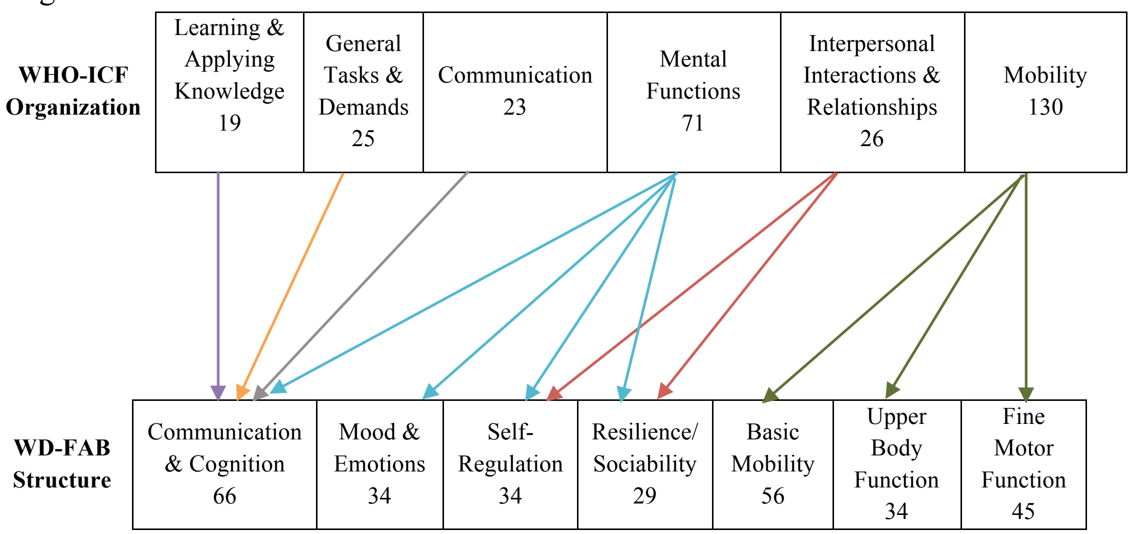
Table 1.3 Cross-tabulation of item classification by WD-FAB scale and ICF Chapter

		WD-FAB Scales (Empirically Derived)							
		Basic Mobility	Upper Body Function	Fine Motor Function	Community Mobility	Cognition & Communication	Self-Regulation	Resilience/ Sociability	Mood & Emotions
	Total Items:	56	34	45	11	66	34	29	34
ICF Chapters									
Learning & Applying Knowledge	19					15	1	1	2
General Tasks & Demands	25				2	12	2	6	3
Communication	23			1	2	20			
Mobility	130	54	27	42	7				
Self-Care	3			1		2			
Domestic Life	10	2	7	1					
Interpersonal Interactions & Relationships	26					1	11	12	2
Community, Social & Civic Life	2							1	1
Mental Functions*	71					16	20	9	26

*Mental Functions ICF Chapter it not within Activities & Participation

Figure 1.2 depicts how the broad activity constructs in the ICF are represented within the structure of the FAB. As shown here, certain chapters converge into one construct, others disperse into multiple different constructs, and one (mobility) is roughly evenly divided into three distinct constructs.

Figure 1.2 Hypothesized Conceptual Organization of Activities (ICF) and the Organization of Activities in the WD-FAB



To develop a modified organization of Activities, the item content and their ICF classifications were examined by each scale, resulting in the model presented in **Figure 1.3**. This model re-organizes the subdomains presented in the Activities domain into distinct and measurable constructs.

Figure 1.3 Alternative Organization of Activities for Measure Development

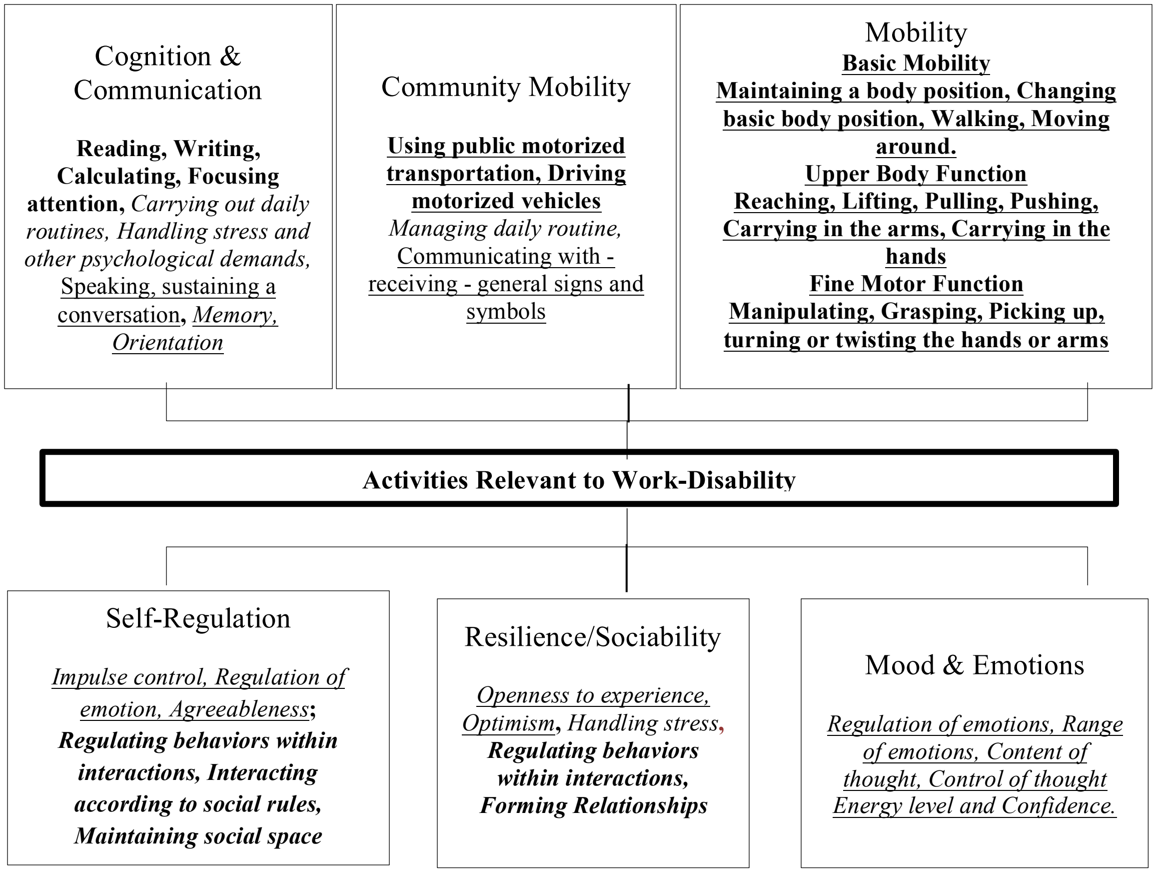


Figure 1.3 Key:

Learning & Applying Knowledge
General Tasks and Demands
Communication,
Mobility
Interpersonal Interactions & Relationships
Mental Functions

Mobility as a conceptual domain rarely overlapped with other Activity subdomains, and there is empirical support for the overarching domain containing three unique and distinct constructs: Basic Mobility, Upper Body Function, and Fine Motor Function. Basic Mobility is defined as the ability to assume, maintain and transfer among various body positions and the ability to move around from one place to another. The constructs include ICF Mobility codes such as maintaining a body position, changing basic body position, walking, and moving around. The construct Upper Body Function includes reaching, lifting, pulling, pushing, and carrying in the arms and hands. All of these constructs are within the carrying, moving and handling objects range of ICF codes. Fine motor function includes manipulation of objects requiring dexterity, specifically the ICF Mobility codes of manipulating, grasping, and picking up, and turning or twisting the hands or arms.

The activity of driving or taking public transportation inherently involves both physical and cognitive components, and the content with the scale reflects this: the items within the Community Mobility scale are almost evenly split between the physical and cognitive aspects of the associated tasks. From a Mobility perspective, it contains items related to using public motorized transportation and driving motorized vehicles. The construct also includes the General Tasks & Demand of managing daily routine and the Communication activity of Communicating with - receiving - general signs and symbols. Due to the small size of the scale, its actual functioning as two smaller short forms, and its other psychometric properties, it was omitted from Figure 1.2.

As a whole, the construct of Cognition & Communication includes items that characterize aspects of function such as organizational skills, attention, following instructions, oral and written communication. Many features of the first three subdomains of the Activities section of the ICF largely overlap into a single subdomain, which WD-FAB refers to as Cognition and Communication. This construct contains the Applying Knowledge tasks such as reading, writing; calculating; and focusing attention, the

General Tasks & Demands of carrying out daily routines and handling stress and other psychological demands, the Communication activities of speaking and sustaining a conversation. This construct also contains select features of the Body Functions subdomain Mental Functions, primarily memory and orientation functions.

Self-Regulation involves the ability to regulate emotions and behaviors based upon those emotions, such as controlling temper, respecting others, following rules, and social abilities. This construct contains elements primarily from the Mental Functions subdomain, and the Interpersonal Interactions and Relationships subdomain of the ICF. Specifically, it addresses the Mental Functions of impulse control, regulation of emotion, and agreeableness; it contains the Interpersonal Interactions of regulating behaviors within interactions and interacting according to social rules. Resilience/Sociability represents a range of content such as handling stress, accomplishing goals, and learning from mistakes. This domain pulls from the Mental Functions subdomain of the Body Functions section of the ICF, and the General Tasks & Demands, and Interpersonal Interactions and relationships subdomains of the Activities section of the ICF. The most common ICF codes in this scale are openness to experience, optimizing, handling stress, and regulating behaviors within interactions. Finally, Mood & Emotions as a construct represents a range of a person's internal emotional state and encompasses feelings such as depression and anxiety. In ICF terms, the Mental Functions of regulation of emotions, range of emotions, content of thought, and control of thought primarily drive the scale based on this construct, with some additional items related to energy level and confidence.

Discussion

The analyses presented above reveal that the content and organization of the items within the physical function scales of the WD-FAB are very similar to the hypothesized Mobility concepts included within the ICF; in contrast, the mental health function results from the WD-FAB analyses diverge from the mental health constructs hypothesized within the ICF. Most notable is the finding that items related to Mental Functions, organized within Body Functions in the ICF, converge with items related to different Activities. These differences can be instructive in future revisions to the ICF.

The physical function scales of the WD-FAB show close alignment with the ICF Mobility domains. Further, this analysis shows that there is empirical justification for the three independent constructs within mobility: Basic Mobility, Upper Body Function, and Fine Motor Function. The first three subdomains of the Activities section of the ICF converge into a single domain—referred to here as Cognition and Communication.

There were very few items in the WD-FAB that were coded as Self-Care, and the WD-FAB lacks a separate distinct Self-Care scale. The large pool of items fielded in the WD-FAB development study included specific task related items in a self-care context and global self-care items. The former tended to load onto their respective physical function scales and the latter did not scale to any other or separate construct. It should be noted that this analysis does not reflect how many of the physical function items take place in a self-care context though they do not specifically tap self-care as a construct. For example, the item “Are you able to stand in the shower to wash your hair?” is a Basic Mobility item that mentions showering to ground the item in a real-world activity, but does not tap directly into the Self-Care activity of "washing oneself."

This study also found that items that are classified as various Mental Functions converge with items classified as various Activities in multiple WD-FAB scales. When attempting to generate items for a self-reported outcome measure, it is important that the items be clearly and reliably interpreted by those who are intended to complete it. Writing items that relate to Mental Health Functioning at the activity level poses a unique challenge that can lead to the blurring of the Mental Functions and Activity. For example, the ICF classifies Memory Functions (b144) squarely in the Mental Functions domain. The item "Are you able to remember a list of 4 or 5 errands without writing it down" is asking about Retrieval of memory (b1442), but in the context of errands. In this sense, it is understandable why items coded as Body Functions converged with items coded as an Activity.

This study also brings up several clarifications to and lessons for the ICF coding methodology process. First, all of the features of 'Changing Basic Body Position' include getting *into* and *out of* each position; when an item or activity mentions both, the coders made a decision as to which ICF activity to code it as. For example, the item "Are you able to move from lying on your back to sitting on the edge of your bed?" clearly involve both getting out of the lying position (ICF code d4100) and getting into a sitting position (ICF code d4103). This study required a single classification of each item, and absent of any clarification rules, we decided to code all such activities as the ending position. Second, once we decided to include Mental Functions, the challenge came to distinguish between a Mental Function item and an Activity item that is *related to* but not inherently about a mental state. The decision rule decided upon was as follows: an 'Activity' is about what the respondent is *doing/not doing;* a 'Mental Function' refers to how people are feeling, thinking, or related to some other internal state. For example, we coded the item "Able to make everyday decisions" as "d177 Making decisions" rather than "b1645 Judgement;" even though the activity requires Judgement, the item itself is about making daily decisions. On the other hand, we coded "I often get angry when I'm told what to do" as "b1521 Regulation of emotion" rather than d7202 "Regulating behaviors within interactions" because the item itself is about the emotion of anger rather than an actual behavior towards the other person.

The study has several limitations. Given the large number of items, we decided on a single code per item for feasibility of coding and analysis. If we had allowed multiple codes it would be possible that more content areas could have been covered (for example, Self-Care) and the organizational structure of the constructs might be different. In addition, the analysis is based upon three-digit ICF codes to allow for comparison across chapters and content areas. This necessarily means a lack of specificity about the four digit codes.

This is one of the first studies to empirically examine the structure of Activities as the ICF proposes them. Other measures based upon the ICF typically examine only a limited number of Activity subdomains at a single time. Previous work has compared the constructs represented in the Patient Reported Outcomes Measurement Information System (PROMIS) to the ICF classification system.[31,32] They found that the unidimensional constructs in some PROMIS scales do primarily map onto ICF concepts;

however for some scales there were items that go beyond the ICF domain and address the impact of a construct on other life areas.[31] For example, the Physical Function scale addresses Mobility (D4) and Self-Care (D5), and the Pain Interference scale addresses the Body Function of Sensation of Pain (b280) and several activities such as mobility, self-care, and domestic life. However, PROMIS scales are developed individually by conceptual area—which limits the opportunity for different ICF subdomains to be combined or separated.

These analyses present several areas for future work. First, in the current organization, the ICF does not have a conceptualization of memory and emotion related functions at the activity level. Future revisions of the ICF may want to consider incorporating these concepts into the Activity domain. Second, although the factor structure of the WD-FAB was established in both a claimant and a general U.S. sample, the items were developed for the use in a specific functional evaluation context. Further work could focus on expanding the scope of item content beyond functional activities relevant to work disability. Finally, the reorganization of activities presented here and other empirical work could be refined through a combination of empirical examination and consensus among a panel of experts to keep with the ICF's standard of being applicable to a range of Health Conditions, Personal Factors, and Environmental Contexts.

The WHO has revised the coding for consideration of diseases (the ICD) several times, yet the structure of the ICF has not been significantly updated since its original inception in 2001. Since 2011 there have been annual updates, including expansions, modifications, of codes and criteria, additions of new codes, and changes in descriptions.[30] This work can be helpful in any potential future revisions of the ICF's structure. These results suggest that the current ICF draws distinctions between concepts that are not empirically different—specifically the first three subdomains of Activities: Learning & Applying Knowledge, General Tasks & Demands, and Communication. This is important for research that is based upon the ICF. Researchers and clinicians looking to measure these constructs may be missing out on important content by relying on the ICF's current organization. For some of these constructs it means that domains can be assessed separately; for example, the three mobility domains, can allow for more targeted measurement in different clinical groups. For other conceptual domains, it can mean a more parsimonious approach to measurement and data collection; if Learning and Applying Knowledge is related to the same underlying construct as Communication it can be captured with a single measure that addresses both, rather than one for each.

The need for a conceptually and empirically defensible model for function and the ability to measure it has been called for in the literature, and the ICF has been proposed as a standard for disability evaluation.[33–36] This paper contributes to the discussion about using the ICF as a foundation for systematically measuring and collecting functional information related to disability. As the ICF continues to be a standard in disability and rehabilitation measurement and evaluation, research such as this project can help improve the structure of the ICF, bringing it more in line with the empirical organization of these important concepts.

References

1. WHO I, World Health Organization. International classification of functioning. *Disability and Health (ICF), endorsed by all*. 2007;191.

2. Cerniauskaite M, Quintas R, Boldt C, et al. Systematic literature review on ICF from 2001 to 2009: Its use, implementation and operationalisation. *Disabil Rehabil*. 2011;33(4):281–309. doi: 10.3109/09638288.2010.529235 [doi].

3. Cieza, A., & Stucki, G. The international classification of functioning disability and health: Its development process and content validity. *European journal of physical and rehabilitation medicine*. 2008;44(3):303–313.

4. Cieza A, Stucki G. Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Quality of Life Research*. 2005;14(5):1225–1237.

5. Cieza A, Ewert T, Ustun TB, Chatterji S, Kostanjsek N, Stucki G. Development of ICF core sets for patients with chronic conditions. *Journal of Rehabilitation Medicine*. 2004 Jul.; (44 Suppl.):9–11.

6. Grill E, Ewert T, Chatterji S, Kostanjsek N, Stucki G. ICF core sets development for the acute hospital and early post-acute rehabilitation facilities. *Disabil Rehabil*. 2005;27(7–8):361–366.

7. Stucki G, Kostanjsek N, Ustun B, Cieza A. ICF-based classification and measurement of functioning. *Eur J Phys Rehabil Med*. 2008;44(3):315–328. doi: R33Y2008N03A0315 [pii].

8. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the international classification of functioning, disability and health. *J Rehabil Med*. 2002;34(5):205–210.

9. Jette AM, Haley SM, Kooyoomjian JT. Are the ICF activity and participation dimensions distinct? *J Rehabil Med*. 2003;35(3):145–149.

10. Jette AM, Tao W, Haley SM, Jette AM, Tao W, Haley SM. Blending activity and participation sub-domains of the ICF. *Disabil Rehabil*. 2007;29(22):1742–1750.

11. Dijkers MP, Whiteneck G, El-Jaroudi R. Measures of social outcomes in disability research. *Arch Phys Med Rehabil*. 2000;81: S63–S80.

12. Stucki G, Cieza A, Ewert T, Kostanjsek N, Chatterji S, ÜstÜn TB. Application of the international classification of functioning, disability and health (ICF) in clinical practice. *Disabil Rehabil*. 2002;24(5):281–282.

13. Lord FM. *Applications of item response theory to practical testing problems.* Routledge; 1980.

14. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care*. 2007;45(5 Suppl 1): S22–31. doi: 10.1097/01.mlr.0000250483.85507.04 [doi].

15. Norris JM. Computer-adaptive testing: A primer. *Language Learning & Technology*. 2001;5(2):23–27.

16. Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika*. 1989;54(3):427–450.

17. Wang S, Wang T. Precision of warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*. 2001;25(4):317–331.

18. Matheson LN, Kaskutas V, McCowan S, Shaw H, Webb C. Development of a database of functional assessment measures related to work disability. *J Occup Rehabil*. 2001;11(3):177–199.

19. Marfeo EE, Haley SM, Jette AM, et al. Conceptual foundation for measures of physical function and behavioral health function for social security work disability evaluation. *Arch Phys Med Rehabil*. 2013;94(9):1645–1652. e2. doi: 10.1016/j.apmr.2013.03.015.

20. Stucki G. ICF linking rules: An update based on lessons learned. *J Rehabil Med*. 2005; 37:212–218.

21. Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF linking rules to strengthen their potential for establishing comparability of health information. *Disabil Rehabil*. 2016:1–10. doi: 10.3109/09638288.2016.1145258 [doi].

22. Fayed N, Cieza A, Edmond Bickenbach J. Linking health and health-related information to the ICF: A systematic review of the literature from 2001 to 2008. *Disabil Rehabil*. 2011;33(21–22):1941–1951.

23. Rivers D. Sample matching: Representative sampling from internet panels. *Polimetrix White Paper Series*. 2006.

24. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling*. 1999;6(1):1–55.

25. Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioral research*. 1990;25(2):173–180.

26. Hoyle RH. *Structural equation modeling: Concepts, issues, and applications.* Sage Publications; 1995.

27. Meterko M, Marfeo EE, McDonough CM, et al. The work disability functional assessment battery (WD-FAB): Feasibility and psychometric properties. *Arch Phys Med Rehabil*. 2014.

28. Ni P, McDonough CM, Jette AM, et al. Development of a computer-adaptive physical function instrument for social security administration disability determination. *Arch Phys Med Rehabil*. 2013;94(9):1661–1669.

29. Marfeo EE, Ni P, Haley SM, et al. Development of an instrument to measure behavioral health function for work disability: Item pool construction and factor analysis. *Arch Phys Med Rehabil*. 2013;94(9):1670–1678.

30. List of official ICF updates. http://www.who.int/classifications/icfupdates/en/. Updated 2016.

31. Tucker CA, Cieza A, Riley AW, et al. Concept analysis of the patient reported outcomes measurement information system (PROMIS((R))) and the international classification of functioning, disability and health (ICF). *Qual Life Res*. 2014;23(6):1677–1686. doi: 10.1007/s11136-014-0622-y [doi].

32. Tucker CA, Escorpizo R, Cieza A, et al. Mapping the content of the patient-reported outcomes measurement information system (PROMIS®) using the international classification of functioning, health and disability. *Quality of Life Research*. 2014;23(9):2431–2438.

33. Escorpizo R, Stucki G. Disability evaluation, social security, and the international classification of functioning, disability and health: The time is now. *J Occup Environ Med*. 2013;55(6):644–651. doi: 10.1097/JOM.0b013e318297ae47 [doi].

34. Anner J, Schwegler U, Kunz R, Trezzini B, de Boer W. Evaluation of work disability and the international classification of functioning, disability and health: What to expect and what not. *BMC Public Health*. 2012;12:470-2458-12-470. doi: 10.1186/1471-2458-12-470 [doi].

35. 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. doi: 10.3109/09638288.2012.720350 [doi].

36. Madden RH, Glozier N, Fortune N, et al. In search of an integrative measure of functioning. *Int J Environ Res Public Health*. 2015;12(6):5815–5832. doi: 10.3390/ijerph120605815 [doi].

CHAPTER TWO: WHO APPLIES TO SOCIAL SECURITY DISABILITY PROGRAMS? DEMOGRAPHICS AND FUNCTIONAL DIFFERENCES AMONG CLAIMANTS

Introduction

In March, 2013 National Public Radio (NPR) released a series on work disability in the United States, focusing on the rise in applicants to the Social Security Disability programs and its potential relationship to the then still recovering economic recession.[1] The series sparked a debate between critics and defenders of the programs. One particular concern was whether the people who were applying for benefits were genuinely too low functioning to work, or if the programs were primarily serving as a safety net for the unemployed—not the truly disabled for whom it was intended. Although many policy makers and analysts have different beliefs about the extent to which this is occurring, [2] it is an under-examined research question.

The purpose of this paper is to examine sociodemographic differences between claimants and the general US working age population, and to, examine within claimants differences in functional status. Establishing a clearer understanding who claimants are sociodemographically and functionally is a necessary first step for monitoring of trends in applications which may help planning for the Social Security Disability programs going forward, and further our understanding of work disability in the US.

