MEASUREMENT ISSUES IN HEALTH-RELATED QUALITY OF LIFE ASSESSMENTS IN PHYSICAL ACTIVITY RESEARCH

by

Peter Donald Hart Jr.

A Dissertation Submitted to the
Faculty of the Graduate School at
Middle Tennessee State University
in Partial Fulfillment
of the Requirements for the Degree of
Doctorate of Philosophy
in Health and Human Performance

Murfreesboro, TN
May 2012

UMI Number: 3514919

UMI 3514919

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

APPROVAL PAGE

MEASUREMENT ISSUES IN HEALTH-RELATED QUALITY OF LIFE ASSESSMENTS IN PHYSICAL ACTIVITY RESEARCH

5/4/12
Date of Final Defense

Dr. Minsoo Kang, Committee Chair

Dr. Thomas M. Brinthaupt, Committee Member

Dr. Yun Soo Lee, Committee Member

Norman Weatherby
Dr. Norman L. Weatherby, Committee Member

Dr. Steven G. Estes, Chair, Department of Health and Human Performance

Dr. Michael D. Allen, Dean of Graduate School

ABSTRACT

In physical activity research, health-related quality of life (HRQOL) is an outcome variable of growing importance. Physical activity is directly associated with HRQOL and intervention-type studies seek to show improvements in HRQOL based on treatment effects. As interest grows in using HRQOL as an outcome measure in physical activity research, the need to investigate the measurement properties of HRQOL assessments increases in importance. The first objective of this project was to explore HRQOL assessments used in physical activity research by examining their instrument characteristics (items, dimensions, scoring, etc.) and their published psychometric properties. The second objective of this project was to investigate the reliability of the most commonly used HRQOL assessment in physical activity research, the Short Form Health Survey (SF-36). The specific aim of the second study was to investigate the extent to which the reliability of the SF-36 generalizes across various types of physical activity studies. The third and final objective of this project was to evaluate the SF-36 for proper measurement functioning using the Rasch model. Results of these studies showed that 10 HRQOL assessments are currently used in physical activity research and recommendations were made relative to different study designs. It was also found, through meta-analytical procedures, that the SF-36 provides strong reliability evidence across a wide range of physical activity research. Finally, the SF-36 met stringent modern measurement criteria using the Rasch model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

Introduction

In 1996, the U.S. Department of Health and Human Services (USDHHS) published a document entitled *Physical Activity and Health: A Report of the Surgeon General*. This document was the first Surgeon General's report specifically addressing physical activity and marked an exciting and historical public service message. The report's finding was that all Americans, young and old, can improve their health status by engaging in moderate physical activity (USDHHS, 1996). This message was motivated by the overwhelming amount of evidence confirming the positive health benefits associated with a physically active lifestyle.

The driving definition of physical activity used in this report was "any bodily movement produced by skeletal muscles that results in energy expenditure" (Caspersen, Powell, & Christenson, 1985, p. 126). This definition allowed for the promotion of both structured and non-structured forms of activity. The specific recommendation from the Surgeon General was that all people should engage in regular moderate physical activity, for at least 30 minutes, on most if not all days of the week. This recommendation included activities such as brisk walking, running, sports, and yard work. The report also stipulated that increasing both intensity and duration above moderate levels would likely provide even greater benefits (USDHHS, 1996).

In 2007, the USDHHS formed the Physical Activity Guidelines Advisory Committee with its goal to review the scientific literature and develop a comprehensive set of physical activity recommendations (USDHHS, 2008). The efforts of this advisory committee culminated with the *2008 Physical Activity Guidelines for Americans*. The guidelines state that adults should accumulate 150 minutes of weekly moderate-intensity physical activity or 75 minutes of weekly vigorous-intensity activity or an equivalent weekly combination of both (USDHHS, 2008).

Since the publication of *Physical Activity and Health: A Report of the Surgeon General*, many studies have shown the positive effects of regular physical activity on specific health outcomes. Such physical activity-related health outcomes have included all-cause mortality (Kampert, Blair, Barlow, & Kohl, 1996; Lee et al., 2010), cause-specific mortality (Kampert et al., 1996; Tanasescu, Leitzmann, Rimm, & Hu, 2003), premature chronic disease (Durand et al., 2011; Franco et al., 2005; Sesso, Paffenbarger, Ha, & Lee, 1999), obesity (Brien, Katzmarzyk, Craig, & Gauvin, 2007; Buchowski et al., 2010), and mental health (Backmand, Kaprio, Kuiala, & Sarna, 2003; Strawbridge, Deleger, Roberts, & Kaplan, 2002; Vallance et al., 2011).

Health-related quality of life (HRQOL) is another such health outcome that has seen a growing interest in physical activity research. HRQOL is a broad latent construct that includes both subjective and objective indicators people's lives that affect their physical and/or mental health status (Centers for Disease Control and Prevention [CDC], 2000). Wilson and Cleary (1995) developed a conceptual path diagram of HRQOL. The path represents causal connections of increasing complexity leading to HRQOL. Therefore, according to their model, HRQOL is a function of general health perceptions, which is a function of functional status, which is a function of symptom status, which is a function of biological and physiological indicators.

Due to its holistic nature, HRQOL has become a standard outcome measure in both intervention and observational studies (CDC, 2000). HRQOL is now more than ever being included in research studies alongside the more conventional and objective measures of health status (Dominick, Ahern, Gold, & Heller, 2004). In addition, HRQOL has been shown to be a strong predictor of physician visits, hospitalization, and mortality (Dominick, Ahern, Gold, & Heller, 2002).

Physical activity has been shown to be directly associated with HRQOL (Heath & Brown, 2009). Specifically, meeting recommended levels of physical activity has shown to be related to superior levels of HRQOL (Brown et al., 2003). Physical activity has been used as a predictor of HRQOL in both prospective (Aoyagi, Park, Park, & Shephard, 2010; Balboa-Castillo et al., 2011; Tessier et al., 2007) and cross-sectional (Heath & Brown, 2009; Luncheon & Zack, 2011) observational studies. Physical activity has also seen major impacts in clinical interventions with links to positive changes in HRQOL (Courneya et al., 2011; Sørensen, Sørensen, Skovgaard, Bredahl, & Puggaard, 2011). In fact, increasing HRQOL has been described as the most important goal in physical activity interventions (Bertheussen et al., 2011).

Statement of the Problem

Despite the abundance of research showing the positive effects of physical activity on health, the majority of Americans remain physically inactive (USDHHS, 2008). The *Healthy People 2020* publication states that more than 80% of adults do not meet the current guidelines for physical activity and has reported 10 objectives aimed at reducing the percentage of adults who engage in no physical activity (USDHHS, 2011). With physical activity interventions serving as the primary tool for achieving a more physically active population, the need to assess HRQOL will be in even greater demand. Given the overwhelming interest in HRQOL as an outcome measure in physical activity research, there is a strong need for a better understanding of the measurement properties of HRQOL assessments commonly used in physical activity research.

Statement of the Purpose

The purpose of this project is to address specific measurement-related research gaps related to HRQOL assessments in physical activity research. This research project specifically addresses these gaps by first gathering information about the most commonly used HRQOL scales in physical activity research, and then progressing to the issue of measurement reliability concerning the most commonly used HRQOL assessment, to finally the issue of measurement validity and item functioning of the SF-36. These three studies combined address some major measurement issues related to HRQOL assessments in physical activity research.

The purpose of the first study was to systematically review the common instruments used to measure generic HRQOL in physical activity research in adults by summarizing the characteristics and scoring options of each instrument as well as the psychometric properties of each HRQOL scale. The overall usefulness of each HRQOL instrument relative to different study designs will be discussed.

The purpose of the second study was to perform a systematic review and meta-analysis of reliability coefficients on the SF-36 so as to assess the generalizability of HRQOL reliability. A secondary purpose of the second study was to examine potential moderators which may account for extra variance associated with the SF-36, such as age, gender, and alternate forms of the SF-36.

The purpose of the third study was to evaluate the measurement properties of the Short Form Health Survey (SF-36) administered to adults by using the Rasch model. The evaluations made will be 1) proper category functioning, 2) model-data fit, 3) item-person map, 4) item difficulty parameters, item separation, and item separation reliability, 5) person ability (θ) fit, 6) convergent validity evidence for the SF-36 ability (θ) scores, and 7) construct validity evidence for the SF-36 ability (θ) scores.

Significance of the Study

Although there are a number of psychometric-based studies that evaluate different HRQOL scales currently in the literature, there are no studies specifically evaluating and comparing the various HRQOL scales which are commonly used in physical activity research. The first study aims to clarify the many nuances in the various HRQOL scales. The final report of the first study will serve as a HRQOL resource guide for those conducting physical activity research. In addition, there are currently no studies evaluating the reliability generalization of the SF-36 scales, the most common HRQOL assessment in physical activity research. The second study aims to address this by conducting a systematic review and meta-analysis of published SF-36 reliability coefficients. This approach will allow for the averaging of the HRQOL reliability and in turn evaluating the strength of its generalization. The results of this study will provide valuable evidence as to the strengths and weaknesses of SF-36-determined HRQOL in terms of measurement reliability.

Furthermore, there are very few studies that use modern test theory (item response theory) to evaluate HRQOL scales. The third study aims to fill this gap by calibrating scale items of the two major domains (physical and mental) of the SF-36 HRQOL assessment. The results of this study will serve as a critical evaluation of the SF-36 assessment and determine whether it functions adequately or is in need of modification. In summary, the completion of these three studies allows for a thorough evaluation of HRQOL assessment in physical activity research.

Chapter I References

Aoyagi, Y., Park, H., Park, S., & Shephard, R. J. (2010). Habitual physical activity and health-related quality of life in older adults: Interactions between the amount and intensity of activity (the Nakanojo Study). *Quality of Life Research, 19*(3), 333-338.

Backmand, H., Kaprio, J., Kuiala, U., & Sarna, S. (2003). Influence of physical activity on depression and anxiety of former elite athletes. *International Journal of Sports Medicine, 24*(8), 609-619.

Balboa-Castillo, T., León-Muñoz, L. M., Graciani, A., Rodríguez-Artalejo, F., & Guallar-Castillón, P., (2011). Longitudinal association of physical activity and sedentary behavior during leisure time with health-related quality of life in community-dwelling older adults. *Health and Quality of Life Outcomes, 9*, 47.

Bertheussen, G. F., Romundstad, P. R., Landmark, T., Kaasa, S., Dale, O., & Helbostad, J. L. (2011). Associations between physical activity and physical and mental health--a HUNT 3 study. *Medicine and Science in Sports and Exercise, 43*(7), 1220-1228.