The United States Social Security Administration's (SSA) disability programs provide financial support to over 10.3 million Americans through the Social Security Disability Insurance (SSDI) program and 8.3 million additional individuals through the SSI program, making these programs the largest federal provider of financial assistance to disabled workers and their families.[3] In 2014 alone over 750,000 recipients were awarded SSDI benefits because of a work limiting health condition.[3] The Social Security Act defines work disability as an inability to engage in "substantial gainful activity due to any medically determinable physical or mental impairment that can be expected to result in death or to last for a continuous period of not less than 12 months."[3] The past two decades have seen a steady, and at times precipitous, rise in the numbers of SSDI and SSI applicants and beneficiaries.

In deciding on whether or not a claimant can do other work at Stage 5 of the disability determination process, the SSA considers the person's medical evidence, age, education, work experience, and any transferable skills. All social security programs are neutral towards gender, race, and many other sociodemographic factors; however, at the final stage of adjudication for many people, the SSA does consider age, education, and the intensity of previous work experience when adjudicating applications for disability benefits. The relationship between these considerations for decision making and an applicant's functioning is poorly understood.

Unlike many other trends in health services, the aging population is not likely a large driver of the current increasing applications, as almost all beneficiaries are under the age of 66. Once individuals reach the age of 66 years old, they are converted to the Social Security retirement program, which has a different funding source. There are, however, some demographic changes that may be driving the growth. The influx of women in the workforce in the late 20th century increased the number of people who could become eligible if a health condition limited their ability to work. Younger people have also been receiving benefits in greater numbers, which means beneficiaries may be in the program longer. There have been shifts in the US economy, including fewer jobs for manufacturing and other traditional ‘blue collar’ workers, which has been exacerbated by the recession that started in 2008. These shifts displaced people from the workforce, making them possibly more likely to apply for SSDI benefits.[4–6]

People who choose to apply for disability benefits through the Social Security Administration *are* a self-selected group and therefore likely to be unique. They have been out of work; they have specific health conditions that they feel limit their ability to work their current or other jobs; they have made the choice to apply to the disability insurance program. It is necessary to understand whether individual characteristics such as age, race, or gender play a role in assessment of functioning. Any relationship shown may reveal how different groups or types of people behave when experiencing a functional limitation. Further, if the Social Security Disability programs are being used as an economic alternative to those not actually disabled, then we might hypothesize that there are groups of applicants who are higher functioning.

When measuring a claimant’s function, differences in functional outcome scores could be due to measurement error; Item Response Theory (IRT) allows for the detection and correction of one particular type of measurement error that may be a concern when looking at diverse populations. Differential Item Functioning (DIF) is when individuals with different characteristics with the *same* ability level have a different probability of endorsing an item on an IRT scale. For example, individuals’ with different demographic characteristics have different experiences that may influence responses to individual questions on self-reported measures unrelated to their actual functioning. *If* this differential response pattern on items impacts scores, then the individuals score is reflective of more than just their level of functioning—which is measurement error. Little is known about differences in how those with similar functional abilities respond to individual items on self-reported measures. If there is a desire to understand how a diverse group of claimants differ functionally by demographics, we first must correct for measurement error that would relate to demographic characteristics and function.

This paper has three aims. The first is to examine how claimants compare sociodemographically to the working age US population. Second, to determine whether gender (men v. women), race (white v. not white), or age (age <55 and ≥55) affected respondents item responses, independent to function (that is, identify Differential Item Functioning (DIF)). Once there is a way to control for DIF, the final aim is to assess differences in claimants' functional status by gender, age, race, ethnicity, urban/rural status, and education.

Conceptual Framework

One of the leading models for characterizing the multifaceted nature of disability is the World Health Organizations (WHO) International Classification of Functioning, Disability and Health (ICF). The ICF calls for research on disability and health to be examined with a "whole person" perspective. The ICF is a biopsychosocial model of disability that characterizes disability as a result of the complex relationship between individual health characteristics, environmental factors, personal factors, and the extent to which those interactions affect a person's ability to perform activities and participate in daily roles, such as work. This study examines how the ICF domain of Personal Factors (gender, age, Gender, Race, Ethnicity, Age, Urban/Rural Status, Education) influence function at the Activity level, as measured by the WD-FAB.

Methods

Sample

This study uses data from the development of the WD-FAB. The study involved two samples, referred to as the "claimant sample," and the "general sample." In the claimant sample, a random sampling from a pool of 16,500 SSA disability claimants who applied for benefits in the last 2 months, stratified by urban/rural status across the ten national SSA regions, yielded a final sample size of 3,793. The general sample includes 2,100 working age adults, obtained using sample matching whereby samples representative of a study-appropriate target population could be constructed from large but unrepresentative pools of opt-in survey respondents.[7] In the general sample, there is a core sample of 1,000 individuals, representative of the US working age population (used for Aim 1), and an additional oversample of several racial and ethnic minority groups: Black or African Americans, Hispanic/Latinos, and Asian Americans (used only within DIF analysis, for power). T-test for continuous variables and Pearson Chi-square tests for categorical variables tested differences between the means and proportions of the demographic groups between the two samples.

Outcome Measure

The Work-Disability Functional Assessment Battery (WD-FAB), is a new self-reported measure of activity limitations relevant to work and work disability.[8–12] It was created to allow the Social Security Administration (SSA) to collect systematic and comprehensive information about claimants' functioning to inform the disability determination process. The WD-FAB uses Item Response Theory (IRT).[13] In measures that employ IRT, each item is ordered in a hierarchy along a unidimensional construct, and scores are based upon probability models that represent the likelihood that a respondent selects a response choice, given their ability on that construct.[8,14, 13] In this analysis, the battery consists of seven scales: three physical factors (Basic Mobility, Upper Body Function, and Fine Motor Function), and four nonphysical factors (Cognition & Communication, Resilience/Sociability, Self-Regulation, and Mood & Emotions)[15–17] Raw scores on the WD-FAB are first scored in logits, converted to Z-scores which are then transformed to T-scores. The T-scores are based upon the score distribution of the general sample of working age adults, with a mean of 50 and a standard deviation of 10, with higher scores representing higher function.

Adjustment for Measurement Error (Differential Item Functioning)

In both samples a two-step IRT based DIF method was applied to assess DIF across three pairs of groups: age (<55 vs. ≥55), gender, and race (white vs. non-white).[18–20] In the first step, the mean and standard deviation of the focal group is estimated by fitting a two-group IRT model with all item parameters set equally across the reference and focal groups. The Wald X^2 test was used to examine DIF by fitting another two-group IRT model with the mean and SD of the focal sample fixed as previous estimates. In the second step, non-DIF item parameters identified in the first step are set equal across reference and focal groups, and the mean and standard deviation of the focal group are free to estimate. Wald X^2 was used to examine the DIF in the rest of the items using a two-group IRT model. Benjamin–Hochberg procedure was used to adjust p-value for Wald-X^2 tests.

For the DIF item(s), the item characteristic curves and response characteristic curves were visually examined and, the wABC (the weighted absolute difference between item characteristic curves) was calculated by integrating the absolute difference between the expected score functions of reference and focal groups over the latent distribution. Applying standard thresholds, 5-category items with wABC> 0.3 and 4-category items with wABC> 0.24 were classified as DIF. For items that are classified as DIF, the content was considered to possibly explain the source of the differential response patterns to an item.

Differences in Claimants' Functional Scores

Differences in WD-FAB scores among claimants were tested across different personal factors: age, gender, race, Hispanic ethnicity, urban/rural status, and education. Age and education were selected because of their incorporation into the disability determination process. Gender was selected because one of the main trends accounting for the increase of applicants is due to the increase of women in the workforce over the past decades. The race and ethnicity variables were selected because there are several societal beliefs about different racial and ethnic group's usage of public assistance such as SSA disability programs. And urban/rural status was selected due to the differential access to resources in those two broad areas, and a general difference in the type of work that people do in those areas as well.

For the race groups, the survey allowed individuals to select multiple race categories. All comparisons were done by those who selected they were a particular race, compared to all of those who did not select that race. Therefore, the groups are not mutually exclusive. Independent t-tests were conducted to assess differences in the functioning scores across key subgroups. The significance level for all analyses were set at alpha=0.05. Where statistical differences existed in mean scores between the groups, additional analysis of the magnitude of the mean differences was performed by calculating Cohen's *d* effect size. Small (0.2), medium (0.5) and large (0.8) were defined for Cohen's *d*.[21]

Results

Aim 1: Demographic Differences between Claimants and Working Age Adults

The demographic characteristics of the claimant sample is compared to the general sample in **Table 2.1**. Compared to the general sample, claimants are slightly more likely to be women, slightly more likely to be over 55 years old, and slightly more likely to live in rural areas. There are disproportionately fewer white claimants, and more black/African American claimants. There are fewer Asian or Latino claimants. In terms of education, there are more claimants with less than a high school diploma, and fewer claimants with any college or a college degree.

Table 2.1. Sample Characteristics

Characteristic	Working Age Adult N=1000		Claimant Sample N=3793	
	N	**Weighted %**	**N**	**%**
Age mean (sd)***	43.67 (13.44)		47.54 (11.77)	
Men*	472	49.81%	1721	45.48%
Age 18–55*	799	69.77%	2476	65.28%
White***	746	74.59%	2489	66.28%
Black***	145	12.74%	1022	27.22%
Asian***	37	8.60%	33	0.88%
Latino***	155	14.21%	298	7.93%
Urban***	849	80.00%	2742	72.64%
Education***				
Less than HS	36	3.61%	597	15.74%
HS or GED	344	34.45%	1214	32.01%
Associates Degree	93	9.64%	320	8.44%
Vocational	52	5.30%	261	6.88%
Some College	188	18.02%	904	23.83%
College or More	284	28.74%	489	12.89%

***p<0.05**
****p<0.01**
*****p<0.001**

Aim 2: Differential Item Functioning: Claimants & Working Age Adults

In total 17 of the 309 items displayed DIF (**Table 2.2**). Most of the DIF was based upon gender. For the one non-uniform DIF item identified, the switch in trend for this item occurred at a very low score, 2.5 standard deviations below the mean. Therefore, it is not likely that many individuals will fall in that score range. The Item Characteristic Curves of all of the DIF items can be found in **Appendix 2.1**.

Table 2.2. WD-FAB Items that Displayed DIF

Item Content	**Scale**	**Sample**	**Characteristic**
Are you able to sit on a stool without back support?	Basic Mobility	Claimant	Age
I can relate to other people's feelings	Cognition & Communication	Claimant	Gender
I don't know why I cry so often	Mood & Emotions	Claimant	Gender
I suddenly become emotional for no reason	Mood & Emotions	Claimant	Gender
Are you able to unload a washing machine?	Basic Mobility	Claimant	Gender
Are you able to walk for at least 30 minutes?	Basic Mobility	Claimant	Gender
Are you able to unload the dishwasher?	Upper Body Function	Claimant	Gender
Are you able to use a nut cracker?	Upper Body Function	Claimant	Gender
Are you able to open a bottle of soda?	Fine Motor Function	Claimant	Gender
Are you able to chop or slice vegetables for a large meal?	Fine Motor Function	Claimant	Gender
Are you able to put on a watch or bracelet?	Fine Motor Function	Claimant	Gender
Are you able to walk 150 feet (45 meters) on flat ground?	Basic Mobility	Claimant	Race
I have to talk very slowly to make myself understood	Cognition & Communication	Claimant	Race
I can hear things that other can't	Self-Regulation	Claimant	Race
People have told me that sometimes I act strange.	Self-Regulation	General	Age
I worry people are criticizing me even when they are not.	Mood & Emotions	General	Age
I feel sick when I have to speak in front of people.	Mood & Emotions	General	Gender

Three items displayed DIF based upon age group. For the one physical item, "Are you able to sit on a stool without back support?" within the Basic Mobility scale, individuals in the older age group were more likely to say able—that is they had "higher" responses on that item compared to the younger group individuals who had the same overall Basic Mobility score. The two mental function items ("People have told me that sometimes I act strange" from Self-Regulation, and "I worry people are criticizing me even when they are not" from Mood & Emotions) also showed the older group displayed higher responses on the questions compared to their same scoring younger counterparts. Three items displayed DIF based upon the race group, and for all three items the white group self-reported higher responses on the individual items compared to those with the same score in the non-white group. The items were one Basic Mobility item, "Are you able to walk 150 feet (45 meters) on flat ground, one Cognition & Communication item "I have to talk slowly to make myself understood," and one Self-Regulation item "I can hear things that others can't."

Most of the DIF items found were in the gender comparisons. The three mental function items were: "I can relate to others people's feelings, "I don't know why I cry so often," and "I suddenly became emotional for no reason." These seem to be reflective of how traits such as empathy and behaviors related to displaying certain types of emotions (specifically sadness and crying) are gendered. Seven items from the physical function scales showed gender based DIF.

Aim 3: Differences in Claimants Functional Scores

Table 2.3 shows the average scores for the entire claimant sample. The WD-FAB scores are based upon the score distribution of the general sample, with a mean of 50 and a standard deviation of 10, with higher scores representing higher function. For all scales, the claimant sample is functioning generally one half to almost one and a half standard deviations below average in the general sample. In total, these results provide no evidence that 'non-disabled' people are applying en masse. The scores on the whole are lower in the physical function scales (36.17–38.81) than the mental function scales (41.13–46.34). This could be due to different prevalence's of different health conditions in the sample; if fewer people with these mental health problems are in the sample, then it would make sense that a large portion of the sample would have nearer to average mental functioning scores. This could also be due to the different demographic makeup of the claimant sample, compared to the general sample.

Table 2.3. Average WD-FAB Claimant Scores

WD-FAB Scale	**n**	**Mean**	**SD**
Basic Mobility	3788	37.39	5.61
Upper Body	3188	36.17	5.74
Fine Motor	3188	38.81	6.51
Cognition & Communication	3792	41.73	7.71
Self-Regulation	3791	45.67	11.50
Resilience-Sociability	3770	46.34	10.77
Mood & Emotions	3189	41.13	13.52

Table 2.4 shows the results of T-test and effect sizes for all personal factor comparisons within claimants where a significant result was found, adjusting for DIF; the effect sizes are visually displayed in **Figure 2.1**. The observed differences and effect

sizes will be summarized together. Most of the differences, and the differences with the largest magnitudes were between claimant men and women, and between older and younger claimants.

For the comparisons based upon gender, men had higher scores than women in Basic Mobility, Upper Body, Cognition & Communication, and Mood & Emotion, while women had slightly higher scores than men in Self-Regulation. Most of the effects were small and some bordered on moderate (0.153–0.351). In the physical function scales, the younger group had higher scores; in the mental health function scales, the older group has higher scores. The largest effect was seen for differences in Mood & Emotions (0.535).

There were few significant differences found for all of the different race group comparisons, and those that were observed had very small effect sizes (0.077–0.237). White claimants had slightly lower scores for Resilience/Sociability, and higher scores for Self-Regulation. Black claimants had slightly higher scores for Basic Mobility, and Resilience/Sociability, and slightly lower scores for Self-Regulation. Hispanic/Latino claimants had lower scores in Self-Regulation and Mood & Emotions. No significant differences were found for Asian claimants.

Claimants from rural areas had lower scores in Basic Mobility, Upper Body, Fine Motor, and Cognition & Communication, compared to those in an urban area. Finally, for the education comparison, claimants with a HS/GED or less had lower scores on all Mental Health function scales, with effect sizes ranging from 0.128–0.366. There were no observed differences in the physical function scales.

Table 2.4. Differences in Claimant WD-FAB Scores by Personal Factors

	Men		Women			
Scale	mean	sd	mean	sd	p	Effect Size (*d*)
Basic Mobility	38.13	5.67	36.78	5.49	<.0001	0.242
Upper Body Function	37.26	5.81	35.27	5.51	<.0001	0.351
Fine Motor Function	39.02	6.35	38.64	6.64	0.0988	
Cognition & Communication	42.39	8.11	41.21	7.32	<.0001	0.153
Self-Regulation	45.22	11.27	46.05	11.67	0.0264	0.072
Resilience-Sociability	46.41	11.03	46.27	10.55	0.6904	
Mood & Emotions	42.96	13.46	39.63	13.39	<.0001	0.247

	Age 18–54		Age 55–66			
Scale	mean	sd	mean	sd	p	Effect Size (*d*)
Basic Mobility	37.77	5.95	36.67	4.83	<.0001	0.196
Upper Body Function	36.45	6.13	35.68	4.92	0.0001	0.133
Fine Motor Function	39.14	6.66	38.22	6.21	0.0001	0.140
Cognition & Communication	40.93	7.66	43.25	7.56	<.0001	0.303
Self-Regulation	44.14	12.15	48.53	9.52	<.0001	0.388
Resilience-Sociability	45.56	11.38	47.80	9.34	<.0001	0.208
Mood & Emotions	38.61	13.74	45.63	11.86	<.0001	0.535

	White		Not White			
Scale	mean	sd	mean	sd	p	Effect Size (*d*)
Basic Mobility	37.34	5.53	37.50	5.78	0.4066	
Upper Body Function	36.18	5.65	36.14	5.89	0.8667	
Fine Motor Function	38.87	6.33	38.68	6.90	0.4573	
Cognition & Communication	41.76	7.60	41.70	7.97	0.8103	
Self-Regulation	46.33	10.67	44.40	12.95	<.0001	0.167
Resilience	46.01	9.98	47.07	12.20	0.0074	0.099
Mood & Emotions	41.02	13.10	41.34	14.44	0.5527	

	Black		Not Black			
Scale	mean	sd	mean	sd	p	Effect Size (*d*)
Basic Mobility	37.71	5.77	37.27	5.55	0.0336	0.077
Upper Body Function	36.29	6.00	36.12	5.63	0.4817	
Fine Motor Function	38.85	6.97	38.79	6.34	0.8430	
Cognition & Communication	41.75	7.96	41.74	7.64	0.9716	
Self-Regulation	44.63	13.00	46.07	10.90	0.0017	0.124
Resilience-Sociability	47.20	12.33	46.05	10.13	0.0085	0.105
Mood & Emotions	41.81	14.80	40.87	13.05	0.1072	

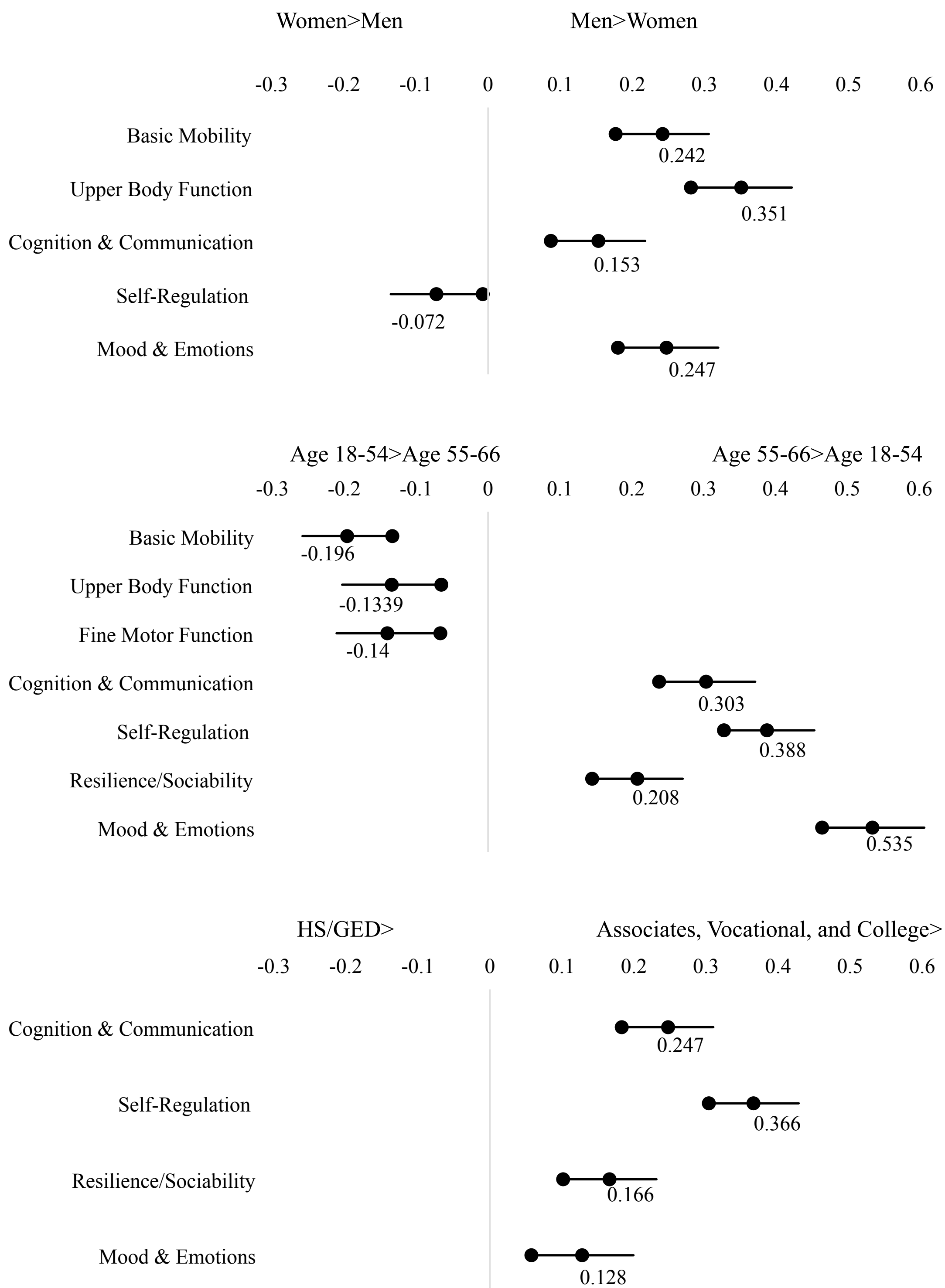
Scale	Asian mean	sd	Not Asian mean	sd	p	Effect Size (*d*)
Basic Mobility	34.46	5.86	37.38	5.61	0.2723	
Upper Body Function	37.09	4.99	36.16	5.73	0.4020	
Fine Motor Function	39.97	4.51	30.80	6.53	0.1911	
Cognition & Communication	41.58	7.45	41.74	7.73	0.9039	
Self-Regulation	42.54	11.34	45.70	11.52	0.1165	
Resilience-Sociability	49.40	10.19	46.34	10.79	0.1051	
Mood & Emotions	40.84	14.25	41.12	13.54	0.9148	

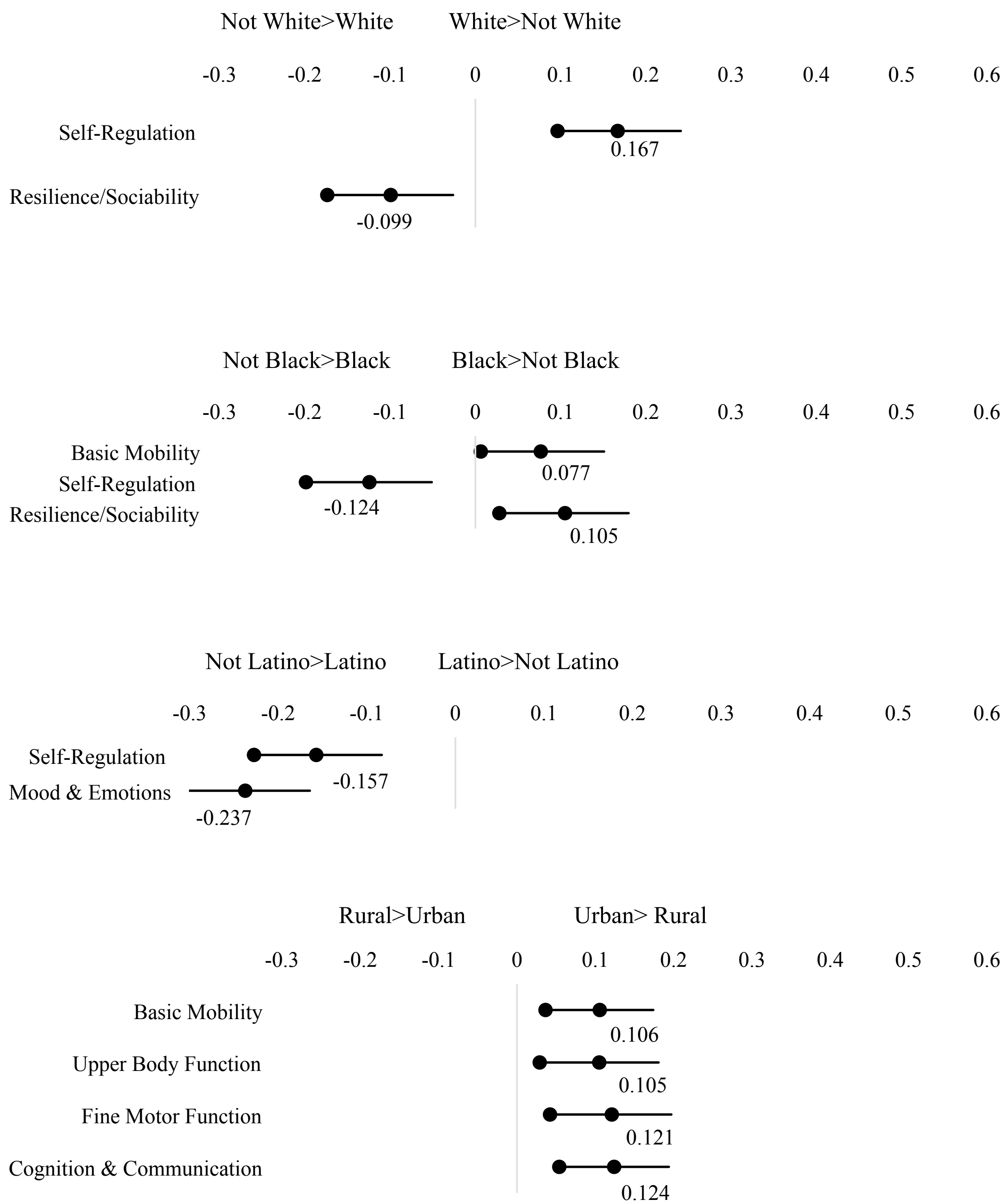
Scale	Latino mean	sd	Not Latino mean	sd	p	Effect Size (*d*)
Basic Mobility	37.75	5.76	37.36	5.60	0.2508	
Upper Body Function	36.33	5.85	36.15	5.72	0.6417	
Fine Motor Function	38.11	6.99	38.87	6.48	0.0803	
Cognition & Communication	41.15	8.07	41.79	7.70	0.1691	
Self-Regulation	44.04	12.98	45.84	11.31	0.0208	0.157
Resilience-Sociability	45.34	12.3473	46.44	10.65	0.1396	
Mood & Emotions	38.17	13.96	41.38	13.48	0.0004	0.237

Scale	Rural mean	sd	Urban mean	sd	p	Effect Size (*d*)
Basic Mobility	36.97	5.27	37.57	5.72	0.0024	0.106
Upper Body Function	35.75	5.45	36.35	5.82	0.0061	0.105
Fine Motor Function	38.25	6.45	39.04	6.51	0.0021	0.121
Cognition & Communication	41.05	7.48	42.01	7.77	0.0007	0.124
Self-Regulation	45.26	11.22	45.87	11.55	0.1421	
Resilience-Sociability	45.81	10.82	46.54	10.72	0.0629	
Mood & Emotions	40.79	13.18	41.28	13.65	0.3649	

Scale	HS/GED or Less mean	sd	Assoc, Voc, College mean	sd	p	Effect Size (*d*)
Basic Mobility	37.54	5.58	37.27	5.63	0.1447	
Upper Body Function	36.15	5.69	36.21	5.77	0.7606	
Fine Motor Function	38.66	6.51	38.93	6.51	0.2490	
Cognition & Communication	40.75	7.47	42.65	7.79	<.0001	0.247
Self-Regulation	43.50	11.62	47.64	10.96	<.0001	0.366
Resilience-Sociability	45.43	11.13	47.21	10.31	<.0001	0.166
Mood & Emotions	40.23	13.73	41.97	13.29	0.0003	0.128

Figure 2.1 Effect Sizes of Differences in Claimant WD-FAB Scores by Personal Factors




Not White>White
White>Not White
-0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6
Self-Regulation
0.167
Resilience/Sociability
-0.099
Not Black>Black
Black>Not Black
-0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6
Basic Mobility
0.077
Self-Regulation
-0.124
Resilience/Sociability
0.105
Not Latino>Latino
Latino>Not Latino
-0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6
Self-Regulation
-0.157
Mood & Emotions
-0.237
Rural>Urban
Urban> Rural
-0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6
Basic Mobility
0.106
Upper Body Function
0.105
Fine Motor Function
0.121
Cognition & Communication
0.124

Discussion

These analyses found that SSA disability claimants are more likely to be older, less educated, and black/African American compared to the working age US population; within claimants the differences in functional scores of greatest magnitude were between age groups and gender groups. Because these differences in scores are accounting for differential item functioning (DIF), we can be confident scores are due to true functional ability, not related to measurement error associated with the personal factors examined.

Aim 1: Demographic Differences between Claimants and Working Age Adults

Compared to a sample of working age adults in the US, SSA claimants tend to be slightly older, slightly more likely to be women, and more rural. In regards to race, claimants are more disproportionately Black or African American, with disproportionately fewer individuals self-identifying as white, Asian, or Latino. There are also more claimants with less than a high school diploma, and fewer claimants with a college degree. There are several possible ways to interpret these sociodemographic differences. In general, any group may be more likely to apply for benefits if they experience health conditions and/or functional limitations in greater frequencies than others. Alternatively, there may be fewer alternative jobs available or economic opportunities for particular groups of people. For example, the differences in urban versus rural status and education may be reflective of limited alterative job opportunities to those who live in rural areas[22] and for individuals with lower education levels.[23] Finally, it could be that there are other factors influencing the choice to apply (beyond functioning and job factors), such as the role a person's income and work plays in a larger family unit. Each of these three general factors will be considered for the two most noted differences observed here: age, and race/ethnicity.

For differences in age, among working age people there may be more physical health conditions that might lead to work disability in the older aged group.[24,25] Many chronic health conditions, such as low back pain, arthritis, and cardiovascular disease, as well as other conditions such as stroke and cancer, tend to start or influence function later in life.[26,27] We could postulate that older working age adults would be more likely to apply for SSA disability benefits. There could be generational trends with younger individuals working in different types of jobs with different functional demands, or occupations that are more accommodating to their functional limitations. Similarly, younger individuals have more potential 'work years' ahead of them, and may be more likely to look for different work in a different field with different functional demands, rather than apply for benefits. Finally, it could be that younger working age individuals may have more dependents (e.g. spouses or young children). It can take a long time to apply for and receive disability benefits—time during which a person cannot afford to not work. Younger workers who are in greater need of an immediate form of income then would be less likely to apply. Similarly, if the actual cash value of the benefits is not perceived to be high enough to support the family unit, younger workers may choose avenues other than applying for work disability benefits.[28,29]

This study found that SSA claimants were more disproportionately black/African American, and less Latino and Asian. For Latino and Asian individuals, it is possible that there are language or cultural reasons that make applying more challenging and therefore those individuals do not apply. For black/African American individuals, this could be a result of this group experiencing a greater *prevalence* of functional limitations—a result of race playing a large role in the social determinants of health and health outcomes.[30–32] It could also be that there are fewer job opportunities for black/African individuals—unemployment rates do tend to be higher among black/African American individuals compared to white individuals. [32–34]

Aim 2: Differential Item Functioning

There are several possible interpretations for the DIF findings described herein, although some items were challenging to interpret. The one physical item that displayed age based DIF ("Are you able to sit on a stool without back support?") could be due to different experiences with sitting on those types of stool: perhaps the older group is less likely to have experience with it (possibly due to avoidance) and they are thinking of in the past when they did (when it may have been an easier task). The two mental function items ("People have told me that sometimes I act strange" from Self-Regulation, and "I worry people are criticizing me even when they are not" from Mood & Emotions) also 'favored' the older group. One possible reason for this may be that the younger group has a heighted sensitivity for criticism from others. On the other hand, it could be that older individuals may be less self-aware or less concerned about these types of social criticisms and worries.

In regards to race-based DIF, the one Basic Mobility item, ("Are you able to walk 150 feet (45 meters) on flat ground") does not contain language that would seem racially coded in any way. However, it is possible that the two groups interpret the distance of

150 feet (45 meters) differently. Finding DIF for the Cognition & Communication item "I have to talk slowing to make myself understood," may be reflective of the different lived experiences of white claimants and black claimants. Regardless of actual Cognition & Communication ability, members of a racial or ethnic minority group may have experienced that they need to speak differently when communicating with the non-group members. This could be related to the phenomenon of code switching.[35] Claimants in the two groups may be differentially interpreting the Self-Regulation item "I can hear things that others can't," because of cultural differences in what 'hearing things' may imply. It is possible that the non-white group is taking a more figurative approach to the phrase, reading it as insight, intuition, or related to spirituality.

The gender based DIF findings generally seem to align with gender stereotypes or activities that may be considered gendered in American culture.[36,37] The three mental function DIF items ("I can relate to others people's feelings, "I don't know why I cry so often," and "I suddenly became emotional for no reason") seem to reflect how traits such as empathy and behaviors related to displaying emotions (specifically sadness, anger, and crying), are very gendered in U.S. culture. Those traits and behaviors could be viewed as "typically feminine," which is why it is not surprising that men who are at the same communication and emotional functioning level are less likely to endorse those activities. On the physical function side, there were several items ("Are you able to unload a washing machine?" "Are you able to unload a dishwasher?" "Are you able to chop or slice vegetables for a large meal?") where the DIF may be related to the fact that many domestic activities remain gendered. If men have less experience with cooking, doing laundry, and cleaning dishes, then it would make sense that they would think these items are more difficult compared to women who are actually at the same physical functioning level.[36] Similarly, gender DIF for the item “Are you able to put on a watch or a bracelet” may be due to the fact that men and women have different exposure to these types of accessories, and that jewelry is differently difficult to put on for the different genders—with women’s bracelets and watches traditionally being smaller and therefore may be more difficult.

The DIF findings in this study are similar to those in the literature.[36,38] The PROMIS Depression item bank, found DIF based upon gender, including items that addressed crying, and general unhappiness, and age based DIF for items that addressed having something to look forward to, and enjoyment of current activities.[39] The PROMIS physical function item bank found gender based DIF for an item involving “light house work.” [40] Studies that examine DIF in patient reported outcomes based upon race/ethnicity are rare, and in that regard, this analysis contributes to the field.[41,42]

These DIF items are retained in the item bank, but with different item parameters (difficulty and discrimination/slope) for the individuals in the different groups. This allows for the retention of important item content, and for maintaining large item pools. Most importantly however, it means that because WD-FAB scores are accounting for DIF, we can be more confident that observed differences in scores between groups (in this analysis and in general use of the measure) are true differences, and not related to this type of measurement error. This means that future researchers using the WD-FAB can be confident that the scores obtained are reflections of *true function* without concerns about measurement error related to a respondent's gender, race/ethnicity, or age.

Aim 3: Differences in Claimants Functional Scores

The prior literature generally suggests that women, non-white individuals, and older individuals tend to perform worse on functional assessments.[43–47] This study found similar results among claimants for gender and somewhat different results for race and ethnicity, although the effects are small. The largest effects seen were for men's higher Upper Body Function compared to women (d=0.351), the older groups higher Self-Regulation (d=0.388) and Mood & Emotions scores (d=0.535), and higher education group's higher scores in Self-Regulation (d=0.366) .The finding that claimants in the older group had lower scores on all physical functional scales, and higher scores on all of the behavioral/mental function scales compared to the younger group has also been observed in other groups with chronic health conditions.[26,48] Differences in mental health function based on education have also been found elsewhere.[49,50] This paper does not attempt to *explain* the sources of these observed differences between claimants, but they present interesting areas for discussion, particularly in light of this paper's first aim.

The prevalence of different types of conditions (such as angina/MI, COPD or arthritis and mental conditions) are possibly associated with functional differences among claimants. For example, first onset of severe mental illness tends to occur in young adulthood.[51] Therefore, if the younger claimants are more likely to be applying due to a mental health condition, then their scores on the mental health scales would be lower compared to older claimants, which is what we found. Conversely, if older claimants are more likely to be applying due to physical reasons (arthritis, low back pain, cancer, etc.), then their scores on the physical health scores would be lower compared to younger claimants, which again is what we found here.

Differences in function could also be reflective of different ‘thresholds’ for applying for benefits in different subgroups. Hypothetically, if a person works in a job that is very physically demanding, then a smaller reduction in his or her physical function may be ‘enough’ to initiate the decision to apply for disability benefits. That might be interpreted as that person having a ‘lower threshold’ than someone who has a job that allows them to sit for most of the workday. Applying this concept to this analysis, we can discuss what group may have different function-related thresholds, and its impact on the program. Finding that there are more women than men applying for disability benefits *might suggest* that women may have a lower threshold—however we found that when there were differences in function based upon gender, it was the men who had higher scores. Therefore, it is unlikely that women are disproportionately applying for benefits due to a lower threshold for those living with functional limitations.

The general lack of differences in functional scores between the different race and ethnicity groups in the claimant sample warrants further discussion. Our first aim found that claimants are disproportionately black/African American, and the current policy and media discussions regarding the Social Security Disability programs that express concern about “not disabled” individuals applying to the program are also often associated with racial and ethnic minorities. *If* this latter explanation were the case, we may have found that the black/African American group of claimants had higher scores compared to other claimants. However, black/African American individual claimants were functionally no different from other claimants. This suggests (1) fears about certain social groups applying for benefits when they do not 'need them' are not supported, and (2) there are likely other reasons to explain the observed race/ethnicity trend. Although this study and dataset does not allow for the investigation into what those reasons are, there are several possibilities which could be investigated in future research. One possible explanation could be different educational and economic opportunities available to individuals in difference race and ethnic groups. Different economic and employment outcomes by race have been described in the literature, with lower labor force participation among black/African American individuals compared to their white counterparts.[33,34,52] This could also be reflective of the effects of systematic racism and discrimination on socioeconomic status and health.[32]

This study has several limitations. First, the analysis was limited in the approach to characterizing race and ethnicity. The race groups examined as part of the third aim were not mutually exclusive and there was not separate capturing of multiracial individuals. Race was self-reported and allowed individuals to select any and all races that they identified as. Therefore, in regards to the three race categories (White, Black, and Asian), there are multiracial individuals in each group. However, there was no multiracial group in analysis due to small sample size of such claimants (n=35). In regards to the Latino comparison, the Hispanic identity question was asked separately from the race question, and therefore there are individuals from all three race categories (and others) in the Latino group. Similarly, grouping all "non-white" individuals together for the DIF analysis has its limitations. Not all non-white claimants have the same social context and experiences and interpretations that tend to be the origin of DIF; by grouping them all together, there may be some DIF that went undetected. Another limitation is that this analyses does not look at differences in job intensity/functional demands or previous work experience between claimant and the general working-age sample, or how claimants with different work demands and job types may be different functionally. Future initiatives should establish a more in depth understanding of who applies to the Social Security disability programs.

There are several areas for future work based upon these analyses. Regarding DIF, qualitative interviews, could be conducted to further examine why the different groups are interpreting items differently, and get recommendations for writing new items that are more neutral. In addition, future studies with large sample sizes can be conducted for analysis of DIF by more specific race groups. Another area of research involves the concepts of "thresholds" for applying for benefits. What is the level of functional limitations that a person meets before he or she applies for benefits, and how do those levels differ by sociodemographic groups? Given any possible differences, what are the other factors that determine a person's "threshold"? For example, there are likely political, cultural, and personal factors that could have an influence on individual's decisions to apply for disability benefits. Any such future research would also have to consider the influence the "economy" that a person is in—in terms of both national trends and local economies of job availability.

Finally, it is known that more people apply to the disability programs when the economy is poor and job opportunities are limited. This study suggests that these individuals do seem to be functionally limited, which implies that when there are a wider range of job opportunities and more jobs available, there are many people with functional limitations who are working. Examining how individuals with functional limitations work when they are employed, can help to keep people in the workforce, and help to inform return to work efforts.

There are also future areas of research related to the disability determination process itself. The SSA is more likely to give benefits to claimants who are older, of lower education levels, and who have previous work experience that are unskilled or with jobs that are non-transferrable. These results suggest that functionally, some of these assumptions are supported and others less so. These observations could be useful to help better refine adjudication decisions in light of other factors not measured here. For example, older claimants are more functionally limited *physically* compared to younger claimants, but not mentally. Perhaps there should be different guidelines based upon age given the different types of health conditions that are more reflective of general functional trends that are also based upon age. Similarly, claimants with less education are more functionally limited mentally compared to claimants with higher education, but no differences in physical function. When considering other work options for an education level, the manner in which the two different groups tend to function should also be considered.

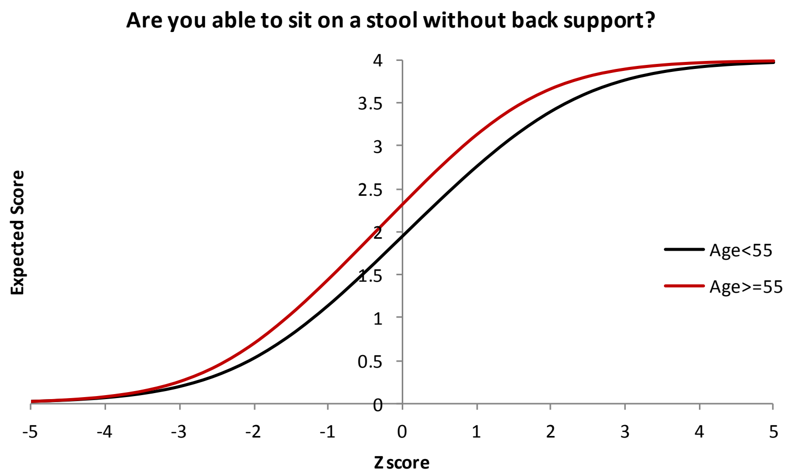
Conclusion

SSA disability claimants are more likely to be older, less educated, and black/African American compared to the working age US population. Seventeen of the

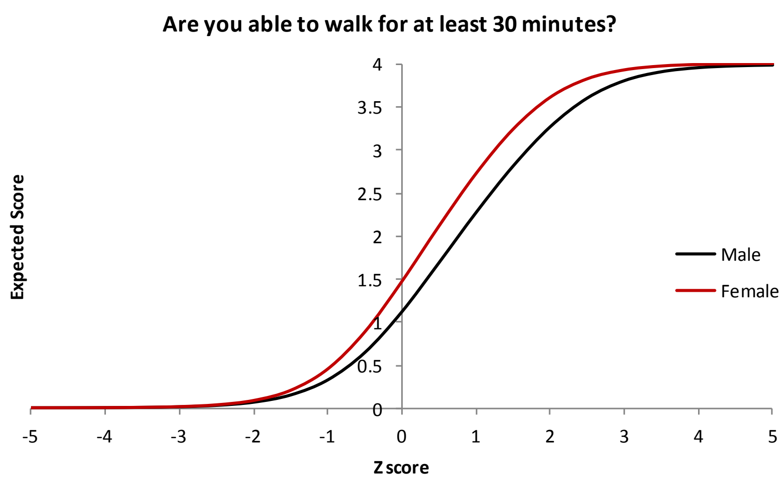
309 of Items in the WD-FAB item bank displayed DIF, mostly gender based. This suggests that DIF appears to have some impact on measuring function at the activity level, and future self-report measures should be sure to include diverse samples to be powered enough to conduct these analyses. We can now be confident that WD-FAB scores are reflections of true function without concerns about measurement error related to a respondent's gender, race/ethnicity, or age.

This paper is the first step towards long-term monitoring of trends in Social Security disability program applicants. Added to this monitoring could be trends in claimants' diagnoses, past work experiences, and current job market factors. Knowing who applies for Social Security Disability programs, when, and why is important for future planning for the disability programs, and to help understand work disability as a whole.

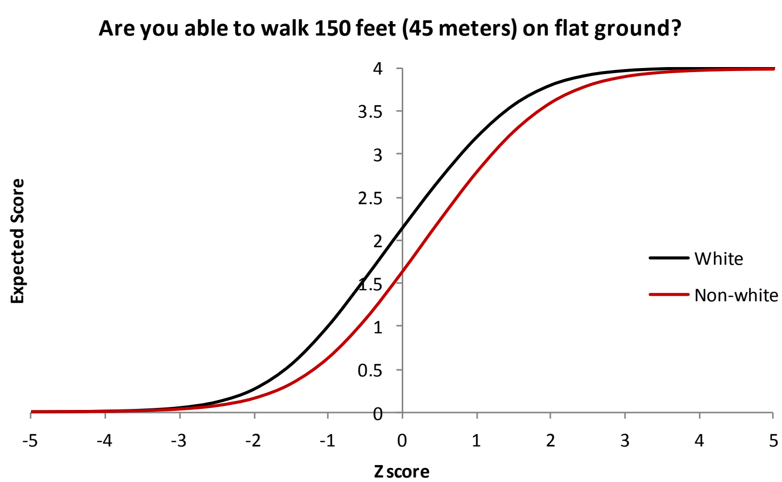
Appendix 2.1 Item Characteristic Curves for WD-FAB items displaying DIF



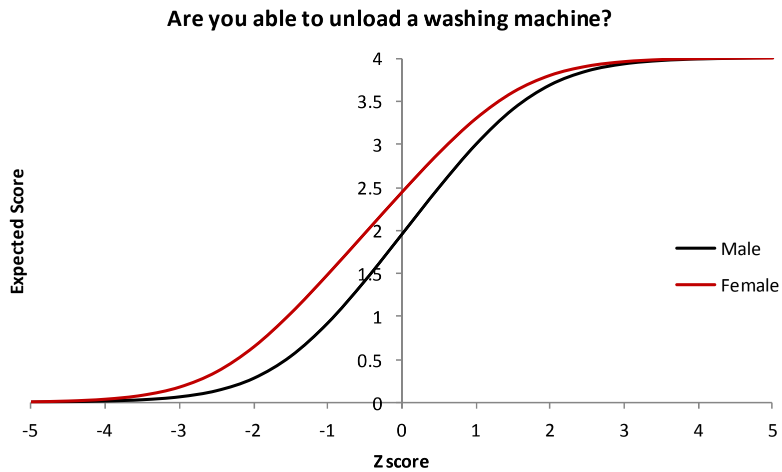
At same score level, claimants age>=55 are more like to say "able" than those age<55



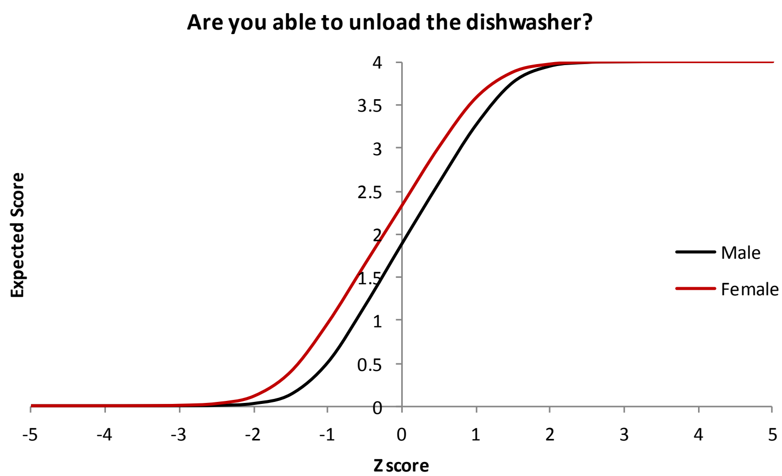
At same score level, female claimants are more like to say "able" than males.