Brien, S. E., Katzmarzyk, P. T., Craig, C. L., & Gauvin, L. (2007). Physical activity, cardiorespiratory fitness and body mass index as predictors of substantial weight gain and obesity: The Canadian physical activity longitudinal study. *Canadian Journal of Public Health, 98*(2), 121-124.

Brown, D. W., Balluz, L. S., Heath, G. W., Moriarty, D. G., Ford, E. S., Giles, W. H., & Mokdad, A. H. (2003). Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. *Preventive Medicine, 37*(5), 520-528.

Buchowski, M. S., Cohen, S. S., Matthews, C. E., Schlundt, D. G., Signorello, L. B., Hargreaves, M. K., & Blot, W. J. (2010). Physical activity and obesity gap between black and white women in the southeastern U.S. *American Journal of Preventive Medicine, 39*(2), 140-147.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126-131.

Centers for Disease Control and Prevention. Measuring healthy days: Population assessment of health-related quality of life. Centers for Disease Control and Prevention, Atlanta, Georgia 2000.

Courneya, K.S., Tamburrini, A.L., Woolcott, C.G., McNeely, M.L., Karvinen, K.H., Campbell, K.L., McTiernan, A., & Friedenreich, C.M. (2011). The Alberta Physical Activity and Breast Cancer Prevention Trial: Quality of life outcomes. *Preventive Medicine, 52*(1), 26-32.

Dominick, K. L., Ahern, F. M., Gold, C. H., & Heller, D. A. (2004). Health-related quality of life among older adults with arthritis. *Health and Quality of Life Outcomes, 13*(2), 5-12.

Dominick, K. L., Ahern, F. M., Gold, C. H., & Heller, D. A. (2002). Relationship of health-related quality to health care utilization and mortality among older adults. *Aging Clinical and Experimental Research, 14*(6), 499-508.

Durand, G., Tsismenakis, A. J., Jahnke, S. A., Baur, D. M., Christophi, C. A., & Kales, S. N. (2011). Firefighters' physical activity: relation to fitness and cardiovascular disease risk. *Medicine and Science in Sports and Exercise*, In Print.

Franco, O. H., de Laet, C., Peeters, A., Jonker, J., Mackenbach, J., & Nusselder, W. (2005). Effects of physical activity on life expectancy with cardiovascular disease. *Archives of Internal Medicine, 165*(20), 2355-2360.

Heath, G. W., & Brown, D. W. (2009). Recommended levels of physical activity and health-related quality of life among overweight and obese adults in the United States, 2005. *Journal of Physical Activity and Health, 6*(4), 403-411.

Kampert, J. B., Blair, S. N., Barlow, C. E., & Kohl, H. W. (1996). Physical activity, physical fitness, and all-cause and cancer mortality: A prospective study of men and women. *Annals of Epidemiology, 6*(5), 452-457.

Lee, D. C., Sui, X., Ortega, F. B., Kim, Y. S., Church, T. S., Winett, U., Ekelund, U., Katzmarzyk, P. T., & Blair, S. N. (2010). Comparisons of leisure-time physical activity and cardiorespiratory fitness as predictors of all-cause mortality in men and women. *British Journal of Sports Medicine, 45*(6), 504-510.

Luncheon, C., & Zack, M. (2011). Health-related quality of life and the physical activity levels of middle-aged women, California Health Interview Survey, 2005. *Prevention of Chronic Disease, 8*(2), A36.

Sesso, H. D., Paffenbarger, R. S., Ha, T., & Lee, I. M. (1999). Physical activity and cardiovascular disease risk in middle-aged and older women. *American Journal of Epidemiology, 150*(4), 408-416.

Sørensen, J., Sørensen, J. B., Skovgaard, T., Bredahl, T., & Puggaard, L. (2011). Exercise on prescription: Changes in physical activity and health-related quality of life in five Danish programmes. *European Journal of Public Health, 21*(1), 56-62.

Strawbridge, W. J., Deleger, S., Roberts, R. E., & Kaplan, G. A. (2002). Physical activity reduces the risk of subsequent depression for older adults. *American Journal of Epidemiology, 156*(4), 328-334.

Tanasescu, M., Leitzmann, M. F., Rimm, E. B., & Hu, F. B. (2003). Physical activity in relation to cardiovascular disease and total mortality among men with type 2 diabetes. *Circulation, 107*(19), 2435-2439.

Tessier, S., Vuillemin, A., Bertrais, S., Boini, S., Le Bihan, E., Oppert, J. M., Hercberg, S., Guillemin, F., & Briançon, S. (2007). Association between leisure-time physical activity and health-related quality of life changes over time. *Preventive Medicine, 44*(3), 202-208.

U.S. Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.

U.S. Department of Health and Human Services. Physical Activity Guidelines Advisory Committee Report. Washington (DC): U.S. Department of Health and Human Services; 2008. p. A-7.

U. S. Department of Health and Human Services. Healthy people 2020. Available at http://www.healthypeople.gov/2020/topicsobjectives2020/default.aspx. Accessed June 7, 2011.

Vallance, J. K., Winkler, E. A., Gardiner, P. A., Healy, G. N., Lynch, B. M., & Owen, N. (2011). Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005-2006). *Preventive Medicine*, In Press.

Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Journal of the American Medical Association, 273*(1), 59-65.

CHAPTER II

A Systematic Review of Health-Related Quality of Life Assessments in Physical Activity Research

Introduction

Physical activity is a strong predictor of health status and should be adopted by individuals of all ages (USDHHS, 2008). Many studies have shown the positive effects of regular physical activity on specific health outcomes. Such physical activity-related health outcomes have included all-cause mortality (Kampert, Blair, Barlow, & Kohl, 1996; Lee et al., 2010), cause-specific mortality (Kampert et al., 1996; Tanasescu, Leitzmann, Rimm, & Hu, 2003), premature chronic disease (Durand et al., 2011; Franco et al., 2005; Sesso, Paffenbarger, Ha, & Lee, 1999), obesity (Brien, Katzmarzyk, Craig, & Gauvin, 2007; Buchowski et al., 2010), and mental health (Backmand, Kaprio, Kuiala, & Sarna, 2003; Strawbridge, Deleger, Roberts, & Kaplan, 2002; Vallance et al., 2011).

Health-related quality of life (HRQOL) is a more personal health outcome that has seen a growing interest in physical activity research. HRQOL is a broad latent construct that includes both subjective and objective indicators of people's lives that affect their physical and/or mental health status (Centers for Disease Control and Prevention [CDC], 2000). Due to its ability to capture overall perceived health, HRQOL has become a standard outcome measure in public health and medical research (CDC, 2000). In addition, because HRQOL usually includes a component of perceived functional status, it has been considered a measure as important to research outcomes as other more objective indicators (Dominick, Ahern, Gold, & Heller, 2004). In addition, HRQOL has been shown to be a strong predictor of physician visits, hospitalization, and mortality (Dominick, Ahern, Gold, & Heller, 2002).

Participating in a physically active lifestyle is linked to greater HRQOL (Heath & Brown, 2009). Specifically, meeting recommended levels of physical activity has shown to be related to superior levels of HRQOL (Brown et al., 2003). Physical activity has been used as a predictor of HRQOL in both prospective (Aoyagi, Park, Park, & Shephard, 2010; Balboa-Castillo et al., 2011; Tessier et al., 2007) and cross-sectional (Heath & Brown, 2009; Luncheon & Zack, 2011) observational studies. Physical activity has also seen major impacts in clinical interventions with links to positive changes in HRQOL (Courneya et al., 2011; Sørensen, Sørensen, Skovgaard, Bredahl, & Puggaard, 2011). In fact, increasing HRQOL has been described as the most important goal in physical activity interventions (Bertheussen et al., 2011).

Despite the empirical evidence confirming the positive effects of physical activity on health, the majority of Americans remain physically inactive (USDHHS, 2008). The *Healthy People 2020* publication states that more than 80% of adults do not meet the current guidelines for physical activity and has reported several national objectives aimed at increasing the percentage of adults who engage in physical activity (USDHHS, 2011). With physical activity interventions serving as the primary tool for achieving a more physically active population, the need to assess HRQOL will be in even greater demand. Furthermore, with the overwhelming interest in HRQOL as an outcome measure in physical activity research, there is a strong need for a better understanding of the measurement properties of HRQOL assessments commonly used in physical activity research.

There are currently no studies that review the scale characteristics, score determination, feasibility issues, and psychometric properties of the common HRQOL instruments used in physical activity research. Therefore, the purpose of this study was to review the most common instruments used to measure generic HRQOL in physical activity research in adults by summarizing the characteristics and scoring options of each instrument as well as the psychometric properties of each HRQOL scale. The overall usefulness of each HRQOL instrument relative to different study designs will be discussed. The study will serve as a HRQOL resource guide for those conducting physical activity research.

Methods

Search strategy

PubMed.gov was systematically searched for published articles of physical activity research containing measures of HRQOL. The following search terms were used: ("physical activity" OR exercise) AND ("health-related quality of life" OR "quality of life"). After pertinent articles were identified, their reference lists were searched for more relevant studies. After all HRQOL instruments used in physical activity research were identified, the assessments were each investigated to determine if they were appropriate for inclusion.

Inclusion and exclusion criteria

An article was included in the study if it 1) was published in English, 2) was available in full text, 3) had a primary objective of evaluating the effects of physical activity on HRQOL as an outcome measure, 4) used a measure of HRQOL assessed via a questionnaire, 5) used adults as participants, and 6) was published on or after January 1, 2000. An HRQOL assessment was excluded from this study if it 1) was not specifically health-related in nature (e.g., life satisfaction), 2) measured a construct other than generic HRQOL (e.g., living with heart failure), 3) did not consist of a set of items measuring the HRQOL construct (e.g., single item or proxy variable), or 4) completely lacked empirical measurement evidence (e.g., researcher developed questions).

Instrument characteristics and properties

For each identified HRQOL assessment tool included in the study, the following characteristics were retrieved: 1) mode of administration, 2) number of items contained in the assessment tool, 3) type of rating scale(s) used, 4) number and types of domains and sub-dimensions, 5) alternate forms, 6) target populations, 7) adopted languages, and 8) scoring methods. The psychometric properties retrieved from each HRQOL assessment were categorized into three domains: validity, reliability, and item response theory.

The validity properties included in this study were: 1) content validity, (2) criterion validity, 3) construct validity, and 4) responsiveness. *Content validity* is the extent to which an assessment tool measures the construct of interest (Allen & Yen, 2002). Appropriate scale construction should include content validity methods such as literature review, expert panel advice and/or ratings, and theme saturation (Kline, 2009). *Criterion validity* is the extent to which measurements from an assessment tool adequately reflect an agreed upon gold standard measurement. HRQOL may have no known gold standard and therefore criterion validity may have limited to no impact on this study. *Construct validity* is the ability of an assessment tool to measure the trait or construct that it was intended to measure (Allen & Yen, 2002). Types of construct validity evidence include known group difference testing, assessment of uni- or multi-dimensionality of scales, and correlation with other measures of hypothesized direction (i.e., convergent or divergent validity). *Responsiveness* is the ability of an assessment tool to detect clinically important changes in the construct of interest (Deyo, Diehr, & Patrick, 1991). Measures of responsiveness should include effect size measures or statistics from receiver operating characteristic (ROC) curves.

The reliability properties included in this study were: 1) internal consistency reliability and 2) test-retest. *Internal consistency* refers to the extent to which items in an assessment tool are inter-correlated (Cronbach, 1951). If such an inter-correlation exists, the items of the scale are said to measure a unidimensional construct. *Test-retest* reliability measures the stability of measurements over repeated trials (Allen & Yen, 2002). Measures of test-retest include limits of agreement, Pearson correlations, and intra-class correlations (ICCs).

Item response theory methods to be reviewed include: 1) item analysis, 2) model data fit, 3) rating scale assessment, 4) scoring, 5) test equating, and 6) differential item functioning (DIF). Item response theory stems from modern psychometric theory and incorporates various scale item and person ability parameters into a statistical probability model (Hambleton, Swaminathan, & Rogers, 1991). That is, the probability of a person's response to an item is a function of the person's trait being measured (i.e., HRQOL) and the characteristics of the item (i.e., difficulty, discrimination, etc.). Through the use of item response theory models, HRQOL assessment tools can be evaluated based on their item's usefulness, the scale's unidimensionality, and the functioning of the chosen rating scale (Hambleton, Swaminathan, & Rogers, 1991). Item response theory can also be used to create an interval level measurement of the construct and equate scores from different assessment tools, as well as determine whether model bias exists across population subgroups (Wood & Zhu, 2006).

Results

A total of 8,263 articles were found using the search terms. After reviewing titles and abstracts, 2,556 articles were identified as meeting inclusion criteria, of which, 1,209 articles were dropped due to exclusion criteria. A total of 1,347 articles were included in the final sample and were examined for their HRQOL assessment. Table 1 displays the characteristics of 10 HRQOL assessments arranged according to their frequency of use in physical activity research. The majority of physical activity studies used the Short Form Health Survey (SF-36) or one of its variants. The next most commonly used HRQOL assessment was the Sickness Impact Profile (SIP) followed by the Euroqol assessment (EQ-5D). Other HRQOL assessments identified (from most common to less common) were the Nottingham Health Profile (NHP), WHO Quality of Life (WHOQOL-BREF), Quality of Well-Being Scale (QWB), Health Utilities Index 3 (HUI3), CDC's Healthy Days Core (HRQOL-4), Assessment of Quality of Life (AQoL), and the Duke Health Profile (DHP).

Short-Form Health Survey (SF-36)

Characteristics. The SF-36 is the most widely used HRQOL instrument in physical activity research. The appeal of the SF-36 is that it is a relatively efficient scale with numerous published sources detailing its psychometric properties. The SF-36 was developed from the Medical Outcomes Study (MOS) conducted by RAND (Ware & Sherbourne, 1992). The SF-36 is a multi-dimensional scale consisting of 36 items, 8 health-related dimensions, and two domains. The dimensions include: 1) vitality, 2) physical functioning, 3) bodily pain, 4) general health, 5) physical role functioning, 6) emotional role functioning, 7) social role functioning, and 8) mental health. The physical domain consists of the physical functioning, bodily pain, general health, and physical role functioning dimensions and the mental domain consists of the vitality, emotional role functioning, social role functioning, and mental health dimensions (Ware, 2004). The latest version (v2) of the SF-36 has 3 different rating scale categories, ranging from 3-point to 6-point.

The SF-36 is intended to measure HRQOL in adults and can be self-administered, administered via computer, with aid of an interviewer, or by telephone. The instrument can be modified to include either a (standard) 4-week recall or a 1-week recall and has been incorporated into both observational and intervention-type studies. The SF-36, due to advances in measurement theory, has made several transformations, and is now referred to as the 2nd version (SF-36v2). The newer version made changes to item wording, item layout, and number of response categories to certain items (Ware, 2004). Three other alternate forms of the SF-36 are available. The SF-12, SF-12v2, and SF-8 are shorter forms of the original that, however, maintain the measurement of all 8 dimensions as well as the two domain-specific summary scores (QualityMetric, 2011).

The scoring of the SF-36 is relatively simple, relying on the assumption that item scores are linearly related to the underlying construct with the scales summated according to the Likert approach (Ware & Sherbourne, 1992). The updated version of SF-36 (SF-36v2) allows scores to be normalized to allow for easy comparisons (Ware, 2004). The normalizing process used national data to allow for standardization of summated scores, followed by T-score conversion.

Psychometric properties. The SF-36 was constructed from a pool of items retrieved from existing instruments used for measuring physical limitations, role functioning, mental health, and perceived general health (Ware & Sherbourne, 1992). The larger pool of 245-items was part of the Medical Outcomes Study (MOS), of which the 36-items of the SF-36 were a subset. Participants in the MOS who completed the lengthy survey and took the follow-up health examination (within 1-month) were used for the validity study. The health examination was required to allow for clinical diagnoses for the construction of contrasting groups (clinical tests of validity).

Table 1

Characteristics of generic HRQOL assessments in adult physical activity research

Instrument	Mode	Items	Scale	Scoring	Dimensions	Forms	Languages
Short-Form Health Survey (SF-36)	Self-Administered Computer Interviewer Telephone	36	Categorical Rating 3 to 6-point	1) Summated Scoring 2) Norm-Based T-scoring 8-dimensions 2-domains	Vitality Physical functioning Bodily pain General health Physical role functioning Emotional role functioning Social role functioning Mental health	SF-36v1* SF-36v2* SF-12v1 SF-12v2 SF-8 VF-36	Multiple
Sickness Impact Profile (SIP)	Self-Administered Interview	136	Yes/No	1) Standardized Weighted 2) Overall 3) 12-dimensions 4) 2-domains	Sleep and rest Emotional behavior Body care and movement Home management Mobility Social interaction Ambulation Alertness behavior Communication Work Recreation and pastimes Eating	SIP-136* SIP-68 SIP-66 SIP-30 SIP-24 SIP-82	Multiple
EuroQol (EQ-5D)	Self-Administered Interview Telephone	6	Categorical Rating 3-point VAS 0-100	1) Descriptive Profile (11111 to 33333) 2) Health Index Score (-0.11 to 1) 3) Self-Reported Health Status (0 to 100)	Mobility Self-Care Usual Activities Pain/Discomfort Anxiety/Depression	EQ-5D-3L* EQ-5D-5L	Multiple

Note. * indicates the common form used in physical activity research.

Table 1 (continued)

Characteristics of generic HRQOL assessments in adult physical activity research

Instrument	Mode	Items	Scale	Scoring	Dimensions	Forms	Languages
Nottingham Health Profile (NHP)	Self-Administered Interview	45	Yes/No	Scaled Weights (0 to 100)	Physical mobility Pain Social isolation Emotional reactions Energy Sleep	NHP*	Multiple
WHO Quality of Life Assessment (WHOQOL-BREF)	Self-Administered Interviewer	26	Categorical Scale 5-point	1) 4 Domain Scores 2) 2 Descriptive Items	Physical Health Psychological Social Relationships Environment	WHOQOL-BREF* WHOQOL-100	Multiple
Quality of Well-being Scale (QWB)	Self-Administered Computer Interviewer Telephone	76	Categorical Scale 2 to 5-point	1) 4 Domain Scores 2) Health Index Score (0 to 1)	Symptoms Mobility Physical Activity Social Activity	QWB QWB-SA*	Multiple
Health Utilities Index Mark 3 (HUI3)	Self-Administered Computer Interviewer Telephone	8	Categorical Scale 5 to 6-point	1) Descriptive Profile 2) Health Index Score (0 to 1)	Emotion Pain Vision Hearing Speech Ambulation Dexterity Cognition	HUI1 HUI2 HUI3*	Multiple

Note. * indicates the common form used in physical activity research.

Table 1 (continued)

Characteristics of generic HRQOL assessments in adult physical activity research

Instrument	Mode	Items	Scale	Scoring	Dimensions	Forms	Languages
CDC Healthy Days (CDC HRQOL)	Interview Telephone	4	1 Categorical Rating Scale 3 Continuous Measures	1) Descriptive Score 2) Summary Index	Physical Mental	HRQOL-4* HRQOL-9 HRQOL-12	English
Assessment of QoL (AQoL)	Self-Administered Interview Mail Telephone	15	Categorical Rating 4-point	1) Overall Score 2) 5-Dimension Scores 3) Utility Score (1 to 0)	Illness Independent Living Social Relationships Physical Senses Psychological Wellbeing	AQoL I AQoL II* AQoL-8	English
Duke Health Profile (DHP)	Self-Administered	17	Categorical Rating 3-point	1) 10 Dimension Scores 2) 1 Summary General Health Score	Physical Mental Social General Perceived Health Self-Esteem Anxiety Depression Pain Disability	DHP* DUHP	English

Note. * indicates the common form used in physical activity research.

Two types of criteria were used for the initial validation of the SF-36. Psychometric criteria were considered by the use of principal components analysis (construct validity) and inspection of correlations among the eight scales. Clinical criteria were considered by comparing the specific scale scores between four distinct groups of subjects: 1) minor chronic conditions only, 2) serious chronic medical conditions only, 3) psychiatric conditions only, and 4) both serious medical and psychiatric conditions (McHorney, Ware, & Sherbourne, 1993).

As hypothesized, the principal components showed high loadings of physical functioning, physical role, and bodily pain on the physical domain. Also, high loadings were seen of mental health, emotional role, and social functioning on the mental domain. Vitality and general health had cross-loaded on both domains. These results provided evidence of both convergent (i.e., physical functioning loading on physical domain) and divergent (i.e., physical functioning not loading on mental domain) validity. As well, results of the contrasting groups analysis provided acceptable validity evidence for the SF-36 scales (McHorney, Ware, & Sherbourne, 1993).

Initial reliability was estimated for the SF-36 using corrected item-test correlations as well as Cronbach's alpha for each scale (McHorney, Ware, & Sherbourne, 1994). Using acceptable cut-point criteria, all eight scales had a 100% success rate. Average item-test correlations ranged from .42 to .74. As well, Cronbach's alpha ranged from .78 to .93.

Since its inception, the SF-36 has undergone hundreds of psychometric-related testings. Some of these studies have focused on validating the SF-36 instrument on different language speaking populations and/or cultures (Laguardia, Campos, Travassos, Najar, Anjos, & Vasconcellos, 2011) or demographic-specific populations (Mishra et al., 2011). Other studies have focused on validating the scales on disease-specific populations (Laosanguanek, Wiroteurairuang, Siritho, & Prayoonwiwat, 2011) or condition-specific populations (Freidheim, Borchgrevin, Saltnes, Kaasa, 2007).