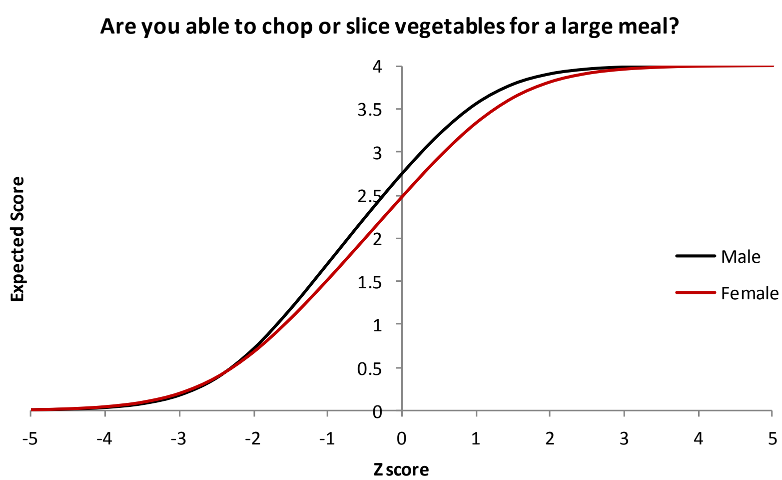
At same score level, white claimants are more like to say "able" than non-white claimants.



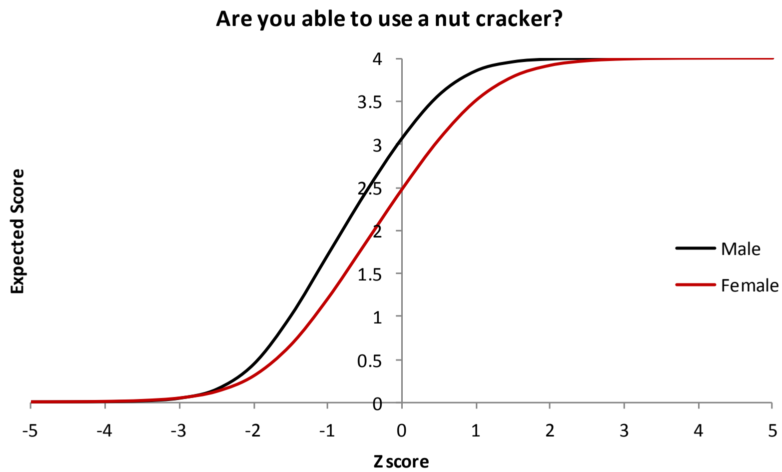
At same score level, female claimants are more like to say "able" than males.



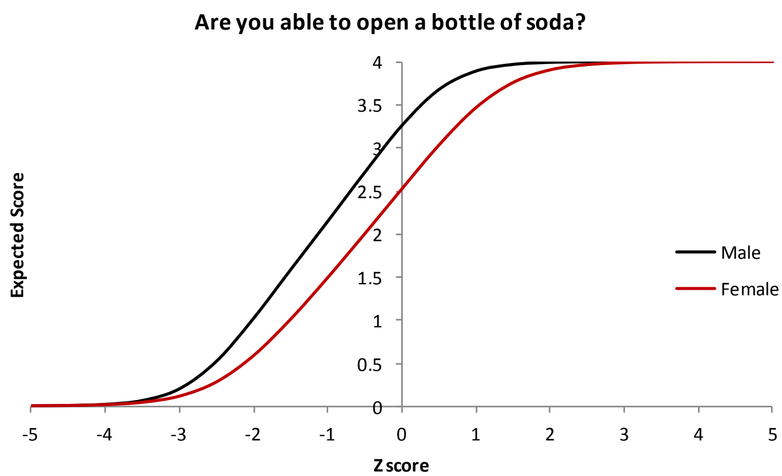
At same score level, female claimants are more like to say "able" than males.



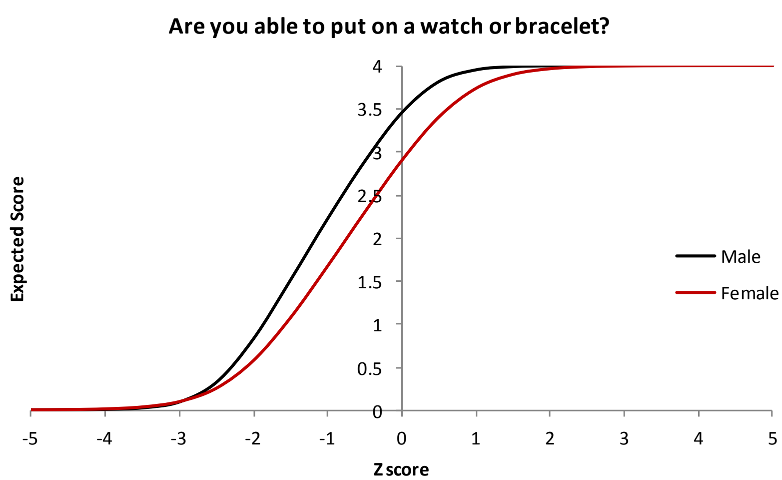
At same score level, male claimants are more like to say "able" than females.



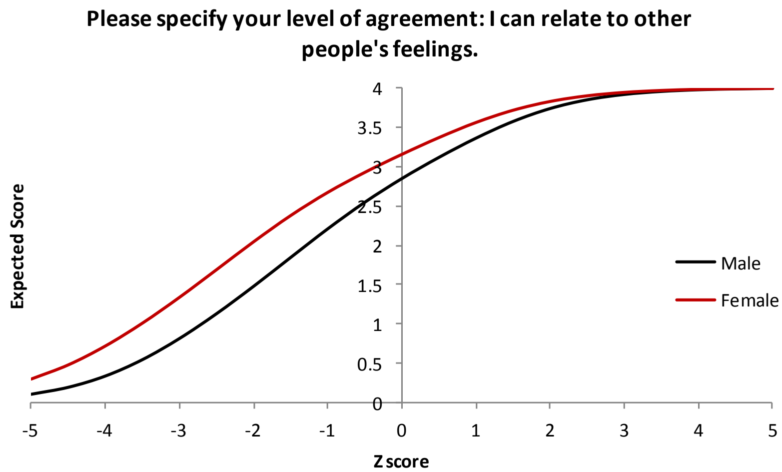
At same score level, male claimants are more like to say "able" than females.



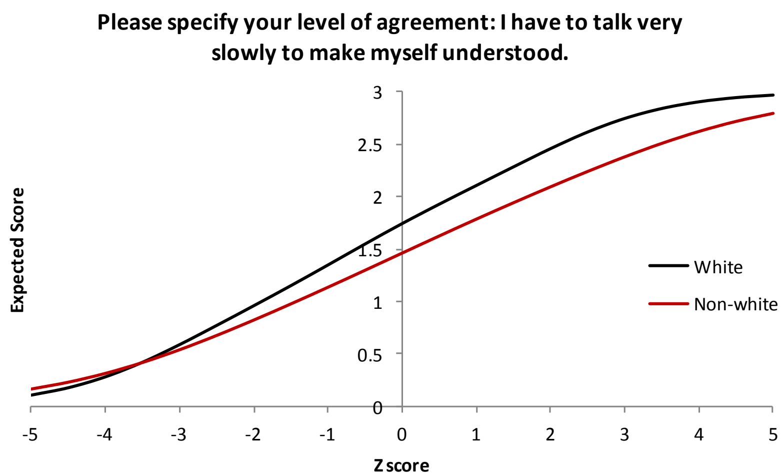
At same score level, male claimants are more like to say "able" than females.



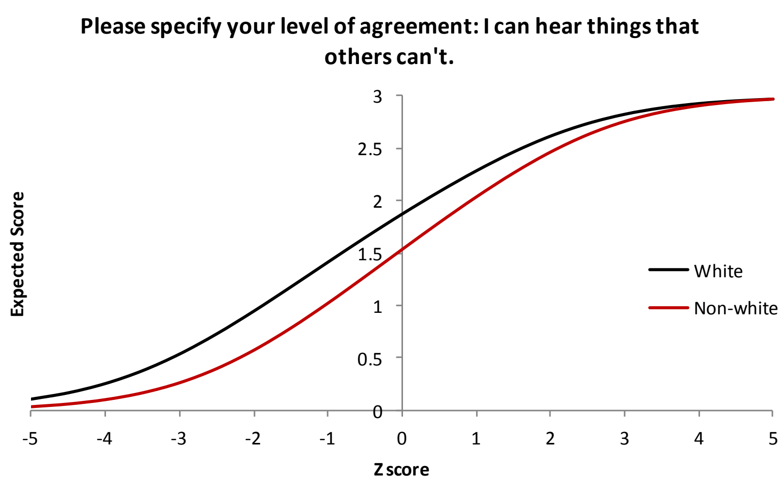
At same score level, male claimants are more like to say "able" than females.



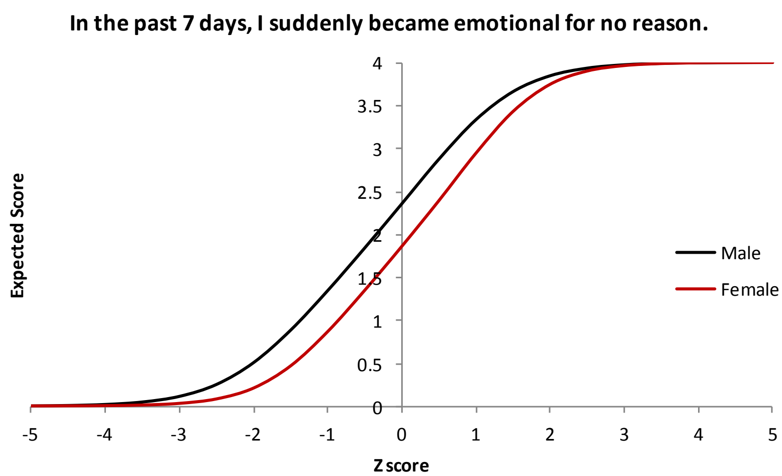
At same score level, male claimants are more like to say "disagree" than females.



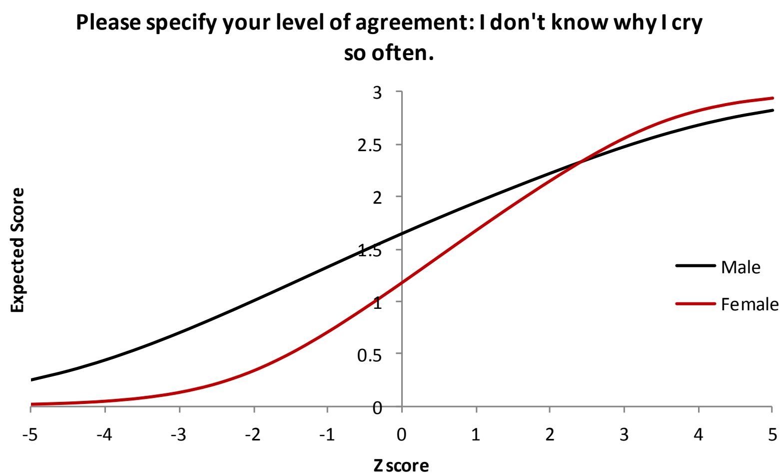
At same score level, white claimants are more like to say "disagree" than non-white claimants.



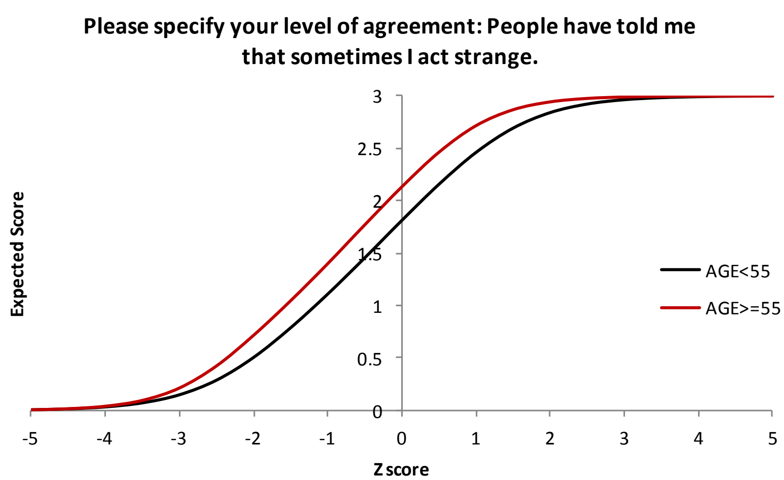
At same score level, white claimants are more like to say "disagree" than non-white claimants.



At same score level, male claimants are more like to say "disagree" than females.



At same score level, male claimants are more like to say "disagree" than females.



At same score level, people age>=55 are more like to say "disagree" than those age<55.

At same score level, people age>=55 are more like to say "disagree" than those age<55.

Score>-2.5 range, at same score level, males are more like to say "disagree" than females. Score<=-2.5 range, at same score level, females are more like to say "disagree" than males.

References

1. Unfit for work: The startling rise of disability in America. http://apps.npr.org/unfit-for-work/. Updated 2013.

2. Former social security commissioners and others respond to our disability story. http://www.npr.org/sections/money/2013/04/18/177745599/former-social-security-commissioners-and-others-respond-to-our-disability-story. Updated 2013.

3. Social Security Administration. Annual statistical report on the social security disability insurance program, 2014.

4. Wu AY. How has the changing nature of work contributed to SSDI growth? 2015.

5. Coe NB, Haverstick K, Munnell AH, Webb A. What explains state variation in SSDI application rates? *Center for Retirement Research at Boston College Working Paper*. 2011(2011-23).

6. Coe NB, Haverstick K, Munnell AH, Webb A. Why do state disability application rates vary over time? *Center for Retirement Research at Boston College Working Paper*. 2012(12-2).

7. Rivers D. Sample matching: Representative sampling from internet panels. *Polimetrix White Paper Series*. 2006.

8. Lord FM. *Applications of item response theory to practical testing problems.* Routledge; 1980.

9. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care*. 2007;45(5 Suppl 1):S22–31. doi: 10.1097/01.mlr.0000250483.85507.04 [doi].

10. Norris JM. Computer-adaptive testing: A primer. *Language Learning & Technology*. 2001;5(2):23–27.

11. Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika*. 1989;54(3):427–450.

12. Wang S, Wang T. Precision of warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*. 2001;25(4):317–331.

13. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*. 1969.

14. Hambleton R, Pitoniak M. Testing and measurement: Advances in item response theory and selected testing practices. 2002.

15. Meterko M, Marfeo EE, McDonough CM, et al. The work disability functional assessment battery (WD-FAB): Feasibility and psychometric properties. *Arch Phys Med Rehabil*. 2014.

16. Ni P, McDonough CM, Jette AM, et al. Development of a computer-adaptive physical function instrument for social security administration disability determination. *Arch Phys Med Rehabil*. 2013;94(9):1661–1669.

17. Marfeo EE, Ni P, Haley SM, et al. Development of an instrument to measure behavioral health function for work disability: Item pool construction and factor analysis. *Arch Phys Med Rehabil*. 2013;94(9):1670–1678.

18. Langer MM. *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation*. 2008.

19. Woods CM, Cai L, Wang M. The langer-improved wald test for DIF testing with multiple groups evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*. 2013;73(3):532–547.

20. Edelen MO, Stucky BD, Chandra A. Quantifying ‘problematic’ DIF within an IRT framework: Application to a cancer stigma index. *Quality of Life Research*. 2013:1–9.

21. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law Hum Behav*. 2005;29(5):615.

22. Abel JR, Gabe TM, Stolarick K. Workforce skills across the urban-rural hierarchy. 2012.

23. Autor D. The polarization of job opportunities in the US labor market: Implications for employment and earnings. *Center for American Progress and The Hamilton Project*. 2010.

24. Lin S, Beck AN, Finch BK, Hummer RA, Master RK. Trends in US older adult disability: Exploring age, period, and cohort effects. *Am J Public Health*. 2012;102(11):2157–2163.

25. Loprest P, Maag E. The relationship between early disability onset and education and employment. *Journal of Vocational Rehabilitation*. 2007;26(1):49–62.

26. Hopman W, Harrison M, Coo H, Friedberg E, Buchanan M, Van Den Kerkhof E. Associations between chronic disease, age and physical and mental health status. *Chronic Dis Can*. 2009;29(3):108–116.

27. Ward BW. Prevalence of multiple chronic conditions among US adults: Estimates from the national health interview survey, 2010. *Preventing chronic disease*. 2013;10.

28. Haveman R, Wolfe B. The economics of disability and disability policy. *Handbook of health economics*. 2000;1:995–1051.

29. Duggan M, Imberman SA. Why are the disability rolls skyrocketing? the contribution of population characteristics, economic conditions, and program generosity. In: *Health at older ages: The causes and consequences of declining disability among the elderly.* University of Chicago Press; 2009:337–379.

30. Hendley AA, Bilimoria NF. Minorities and social security: An analysis of ethnic differences in the current program. *Soc. Sec. Bull.* 1999;62:59.

31. Kington RS, Smith JP. Socioeconomic status and racial and ethnic differences in functional status associated with chronic diseases. *Am J Public Health*. 1997;87(5):805–810.

32. Williams DR. Race, socioeconomic status, and health the added effects of racism and discrimination. *Ann N Y Acad Sci*. 1999;896(1):173–188.

33. Nichols A, Simms M. Racial and ethnic differences in receipt of unemployment insurance benefits during the great recession. *Unemployment and Recovery Project Brief*. 2012;4.

34. Kao G, Thompson JS. Racial and ethnic stratification in educational achievement and attainment. *Annual review of sociology*. 2003;29(1):417–442.

35. DeBose CE. Codeswitching: Black English and standard English in the African-American linguistic repertoire. *Journal of Multilingual & Multicultural Development*. 1992;13(1–2):157–167.

36. Fleishman JA, Spector WD, Altman BM. Impact of differential item functioning on age and gender differences in functional disability. *J Gerontol B Psychol Sci Soc Sci*. 2002;57(5):S275–84.

37. Sullivan O. Time waits for no (wo) man: An investigation of the gendered experience of domestic time. *Sociology*. 1997;31(2):221–239.

38. Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Quality of Life Research*. 2007;16(1):33–42.

39. Teresi JA, Ocepek-Welikson K, Kleinman M, et al. Analysis of differential item functioning in the depression item bank from the patient reported outcome measurement information system (PROMIS): An item response theory approach. *Psychol Sci Q*. 2009;51(2):148–180.

40. Rose M, Bjorner JB, Becker J, Fries J, Ware J. Evaluation of a preliminary physical function item bank supported the expected advantages of the patient-reported outcomes measurement information system (PROMIS). *J Clin Epidemiol*. 2008;61(1):17–33.

41. Iwata N, Turner RJ, Lloyd DA. Race/ethnicity and depressive symptoms in community-dwelling young adults: A differential item functioning analysis. *Psychiatry Res*. 2002;110(3):281–289.

42. Gao Y, Zhu W. Differential item functioning analysis of the 2003–04 NHANES physical activity questionnaire. *Res Q Exerc Sport*. 2011;82(3):381–390.

43. Hafner-Eaton C. Physician utilization disparities between the uninsured and insured: Comparisons of the chronically III, acutely III, and well nonelderly populations. *JAMA*. 1993;269(6):787–792.

44. Williams DR, Collins C. US socioeconomic and racial differences in health: Patterns and explanations. *Annual review of sociology*. 1995:349–386.

45. Willimas D, Yu Y, Jackson J. Racial differences in physical and mental health. *Journal of Health Psychology*. 1997;2:335–351.

46. Yee S. Health and health care disparities among people with disabilities. *Washington, DC: Disability Rights Education & Defense Fund*. 2011.

47. Fremstad S. Half in ten. *Why taking disability into account is essential to reducing income poverty and expanding economic inclusion. Washington, DC: Center for Economic and Policy Research*. 2009.

48. Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. *Medical Decision Making*. 2006;26(4):391–400.

49. Milner A, LaMontagne A, Aitken Z, Bentley R, Kavanagh A. Employment status and mental health among persons with and without a disability: Evidence from an Australian cohort study. *J Epidemiol Community Health*. 2014:jech-2014-204147.

50. McFarland MJ, Wagner BG. Does a college education reduce depressive symptoms in American young adults? *Soc Sci Med*. 2015;146:75–84.

51. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustun TB. Age of onset of mental disorders: A review of recent literature. *Curr Opin Psychiatry*. 2007;20(4):359–364. doi: 10.1097/YCO.0b013e32816ebc8c [doi].

52. Lee M, Mather M. *US labor force trends.* Vol 63. Population Reference Bureau; 2008.

CHAPTER THREE: DETERMINING FUNCTIONAL PROFILES OF COMMON CONDITIONS

Introduction

The United States Social Security Administration's (SSA) disability programs provide financial support to over 10.3 million Americans through the SSDI program and 8.3 million additional individuals through the SSI program, making this program the largest federal provider of financial assistance to disabled workers and their families.[1] In 2014 alone, over 750,000 recipients were awarded Social Security Disability Insurance (SSDI) benefits because of a work limiting health condition.[1] The Social Security Act defines work disability as an inability to engage in "substantial gainful activity due to any medically determinable physical or mental impairment that can be expected to result in death or to last for a continuous period of not less than 12 months."[1] The past two decades have seen a steady and at times precipitous rise in the number of SSDI applicants and beneficiaries.