The widespread popularity and use of the SF-36 has drawn the attention of a few investigators trained in item response theory. Specifically, Rasch analysis has been used to compare its measurement results with the traditional Likert summation and (Raczek et al., 1998). Results showed that SF-36 scores from the Rasch analysis displayed stronger relative validity evidence as compared to the traditional summation approach. Rasch measurement has also been used to compare the two methods in relative precision (McHorney, Haley, Ware, 1997) and confirm the unidimensionality and reproducibility of the instrument (Haley, McHorney, Ware, 1994).

Sickness Impact Profile (SIP)

Characteristics. The SIP is another instrument used to measure HRQOL. The SIP was designed specifically as a measure of behavioral dysfunction in usual daily activities (Gilson et al., 1975). The final version consisted of 136-items of 12 categories: 1) sleep and rest, 2) emotional behavior, 3) body care and movement, 4) home management, 5) mobility, 6) social interaction, 7) ambulation, 8) alertness behavior, 9) communication, 10) work, 11) recreation and pastimes, and 12) eating (Bergner, Bobbitt, Carter, & Gilson, 1981).

The SIP is designed to be self-administered or given by face-to-face interview. The instrument is intended for generally healthy adults as well as adults with specific health conditions. The SIP is a relatively long instrument, as compared to SF-36 and EQ-5D. Like other popular HRQOL instruments, SIP has been used internationally and therefore has been translated into several language specific versions (Coons, Rao, Keininger, & Hays, 2000).

Several different scores can be obtained through SIP use: overall score, 12 different dimension scores, and 2 domain scores (physical and psychosocial). The response scale is a simple dichotomous yes or no type and the scoring is derived from a standardized weighting scheme (Bergner et al., 1981).

Psychometric properties. The development of the SIP was driven by strong content validity (Gilson et al., 1975). The investigators developing the SIP began with an open-ended request form to elicit statements from individuals describing sickness-related changes in behavior. This procedure produced 1,250 statements of sickness-related behavior, which then resulted in 312 unique statements comprising 14 different dimensions. Using 25 judges and their ratings of the 312 items, content validity was affirmed by showing the correlations of each judge's rating of an item with the mean of the 25 judge's ratings.

Over the course of a few years, other psychometric data appeared regarding the SIP (Bergner et al., 1981). Construct validity, in experimental (clinical) format, was demonstrated using the differing health status and severity approach (Bergner, Bobbitt, Pollard, Martin, & Gilson, 1976). Reliability was also tested extensively for the SIP (Pollard, Bobbitt, Bergner, Martin, & Gilson, 1976). Test-retest reliability, internal consistency, and inter-rater reliability were tested on 119 respondents. Also, tests were carried out with two different forms (long and short), two different modes of administration (interviewer and self-administration), and with the sample stratified by disease severity. Overall, the reliability of the SIP was moderate to high in all circumstances.

There has only been one published study utilizing item response theory on the SIP (Lindeboom et al., 2004). The extended Rasch model was used to calibrate the SIP items, assess item bias, and create a shorter form via test equating. Results showed that 82 items fit the Rasch model. Item bias was seen in age, gender, and diagnosis groups, and the Rasch calibrated shorter 82-item form showed a moderate correlation with the SIP full form. Several shorter forms of the SIP have been developed (see table 1), but their use in physical activity research is sparse.

Euroqol (EQ-5D)

Characteristics. The EQ-5D questionnaire is a standardized instrument used to measure HRQOL (Brazier, Jones, & Kind, 1993). The EQ-5D is a very simple and short instrument that has two distinct parts (AHRQ, 2005). The first part is a set of five items, each serving as a separate dimension: 1) mobility, 2) self-care, 3) usual activities, 4) pain/discomfort, and 5) anxiety/depression. Each item has a 3-category response: 1 = *no problems*, 2 = *some problems* and 3 = *extreme problems*. The second part is a visual analog scale (VAS) representing self-assessed health status. The scale ranges from *Best imaginable health state* (100) to the *Worst imaginable health state* (0). Respondents mark the vertical scale (which resembles a thermometer) at their perceived level of health.

The EQ-5D is designed to be self-administered or given by face-to-face interview. The instrument is intended for the general adult population as well as adults with specific health conditions. The EQ-5D is most efficiently used in large population-based surveys (Euroqol, 2011) but has also been widely used in clinical settings (Vestergaard, Kronborg, & Puggaard, 2008). Like the SF-36 instrument, the EQ-5D has been used internationally and therefore has been adapted to several language specific versions (Euroqol, 2011). Also, with advances in psychometric theory, researchers have found a possible benefit of having a 5-category response scale as opposed to the 3-category response scale (Herdman et al., 2011). These psychometric-based changes have resulted in separate EQ-5D-3L (3-level) and EQ-5D-5L (5-level) forms.

The EQ-5D has three different scoring methods (AHRQ, 2005). The first is just the simple scoring profile of the five items (i.e., 32124). There are 243 possible combinations of these five components and therefore this scoring method yields 243 different *health states*. The second method is a population preference-weighted index score based on the five items. The index ranges from -0.11 (if scores are all 3s for each item) to 1.0 (if scores are all 1s for each item). Given that a score of 0.0 equates to death and a score of 1.0 equates to perfect health, it can be seen that the index can assume a quality of health worse than death itself. The last scoring method simply comes straight from the VAS instrument and serves as a measure of self-reported health.

Psychometric properties. The EQ-5D was developed by a multi-disciplinary group of researchers to measure HRQOL (Brazier, Jones, & Kind, 1993). The most notable psychometric data on the EQ-5D are from a performance study of its construct, convergent, and divergent validity (Brazier et al., 1993). Construct validity was tested according to hypothesized relationships between special groups of people and the noted difference in their scoring profile. For example, older adults, women, professional workers, recent users of health services, and those diagnosed with a chronic health condition were hypothesized (and shown) to have lower scoring profiles, compared to their respective counterparts. Convergent and divergent validity were tested by comparing EQ-5D scores to dimensional scores of the SF-36. Convergent validity was established by showing that the EQ-5D anxiety/depression dimension was highly correlated with the mental health dimension of the SF-36. Likewise, divergent validity was established by showing that the EQ-5D anxiety/depression dimension was not highly correlated with the physical functioning dimension of the SF-36.

Another useful study that provided psychometric data for the EQ-5D investigated its construct validity and discriminant ability (Essink-Bot, Krabbe, Bonsel, & Aaronson, 1997). The EQ-5D reliability (internal consistency) could not be assessed because each dimension (scale) had only a single item. The construct validity was tested using polychoric correlation coefficients (PCCs) between its scales and those of the

COOP/WONCA instrument. PCC results showed strong correlations with like scales and low correlations with unlike scales. Construct validity was further tested using common factor analysis. Results showed a two-factor model: mental and physical health. Also, as suspected, the anxiety/depression scale loaded on the "mental" factor and mobility, self-care, usual activities, and pain/discomfort scales loaded on the "physical" factor, providing adequate construct validity. The discriminant ability of the EQ-5D was tested using receiver operating characteristic (ROC) curves. The grouping variables in the study were migraine headache status and reporting an absence from work due to illness. EQ-5D successfully and significantly distinguished between both grouping variables, providing evidence for discriminant ability.

Item response theory has been used, in a state-of-the-art fashion, to determine the equivalency of EQ-5D measures between the 3-level response form and the 5-level response form (Pickard, Kohlmann, Janssen, Bonsel, Rosenbloom, & Cella, 2007). Another study, using a Rasch measurement model, showed the benefits of having a 5-category response scale as opposed to the 3-category response scale (Herdman et al., 2011).

Nottingham Health Profile (NHP)

Characteristics. The NHP is a generic HRQOL instrument with physical, emotional, and social domains of health (Hunt, McEwan, & McKenna, 1985). The NHP has a total of 45 items, all of which are dichotomous response. Two parts make up the instrument. The first part contains six different health dimensions: 1) physical mobility,

2) pain, 3) social isolation, 4) emotional reactions, 5) energy, and 6) sleep. The second part includes specific health status questions.

The NHP is designed for adults (16+ years) and to be self-administered or interviewer-administered. It was originally designed as an instrument for epidemiological research (Coons et al., 2000) but has since been used in several different arenas. No alternate forms (to date) have been found in the published literature. However, translated versions have been created in several languages.

Only the first part of NHP is considered in its scoring (Hunt, McKenna, McEwen, Williams, & Papp, 1981). Scores can be obtained using weights associated with subject responses and yield a single value ranging from 0 (no health problems) to 100 (severe health problems). If all weights are summed, in part I, a score of 100 will occur.

Psychometric properties. The NHP was developed using methods of content validity, beginning with a large pool of statements (over 2,200) from approximately 700 people regarding their typical feelings about poor health (Hunt, McEwan, & McKenna, 1985). The resulting instrument took on 45 items, 38 of which were part of the overall scoring profile, and six dimensions (Hunt et al., 1985). Construct validity evidence was published on the NHP by testing the instrument's ability to distinguish between different levels of pain severity (Mauskopf, Austin, Dix, & Berzon, 1994). Results successfully showed that the pain, energy, and sleep dimensions were highly correlated with pain severity (convergent validity evidence), whereas the other three dimensions were not highly correlated with pain severity (divergent validity).

Another study providing psychometric data for the NHP, investigated its construct validity and discriminant ability (Essink-Bot et al., 1997). The NHP reliability (internal consistency) was determined by Cronbach's alpha. The construct validity was tested using intraclass correlation coefficients (ICCs) between its scales and those of the SF-36 instrument. Reliability results showed good internal consistency. ICC results showed strong correlations with like scales and low correlations with unlike scales. Construct validity was further tested using common factor analysis. Results showed a two-factor model: mental and physical health. Also, as suspected, the energy, emotional reactions, and social isolation scales loaded on the "mental" factor and energy, pain, and physical mobility scales loaded on the "physical" factor, providing adequate construct validity. The discriminative ability of the NHP was tested using ROC curves. Groups were formed by migraine headache status and reporting an absence from work due to illness. NHP significantly distinguished between both grouping variables, providing evidence for discriminant ability.

Item response theory has been used to assess the psychometric properties of the NHP (Hagell, Whalley, McKenna, & Lindvall, 2003). Results showed adequate fit to the model, however, differential item functioning (DIF) was found in age and gender groups. A Rasch study was performed to reduce the number of items in the NHP from 38 to 22, while maintaining the new scale's validity (Prieto, Alonso, Lamarca, & Wright, 1998). Finally, the Rasch model was used to assess unidimensionality and item-fit of the Brazilian version of the NHP (Teixeira-Salmela et al., 2004). Despite adequate fit, some items were found to be too easy for the population under study.

WHO Quality of Life Assessment (WHOQOL-BREF)

Characteristics. The WHOQOL-BREF is a generic HRQOL tool developed from the larger WHOQOL-100 (Saxena, Carlson, & Billington, 2001). The assessment consists of 26 items which make up four HRQOL domains: 1) physical health, 2) psychological, 3) social relationships, and 4) environment. Additionally, two of the items are included to assess self-perceived general health and are for descriptive purposes. The WHOQOL-BREF was developed for adult use and has been designed to be an international HRQOL assessment. The scoring of the WHOQOL-BREF results in a single score for each domain.

Psychometric properties. The development of the WHOQOL-BREF stemmed from the larger WHOQOL version (WHOQOL-100). The initial WHOQOL-BREF project showed adequate validity with strong correlations between the WHOQOL-BREF and WHOQOL-100 domain scores ("Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group," 1998). The same project also showed strong evidence for content validity, discriminant validity, internal consistency, and stability in the WHOQOL-BREF scales. A more recent validation study of the WHOQOL-BREF used a large sample of participants from 23 countries. Participants were very diverse, consisting of people of various health ranges and diseases, and various sociodemographic characteristics. Results of the study showed strong evidence of internal consistency, discriminant validity, and construct validity (Skevington, Lotfy, & O'Connell, 2004).

The construct validity was evaluated in the WHOQOL-BREF using item response theory among a general population of adults (Noerholm et al., 2004). Using a mail survey format and a random sample of Danish adults, the WHOQOL-BREF was administered to 1,101 respondents. Results indicated that each of the four domains of the WHOQOL-BREF fit a 2-parameter item response model. However, the total scale did not fit either a 2-parameter model or a Rasch model. The conclusion was that domain specific scores should be used when administering the WHOQOL-BREF and that the total scores of the WHOQOL-BREF may not be sufficiently valid.

Quality of Well-Being Scale (QWB)

Characteristics. The QWB is a generic HRQOL assessment tool that measures 3 different dimensions (Coons et al., 2000). A recently updated version of the QWB scale has been developed specifically for participant self-administration (QWB-SA). The QWB-SA does not require a trained interviewer, as does the QWB, and therefore is easier and less expensive to use in research and practice. The different dimensions attempt to assess HRQOL in relation to daily functioning with scales in 1) mobility, 2) physical activity, and 3) social activity. The functioning scales ask questions about certain activities and ask respondents to respond using the past 3 days as their reference. A second component of health problems is assessed by asking questions about 26 (QWB) or 58 symptoms (QWB-SA). Four different domain scores can be generated which can also be combined to form a total utility score representing HRQOL, ranging from 0 to 1 (death to optimal health, respectively).

Psychometric properties. Test-retest reliability was provided for both forms (QWB & QWB-SA) of the QWB scale (Kaplan, Sieber, & Ganiats, 1997). English speaking primary care patients were used for the reliability study. Participants were randomized to receive either the QWB or the QWB-SA and were administered their respective forms twice with a one month interval. Results showed that the two forms were equivalent in terms of HRQOL scores. Also, results indicated that both QWB forms were stable in assessing HRQOL over time.

Construct validity has been established for the QWB scale by showing strong relationships between its HRQOL scores and various health outcomes among patients with chronic obstructive pulmonary disease (Kaplan, Atkins, & Timms, 1984). Further validity evidence was presented when QWB scores were found to be associated with four health outcomes among HIV-infected adults (Hughes et al., 1997). Finally, evidence was also provided for the QWB's construct validity when HRQOL scores were significantly related to dementia ratings and behavioral problems among patients with and without Alzheimer's disease (Hughes et al., 1997).

Although not specifically evaluated for its functioning, the QWB has been analyzed using item response theory, in comparison to four other HRQOL asseessments (Cherepanov, Palta, & Fryback, 2010). As part of the National Health Measurement Study, the QWB was administered to 3,844 adults along with the EQ-5D, HUI2, HUI3, and SF-6D assessments. Findings showed that the five assessments combined contributed to 3 domains consisting of physical, psychosocial, and pain. However, the QWB only contributed to 2 of these domains, physical and psychosocial.

Health Utility Index Mark 3 (HUI3)

Characteristics. The HUI3 is another tool used often in economics research. Its development was driven by the need to describe 1) experiences of medical patients, 2) outcomes associated with therapy and disease, 3) the effectiveness of medical and health-related interventions, and 4) health status in large population studies (Horsman, Furlong, Feeny, & Torrance, 2003). The HUI3 is the most recent version of the Health Utility Index series, starting with HUI1 and then HUI2 (Grootendorst, Feeny, & Furlong, 2000). The HUI3 consists of 8 attributes: 1) vision, 2) hearing, 3) speech, 4) ambulation, 5) dexterity, 6) emotion, 7) cognition, and 8) pain. With these attributes and a multi-attribute utility algorithm, the HUI3 can yield HRQOL values covering over 900,000 unique health states (Coons, Rao, Keininger, & Hays, 2000). Scores can also be computed using a different set of algorithms to yield either single-attribute or multi-attribute values ranging from -0.36 (*worse than dead*) to 0.00 (*dead*) to 1.00 (*perfect health*).

Psychometric properties. The HUI3 is a third generation HRQOL instrument that began from earlier work with the HUI1 (Torrance, Furlong, Feeny, & Boyle, 1995). The rationale for the original items and dimensions came from a population perspective of health outcomes (Coons et al., 2000). The evolution of the index to the HUI3 was driven by the desire to make the assessment practical for both clinical use as well as population studies. Each dimension of the HUI3 is assessed with a single item; therefore, internal consistency reliability has not been examined with this assessment. Stability has been evaluated in the HUI3 using the Kappa statistic of agreement. The HUI3 was administered to a large sample at two different time periods, one month apart. Results showed that 6 of the 8 dimensions had acceptable reliability (Boyle, Furlong, Feeny, Torrance, & Hatcher, 1995).

Construct validity was evaluated in the HUI3 by comparing HRQOL scores between groups with known differences. Participants were used from the 1990 Ontario Health Survey. Adequate validity was shown as participants with stroke, arthritis, and both stroke and arthritis had significantly lower HUI3 scores (Grootendorst et al., 2000). Convergent validity evidence was evaluated on the HUI3 by comparing the scoring patterns between the HUI3, EQ-5D, and SF-36 HRQOL assessments (Luo et al., 2003). Participants for this validity study were outpatients with rheumatic disease. Results provided adequate evidence for convergent validity. Those patients with higher SF-36 scores also had significantly higher EQ-5D and HUI3 scores. Total scores on EQ-5D and HUI3 were not significantly different from each other.

To compare 5 different HRQOL assessments for their interrelationships, item response theory was used on the HUI3 (Fryback, Palta, Cherepanov, Bolt, & Kim, 2010). As part of the National Health Measurement Study, the HUI3 was administered to 3,844 adults along with the EQ-5D, HUI2, QWB-SA, and SF-6D. Results indicated, that the

HUI3 was linearly related to the EQ-5D and the HUI2 scales. Although a linear relationship was shown, it was stated that the relationship was simplistic and that the different scales were in actuality measuring different aspects of generic HRQOL.

CDC Health-Related Quality of Life (HRQOL-4) Scale

Characteristics. The HRQOL-4 Scale consists of four items and was developed as a surveillance tool to be used in the U.S. Behavioral Risk Factor Surveillance System (BRFSS) (Taylor & National Center for Chronic Disease Prevention and Health Promotion (U.S.). Division of Adult and Community Health., 2000). The four items were created through the CDC's definition of HRQOL which includes perceived physical and mental health over time. The first item asks participants to rate their own general health on a 5-point scale starting with *excellent* and ending with *poor*. The second and third questions were specifically geared toward *physical health* (physical illness and injury) and *mental health* (stress, depression, and emotional problems), respectively. These questions ask respondents to report the number of days (out of the previous 30 days) that their physical (or mental) health was not good. The last question specifically addresses the amounts of *usual activity* (self-care, work, or recreation) influenced by physical and/or mental health. Respondents are asked to report the number of days (out of the previous 30 days) that poor physical or mental health kept them from their usual activities (Hennessy, Moriarty, Zack, Scherr, & Brackbill, 1994).

The scoring methods for the CDC HRQOL-4 scale are twofold. The first option is a descriptive scoring method. This can be done by creating dichotomous categories for each item (Heath & Brown, 2009). For example, for the first item, those reporting either *fair* or *poor* general health can be considered to exhibit poor HRQOL and those reporting *excellent, very good,* or *good* general health can be considered to exhibit good HRQOL. For the second and third items, those reporting 14 days of poor health or more can be considered to exhibit poor physical (or mental) health. For the fourth item, those reporting 14 days or more can be considered to be inactive due to poor health. The second scoring option is to create a summary index of unhealthy (or healthy) days. The index can be constructed from the physical and mental health items and used to assess the overall number of unhealthy days due to physical and/or mental health, not to exceed 30 days (Taylor & National Center for Chronic Disease Prevention and Health Promotion (U.S.). Division of Adult and Community Health., 2000).

Psychometric properties. The CDC HRQOL-4 scale was developed using a strong conceptual framework (Hennessy et al., 1994). Items were specifically constructed to be 1) individual-oriented, 2) subjective in nature, 3) non disease-specific, 4) sensible to the general public, 5) non-biased toward various ethnic groups, and 6) practical. The time frame was also considered to capture an adequate reflection of an individual's health. Developers of the CDC HRQOL-4 scale used early BRFSS data to test the scale's validity. This was accomplished first by showing the relationship between the first core HRQOL item (perceived general health rating) and the second core item (number of days respondents said their physical health was not good). The relationship provided convergent validity evidence as those reporting better general health had significantly fewer days of poor physical health and those reporting poor general health reported significantly more days of poor physical health. This relationship was also found between the first item and the third (number of days poor mental health) and fourth (number of days limited by physical and/or mental health) items.

The retest reliability was assessed for the CDC HRQOL-4 using a random sample of BRFSS respondents approximately two weeks after their initial survey (Andresen, Catlin, Wyrwich, & Jackson-Thompson, 2003). The Kappa coefficient and proportion of agreement were used for the first core item and the intra-class correlation coefficient was used for the other three items as well as the healthy days index. Reliability coefficients for the general health item and the healthy days index were both acceptable. Reliability was moderate for the other three (number of days) items.