The evaluation process for determining work disability includes five sequential steps.[2,3] First, the SSA verifies that a person is not currently employed in what is referred to as a substantially gainful activity (SGA). Second, the person applying for disability benefits (referred to as a claimant) must demonstrate that a condition is present that interferes with what would be basic work activities. These first two steps are to verify the basic eligibility criteria for meeting the statutory definition of work disability. Then, the applicant's list of conditions—referred to as allegations—along with their associated medical documentation, are compared against the SSA's "Listing of Impairments," which are grouped into fourteen major body systems. SSA's 14 "listings" are: Musculoskeletal System, Special Senses and Speech, Respiratory System, Cardiovascular System, Digestive System, Genitourinary Disorders, Hematological Disorders, Skin Disorders, Endocrine Disorders, Congenital Disorders that Affect Multiple Body Systems, Neurological, Mental Disorders, Cancer, Immune System Disorders. If the claimant's impairment does not meet specified severity criteria in the listings, then he/she advances to the next steps: previous and potential work capacity. At the fourth step, the SSA uses medial evidence and other relevant documents to complete what is called a Residual Functional Capacity (RFC) Assessment. There is both a Physical and Mental RFC, and each is filled out on behalf of the claimant, with instructions that each activity is to be evaluated within the context of the claimant's ability to sustain that activity over a normal workday. For the Physical RFC, activities include limitations in exerting physical energy, posture, manipulation, visual acuity, communication, and the environment. Mental RFC activities relate to understanding and memory, concentration, social interaction, and adaptation. Finally, the SSA will examine whether a claimant can perform any past work or other substantially gainful work available in the national economy based on the RFC assessments.

The basis for the SSA listings are anatomic, diagnostic, and functional, the mix of which varies from listing to listing, of which there are more than one hundred. For each body system, there are descriptions of diagnoses and impairments considered severe enough to prevent an individual from doing any gainful activity. The body systems and criteria were initially developed through expert consensus for the purpose of expediting claims. The rationale was if there were a set of conditions and severity level that the Administration could agree would qualify a claimant as definitely work-disabled, checking if a claimant met those criteria would enhance efficiency of the determination process. The SSA would save on resources spent on subsequent stages, and people who are in need of financial assistance would receive their benefits in a timely fashion. Initially, the listing process was predominately used as determination justification—once accounting for 90% of awards. Currently, however, only 52% of beneficiaries received their determination at the listing stage.[4] The cause of this drop is not known. However, one potential explanation is the change in the types of conditions claimants have over the past several decades. Historically, cancer and cardiovascular conditions accounted for more than half of the benefits awarded; currently musculoskeletal system and connective tissue, and mental disorders are the most common categories of diseases.[5] It is possible that this change is related to the decreased efficiency or possibly relevance of the listing stage. Musculoskeletal and mental condition severity are more difficult to characterize with traditional biomedical criteria, and the symptoms can fluctuate over time.

In 2007, the Institute of Medicine Committee on Improving the Disability Decision Process issued their final report and recommendations regarding the SSA's medical procedures and criteria for disability determination.[4] The listings and body system groupings, their validity, and utility were a large focus of the report's findings and recommendations. The IOM committee pointed out that use of the listings relies on an assumption there are groups of individuals identifiable based upon medical impairments, and these impairments are related to functional impairments, however "little work has been done to establish the extent to which the Listings are a valid measure of work disability."[4] Beyond the SSA context, the link between diagnosis and function has not been established for a wide range of conditions.[6,7] The report also mentions the increasing number of claimants who report multiple impairments, and suggest that rather than trying to match multiple conditions to a single listing, common comorbidities could possibly be added to the listings, instead of waiting to assess them at steps 4 and 5. Finally, the committee emphasized the importance of robust functional assessments at the listing stage, and recommended improvements in this area.

Functional profiles could help present a whole picture of a person's functional ability across a range of content areas—referred to as domains—that are considered potentially important for work-related functioning. A functional profile is a display of a person's scores on a functional measure, across multiple domains, rather than a singular score. A profile approach can examine if there is a unique constellation of functional abilities for individuals with different conditions.

The main challenge for the Listing process is the untested assumption that individuals in a single listing diagnostic group will have similar levels of functioning in similar ways. This paper attempted to test that assumption, and has two aims. The first is to identify common patterns of allegations among SSI and SSDI claimants, as a means of assessing if the current fourteen body system categories are distinct. The second aim is to develop unique functional profiles for the different diagnostic categories for the purpose of exploring if there is indeed a link between diagnosis and function. The hypotheses for this aim are (1) claimants with a physical health allegation will have lower scores on all physical function scales compared to all other claimants, and (2) claimants with a mental health allegation will have lower scores on all mental health function scales compared to all other claimants.

Conceptual Framework

One of the leading models for characterizing the multifaceted nature of disability is the World Health Organizations (WHO) International Classification of Functioning, Disability and Health (ICF). The ICF calls for research on disability and health to be examined with a "whole person" perspective. The ICF is a biopsychosocial model of disability that characterizes disability as a result of the complex relationship between individual health characteristics, environmental factors, personal factors, and the extent to which those interactions affect a person's ability to perform activities and participate in daily roles such as work. This study examines the relationship between Health Conditions, measured by claimant's allegations and function at the Activity level, as measured by the WD-FAB.

Methods

Outcome Measure

The Work-Disability Functional Assessment Battery (WD-FAB), is a new self-reported measure of activity limitations relevant to work and work disability.[8–12] The WD-FAB was created so that the Social Security Administration (SSA) can collect systematic and comprehensive information about claimant functioning to inform the disability determination process. The WD-FAB uses Item Response Theory (IRT).[13] In measures that employ IRT, each item is ordered in a hierarchy along a unidimensional construct, and scores are based upon probability models that represent the likelihood that a respondent selects an answer choice, given the difficulty level of the item and their ability on that construct.[8,14, 13] In this analysis, the battery consists of seven scales: three physical factors (Basic Mobility, Upper Body Function, and Fine Motor Function), and four nonphysical factors (Cognition & Communication, Resilience/Sociability, Self-Regulation, and Mood & Emotions).[15–17] Raw scores on the WD-FAB are first scored in logits, converted to Z-scores which are then transformed to T-scores. The T-scores are based upon the score distribution of a general sample of working age adults, with a mean of 50 and a standard deviation of 10, with higher scores representing higher function. WD-FAB scores are examined for the claimant sample only in this study, but scores are referenced to a 50 from the general sample of working age adults.

Sample

This study used data from the development of the WD-FAB. The study involved two samples, referred to as the “claimant sample,” and the “general sample.” In the claimant sample, a random sampling from a pool of 16,500 SSA disability claimants who applied for benefits in the last 2 months, stratified by urban/rural status across the ten national SSA regions, yielded a final sample size of 3,793. The general sample contains a total of 2,100 working age adults, obtained using sample matching whereby samples representative of a study-appropriate target population could be constructed from large but unrepresentative pools of opt-in survey respondents from the online research organization YouGov.[18] This study uses scores from the claimant sample, but scores on the WD-FAB are scaled to the general sample, and therefore scores can be interpreted relative to working age sample.

Allegation Coding

When claimants applied for benefits, they list a series of conditions that make it difficult for them to work, referred to as allegations. Based upon the definitions in the SSA's "Blue Book," each claimant's allegations were classified into a potential listing/body system group. The result was fourteen new dichotomous variables (one for each body system). For example, if a claimant listed "anxiety, mitral valve prolapse, arthritis; heart" the listings for this person would be Mental Conditions, Cardiovascular, Musculoskeletal. This process did not require parsing out if the "heart" description referred to the mitral valve prolapse or not, because they had already been classified as Cardiovascular.

Latent Class Analysis

The 14 listing/body system group variables created in the previous step are not mutually exclusive, which presents challenges for comparing scores across listing/body system groups; in addition, making comparisons across fourteen groups was not particularly feasible or practical. Because of this, a Latent Class Analyses (LCA) was conducted based upon the 14 dichotomous listing/body system group variables.[19–21] The goal of LCA is to identify common subgroups/typologies within a latent variable. In this case the goal is to identify common allegations, and patterns of 'co-allegations.' LCA was conducted for 1–7 class models. Model choice was determined based upon a series of fit criteria. The best fit model was the one that showed the lowest Akaike information criteria (AIC), Bayesian information criteria (BIC), and sample size adjusted Bayesian information criteria (adj-BIC). In the interest of also selecting the most parsimonious model (model with fewest number of classes), three different likelihood ratio tests (Vuong-Lo –Mendell-Rubin LRT, Lo-Mendell-Rubin Adjusted LRT, and Parametric Bootstrapped LRT) were examined. Significant p values for these tests indicate that the current model has better fit compared to the model with one fewer class.[22]

Developing Functional Profiles

Independent t-tests and ANOVA with post-hoc Tukey tests were used to assess differences in the average functioning scores across the latent classes identified. The significance level for all analyses was set at alpha 0.05. Where statistical differences existed in mean scores between the groups, additional analysis of magnitude of the mean differences was performed by calculating Cohen's *d* effect size. Small (<.2 to <.4), medium (0.4–0.8), and large effects (>0.8) were defined for Cohen's *d.*[23] Radar plots are a way to graphically display multiple continuous variables each on its own axis, all starting from the same point; the result resembles as spider web. Radar plots were generated to display the mean scores for each group to illustrate the unique functional phenotype of each group.

Results

The claimant sample was 45.48% male, average age 47.54 years. 66.28% of the sample reported that they identified as white, and 27.22% identified as black/African American (**Table 3.1**). WD-FAB scores for the entire claimant sample are also presented in **Table 3.1**. The physical function scores range from 36.17–38.81, the mental health function scores range from 41.13–46.34. Overall, claimants are functioning at a level lower than general working age adults in the U.S.

Table 3.1 Sample Characteristics

Characteristic	**Claimant Sample N=3,793**			
	N		**%**	
Age mean (sd)	47.54 (11.77)			
Men	1,721		45.48%	
White	2,489		66.28%	
Black	1,022		27.22%	
Asian	33		0.88%	
Latino	298		7.93%	
Urban	2,742		72.64%	
Education				
Less than HS	597		15.74%	
HS or GED	1,214		32.01%	
Associates Degree	320		8.44%	
Vocational Training	261		6.88%	
Some College No Degree	904		23.83%	
College or More	489		12.89%	
WD-FAB Scale	**Mean**	**SD**	**95% LCL**	**95% UCL**
Basic Mobility	37.39	5.61	37.22	37.58
Upper Body	36.17	5.74	35.98	36.38
Fine Motor	38.81	6.51	38.59	39.04
Cognition & Communication	41.73	7.71	41.49	41.98
Self-Regulation	45.67	11.50	45.31	46.038
Resilience	46.34	10.77	46.00	46.69
Mood & Emotions	41.13	13.52	40.67	41.61

Allegation Coding

The results of the initial allegation coding are presented in **Table 3.2**. The most common listing/body system group categories were Musculoskeletal (61.01%), Mental

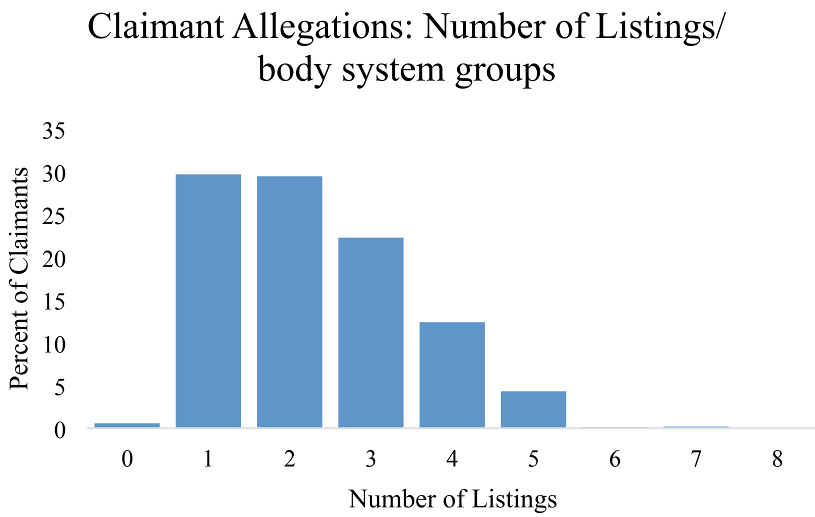
(49.35%), and Cardiovascular conditions (31.51%). Least common were Hematological (2.27%), Skin (1.48%), and Congenital conditions (0.11%).

Table 3.2 Frequencies of Allegations by Listing/body system group

Group	n	%
Musculoskeletal	2314	61.01
Mental	1872	49.35
Cardiovascular	1195	31.51
Endocrine	711	18.75
Respiratory	670	17.66
Neurological	638	16.82
Digestive	467	12.31
Special Senses & Speech	406	10.70
Cancer	189	4.98
Genitourinary	155	4.09
Immune	146	3.85
Hematological	86	2.27
Skin	56	1.48
Congenital	4	0.11

Multiple claimants had allegations that were classified into multiple listing groups. **Figure 3.1** shows the distribution of the frequency of listing groups. The average number listing group classifications was 2.35 (1.24), the median was 2, and the mode was 1. As reflected in the figure, most claimants had allegations that went into one or two listing/body system group. However, many individuals had more allegations, including as many as eight different listing groups. There were also a few individuals whose allegations did not fit into any listing group; many of those individuals had allegations such as fibromyalgia and chronic fatigue syndrome. These conditions are not classified into a body system, which is the basis for many of the listings. The WD-FAB scores by all 14 Listing/body system group is available in **Appendix 3.1**.

Figure 3.1 Number of possible listings for claimants based upon coding of allegations.



Latent Class Analysis

Seven different latent class models were tested for comparison, the results of which are summarized in **Table 3.3**. The Akaike information criteria (AIC) was the lowest for the 7-class model; the BIC and sample-adjusted BIC were lowest for the 5-class model. The three likelihood ratio tests suggested that the 5-class model was optimal. The 5-class model was then selected for the rest of the analyses.

Table 3.3. Latent Class Analysis Results

Test	1 Class	2 Classes	3 Classes	4 Classes	5 Classes	6 Classes	7 Classes
AIC	36632.70	36214.83	36052.70	35909.84	35840.46	35798.28	35760.05
BIC	36720.08	36395.82	36327.30	36278.05	36302.29	36353.72	36409.11
Adjusted BIC	36675.59	36303.67	36187.49	36090.58	36067.16	36070.92	36078.64
VLMR LRT		0.0000	0.0000	0.0000	0.0084	0.3958	0.0908
VLMR Adjusted LRT		0.0000	0.0000	0.0000	0.0087	0.3988	0.0926
Parametric Bootstrapped LRT		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

A table of the probabilities of a claimant belonging to each listing/body system group occurring was generated to describe each of the latent classes (**Table 3.4**). As a general rule, listing/body system groups with a greater than 0.5 representation in the class were considered in the class naming; in addition, listings/body system groups that occur at a relatively low frequency in the dataset, but were disproportionately highly probable in a class were also considered in describing them. All claimants assigned to the first class had a musculoskeletal allegation, and all claimants assigned to the second class had a cancer allegation thus the classes were named “Musculoskeletal,” and “Cancer,” respectively. The third class had no 1.00 probabilities, but rather several high probability allegations, specifically musculoskeletal and cardiovascular. This class was named Multisystem. Claimants in Class 4 had a 0.660 probability of being in the neurological listing/body system group, and a 0.293 probability of being in the special senses and speech listing/body system group (compared to 10.70% frequency in the entire sample), and was named the Neurological & Sensory class. Finally, all claimants in the last class have a mental listing, and there is a very low probability of any other listing in the class, leading this class to be named Mental.

The same table can also be examined by looking at the listings rows. For example, people in the Musculoskeletal, Multisystem, and Neurological & Sensory class all have fair representation of a musculoskeletal listing. Similarly, a mental listing is likely in most groups except the Cancer class. Also, worth noting is that the probability of having an endocrine listing and a respiratory listing is at its highest (0.346 and 0.283) in the Multisystem class.

Table 3.4: Probabilities of Each Listing Group by Class

Listing	Musculoskeletal	Cancer	Multisystem	Neurological & Sensory	Mental
Musculoskeletal	**1.000**	0.081	**0.577**	0.300	0.000
Special Senses & Speech	0.056	0.000	0.129	0.293	0.025
Respiratory	0.105	0.042	0.283	0.046	0.031
Cardiovascular	0.081	0.000	**0.567**	0.190	0.000
Digestive	0.088	0.042	0.179	0.014	0.071
Genitourinary	0.001	0.062	0.080	0.000	0.005
Hematological	0.004	0.000	0.043	0.000	0.007
Skin	0.008	0.000	0.022	0.007	0.011
Endocrine	0.045	0.034	0.346	0.036	0.024
Congenital	0.002	0.000	0.000	0.007	0.000
Neurological	0.114	0.089	0.137	**0.660**	0.085
Mental	0.443	0.085	0.458	0.347	**1.000**
Cancer	0.018	**1.000**	0.033	0.050	0.025
Immune	0.021	0.027	0.056	0.015	0.029

Latent Class Analysis is based upon a model, but it also assigns each observation (claimant) to a class. **Table 3.5** presents a comparison of what the model estimates the frequency of class 'membership' is to what the actual assignment is. The two largest classes are Musculoskeletal (30.35% vs. 33.19%) and Multisystem (48.37% vs. 46.53%). The similarities between the model-based and actual frequencies suggest that there was minimal error in class assignment. **Table 3.6** presents a brief summary of the demographics of each class. The Cancer class includes older claimants, with a mean age of 52.29 years old, and the mental class is much younger than the other classes, with a mean age of 37.89 years old. There also appears to be proportionally fewer men in the Cancer and Mental class (40.26% and 40.33%, respectively). The Cancer class also appears to be less white, with 59.74% of that group identifying as white, where the other class ranges from 63.31%–70.76%.

Table 3.5. Final Class Counts and Proportions for the Latent Classes

	Musculoskeletal	Cancer	Multisystem	Neurological & Sensory	Mental	Total
Model(n)	1151.34	80.21	1835.01	323.46	402.94	3792.96
%	30.35	2.11	48.37	8.52	10.62	
Sample (n)	1259	78	1765	203	488	3793
%	33.19	2.05	46.53	5.35	12.86	

Table 3.6. Demographics of Each Latent Class

	Musculoskeletal		Cancer		Multisystem		Neurological & Sensory		Mental	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Age mean(sd)	47.37 (11.32)		52.29 (9.76)		50.12 (10.57)		47.49 (12.38)		37.89 (11.94)	
Men	583	46.34	31	40.26	815	46.31	96	47.29	196	40.33
White	881	70.76	46	59.74	1108	63.31	130	65.00	324	67.08
Latino	113	9.07	3	3.90	125	7.16	12	5.97	45	9.28
Urban	911	72.71	57	73.08	1249	71.13	151	75.12	374	76.80

There were several significant differences in average WD-FAB scale scores between the different listing/body system group displayed in **Figure 3.2** (additional information available in **Appendix 3.2**). On the physical function scales (**Figure 2a)**, the Musculoskeletal and Multisystem classes tended to have significantly lower scores than the other groups with the exception of fine motor. The Mental class had higher scores on the physical function scales compared to all groups, except for the Fine Motor scale, where there was no difference between the Mental and Cancer Class. Tukey tests for the Basic Mobility and Upper Body showed that the Mental class was its own group, Cancer and Neurological & Sensory formed a group, and Musculoskeletal, and Multisystem were grouped. For the Fine Motor scale, there were two groups, Mental and Cancer formed one; and Musculoskeletal, Multisystem, and Neurological & Sensory formed the other.

The mental function scales (**Figure 2b**), showed a similar but reversed trend. The Mental class by far had the lowest scores on all scales, ranging 34.06–41.74. On the other hand, the Cancer class seemed fairly comparable to the general US working sample, with scores ranging from 46.67–52.61. Tukey tests for the Cognition & Communication, Self-Regulation, and Resilience/Sociability scales showed that the Mental and Cancer class were each their own group and Musculoskeletal, and Multisystem Neurological & Sensory, formed the third... For the Mood & Emotions scale, the Mental class was its own group, Cancer and Neurological & Sensory formed a group, and Musculoskeletal, and Multisystem were grouped.

Figure 3.2a. Average WD-FAB Physical Function Scores by Class with Tukey Grouping