Another validation study of the CDC HRQOL-4 was conducted with a sample of Dutch adults (Toet, Raat, & van Ameijden, 2006). First, reliability was evaluated by computing the Cronbach alpha on the three number of days core items. The reliability of the three items was deemed acceptable (alpha = .77). Second, criterion validity was assessed by comparing the HRQOL-4 items with three other well-respected HRQOL assessments: SF-36, WHOQoL-BREF, and GHQ-12. Spearman correlations confirmed that HRQOL-4 items of similar domain were highly related across instruments. As well, HRQOL-4 items of different domains were not correlated across instruments. Finally, construct validity was examined by comparing HRQOL-4 scores between groups of adults with known differences in health status. Those respondents reporting a chronic condition, depression, use of prescription drugs, and visiting a doctor, had significantly lower scores of HRQOL as assessed by the CDC HRQOL-4. To date, no item response theory studies have been published on the CDC HRQOL-4 assessment.

Assessment of Quality of Life (AQoL)

Characteristics. The AQoL instrument is a generic HRQOL assessment developed by Australian researchers (Hawthorne, Richardson, & Osborne, 1999). The AQoL consists of five dimensions covering the HRQOL construct and contains questions specifically targeted for economic evaluation. The AQoL has a total of 15 items each measured on a four point categorical scale ranging from A (*Good HRQOL*) to D (*poor HRQOL*). The five dimensions consist of: 1) illness, 2) independent living, 3) social relationships, 4) physical senses, and 5) psychological wellbeing.

The AQoL is designed to be administered by self, interviewer, mail, or telephone. It was designed as a multi-attribute health utility index, however, it is also used as a health states assessment. The original AQoL was replaced by its developers and referred to as AQoL-II. A shorter version has been developed (Hawthorne, 2009) consisting of only eight items (AQoL-8); however, this version has not been used in physical activity research.

There are three scoring options for AQoL users (Hawthorne et al., 1999). The first is an overall HRQOL score. This is computed by assigning a zero to an 'A' response, a one to a 'B' response, a two to a 'C' response, and a three to a 'D' response. Therefore, a low overall score of zero is possible and a high score of 45 is possible. The second option is to sum the same scale by subdomains. Therefore, each subdomain can range in score from zero to nine. Finally, an algorithm can be used to transform the raw AQoL scores to preference weighted utility scores ranging from -0.04 (*worse than death*) to zero (*death*) to one (*complete health*).

Psychometric properties. The AQoL was developed using a content validation procedure (Hawthorne et al., 1999). The development began using a strong conceptual framework based on the World Health Organizations's (WHO) definition of health. With this framework in mind, researchers and professionals constructed appropriate items, reviewed the items for clarity and simplicity, and administered the selected items to both hospital patients and community members. After data collection, an item analysis was performed using 100% range criteria (all categories of an item being selected) for item sensitivity and a standard deviation of .50 cutoff as the criterion for item discrimination. Items surviving the preliminary analysis were further tested for construct validity. First, principal components analysis was run, dropping items that did not load on the underlying HRQOL construct. Second, exploratory factor analysis was performed, dropping items that failed to load on a single factor only. Results indicated a five factor structure and measures of internal consistency confirmed its reliability. Re-analysis of each factor separately by principal components analysis provided evidence of the unidimensionality of each factor. Finally, structural equation modeling was performed to assess the explanatory power of the AQoL in providing HRQOL information.

Another study providing psychometric information for the AQoL investigated its stability across different methods of administration. The developer of the instrument showed that administering the AQoL via mail or telephone resulted in statistically equivalent HRQOL scores (Hawthorne, 2003). Furthermore, the stability of scores was maintained as well for each set of subscale scores.

Item response theory has been used on the AQoL with a specific purpose to find the most parsimonious scale (Hawthorne, 2009). First, subscale unidimensionality was determined followed by full scale unidimensionality, using Mokken item response theory. Items which were not considered unidimensional (homogenous) were candidates for deletion. Second, a Rasch partial credit model was used to determine each item's set of category thresholds. Items with disordered thresholds (i.e., persons with low HRQOL endorsing categories representing higher HRQOL levels) were also candidates for deletion. The goal of the study was to reduce each of the four subscales by one item resulting in an 8-item AQoL scale using the two criteria of unidimensionality and ordered category thresholds. The resulting AQoL-8 correlated well (intraclass correlation coefficient = .95) with the full AQoL scale and showed 97% of a validation sample within +/- 2 SD limit of agreement and was therefore considered a more parsimonious measure of HRQOL.

Duke Health Profile (DHP)

Characteristics. The DHP is a generic self-report HRQOL assessment tool that contains 10 different measures of health (Parkerson, Broadhead, & Tse, 1990). Six of the measures are considered positive health measures (physical, mental, social, general, perceived health, and self-esteem) and the other four are considered measures of dysfunction (anxiety, depression, pain, and disability). The scale consists of only 17 items, each measured on a 3-point categorical rating scale. Scoring for the DHP is relatively simple, summing each separate dimension and multiplying (or dividing for the general health score) by a constant. Each dimension has a score range from 0 to 100 where 100 represents the best health for the positive health measures and the worst health for the measures of dysfunction.

Another unique characteristic of this assessment is that the general health dimension is a composite of the physical, mental, and social dimensions. Combining these three major dimensions of HRQOL allows for a more realistic measure of general health. The DHP was developed for adults but has been revised and validated for adolescents (Vo, Guillemin, & Deschamps, 2005). The DHP HRQOL assessment is primarily used in English speaking countries but has recently been validated in France (Baumann et al., 2011).

Psychometric properties. The DHP was developed from a slightly larger (63-item) Duke-UNC Health Profile (DUHP) assessment (Parkerson et al., 1990). Items were selected from the larger pool of items using content validity (or face validity) and item-remainder (item score and dimension score with item removed) correlations. Cronbach alphas provided evidence of internal consistency reliability with multi-item dimensions having alphas ranging from .55 to .78. Test-retest provided evidence of stability with all dimensions having reliability greater than .50, except disability and pain. Spearman correlations were used for item-convergent and item-divergent evidence against three other assessment tools: DUHP, SIP, and the Tennessee Self-Concept Scale. Validity was established for the DHP with strong positive correlations among similar constructs and strong negative correlations among different constructs. Finally, mean comparisons were used to provide construct validity evidence by showing DHP score differences between groups with known health problems. Validity was established when results showed that groups with lower levels of health had significantly lower DHP scores compared to groups with better health (Parkerson et al., 1990).

Discussion

The purpose of this study was to systematically review assessments used to measure generic HRQOL in physical activity research in adults. The review included summarizing the characteristics, scoring options, and psychometric properties of each HRQOL assessment. A total of 10 instruments were found and examined. By far, the SF-36 along with its variants was the most commonly used HRQOL assessment in physical activity research. Table 2 displays the recommendation for each assessment based on whether a researcher's reason for selecting it was its psychometric properties, amount of HRQOL information (scores from dimensions), or its length.

In terms of participant burden, the CDC HRQOL-4 and the EQ-5D both provide a valid HRQOL score given they contain only 4 and 6 items, respectively. The AQoL, DHP, and WHOQL-BREF, however, also allow for low participant strain (15, 17, and 26 items, respectively) and provide slightly more information. The AQoL provides a single HRQOL score along with 5 subdomain scores (illness, independent living, social relationships, physical senses, and psychological wellbeing). The DHP provides 10 dimensional scores (physical, mental, social, general perceived health, self-esteem, anxiety depression, pain, and disability), one of which is a general health score. The WHOQOL-BREF measures HRQOL with 4 separate dimensions (physical health, psychological, social relationships, and environment). These 3 mid-sized assessments may be useful to physical activity researchers who seek to investigate very specific HRQOL changes (i.e., pain or social relations) without overwhelming their subjects with several items or forms. In terms of psychometric properties, the SF-36 leads in both amounts and quality of supporting information. The evidence backing the SF-36's validity includes both classical test theory as well as modern test theory. The other 9 assessments all have several studies validating their scales using both psychometric approaches, with the exception of the CDC HRQOL-4 and DHP which have no published data (to date) using item response theory.

In conclusion, 10 HRQOL assessments were found to be used in physical activity research. The SF-36, the most commonly used and validated assessment, provides the most information given its size. Other HRQOL assessments with good potential include AQoL, DHP, and WHOQOL-BREF. If time is the most important factor, the EQ-5D and CDC HRQOL-4 are useful and valid scales.

Table 2

Recommendation for HRQOL assessment based on psychometric properties, amount of HRQOL information, and length

Form	Psychometric	Information	Short Length
SF-36	Definitely	Definitely	SF-12/8
SIP	Yes	Yes	No
EQ-5D	Yes	Maybe	Definitely
NHP	Yes	No	No
WHOQOL	Yes	Maybe	Maybe
QWB	Maybe	Maybe	No
HUI3	Yes	Yes	Yes
HRQOL-4	Maybe	No	Definitely
AQoL	Maybe	Yes	Yes
DHP	Maybe	Definitely	Yes

Chapter II References

Agency for Healthcare Research and Quality (AHRQ). (2005). Calculating the U.S. Population-based EQ-5D Index Score. Retrieved August 15, 2011, from http://www.ahrq.gov/rice/EQ5Dscore.htm

Allen, M. J., & Yen, W. M. (2002). Introduction to Measurement Theory. Long Grove, IL: Waveland Press.

Andresen, E. M., Catlin, T. K., Wyrwich, K. W., & Jackson-Thompson, J. (2003). Retest reliability of surveillance questions on health related quality of life. *Journal of Epidemiology and Community Health, 57*(5), 339-343.

Aoyagi, Y., Park, H., Park, S., & Shephard, R. J. (2010). Habitual physical activity and health-related quality of life in older adults: Interactions between the amount and intensity of activity (the Nakanojo Study). *Quality of Life Research, 19*(3), 333-338.

Backmand, H., Kaprio, J., Kuiala, U., & Sarna, S. (2003). Influence of physical activity on depression and anxiety of former elite athletes. *International Journal of Sports Medicine, 24*(8), 609-619.

Balboa-Castillo, T., León-Muñoz, L. M., Graciani, A., Rodríguez-Artalejo, F., & Guallar-Castillón, P., (2011). Longitudinal association of physical activity and sedentary behavior during leisure time with health-related quality of life in community-dwelling older adults. *Health and Quality of Life Outcomes, 27*(9), 47.

Baumann, C., Erpelding, M. L., Perret-Guillaume, C., Gautier, A., Regat, S., Collin, J. F., & Briancon, S. (2011). Health-related quality of life in French adolescents and adults: Norms for the DUKE Health Profile. *BMC Public Health, 27*(11), 401. doi: 10.1186/1471-2458-11-401

Bergner, M, Bobbitt, R. A., Pollard, W. E., Martin, D. P., & Gilson, B.S. (1976). The sickness impact profile: Validation of a health status measure. *Medical Care, 14*(1), 57.

Bergner, M., Bobbitt, R. A., Kressel, S., Pollard, W. E., Gilson, B. S., & Morris, J. R.. (1976). The sickness impact profile: Conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services*, *6*(3), 393.

Bergner, M., Bobbitt, R. A., Carter, W. B., & Gilson, B. S. (1981). Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care, 19*(8), 787.

Bertheussen, G. F., Romundstad, P. R., Landmark, T., Kaasa, S., Dale, O., & Helbostad, J. L. (2011). Associations between physical activity and physical and mental health--a HUNT 3 study. *Medicine and Science in Sports and Exercise*, *43*(7), 1220-1228.

Boyle, M. H., Furlong, W., Feeny, D., Torrance, G. W., & Hatcher, J. (1995). Reliability of the Health Utilities Index--Mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. *Quality of Life Research, 4*(3), 249-257.

Brazier, J., Jones, N., & Kind, P. (1993). Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Quality of Life Research, 2*(3), 169.

Brien, S. E., Katzmarzyk, P. T., Craig, C. L., & Gauvin, L. (2007). Physical activity, cardiorespiratory fitness and body mass index as predictors of substantial weight gain and obesity: The Canadian physical activity longitudinal study. *Canadian Journal of Public Health, 98*(2), 121-124.

Brown, D. W., Balluz, L. S., Heath, G. W., Moriarty, D. G., Ford, E. S., Giles, W. H., & Mokdad, A. H. (2003). Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. *Preventive Medicine, 37*(5), 520-528.

Buchowski, M. S., Cohen, S. S., Matthews, C. E., Schlundt, D. G., Signorello, L. B., Hargreaves, M. K., & Blot, W. J. (2010). Physical activity and obesity gap between black and white women in the southeastern U.S. *American Journal of Preventive Medicine, 39*(2), 140-147.

Centers for Disease Control and Prevention. *Measuring healthy days: Population assessment of health-related quality of life*. Centers for Disease Control and Prevention, Atlanta, Georgia 2000.

Cherepanov, D., Palta, M., & Fryback, D. G. (2010). Underlying dimensions of the five health-related quality-of-life measures used in utility assessment: Evidence from the National Health Measurement Study. *Medical Care, 48*(8), 718-725.

Coons, S. J., Rao, S., Keininger, D. L., & Hays, R. D. (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics, 17*(1), 13-35.

Courneya, K.S., Tamburrini, A.L., Woolcott, C.G., McNeely, M.L., Karvinen, K.H., Campbell, K.L., McTiernan, A., & Friedenreich, C.M. (2011). The Alberta Physical Activity and Breast Cancer Prevention Trial: Quality of life outcomes. *Preventive Medicine, 52*(1), 26-32.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures: Statistics and strategies for evaluation. *Controlled Clinical Trial, 12*(4 Suppl), 142S-158S.

Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. (1998). *Psychological Medicine, 28*(3), 551-558.

Dominick, K. L., Ahern, F. M., Gold, C. H., & Heller, D. A. (2004). Health-related quality of life among older adults with arthritis. *Health and Quality of Life Outcomes, 13*(2), 5-12.

Dominick, K. L., Ahern, F. M., Gold, C. H., & Heller, D. A. (2002). Relationship of health-related quality to health care utilization and mortality among older adults. *Aging Clinical and Experimental Research, 14*(6), 499-508.

Durand, G., Tsismenakis, A. J., Jahnke, S. A., Baur, D. M., Christophi, C. A., & Kales, S. N. (2011). Firefighters' physical activity: Relation to fitness and cardiovascular disease risk. *Medicine and Science in Sports and Exercise*, In Print.

Essink-Bot, M. L., Krabbe, P. F., Bonsel, G. J., Aaronson, N. K. (1997). An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Medical Care*, *35*(5), 522-537.

Franco, O. H., de Laet, C., Peeters, A., Jonker, J., Mackenbach, J., & Nusselder, W. (2005). Effects of physical activity on life expectancy with cardiovascular disease. *Archives of Internal Medicine*, *165*(20), 2355-2360.

Freidheim, O. M. S., Borchgrevin, P. C., Saltnes, T., & Kaasa, S. (2007). Validation and comparison of the health related quality of life instrument EORTC QLQ C30 and SF36 in assessment of patients with non-malignant pain. *Journal of Pain and Symptom Management*, *34*(6), 657.

Fryback, D. G., Palta, M., Cherepanov, D., Bolt, D., & Kim, J. S. (2010). Comparison of 5 health-related quality-of-life indexes using item response theory analysis. *Medical Decision Making, 30*(1), 5-15.

Gilson BS, Gilson JS, Bergner M, Bobbit RA, Kressel S, Pollard WE, Vesselago M. (1975). The sickness impact profile. Development of an outcome measure of health care. *American Journal of Public Health*, *65*(12), 1304-1310.

Grootendorst, P., Feeny, D., & Furlong, W. (2000). Health Utilities Index Mark 3: Evidence of construct validity for stroke and arthritis in a population health survey. *Medical Care, 38*(3), 290-299.

Hagell P, Whalley D, McKenna SP, Lindvall O. (2003). Health status measurement in Parkinson's disease: Validity of the PDQ-39 and Nottingham Health Profile. *Movement Disorders, 18*(7), 773-783.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. New Deli, India: Sage Publications.

Hawthorne, G. (2003). The effect of different methods of collecting data: Mail, telephone and filter data collection issues in utility measurement. *Quality of Life Research, 12*(8), 1081-1088.

Hawthorne, G. (2009). Assessing utility where short measures are required: Development of the short Assessment of Quality of Life-8 (AQoL-8) instrument. *Value Health, 12*(6), 948-957.

Hawthorne, G., Richardson, J., & Osborne, R. (1999). The Assessment of Quality of Life (AQoL) instrument: A psychometric measure of health-related quality of life. *Quality of Life Research, 8*(3), 209-224.

Heath, G. W., & Brown, D. W. (2009). Recommended levels of physical activity and health-related quality of life among overweight and obese adults in the United States, 2005. *Journal of Physical Activity and Health, 6*(4), 403-411.

Heath, G. W., & Brown, D. W. (2009). Recommended levels of physical activity and health-related quality of life among overweight and obese adults in the United States, 2005. *Journal of Physical Activity and Health,* 6, 403-411.

Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research,* In Press.

Hennessy, C. H., Moriarty, D. G., Zack, M. M., Scherr, P. A., & Brackbill, R. (1994). Measuring health-related quality of life for public health surveillance. *Public Health Reports, 109*(5), 665-672.

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI): Concepts, measurement properties and applications. *Health and Quality of Life Outcomes, 1*, 54. doi: 10.1186/1477-7525-1-54

Hughes, T. E., Kaplan, R. M., Coons, S. J., Draugalis, J. R., Johnson, J. A., & Patterson, T. L. (1997). Construct validities of the Quality of Well-Being Scale and the MOS-HIV-34 Health Survey for HIV-infected patients. *Medical Decision Making, 17*(4), 439-446.

Hunt, S. M., McEwan, J., & McKenna, S. P. (1985). Measuring health status: A new tool for clinicians and epidemiologists. *Journal of the Royal College of General Practitioners*, 35(273), 185.

Hunt, S. M., McKenna, S. P., McEwen, J., Williams, J., & Papp, E. (1981). The Nottingham Health Profile: Subjective health status and medical consultations. *Social Science Medicine*, 15(3), 221.

Kampert, J. B., Blair, S. N., Barlow, C. E., & Kohl, H. W. (1996). Physical activity, physical fitness, and all-cause and cancer mortality: A prospective study of men and women. *Annals of Epidemiology*, 6, 452-457.

Kaplan, R. M., Atkins, C. J., & Timms, R. (1984). Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. *Journal of Chronic Disease, 37*(2), 85-95.

Kaplan, R. M., Sieber, W. J., & Ganiats, T. G. (1997). The quality of well-being scale: Comparison of the interviewer-administered version with a self-administered questionnaire. *Psychology & Health, 12*(6), 783-791.

Kline, T. (2009). *Psychological testing: A practical approach to design and evaluation.* (3rd ed.). New Deli, India: Vistaar Publications.

Laguardia, J., Campos. M,R,, Travassos. C.M., Najar, A.L., Anjos, L.A., Vasconcellos, M.M. (2011). Psychometric evaluation of the SF-36 (v.2) questionnaire in a probability sample of household: Results of the survey Pesquisa Dimensoes Sociais das Desigualdades (PDSD), Brazil, 2008. *Health and Quality of Life Outcomes, 9*(1), 61.

Laosanguanek, N., Wiroteurairuang, T., Siritho, S., Prayoonwiwat, N. (2011). Reliability of the Thai version of SF-36 questionnaire for an evaluation of quality of life in multiple sclerosis patients in multiple sclerosis clinic at Siriraj Hospital. *Journal of the Medical Association of Thailand, 94* Suppl 1, S84.

Lee, D. C., Sui, X., Ortega, F. B., Kim, Y. S., Church, T. S., Winett, U., Ekelund, U., Katzmarzyk, P. T., & Blair, S. N. (2010). Comparisons of leisure-time physical activity and cardiorespiratory fitness as predictors of all-cause mortality in men and women. *British Journal of Sports Medicine*, *45*(6), 504-510.

Lindeboom R, Holman R, Dijkgraaf MG, Sprangers MA, Buskens E, Diederiks JP, De Haan RJ. (2004). Scaling the sickness impact profile using item response theory: An exploration of linearity, adaptive use, and patient driven item weights. *Journal of Clinical Epidemiology*, *57*(1), 66.

Luncheon, C., & Zack, M. (2011). Health-related quality of life and the physical activity levels of middle-aged women, California Health Interview Survey, 2005. *Prevention of Chronic Disease*, *8*(2), A36.

Luo, N., Chew, L. H., Fong, K. Y., Koh, D. R., Ng, S. C., Yoon, K. H., . . . Thumboo, J. (2003). A comparison of the EuroQol-5D and the Health Utilities Index mark 3 in patients with rheumatic disease. *Journal of Rheumatology, 30*(10), 2268-2274.

Mauskopf J, Austin R, Dix L, Berzon R. (1994). The Nottingham Health Profile as a measure of quality of life in zoster patients: Convergent and discriminant validity. *Quality of Life Research*, *3*(6), 43.

McHorney CA, Haley SM, Ware JE Jr. (1997). Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, *50*(4), 451.

McHorney, C.A., Ware, J.E., Sherbourne, C.D. (1993). The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*, *31*(3), 247.

McHorney, C.A., Ware, J.E., Sherbourne, C.D. (1994). The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*, *32*(1), 40.

Mishra, G.D., Gale, C.R., Sayer, A.A., Cooper, C., Dennison, E.M., Whalley, L.J., Craig, L., Kuh, D., Deary, I.J. (2011). How useful are the SF-36 sub-scales in older people? Mokken scaling of data from the HALCyon programme. *Quality of life Research*, *20*(7), 1005.