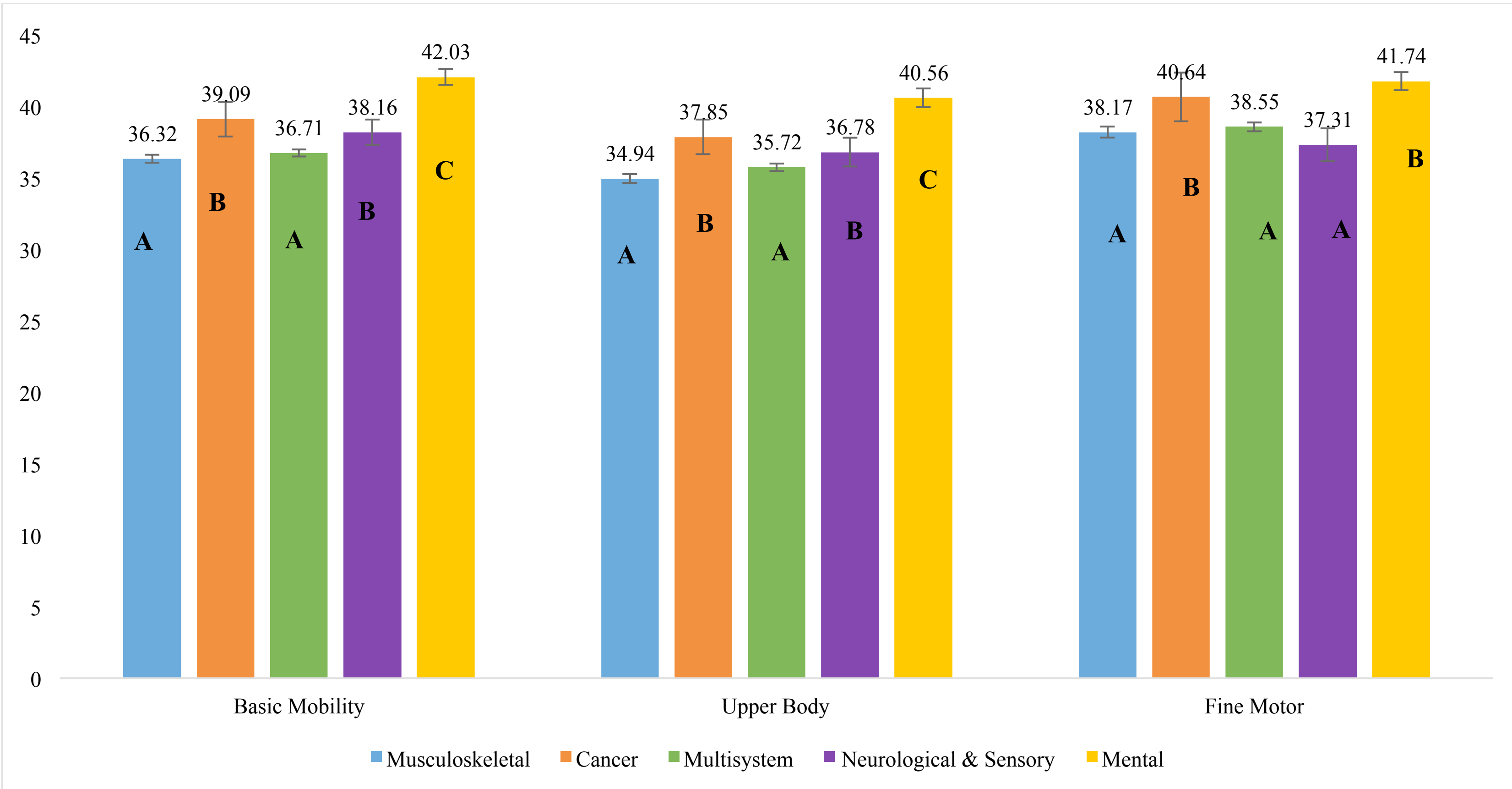
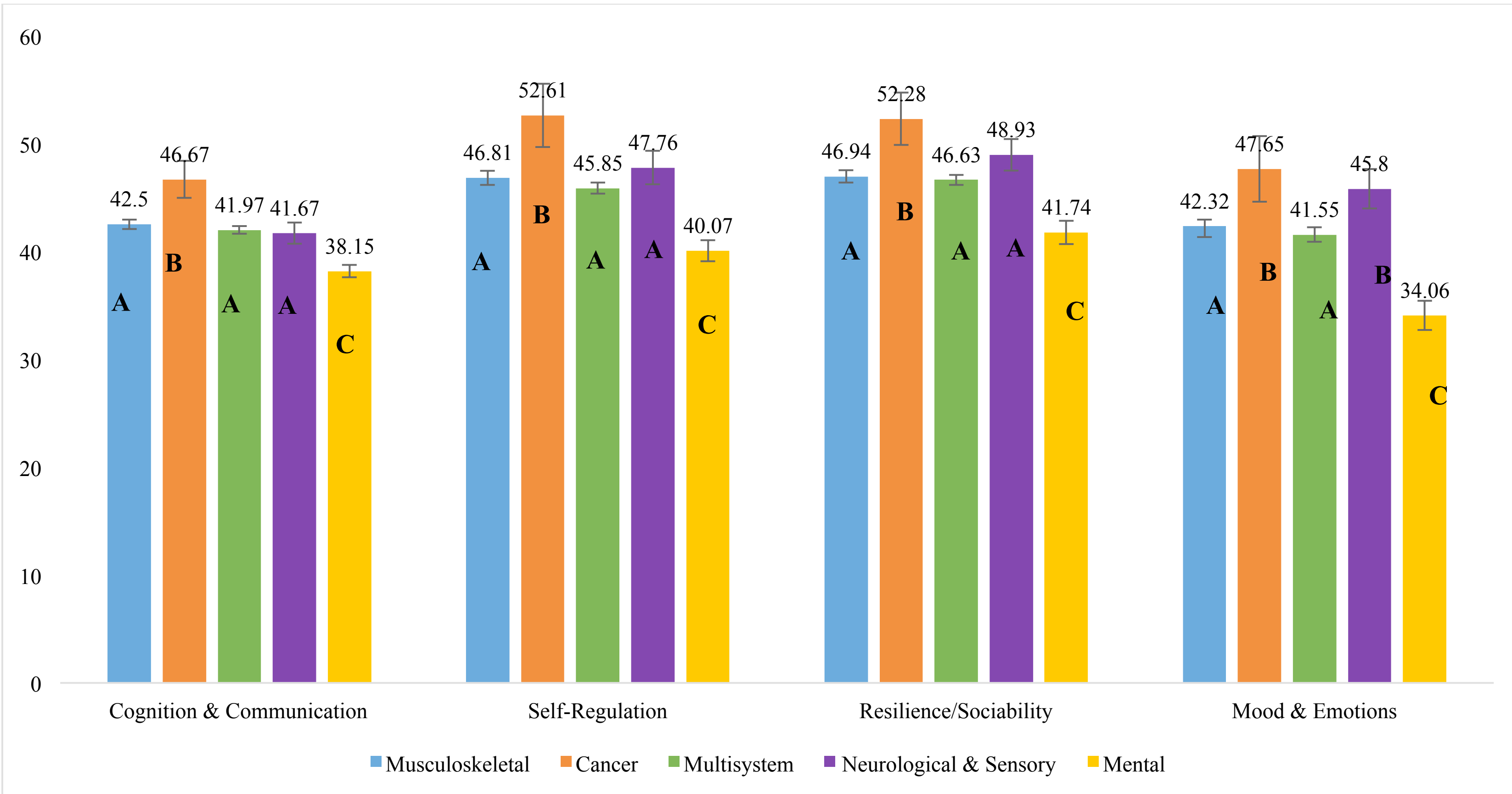


Figure 3.2b. Average Mental Function WD-FAB Scores by Class with Tukey Grouping



Developing Functional Profiles

Figures **3.3a** and **3.3b** present the WD-FAB scores for all classes on two radar (plots, physical function scales, and mental function scales). Scores closer to the center denote lower functioning, and scores closer to the perimeter denote better functional status. The maximum score on these figures is 50, the average score of the general US working age sample. There are two ways to visually examine how 'unique' a functional profile is: scale (the overall size of the plot—larger indicating higher functioning), and shape (the less similar the shapes are between profiles, the more unique the profiles are.

This analysis found three unique physical function profiles, one shared by the Musculoskeletal, Multisystem, and Neurological & Sensory classes, one for the Cancer class, and one for the Mental class. This analysis also identified four unique mental function profiles, one shared by the Musculoskeletal and Multisystem classes, and one for each of the Cancer, Neurological, and Mental classes.

For both the physical and mental function profiles, the Musculoskeletal and Multisystem classes are nearly overlapping each other, which is also reflected in their same Tukey groupings for all scales. The Neurological & Sensory has a similar shape to the Musculoskeletal and Multisystem group for the physical function scores, but the scale is *larger* for the Basic Mobility and Upper Body function, and lower for Fine motor function. For the mental function scales, the Neurological & Sensory group presents a unique shape, diverging from the previous three classes in shape with higher scores in the Mood & Emotions scale.

Figure 3.3a. Profiles of WD-FAB Physical Function scores by diagnosis class

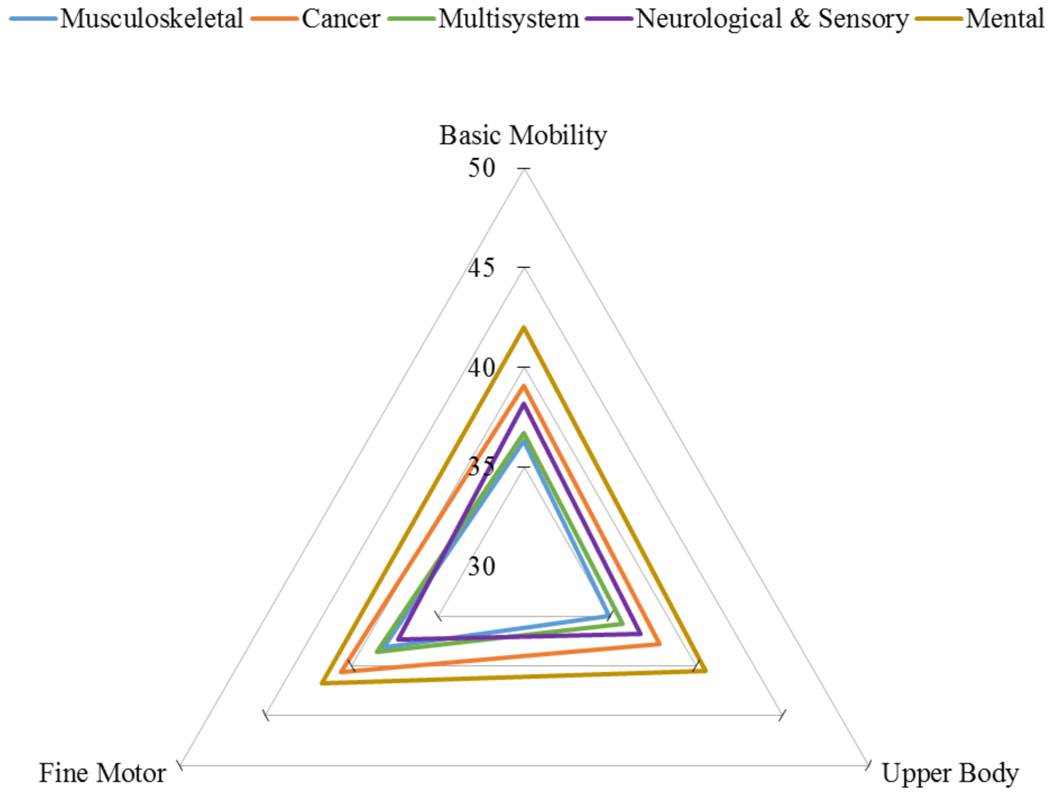
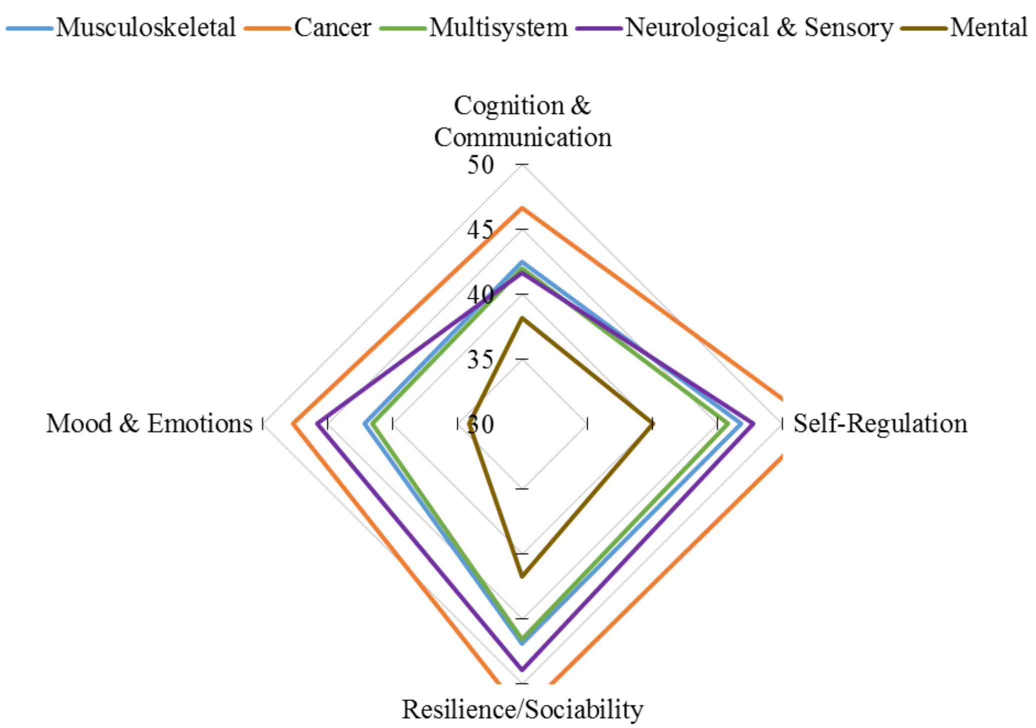


Figure 3.3b. Profiles of WD-FAB Mental Function scores by diagnosis class



The Cancer class presents a similar shape but larger scale physical function profile, compared to the Musculoskeletal, Multisystem, and Neurological & Sensory classes, indicating that physical function in this group is higher in this class, but follows a similar pattern. In Mental Health, however, the Cancer class is quite unique in scale and shape. The scores on two of the scales, Self-Regulation and Resilience/Sociability, go beyond the figures scale itself (above 50). For the Mental class, the physical function scores are a slightly more symmetrical shape, but overall quite similar. The scale, however is visibly larger than most of the other classes. The one notable exception is the closeness of fine motor scores with the Cancer class—which is also reflected in their Tukey grouping for this scale. The mental functioning scores for this class is the most unique profile observed. Not only it is much smaller in scale, its shape is also compressed on the left—at the Mood & Emotions scale.

The radar plots in **Figure 3.4** present a comparison of the average scores of each class compared to the *rest of the claimant sample*. The actual scores, the results of the T-tests between the pairwise comparisons, and effect sizes for significant results are presented in **Table 3.7**. The results from both the figures and the table with be discussed together.

Claimants in the Musculoskeletal class (**Figures 3.4a–b**) were physically functioning poorer than other claimants (effect size 0.144–0.321), and mentally functioning slightly better than other claimants (effect size 0.083–0.149). The scores were lowest for upper body function (34.94), just over one and a half standard deviations below the scale mean. The shapes of the two plots are similar, indicating that the *pattern* of different scores in this class are similar to other claimants, but that they were physically on average 2–3 points lower on physical functioning and 1–2 points higher physically.

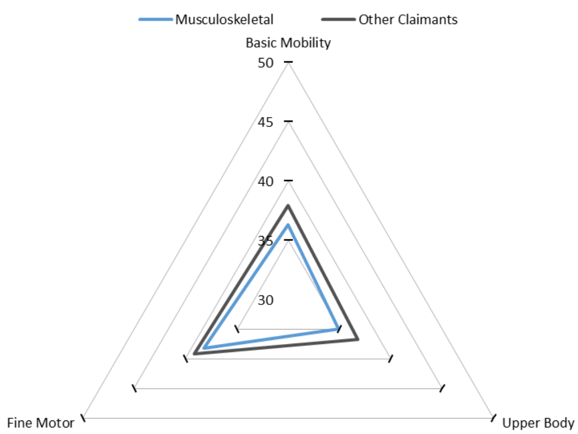
The functional phenotype for claimants in in the Multisystem class (**Figures 3.4c–d**) displayed a pattern of physical function scores that was similar to the Musculoskeletal class, functioning slightly lower than the rest of the sample (effect size 0.076–0.227). However, there were no differences in scores on the mental function sales. The Neurological & Sensory class functional profile (**Figures 3.4e–f**) is also unique in contrast to the rest of the claimants, and the profile of the other classes. The only difference in physical function scores was for Fine Motor function, (effect size 0.242). For the mental health function, claimants in this class were functioning better than other claimants on all scales except Cognition & Communication. Interestingly, this scale is also where the claimants had the *lowest* scores 41.67(6.79).

Claimants in the Cancer class (**Figures 3.4g–h**) showed a unique functional phenotype. On the physical function scales the pattern of the scores was similar to all other claimants, with the Cancer class scores slightly higher (effect size 0.287–0.308), though still lower than the scale's average of 50. For mental functioning, on the other hand, claimants in the Cancer class were generally functioning with average scores ranging from 46.67–52.61, which was significantly better than other claimants (effect size 0.493–0.656).

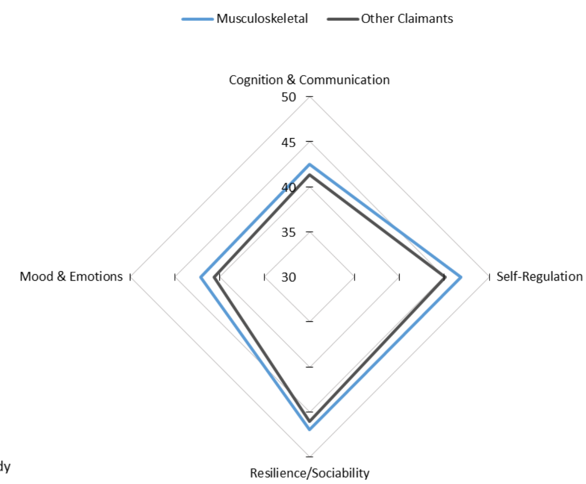
Finally, the Mental class also showed a unique functional profile (**Figures 3.4i–j**). The group was functioning significantly better than other claimants on the physical function scales (effect size 0.522–0.998), nevertheless the scores were still almost a full standard deviation lower than the *scale's* mean (40.56–42.03). On the mental function scales, claimants in the mental class were functioning much lower than other claimants (effect size 0.496–0.610). The largest effect size was also where the scores were lowest: the average Mood & Emotions score was 34.06, more than one and a half standard deviations lower than the scale mean.

Figure 3.4 Functional profile of WD-FAB scores for each diagnosis class compared to all other classes

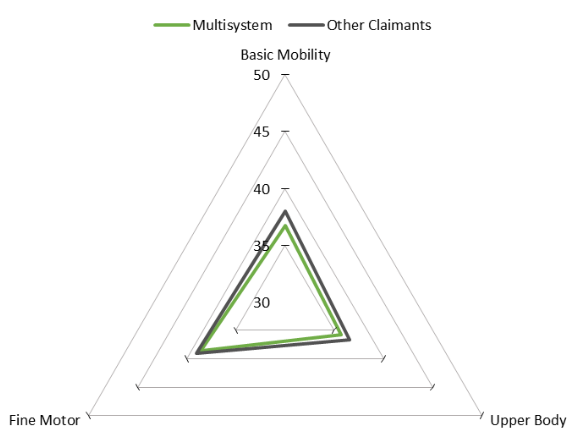
a. Musculoskeletal- Physical Functional



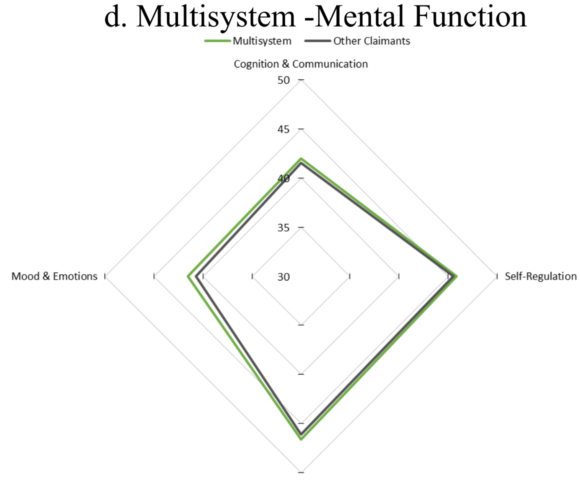
b. Musculoskeletal-Mental Function



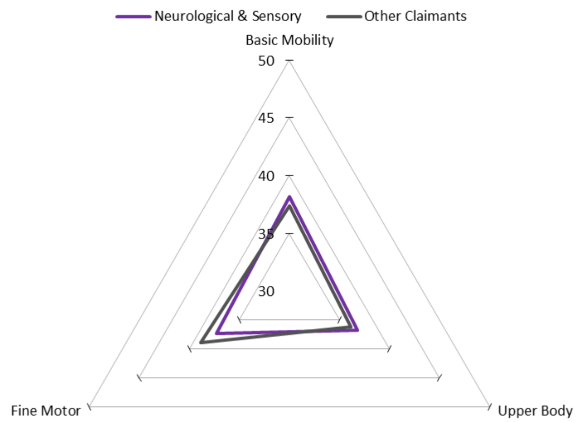
c. Multisystem- Physical Functional



d. Multisystem -Mental Function

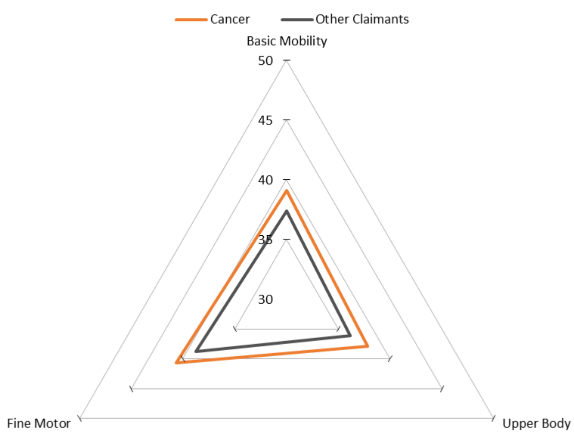


g. Neurological & Sensory- Physical Functional Function

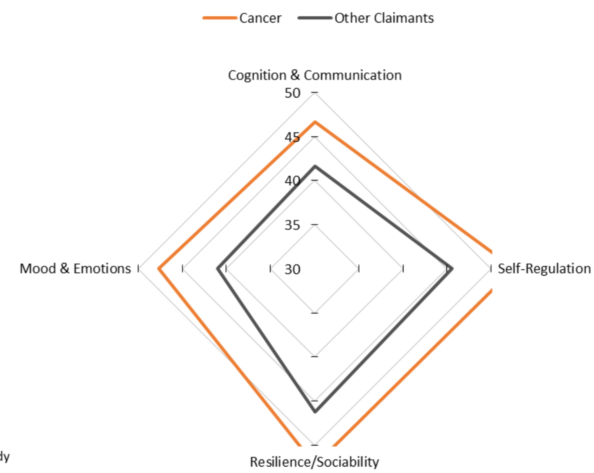


h. Neurological & Sensory -Mental

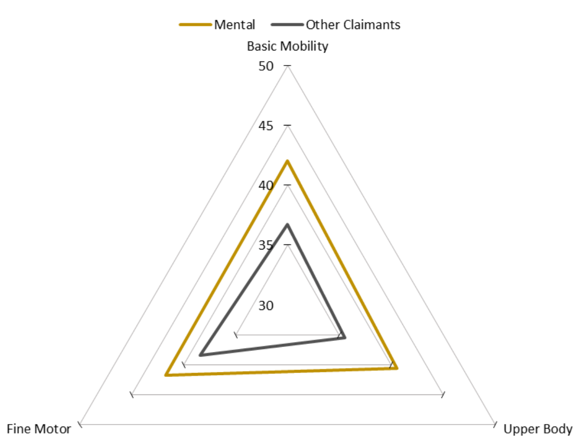
e. Cancer- Physical Functional



f. Cancer -Mental Function



i. Mental- Physical Functional



j. Mental-Mental Function

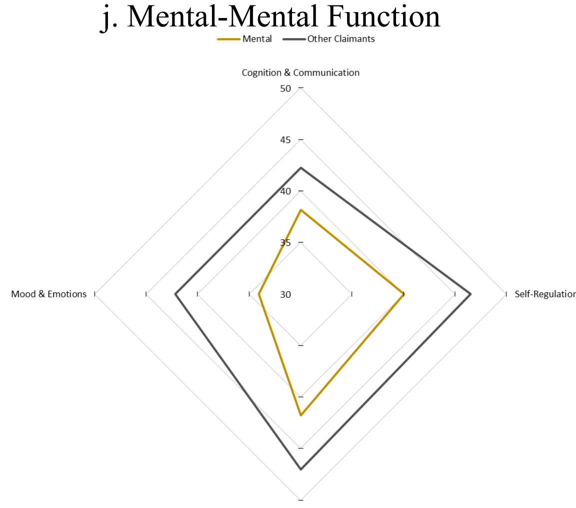


Table 3.7 Differences in WD-FAB Scores by Latent Class

Scale	**mean**	**sd**	**mean**	**sd**	**p**	**Effect Size (*d*)**
	Musculoskeletal (Y)		**Rest**			
Basic Mobility	36.32	5.01	37.93	5.82	<.0001	0.288
Upper Body	34.94	5.15	36.76	5.91	<.0001	0.321
Fine Motor	38.17	6.18	39.11	6.65	<.0001	0.144
Cognition & Communication	42.50	7.79	41.35	7.64	<.0001	0.148
Self-Regulation	46.81	11.94	45.10	11.23	<.0001	0.149
Resilience/Sociability	46.94	10.45	46.04	10.92	0.0164	0.083
Mood & Emotions	42.13	13.16	40.66	13.67	0.0041	0.108
	Cancer (Y)		**Rest**			
Basic Mobility	39.09	5.39	37.36	5.61	0.0071	0.308
Upper Body	37.85	4.99	36.14	5.75	0.0167	0.297
Fine Motor	40.64	6.95	38.77	6.50	0.0211	0.287
Cognition & Communication	46.67	7.72	41.63	7.67	<.0001	0.656
Self-Regulation	52.61	13.05	45.52	11.42	<.0001	0.618
Resilience/Sociability	52.28	10.75	46.21	10.74	<.0001	0.564
Mood & Emotions	47.65	12.42	41.00	13.51	<.0001	0.493
	Multisystem (Y)		**Rest**			
Basic Mobility	36.71	5.14	37.98	5.93	<.0001	0.227
Upper Body	35.72	5.22	36.59	6.15	<.0001	0.152
Fine Motor	38.55	6.37	39.05	6.63	0.0314	0.076
Cognition & Communication	41.97	7.73	41.53	7.68	0.0819	
Self-Regulation	45.85	10.71	45.51	12.14	0.3568	
Resilience/Sociability	46.63	10.30	46.09	11.16	0.1245	
Mood & Emotions	41.55	13.24	40.75	13.76	0.0949	
	Neurological & Sensory (Y)		**Rest**			
Basic Mobility	38.16	6.40	37.35	5.56	0.0784	
Upper Body	36.78	6.52	36.14	5.69	0.2222	
Fine Motor	37.31	7.43	38.89	6.45	0.0087	0.242
Cognition & Communication	41.67	6.97	41.74	7.75	0.9000	
Self-Regulation	47.76	11.13	45.55	11.51	0.0078	0.192
Resilience/Sociability	48.93	10.37	46.20	10.77	0.0005	0.393
Mood & Emotions	45.80	11.84	40.88	13.56	<.0001	0.364
	Mental (Y)		**Rest**			
Basic Mobility	42.03	6.02	36.71	5.21	<.0001	0.998
Upper Body	40.56	6.65	35.54	5.30	<.0001	0.912
Fine Motor	41.74	6.58	38.39	6.40	<.0001	0.522
Cognition & Communication	38.15	6.42	42.26	7.74	<.0001	0.541
Self-Regulation	40.07	11.06	46.49	11.33	<.0001	0.568
Resilience/Sociability	41.74	11.96	47.02	10.41	<.0001	0.496
Mood & Emotions	34.06	13.79	42.16	13.17	<.0001	0.610

Discussion

This study found that Musculoskeletal, Mental, and Cardiovascular conditions are the most common allegation listing/body system group among claimants, which most claimants had allegations that fit in more than one body system, that these 14 systems cluster into five diagnostic classes, and within these clusters there are three unique physical functional profiles, and 4 unique mental functional profiles. Taken together the findings in this paper support the assumption that there is some alignment between the conditions claimants' report, body systems included in the SSA's Listings, and physical and mental functioning. The identification of functional profiles for different condition groups that align with the SSA 14 Listing groups suggests that the functional information the WD-FAB provides could be meaningfully incorporated into the Listing Process in of disability determination, which may facilitate reviews and decisions.

Allegations, Listings, and Latent Class Analyses

The most common listing/body system group category that claimants reported were musculoskeletal, mental, and cardiovascular conditions, which aligns with these listing/body system groups comprising the largest percentage of benefits awarded.[5] Although there were separate Musculoskeletal, and Mental classes, these listing/body system groups were fairly represented in multiple classes. Every claimant in the Musculoskeletal class, more than half of the individuals in the Multisystem class and almost a third of the individuals in the Neurological & Sensory class have a musculoskeletal condition. Similarly, every claimant in the Mental class, just under half of the claimants in the Musculoskeletal class, and the Multisystem class, and around one third of the claimants in the Neurological & Sensory class have a mental health condition. This suggests that mental health conditions are very common among disability claimants, although there is a unique mental class that seems to have mental health issues as their only type health problem.

Also noteworthy was that most claimants could potentially be classified in more than one body system. This brings up questions of how multimorbidity may be related to functioning in a way that the listing process (which requires claimants to consider each condition in isolation) does not currently capture. Furthering this line of thinking was the identification of a Multisystem latent class. It seemed that these individuals—which comprised 46.53% of the claimant sample—do not have a defining or primary condition that leads to limitations. Rather, there is a relatively high prevalence of musculoskeletal, respiratory, cardiovascular, endocrine (such as diabetes), and mental conditions in various combinations. Asking these claimants to attribute functional limitations they are experiencing to one specific condition is challenging.[24–26] With such a high proportion of claimants presenting with this pattern of comorbidities, we might postulate this as one of the reasons that fewer awards are now given at the listing stage.

Functional Profiles

This analysis identified three unique physical function profiles, one shared by the Musculoskeletal, Multisystem, and Neurological & Sensory classes, one for the Cancer class, and one for the Mental class. and four unique mental function profiles, one shared by the Musculoskeletal and Multisystem classes, and one for each of the Cancer, Neurological, and Mental classes.

The average physical function WD-FAB scores for claimants in the Musculoskeletal class were significantly lower than other claimants, even though many other claimants in other classes also have a musculoskeletal condition. On the other hand, despite the common occurrence of a mental condition in both the Musculoskeletal Class and Multisystem Class, mental health function scores for this group were *similar* to other claimants, though still poorer than the general working age US sample. This could point to a way of teasing out people with a mental health "primary" allegation compared to others.

Another interesting finding was in the physical function scores for the Mental class claimants. No claimants in this class have a musculoskeletal condition, yet the physical function for this group was low (though not lower than other claimants). In fact, the claimants in this class had a very low prevalence of other conditions/reported comorbidities. It is clear that, in this sample of claimants, individuals with a mental health condition and few other non-mental comorbidities are also physically limited. The literature has described the relationship between physical activity and mental health,[27] and the dynamics between depression, anxiety, and physical function.[28,29] The physical functional consequences of mental health conditions are not typically captured in the symptoms assessed at the listing stage beyond conditions such as eating disorders, substance addiction disorders, and side effects from medications.

There are several policy implications of this research. First, there is a concern in the field about whether self-reported functioning is reliable in a claimant population. The fears are that claimants may not be insightful enough to answer truthfully on some scales (particularly mental function scale), and that claimants will respond in such a way as to make themselves universally "look poor," functionally speaking, in order to have a greater chance of obtaining benefits—a potential form of gaming. Finding interpretable differences and variation in scores indicates that the WD-FAB is obtaining meaningful responses. For example, there are many individuals who obtained scores in the mental function scales similar to the general working age adult US sample. If claimants were systematically responding that their functioning was poor regardless of their actual functioning, such variability would not have been observed. Although this was not an aim of this these findings are encouraging for the systematic collection of objective self-reports of functioning in the future.

These findings also suggest that there is meaning to the listing/body system groups—both in an empirical basis for the groupings and that some of the groupings are functionally unique—specifically for Musculoskeletal conditions and Mental conditions. This supports continuing use of the listings and to further "evolve" them by incorporating functional criteria and information into the process, particularly for the frequent conditions that formed distinctive classes in this analysis. The functional profiles of the Musculoskeletal class and Mental class are key examples of this. These two condition groups comprise half of eventual beneficiaries, and both are often difficult to classify with traditional biomedical criteria. In addition, symptoms can fluctuate over time. For the listings/body system groups that did not emerge as a distinctive class with a unique functional profile, these findings do *not* recommend removing them as a listing. For example, it may be that for chronic kidney disease there is a clear point, identifiable by lab test and other medical criteria, beyond which a person is not going to be able to engage in substantially gainful activity. Regardless of scores on a measure such as the WD-FAB, the listing process could still be effective for adjudicating the application for these types of claimants.

Limitations

This study has several limitations. First, the allegation classification into the listings/body system groups was not independently validated. Also, all allegations were grouped into body system regardless of whether the reported condition would likely actually meet a listing criteria. This was done because there was no severity or other medical data to facilitate the classification. Severity within diagnoses is a key aspect of how the Listing of Impairments are used at Stage 3 programmatically, and therefore this study is not a direct replication of how the SSA classifies claimants and uses the listings. Another limitation is that because all of the claimant's allegations are listed together, there is no way of identifying what a claimant considers to be his or her *primary* allegation. As previously discussed, claimants may not be able to identify accurately which condition is the largest contributor to why he or she cannot work, but knowing what individuals would consider their primary condition, or the largest reason could help to better understand the relationship between function, perceived attribution, and diagnoses.

This study suggests that there is some relationship between certain diagnostic groups and their self-reported function; what has yet to be established is a link to actual work or Substantially Gainful Activity. Future studies could explore the relationship between functional scores and the different functional demands at different kinds of work. Once the ability to functionally describe work is established, claimants' scores on the WD-FAB can be used at stages 4 and 5 to consider if their functional profile is commensurate with the demands of past work, or any other potentially available jobs. If functional scores and profiles were to be further studied to the point of establishing a standard profile for certain conditions/classes, these profiles can then be used to facilitate claimant classification. For example, if a person has several allegations and potential listings to classify, the SSA could start with the listing category whose group functional profile most closely matches the claimant's individual functional profile. Incorporating these measures into the disability determination process could hopefully expedite the process.

Beyond including the WD-FAB into the disability determination for purposes of improving the process, the measure can also be used to study the process itself. Again, if cutoff scores for the WD-FAB were to be established, they could then serve as a standard against which to assess the current Listing process as it is. If claimants below a certain function score level are determined "truly work disabled," then scores for people who are granted benefits at stage 3 can be compared to those who get passed onto stages 4 and 5. This could establish the current listing process's sensitivity (granting benefits to those who are truly work disabled), and specificity (denying benefits at stage 3 to those who are not).

The SSA has been using the Listing of Impairment for decades to screen, classify, and facilitate disability determinations. Recent evolutions in the conceptualization of disability and disability measurement, and long term trends in the types of conditions claimants are reporting has brought many, including the SSA themselves, to consider the validity of the listings and their usefulness as a tool. This paper suggests that there are connections that can be made between a traditional diagnosis based model (the listings) and the modern function based model of disability (measured by the WD-FAB), allowing for the opportunity to retain a long-standing process within a large institution, but still incorporate contemporary theory and measurement. Use of a functional assessment measure and functional profiles in the disability determination process may improve the process for both the Administration and claimants.

Appendix 3.1 Differences in FAB Scores by Listing/Body System Group

	Musculoskeletal(Y)		Musculoskeletal(N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	35.93	4.86	39.68	5.94	<.0001	0.704
Upper Body	34.74	4.97	38.44	6.13	<.0001	0.677
Fine Motor	37.86	6.18	40.31	6.74	<.0001	0.382
Cognition & Communication	42.04	7.76	41.25	7.61	0.0022	0.170
Self-Regulation	46.29	11.35	44.70	11.66	<.0001	0.138
Resilience/Sociability	46.69	10.37	45.79	11.34	0.0142	0.083
Mood & Emotions	41.59	13.20	40.41	13.99	0.0170	0.087
	Sense/Speech (Y)		Sense/Speech (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	37.54	5.46	37.37	5.63	0.5833	
Upper Body	36.65	5.47	36.12	5.77	0.1074	
Fine Motor	38.08	6.06	38.89	6.56	0.0291	0.125
Cognition & Communication	41.67	7.38	41.74	7.75	0.8570	
Self-Regulation	46.65	10.69	45.55	11.59	0.0527	
Resilience/Sociability	47.23	10.24	46.23	10.83	0.0785	
Mood & Emotions	43.71	12.87	40.83	13.57	0.0002	0.213
	Respiratory (Y)		Respiratory (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	36.25	4.71	37.64	5.76	<.0001	0.247
Upper Body	35.29	4.84	36.36	5.90	<.0001	0.186
Fine Motor	38.61	6.06	38.85	6.61	0.4117	
Cognition & Communication	41.22	7.18	41.84	7.81	0.0559	
Self-Regulation	44.53	10.36	45.91	11.71	0.0024	0.119
Resilience/Sociability	46.07	10.29	46.40	10.87	0.4747	
Mood & Emotions	40.93	13.60	41.18	13.51	0.6990	
	Cardiovascular (Y)		Cardiovascular (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	36.38	4.97	37.86	5.82	<.0001	0.264
Upper Body	35.54	5.08	36.48	6.01	<.0001	0.164
Fine Motor	38.33	6.43	39.04	6.54	0.0038	0.109
Cognition & Communication	42.06	7.70	41.58	7.71	0.0760	
Self-Regulation	46.06	10.30	45.49	12.01	0.1337	
Resilience/Sociability	46.76	10.30	46.15	10.98	0.0970	
Mood & Emotions	42.14	12.83	40.64	13.82	0.0026	0.110
	Digestive (Y)		Digestive (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	36.39	4.72	37.53	5.71	<.0001	0.203
Upper Body	35.33	5.14	36.29	5.81	0.0006	0.168
Fine Motor	37.96	6.23	38.93	6.54	0.0054	0.149
Cognition & Communication	40.73	7.57	41.87	7.72	0.0027	0.148
Self-Regulation	45.42	10.67	45.70	11.61	0.6007	
Resilience	45.99	9.85	46.39	10.89	0.4260	
Mood & Emotions	40.68	12.51	41.20	13.66	0.4474	

	Genitourinary (Y)		Genitourinary (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	38.03	5.47	37.36	5.62	0.1475	
Upper Body	37.28	4.99	36.13	5.76	0.0119	0.200
Fine Motor	39.61	6.77	38.77	6.50	0.1536	
Cognition & Communication	43.66	8.06	41.65	7.68	0.0014	0.261
Self-Regulation	47.07	10.35	45.61	11.54	0.1212	
Resilience/Sociability	49.09	10.76	46.22	10.76	0.0012	0.266
Mood & Emotions	44.31	14.44	41.00	13.47	0.0065	0.244
	Hematological (Y)		Hematological (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	36.26	5.19	37.42	5.62	0.0579	
Upper Body	35.07	4.98	36.20	5.75	0.0919	
Fine Motor	38.21	7.02	38.82	6.50	0.4240	
Cognition & Communication	41.06	7.43	41.75	7.71	0.4106	
Self-Regulation	45.76	10.31	45.66	11.53	0.9362	
Resilience/Sociability	46.46	10.60	46.34	10.77	0.9190	
Mood & Emotions	39.51	14.46	41.17	13.50	0.2920	
	Skin (Y)		Skin (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	36.24	4.75	37.41	5.62	0.1223	
Upper Body	35.50	5.84	36.18	5.74	0.4260	
Fine Motor	37.14	6.68	38.83	6.51	0.0804	
Cognition & Communication	42.08	7.46	41.73	7.71	0.7341	
Self-Regulation	46.29	8.40	45.66	11.54	0.5792	
Resilience/Sociability	45.05	10.82	46.36	10.77	0.3668	
Mood & Emotions	43.00	13.50	41.11	13.52	0.3462	
	Endocrine (Y)		Endocrine (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	36.45	5.17	37.61	5.69	<.0001	0.206
Upper Body	35.51	5.16	36.33	5.86	0.0005	0.143
Fine Motor	38.08	6.18	38.98	6.58	0.0019	0.138
Cognition & Communication	41.84	7.52	41.71	7.75	0.6723	
Self-Regulation	45.93	10.41	45.61	11.73	0.4645	
Resilience/Sociability	46.56	9.66	46.29	11.01	0.5136	
Mood & Emotions	41.47	13.06	41.05	13.63	0.4864	
	Congenital (Y)		Congenital (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	38.14	6.39	37.39	5.61	0.7911	
Upper Body	37.12	3.35	36.17	5.74	0.8151	
Fine Motor	41.24	6.73	38.81	6.51	0.5978	
Cognition & Communication	39.87	5.77	41.74	7.71	0.6291	
Self-Regulation	44.67	8.97	45.67	11.50	0.8628	
Resilience/Sociability	48.66	11.72	46.34	10.77	0.6669	
Mood & Emotions	29.20	28.85	41.14	13.51	0.2120	

	Neurological (Y)		Neurological (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	37.08	5.78	37.46	5.58	0.1224	
Upper Body	35.85	5.85	36.24	5.71	0.1545	
Fine Motor	37.42	6.83	39.09	6.41	<.0001	0.257
Cognition & Communication	40.55	7.12	41.97	7.80	<.0001	0.184
Self-Regulation	46.02	9.49	45.59	11.86	0.3210	
Resilience/Sociability	46.21	10.41	46.37	10.84	0.7331	
Mood & Emotions	41.57	13.27	41.04	13.57	0.4139	
	Mental (Y)		Mental (N)			
Scale	mean	sd	mean	sd	p	*d*
Basic Mobility	37.91	5.90	36.89	5.26	<.0001	0.181
Upper Body	36.63	6.10	35.72	5.33	<.0001	0.159
Fine Motor	38.85	6.51	38.77	6.52	0.7418	
Cognition & Communication	38.89	6.77	44.50	7.56	<.0001	0.781
Self-Regulation	42.31	10.88	48.93	11.14	<.0001	0.600
Resilience/Sociability	43.22	10.85	49.38	9.78	<.0001	0.596
Mood & Emotions	35.70	13.21	46.45	11.57	<.0001	0.865
	Cancer (Y)		Cancer (N)			
Scale	mean	sd	mean	sd	p	ES
Basic Mobility	37.31	5.30	37.40	5.63	0.8390	
Upper Body	36.24	5.26	36.17	5.76	0.8838	
Fine Motor	39.11	6.62	38.79	6.51	0.5417	
Cognition & Communication	43.24	7.89	41.65	7.69	0.0060	0.205
Self-Regulation	48.51	11.21	45.52	11.49	0.0005	0.261
Resilience/Sociability	48.42	10.19	46.23	10.79	0.0065	0.203
Mood & Emotions	43.38	13.02	41.01	13.54	0.0304	0.175
	Immune (Y)		Immune (N)			
Scale	mean	sd	mean	sd	p	ES
Basic Mobility	37.35	5.72	37.39	5.61	0.9218	
Upper Body	35.79	5.77	36.19	5.74	0.4417	
Fine Motor	37.63	6.43	38.86	6.51	0.0372	0.188
Cognition & Communication	41.06	8.28	41.76	7.68	0.2847	
Self-Regulation	47.48	13.64	45.59	11.40	0.1020	
Resilience/Sociability	46.43	10.93	46.34	10.76	0.9160	
Mood & Emotions	40.87	12.70	41.14	13.56	0.8189	

Appendix 3.2 Average WD-FAB Scores by Class

	Musculoskeletal		**Cancer**		**Multisystem**		**Neurological & Sensory**		**Mental**		**F**	**p**
	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd		
Basic Mobility	36.32	5.01	39.09	5.39	36.71	5.14	38.16	6.40	42.03	6.02	116.12	<.0001
Upper Body	34.94	5.15	37.85	4.99	35.72	5.22	36.78	6.52	40.56	6.65	82.23	<.0001
Fine Motor	38.17	6.18	40.64	6.95	38.55	6.37	37.31	7.43	41.74	6.58	27.73	<.0001
Cognition & Communication	42.50	7.79	46.67	7.72	41.97	7.73	41.67	6.97	38.15	6.42	39.32	<.0001
Self-Regulation	46.81	11.94	52.61	13.05	45.85	10.71	47.76	11.13	40.07	11.06	42.76	<.0001
Resilience/Sociability	46.94	10.45	52.28	10.75	46.63	10.30	48.93	10.37	41.74	11.96	33.29	<.0001
Mood & Emotions	42.32	13.16	47.65	12.42	41.55	13.24	45.80	11.84	34.06	13.79	39.81	<.0001

References

1. Social Security Administration. Annual statistical report on the social security disability insurance program, 2014.

2. Brandt DE, Houtenville AJ, Huynh MT, Chan L, Rasch EK. Connecting contemporary paradigms to the social security administration's disability evaluation process. *Journal of Disability Policy Studies*. 2011:1044207310396509.

3. Hu J, Lahiri K, Vaughan DR, Wixon B. A structural model of social security's disability determination process. *Rev Econ Stat*. 2001;83(2):348–361.

4. Barnes DK, McGeary M, Stobo JD. *Improving the social security disability decision process.* National Academies Press; 2007.

5. Social Security Administration. Trends in the social security and supplemental security income disability programs. 2007.

6. Ganellen RJ. Assessing normal and abnormal personality functioning: Strengths and weaknesses of self-report, observer, and performance-based methods. *J Pers Assess*. 2007;89(1):30–40.

7. Van der Werff E, Verboom C, Penninx B, Nolen W, Ormel J. Explaining heterogeneity in disability associated with current major depressive disorder: Effects of illness characteristics and comorbid mental disorders. *J Affect Disord*. 2010;127(1):203–210.

8. Lord FM. *Applications of item response theory to practical testing problems.* Routledge; 1980.

9. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care*. 2007;45(5 Suppl 1): S22–31. doi: 10.1097/01.mlr.0000250483.85507.04 [doi].

10. Norris JM. Computer-adaptive testing: A primer. *Language Learning & Technology*. 2001;5(2):23–27.

11. Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika*. 1989;54(3):427–450.

12. Wang S, Wang T. Precision of warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*. 2001;25(4):317–331.

13. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*. 1969.

14. Hambleton R, Pitoniak M. Testing and measurement: Advances in item response theory and selected testing practices. 2002.

15. Meterko M, Marfeo EE, McDonough CM, et al. The work disability functional assessment battery (WD-FAB): Feasibility and psychometric properties. *Arch Phys Med Rehabil*. 2014.

16. Ni P, McDonough CM, Jette AM, et al. Development of a computer-adaptive physical function instrument for social security administration disability determination. *Arch Phys Med Rehabil*. 2013;94(9):1661–1669.

17. Marfeo EE, Ni P, Haley SM, et al. Development of an instrument to measure behavioral health function for work disability: Item pool construction and factor analysis. *Arch Phys Med Rehabil*. 2013;94(9):1670–1678.

18. Rivers D. Sample matching: Representative sampling from internet panels. *Polimetrix White Paper Series*. 2006.

19. McCutcheon AL. *Latent class analysis.* Sage; 1987.

20. Hagenaars JA, McCutcheon AL. *Applied latent class analysis.* Cambridge University Press; 2002.

21. Vermunt JK, Magidson J. Latent class analysis. *The sage encyclopedia of social sciences research methods*. 2004:549–553.

22. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*. 2007;14(4):535–569.

23. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law Hum Behav*. 2005;29(5):615.

24. McDermott S, Turk MA. The myth and reality of disability prevalence: Measuring disability for research and service. *Disability and Health Journal*. 2011;4(1):1–5.

25. Lenderink AF, Zoer I, van der Molen, Henk F, Spreeuwers D, Frings-Dresen MH, van Dijk FJ. Review on the validity of self-report to assess work-related diseases. *Int Arch Occup Environ Health*. 2012;85(3):229–251.

26. Marx RG, Hogg-Johnson S, Hudak P, et al. A comparison of patients' responses about their disability with and without attribution to their affected area. *J Clin Epidemiol*. 2001;54(6):580–586.

27. Paluska SA, Schwenk TL. Physical activity and mental health. *Sports medicine*. 2000;29(3):167–180.

28. Stegenga BT, Nazareth I, Torres-Gonzalez F, et al. Depression, anxiety and physical function: Exploring the strength of causality. *J Epidemiol Community Health*. 2012;66(7):e25. doi: 10.1136/jech.2010.128371 [doi].

29. Lever-van Milligen BA, Lamers F, Smit JH, Penninx BW. Six-year trajectory of objective physical function in persons with depressive and anxiety disorders. *Depress Anxiety*. 2016.

CONCLUSION

These studies contribute to the work disability and measurement literature in a number of ways. Study one is among the first to empirically examine the structure of Activities as the ICF proposes them. The comparison found good alignment between the physical function scales of the WD-FAB and the Mobility subdomain of the ICF, and several of the Activities subdomains converging into an empirical domain named Cognition & Communication in the WD-FAB. Further, it found that the distinction between Body Functions—specifically Mental Functions—and certain Activities such as Interpersonal Interactions, is not established. The reorganized framework presented should be considered in future work that uses the ICF as a foundation for systematically measuring and collecting functional information related to disability. Based upon these findings and the findings of others, this study suggest that there is a need to revise and update the structure of the ICF, similar to how the World Health Organization has updated the ICD throughout the years. Future work should focus on this reorganization, with particular emphasis on incorporating memory and emotional related function at the activity level.

Study two found that SSA disability claimants were different from general working age US population sociodemographically (especially based upon race and ethnicity), and that they are functioning on average half a standard deviation to a full standard deviation poorer than the general sample. When looking at only the claimants, differences in functional scores observed were primarily based on age and gender, but not race. Current policy and media discussions regarding the Social Security Disability programs tend toward a concern about individuals applying to the program who are not "truly disabled." These concerns are also often associated with specific sociodemographic groups—particularly low income individuals, and racial and ethnic minorities. The findings in the second study do not support that belief, but rather suggest that those who apply for Social Security disability benefits are indeed functionally limited. Further work is needed to understand the factors that lead a person to apply for disability benefits—functional and others—and how those factors vary among different groups of people.

This study is also one of the few in the patient reported outcomes field to look at Differential Item Functioning by race and ethnicity, in addition to age and gender. Seventeen items displayed DIF, primarily gender-based DIF. Although the presence of DIF items was not extensive, it occurred for many items that relate to domestic tasks. When asking self-reported questions about physical functioning, grounding items in an activity or a daily context and avoiding clinical jargon is important for item interpretation. Therefore, this analysis suggests that DIF is an important concern to address when measuring function at the activity level. The ability to present WD-FAB scores *after* controlling for DIF allows for measuring and detecting true differences in function, with minimal measurement error. This is particularly advantageous to researchers or administrators looking to measure functioning in a diverse group of individuals, such as SSA disability applicants.

The third and final study presented here identified patterns among claimant-reported health problems and diagnoses, and explored the link between diagnosis/comorbidity groups and functional scores on the WD-FAB. Latent Class Analyses found 5 classes of listings/body system groups, named Musculoskeletal, Cancer, Multisystem, Neurological & Sensory, and Mental. Comparisons in functional scores between the groups found unique functional phenotypes for many of the groups. The Musculoskeletal class and Multisystem class displayed very similar functional profiles. The Neurological & Sensory class shows similar physical function profiles to the Musculoskeletal and Multisystem group, but a different mental functional profile—showing considerably better Mood & Emotions & Resilience/Sociability scores. The Cancer class was functioning physically better than the previous three classes and showed mental function scores close to or at the average of the general sample—indicating no substantial limitations in those domains. Finally, the Mental class had the highest functional scores on all scales, and by far the lowest on the mental function scales, presenting the most unique profile compared to the others.

Differences between the profiles supports incorporating functional information from the WD-FAB in into the disability determination process—particularly for the Listing groups that were identified as distinct classes in this analysis. Future work should focus on establishing severity criteria within each diagnostic profile, to facilitate the Listings' primary objective of identifying those most likely to obtain benefits. For example, the two most common condition groups that comprise half of eventual beneficiaries are reflected in the Musculoskeletal and Mental class. Both of those groups of conditions are often difficult to classify with traditional biomedical criteria, and involve symptoms that can fluctuate over time. Knowing how individuals in this group look *functionally* can help understand whether a claimant fits into a listing group or not. If severity criteria were established and the WD-FAB was incorporated into the process, WD-FAB scores may provide the SSA with information that may facilitate reviews and decisions.

The goal of this thesis was to advance the overall field of measuring work disability. Conceptually, the findings revealed that the WHO's ICF is a comprehensive basis for disability measurement, although its structure should be reorganized to reflect the empirical organization of the concepts it presents. Methodologically, DIF does appear to be having some impact on measuring function at the activity level. Future self-report measures should be sure to include gender, age, and racial diversity in cognitive testing samples, and conduct DIF analyses if possible. The Social Security Disability programs are a key area where work disability measurement is of great concern. The WD-FAB appears to measure function in an interpretable manner within several of the SSA's most common diagnostic groups. Based upon this work, the WD-FAB is a potentially valuable tool to supplement the information currently considered in the disability determination process.

There is no single gold standard or litmus test for measuring work disability. Approaching disability from a functional rather than a diagnostic or deficit perspective has been an important focus for disability research. Incorporating individual's self-reported experiences and function is also becoming an increasingly more common field of research. These three studies demonstrate how a conceptually grounded self-reported measure of functioning at the activity level can be used to understand the condition of individuals whose health limits their ability to work. A more complete understanding of how individuals are functioning in the context of their work and other life areas is vital for accurately measuring work disability and providing the health care, financial, and social supports that individuals need.

CUMULATIVE BIBLIOGRAPHY

Abel, J. R., Gabe, T. M., & Stolarick, K. (2012). Workforce skills across the urban-rural hierarchy. Federal Reserve Bank of New York Staff Report No. 552.

Anner, J., Schwegler, U., Kunz, R., Trezzini, B., & de Boer, W. (2012). Evaluation of work disability and the international classification of functioning, disability and health: What to expect and what not. *BMC Public Health, 12*, 470-2458-12-470. doi:10.1186/1471-2458-12-470 [doi]

Autor, D. (2010). The polarization of job opportunities in the US labor market: Implications for employment and earnings. *Center for American Progress and the Hamilton Project,*

Barnes, D. K., McGeary, M., & Stobo, J. D. (2007). *Improving the social security disability decision process* National Academies Press.

Brandt, D. E., Houtenville, A. J., Huynh, M. T., Chan, L., & Rasch, E. K. (2011). Connecting contemporary paradigms to the social security administration's disability evaluation process. *Journal of Disability Policy Studies, 22*(2), doi:1044207310396509.

Cerniauskaite, M., Quintas, R., Boldt, C., Raggi, A., Cieza, A., Bickenbach, J. E., & Leonardi, M. (2011). Systematic literature review on ICF from 2001 to 2009: Its use, implementation and operationalisation. *Disability and Rehabilitation, 33*(4), 281–309. doi:10.3109/09638288.2010.529235 [doi]

Cieza, A., & Stucki, G. (2008). The international classification of functioning disability and health: Its development process and content validity. *European Journal of Physical and Rehabilitation Medicine, 44*(3), 303–313.

Cieza, A., Brockow, T., Ewert, T., Amman, E., Kollerits, B., Chatterji, S., . . . Stucki, G. (2002). Linking health-status measurements to the international classification of functioning, disability and health. *Journal of Rehabilitation Medicine, 34*(5), 205–210.

Cieza, A., Ewert, T., Ustun, T. B., Chatterji, S., Kostanjsek, N., & Stucki, G. (2004). Development of ICF core sets for patients with chronic conditions. *Journal of Rehabilitation Medicine,* 2004 Jul:(44 Suppl.), 9–11.

Cieza, A., & Stucki, G. (2005). Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Quality of Life Research, 14*(5), 1225–1237.

Cieza, A., Fayed, N., Bickenbach, J., & Prodinger, B. (2016). Refinements of the ICF linking rules to strengthen their potential for establishing comparability of health information. *Disability and Rehabilitation,* 2016 Mar. 17, 1–10. doi:10.3109/09638288.2016.1145258 [doi] (EPub ahead of print)

Coe, N. B., Haverstick, K., Munnell, A. H., & Webb, A. (2011). What explains state variation in SSDI application rates? *Center for Retirement Research at Boston College Working Paper,* (2011-23).

Coe, N. B., Haverstick, K., Munnell, A. H., & Webb, A. (2012). Why do state disability application rates vary over time? *Center for Retirement Research at Boston College Working Paper,* (12-2).

D'Amato, A., & Zijlstra, F. (2010). Toward a climate for work resumption: The nonmedical determinants of return to work. *Journal of Occupational and Environmental Medicine, 52*(1), 67–80. doi:10.1097/JOM.0b013e3181c75291 [doi]

DeBose, C. E. (1992). Codeswitching: Black english and standard english in the African-American linguistic repertoire. *Journal of Multilingual & Multicultural Development, 13*(1–2), 157–167.

Dijkers, M. P., Whiteneck, G., & El-Jaroudi, R. (2000). Measures of social outcomes in disability research. *Archives of Physical Medicine and Rehabilitation, 81*, S63–S80.

Duggan, M., & Imberman, S. A. (2009). Why are the disability rolls skyrocketing? the contribution of population characteristics, economic conditions, and program generosity. *Health at older ages: The causes and consequences of declining disability among the elderly* (pp. 337–379) University of Chicago Press.

Edelen, M. O., Stucky, B. D., & Chandra, A. (2013). Quantifying ‘problematic’ DIF within an IRT framework: Application to a cancer stigma index. *Quality of Life Research, 24*(1), 95–103.

Escorpizo, R., & Stucki, G. (2013). Disability evaluation, social security, and the international classification of functioning, disability and health: The time is now. *Journal of Occupational and Environmental Medicine, 55*(6), 644–651. doi:10.1097/JOM.0b013e318297ae47 [doi]

Fayed, N., Cieza, A., & Edmond Bickenbach, J. (2011). Linking health and health-related information to the ICF: A systematic review of the literature from 2001 to 2008. *Disability and Rehabilitation, 33*(21–22), 1941–1951.

Field, M. J., & Jette, A. (2007). *The future of disability in America.* National Academies Press.

Fleishman, J. A., Spector, W. D., & Altman, B. M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 57*(5), S275–84.

Former social security commissioners and others respond to our disability story. (2013). Retrieved from http://www.npr.org/sections/money/2013/04/18/177745599/former-social-security-commissioners-and-others-respond-to-our-disability-story

Fremstad, S. (2009). Half in ten. *Why Taking Disability into Account is Essential to Reducing Income Poverty and Expanding Economic Inclusion.* Washington, DC: Center for Economic and Policy Research,

Ganellen, R. J. (2007). Assessing normal and abnormal personality functioning: Strengths and weaknesses of self-report, observer, and performance-based methods. *Journal of Personality Assessment, 89*(1), 30–40.

Gao, Y., & Zhu, W. (2011). Differential item functioning analysis of the 2003–04 NHANES physical activity questionnaire. *Research Quarterly for Exercise and Sport, 82*(3), 381–390.

Grill, E., Ewert, T., Chatterji, S., Kostanjsek, N., & Stucki, G. (2005). ICF core sets development for the acute hospital and early post-acute rehabilitation facilities. *Disability and Rehabilitation, 27*(7–8), 361–366.

Hafner-Eaton, C. (1993). Physician utilization disparities between the uninsured and insured: Comparisons of the chronically III, acutely III, and well nonelderly populations. *JAMA: The Journal of the American Medical Association, 269*(6), 787–792.

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied latent class analysis.* Cambridge University Press.

Hambleton, R., & Pitoniak, M. (2002). Testing and measurement: Advances in item response theory and selected testing practices. In J. Wixted and H. Pashler (eds.), *Stevens' Handbook of Experimental Psychology* (3rd Ed., Vol. 4, pp. 517–561). New York: Wiley.

Hanmer, J., Lawrence, W. F., Anderson, J. P., Kaplan, R. M., & Fryback, D. G. (2006). Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. *Medical Decision Making, 26*(4), 391–400.

Haveman, R., & Wolfe, B. (2000). The economics of disability and disability policy. *Handbook of Health Economics, 1*, 995–1051.

Hendley, A. A., & Bilimoria, N. F. (1999). Minorities and social security: An analysis of ethnic differences in the current program. *Social Security Bulletin, 62*, 59.

Hopman, W., Harrison, M., Coo, H., Friedberg, E., Buchanan, M., & Van Den Kerkhof, E. (2009). Associations between chronic disease, age and physical and mental health status. *Chronic Diseases in Canada, 29*(3), 108–116.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications* Sage Publications.

Hu, J., Lahiri, K., Vaughan, D. R., & Wixon, B. (2001). A structural model of social security's disability determination process. *Review of Economics and Statistics, 83*(2), 348–361.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Iwata, N., Turner, R. J., & Lloyd, D. A. (2002). Race/ethnicity and depressive symptoms in community-dwelling young adults: A differential item functioning analysis. *Psychiatry Research, 110*(3), 281–289.

Jette, A. M., Haley, S. M., & Kooyoomjian, J. T. (2003). Are the ICF activity and participation dimensions distinct? *Journal of Rehabilitation Medicine, 35*(3), 145–149.

Jette, A. M., Tao, W., Haley, S. M., Jette, A. M., Tao, W., & Haley, S. M. (2007). Blending activity and participation sub-domains of the ICF. *Disability and Rehabilitation, 29*(22), 1742–1750.

Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology, 29*(1), 417–442.

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustun, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry, 20*(4), 359–364. doi:10.1097/YCO.0b013e32816ebc8c [doi]

Kington, R. S., & Smith, J. P. (1997). Socioeconomic status and racial and ethnic differences in functional status associated with chronic diseases. *American Journal of Public Health, 87*(5), 805–810.

Krause, N., Frank, J. W., Dasinger, L. K., Sullivan, T. J., & Sinclair, S. J. (2001). Determinants of duration of disability and return-to-work after work-related injury and illness: Challenges for future research. *American Journal of Industrial Medicine, 40*(4), 464–484.

Langer, M. M. (2008). *A Reexamination of Lord's Wald Test for Differential Item Functioning using Item Response Theory and Modern Error Estimation,* Unpublished doctoral dissertation, University of North Carolina–Chapel Hill

Lee, M., & Mather, M. (2008). *US labor force trends* Population Reference Bureau.

Lenderink, A. F., Zoer, I., van der Molen, Henk F, Spreeuwers, D., Frings-Dresen, M. H., & van Dijk, F. J. (2012). Review on the validity of self-report to assess work-related diseases. *International Archives of Occupational and Environmental Health, 85*(3), 229–251.

Lever-van Milligen, B. A., Lamers, F., Smit, J. H., & Penninx, B. W. (2016). Six-year trajectory of objective physical function in persons with depressive and anxiety disorders. *Depression and Anxiety, 34*(2), 188–197.

Lin, S., Beck, A. N., Finch, B. K., Hummer, R. A., & Master, R. K. (2012). Trends in US older adult disability: Exploring age, period, and cohort effects. *American Journal of Public Health, 102*(11), 2157–2163.

List of official ICF updates. (2016). Retrieved from http://www.who.int/classifications/icfupdates/en/

Loprest, P., & Maag, E. (2007). The relationship between early disability onset and education and employment. *Journal of Vocational Rehabilitation, 26*(1), 49–62.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems* Routledge.

Madden, R., Fortune, N., Cheeseman, D., Mpofu, E., & Bundy, A. (2013). Fundamental questions before recording or measuring functioning and disability. *Disability and Rehabilitation, 35*(13), 1092–1096. doi:10.3109/09638288.2012.720350 [doi]

Madden, R. H., Glozier, N., Fortune, N., Dyson, M., Gilroy, J., Bundy, A., . . . Madden, R. (2015). In search of an integrative measure of functioning. *International Journal of Environmental Research and Public Health, 12*(6), 5815–5832. doi:10.3390/ijerph120605815 [doi]

Marfeo, E. E., Haley, S. M., Jette, A. M., Eisen, S. V., Ni, P., Bogusz, K., . . . Brandt, D. E. (2013). Conceptual foundation for measures of physical function and behavioral health function for social security work disability evaluation. *Archives of Physical Medicine and Rehabilitation, 94*(9), 1645–1652. e2. doi:10.1016/j.apmr.2013.03.015

Marfeo, E. E., Ni, P., Haley, S. M., Jette, A. M., Bogusz, K., Meterko, M., . . . Rasch, E. K. (2013). Development of an instrument to measure behavioral health function for work disability: Item pool construction and factor analysis. *Archives of Physical Medicine and Rehabilitation, 94*(9), 1670–1678.

Marx, R. G., Hogg-Johnson, S., Hudak, P., Beaton, D., Shields, S., Bombardier, C., & Wright, J. G. (2001). A comparison of patients' responses about their disability with and without attribution to their affected area. *Journal of Clinical Epidemiology, 54*(6), 580–586.

Matheson, L. N., Kaskutas, V., McCowan, S., Shaw, H., & Webb, C. (2001). Development of a database of functional assessment measures related to work disability. *Journal of Occupational Rehabilitation, 11*(3), 177–199.

McCutcheon, A. L. (1987). *Latent class analysis* Sage.

McDermott, S., & Turk, M. A. (2011). The myth and reality of disability prevalence: Measuring disability for research and service. *Disability and Health Journal, 4*(1), 1–5.

McFarland, M. J., & Wagner, B. G. (2015). Does a college education reduce depressive symptoms in american young adults? *Social Science & Medicine, 146*, 75–84.

Meterko, M., Marfeo, E. E., McDonough, C. M., Jette, A. M., Ni, P., Bogusz, K., . . . Chan, L. (2015). The work disability functional assessment battery (WD-FAB): Feasibility and psychometric properties. *Archives of Physical Medicine and Rehabilitation, 96*(6), 1028–1035.

Milner, A., LaMontagne, A., Aitken, Z., Bentley, R., & Kavanagh, A. (2014). Employment status and mental health among persons with and without a disability: Evidence from an australian cohort study. *Journal of Epidemiology and Community Health,* jech-2014-204147.

Ni, P., McDonough, C. M., Jette, A. M., Bogusz, K., Marfeo, E. E., Rasch, E. K., . . . Chan, L. (2013). Development of a computer-adaptive physical function instrument for social security administration disability determination. *Archives of Physical Medicine and Rehabilitation, 94*(9), 1661–1669.

Nichols, A., & Simms, M. (2012). Racial and ethnic differences in receipt of unemployment insurance benefits during the great recession. *Unemployment and Recovery Project Brief, 4.*

Norris, J. M. (2001). Computer-adaptive testing: A primer. *Language Learning & Technology, 5*(2), 23–27.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535–569.

Paluska, S. A., & Schwenk, T. L. (2000). Physical activity and mental health. *Sports Medicine, 29*(3), 167–180.

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., . . . PROMIS Cooperative Group. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Medical Care, 45*(5 Suppl. 1), S22–31. doi:10.1097/01.mlr.0000250483.85507.04 [doi]

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615.

Rivers, D. (2006). Sample matching: Representative sampling from internet panels. *Polimetrix White Paper Series,*

Rose, M., Bjorner, J. B., Becker, J., Fries, J., & Ware, J. (2008). Evaluation of a preliminary physical function item bank supported the expected advantages of the patient-reported outcomes measurement information system (PROMIS). *Journal of Clinical Epidemiology, 61*(1), 17–33.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores.* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved from http://www.psychometrika.org/journal/online/MN17.pdf

Social Security Administration. Annual statistical report on the social security disability insurance program, 2014.

Social Security Administration. (2007). Trends in the social security and supplemental security income disability programs.

Stegenga, B. T., Nazareth, I., Torres-Gonzalez, F., Xavier, M., Svab, I., Geerlings, M. I., . . . King, M. (2012). Depression, anxiety and physical function: Exploring the strength of causality. *Journal of Epidemiology and Community Health, 66*(7), e25. doi:10.1136/jech.2010.128371 [doi]

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173–180.

Stucki, G. (2005). ICF linking rules: An update based on lessons learned. *Journal of Rehabilitation Medicine, 37*, 212–218.

Stucki, G., Cieza, A., Ewert, T., Kostanjsek, N., Chatterji, S., & ÜstÜn, T. B. (2002). Application of the international classification of functioning, disability and health (ICF) in clinical practice. *Disability and Rehabilitation, 24*(5), 281–282.

Stucki, G., Kostanjsek, N., Ustun, B., & Cieza, A. (2008). ICF-based classification and measurement of functioning. *European Journal of Physical and Rehabilitation Medicine, 44*(3), 315–328. doi:R33Y2008N03A0315 [pii]

Sullivan, O. (1997). Time waits for no (wo) man: An investigation of the gendered experience of domestic time. *Sociology, 31*(2), 221–239.

Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research, 16*(1), 33–42.

Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Eimicke, J. P., Crane, P. K., Jones, R. N., . . . Cella, D. (2009). Analysis of differential item functioning in the depression item bank from the patient reported outcome measurement information system (PROMIS): An item response theory approach. *Psychology Science Quarterly, 51*(2), 148–180.

Tucker, C. A., Escorpizo, R., Cieza, A., Lai, J. S., Stucki, G., Ustun, T. B., . . . Forrest, C. B. (2014a). Mapping the content of the patient-reported outcomes measurement information system (PROMIS®) using the international classification of functioning, health and disability. *Quality of Life Research, 23*(9), 2431–2438.

Tucker, C. A., Cieza, A., Riley, A. W., Stucki, G., Lai, J. S., Bedirhan Ustun, T., . . . Forrest, C. B. (2014b). Concept analysis of the patient reported outcomes measurement information system (PROMIS((R))) and the international classification of functioning, disability and health (ICF). *Quality of Life Research, 23*(6), 1677–1686. doi:10.1007/s11136-014-0622-y [doi]

Unfit for work: The startling rise of disability in America. (2013). Retrieved from http://apps.npr.org/unfit-for-work/

Van der Werff, E., Verboom, C., Penninx, B., Nolen, W., & Ormel, J. (2010). Explaining heterogeneity in disability associated with current major depressive disorder: Effects of illness characteristics and comorbid mental disorders. *Journal of Affective Disorders, 127*(1), 203–210.

Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. In M.S. Lewis-Beck, A. Bryman, and T.F. Liao (eds.), *The Sage Encyclopedia of Social Sciences Research Methods*, (pp. 549-553). Thousand Oaks, CA: Sage Publications

Von Wachter, T., Song, J., & Manchester, J. (2011). Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program. *The American Economic Review, 101*(7), 3308–3329.

Wang, S., & Wang, T. (2001). Precision of warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement, 25*(4), 317–331.

Ward, B. W. (2013). Prevalence of multiple chronic conditions among US adults: Estimates from the national health interview survey, 2010. *Preventing Chronic Disease, 10,* 120203.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450.

WHO, I., & World Health Organization. (2007). International classification of functioning. *Disability and Health (ICF), Endorsed by All, 191*

Williams, D. R. (1999). Race, socioeconomic status, and health the added effects of racism and discrimination. *Annals of the New York Academy of Sciences, 896*(1), 173–188.

Williams, D. R., & Collins, C. (1995). US socioeconomic and racial differences in health: Patterns and explanations. *Annual Review of Sociology, 21*, 349–386.

Willimas, D., Yu, Y., & Jackson, J. (1997). Racial differences in physical and mental health. *Journal of Health Psychology, 2*, 335–351.

Wilson, K. B. (2000). Predicting vocational rehabilitation acceptance based on race, education, work status, and source of support at application. *Rehabilitation Counseling Bulletin, 43*(2), 97–105.

Woods, C. M., Cai, L., & Wang, M. (2013). The langer-improved wald test for DIF testing with multiple groups evaluation and comparison to two-group IRT. *Educational and Psychological Measurement, 73*(3), 532–547.

Wu, A. Y. (2015). How has the changing nature of work contributed to SSDI growth? Paper presented at the *2015 Fall Conference: The Golden Age of Evidence-Based Policy,*

Yee, S. (2011). Health and health care disparities among people with disabilities. *Washington, DC: Disability Rights Education & Defense Fund,*

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]