Noerholm, V., Groenvold, M., Watt, T., Bjorner, J. B., Rasmussen, N. A., & Bech, P. (2004). Quality of life in the Danish general population--normative data and validity of WHOQOL-BREF using Rasch and item response theory models. *Quality of Life Research*, *13*(2), 531-540.

Parkerson, G. R., Jr., Broadhead, W. E., & Tse, C. K. (1990). The Duke Health Profile. A 17-item measure of health and dysfunction. *Medical Care*, *28*(11), 1056-1072.

Pickard AS, Kohlmann T, Janssen MF, Bonsel G, Rosenbloom S, Cella D. (2007). Evaluating equivalency between response systems: Application of the Rasch Model to a 3-level and 5-level EQ-5D. *Medical Care*, *45*(9), 812.

Pollard WE, Bobbitt RA, Bergner M, Martin DP, Gilson BS. (1976). The Sickness Impact Profile: Reliability of a health status measure. *Medical Care*, *14*(2), 146.

Prieto, L., Alonso J, Lamarca R, Wright BD, (1998). Rasch Measurement for Reducing the Items of the Nottingham Health Profile. *Journal of Outcome Measurement*, *2*(4), 285.

QualityMetric. (2011). SF Health Surveys. Retrieved August 15, 2011, from http://www.qualitymetric.com/WhatWeDo/GenericHealthSurveys/tabid/184/Default.aspx

Raczek AE, Ware JE, Bjorner JB Gandek B, Haley SM, Aaronson NK, Apolone G, Bech P, Brazier JE, Bullinger M, Sullivan M. (1998). Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: Results from the IQOLA Project. *Journal of Clinical Epidemiology, 51*(11), 1203.

Saxena, S., Carlson, D., & Billington, R. (2001). The WHO quality of life assessment instrument (WHOQOL-Bref): The importance of its items for cross-cultural research. *Quality of Life Research, 10*(8), 711-721.

Sesso, H. D., Paffenbarger, R. S., Ha, T., & Lee, I. M. (1999). Physical activity and cardiovascular disease risk in middle-aged and older women. *American Journal of Epidemiology, 150*(4), 408-416.

Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research, 13*(2), 299-310.

Sørensen, J., Sørensen, J. B., Skovgaard, T., Bredahl, T., & Puggaard, L. (2011). Exercise on prescription: Changes in physical activity and health-related quality of life in five Danish programmes. *European Journal of Public Health, 21*(1), 56-62.

Strawbridge, W. J., Deleger, S., Roberts, R. E., & Kaplan, G. A. (2002). Physical activity reduces the risk of subsequent depression for older adults. *American Journal of Epidemiology, 156*(4), 328-334.

Tanasescu, M., Leitzmann, M. F., Rimm, E. B., & Hu, F. B. (2003). Physical activity in relation to cardiovascular disease and total mortality among men with type 2 diabetes. *Circulation, 107*(19), 2435-2439.

Taylor, V. R., & National Center for Chronic Disease Prevention and Health Promotion (U.S.). Division of Adult and Community Health. (2000). *Measuring healthy days : Population assessment of health-related quality of life*. Atlanta, Ga.: U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Adult and Community Health.

Teixeira-Salmela LF, Magalhães Lde C, Souza AC, Lima Mde C, Lima RC, Goulart F. (2004). Adaptation of the Nottingham Health Profile: A simple measure to assess quality of life. *Cadernos de saúde pública, 20*(4), 905.

Tessier, S., Vuillemin, A., Bertrais, S., Boini, S., Le Bihan, E., Oppert, J. M., Hercberg, S., Guillemin, F., & Briançon, S. (2007). Association between leisure-time physical activity and health-related quality of life changes over time. *Preventive Medicine, 44*(3), 202-208.

Toet, J., Raat, H., & van Ameijden, E. J. (2006). Validation of the Dutch version of the CDC core healthy days measures in a community sample. *Quality of Life Research, 15*(1), 179-184.

Torrance, G. W., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics, 7*(6), 503-520.

U.S. Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.

U.S. Department of Health and Human Services. Physical Activity Guidelines Advisory Committee Report. Washington (DC): U.S. Department of Health and Human Services; 2008. p. A-7.

U. S. Department of Health and Human Services. Healthy people 2020. Available at http://www.healthypeople.gov/2020/topicsobjectives2020/default.aspx. Accessed June 7, 2011.

Vallance, J. K., Winkler, E. A., Gardiner, P. A., Healy, G. N., Lynch, B. M., & Owen, N. (2011). Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005-2006). *Preventive Medicine*, In Press.

Vo, T. X., Guillemin, F., & Deschamps, J. P. (2005). Psychometric properties of the DUKE Health Profile-adolescent version (DHP-A): A generic instrument for adolescents. *Quality of Life Research, 14*(10), 2229-2234.

Ware, J.E. (2004). SF-36 Health Survey Update. Retrieved August 15, 2011, from http://www.sf-36.org/announcements/Updated_SF36_bookChapter_Sept04.pdf

Ware, J.E., Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care, 30*(6), 473.

Wood, T. M., Zhu, W. (2006). *Measurement theory and practice in kinesiology.* Champaign, IL: Human Kinetics.

CHAPTER III

Reliability of the Short-Form Health Survey (SF-36) in Physical Activity Research Using Meta-Analysis

Introduction

Health-related quality of life (HRQOL) is a health outcome that has seen a growing interest in physical activity research. The United States Department of Health and Human Services (USDHHS), in its publication entitled *Healthy People 2020*, includes the enhancement of quality of life as a major public health service goal (USDHHS, 2011). HRQOL is a broad concept that includes both subjective and objective indicators of a people's lives that affect their physical and/or mental health status (Centers for Disease Control and Prevention [CDC], 2000). A more comprehensive description of HRQOL includes several dimensions including but not limited to physical functioning, psychological well-being, social functioning, role functioning, and health perceptions (Hennessy, Moriarty, Zack, Scherr, & Brackbill, 1994).

HRQOL has become a standard outcome measure in both intervention and observational studies (CDC, 2000). A search on Pubmed.gov showed only 1,410 publication hits from the years 1980 to 1999 using the key word term "health-related quality of life." The same search resulted in 15,180 hits from the years 2000 to 2011. HRQOL is also being included in research studies alongside the more conventional and objective measures of health status (Dominick, Ahern, Gold, & Heller, 2004). Given the overwhelming interest in HRQOL as an outcome measure in physical activity research, there is a strong need for a better understanding of the measurement properties of HRQOL assessments commonly used in physical activity research.

The *Short-Form Health Survey (SF-36)* is the most widely used HRQOL instruments in physical activity research. The SF-36 was developed from the Medical Outcomes Study (MOS) conducted by RAND (Ware & Sherbourne, 1992). The SF-36 is a multi-dimensional scale consisting of 36 items, 8 health-related dimensions, and two domains. The dimensions include: 1) vitality, 2) physical functioning, 3) bodily pain, 4) general health, 5) physical role functioning, 6) emotional role functioning, 7) social role functioning, and 8) mental health. The physical domain consists of the physical functioning, bodily pain, general health, and physical role functioning dimensions and the mental domain consists of the vitality, emotional role functioning, social role functioning, and mental health dimensions (Ware, 2004). The SF-36 is intended to measure HRQOL in adults and is easily self-administered in physical activity research. The *SF-12* and *SF-8* are both shorter versions of the original form with 12 and 8 items, respectively. Both shorter versions maintain the measurement of the 8 dimensions as well as the two domain-specific summary scores (QualityMetric, 2011).

A *Reliability* index can be thought of as a percentage of variance in a set of scores that is true or from non-error factors (Allen & Yen, 2002). Reliability also refers to the consistency or stability of test scores. This can be consistency of scores over time, consistency of scores across items (internal consistency) or consistency of scores across raters (Kline, 2005). Of specific interest to this study are the internal consistency and test-retest reliability coefficients. *Internal consistency* refers to the extent to which items in an assessment tool are inter-correlated (Cronbach, 1951). If such an inter-correlation exists, the items of the scale are said to measure a unidimensional construct. *Test-retest* reliability measures the stability of measurements over repeated trials (Allen & Yen, 2002). Since reliability is not a property of an assessment tool itself but rather a property of the assessment scores, it should be common practice for researchers to report reliability of their scores at hand (Ragan & Kang, 2005). Despite the inability to infer an assessment's reliability from a single study to all future studies, it is still possible to strengthen the reliability generalization using meta-analysis (Vacha-Haase, 1998).

To date, there are no studies which assess the reliability generalization of the SF-36 HRQOL assessment in physical activity research using meta-analysis. The purpose of this study was to perform a systematic review and meta-analysis of reliability coefficients on the SF-36 HRQOL assessment applied in physical activity research so as to assess the generalizability of HRQOL measurement reliability. A secondary purpose of the study was to examine potential moderators which may account for extra variance associated with the HRQOL measurements, such as age, gender, and even HRQOL assessment tool. This approach will provide valuable evidence as to the strengths and weaknesses of HRQOL scales in terms of measurement reliability.

Methods

Search Strategy

Using PubMed.gov, a systematic search was conducted to identify all physical activity studies using the SF-36 assessment or one of its variants (i.e., SF-12 or SF-8). The following search terms were used: ("physical activity" OR exercise) AND ("short-form 36" OR "sf-36" OR "MOS 36" OR "rand 36"). The same search strategy was used for identifying studies using the SF-12 and SF-8 forms. After all potential studies were identified; inspection of titles and abstracts ensued to identify exclusion criteria.

Studies were excluded if 1) they were published in a non-English language, 2) they were non-research based, 3) they were from a non-peer-reviewed journal, 4) HRQOL was not measured in the study using the SF-36 or one of its variants, 5) the study was a validation study where reliability measures were repeatedly computed due to changes made to the scale, 6) the HRQOL assessment tool was modified from its original form prior to its administration, 7) the study was not physical activity oriented, or 8) the study population involved youth under the age of 18 years.

By searching the remaining articles surviving exclusion criteria, reported measures of reliability were retrieved. Reliability measures were defined as Cronbach's alpha, Kuder-Richardson coefficient, Split-half reliability coefficient, or test-retest coefficients. When possible, reliability measures were extracted for each of the SF-36 domains (physical and mental). If a situation occurred where an article published reliability in the form of a range (i.e., .78 to .88), the midpoint of that range was used as the effect size. If a study reported a lower-bound value (i.e., > .80), the lower-bound was used as the effect size. Precaution was taken to not include redundant coefficients from multiple publications on the same study population, by matching author names and study characteristics.

Data extraction and coding

Data were extracted from each identified study using the following strategy. Reliability estimates along with the number of items for each assessment tool were recorded. If reliability estimates were reported separately for different populations (i.e., gender) then each population was considered a separate sample. Common demographic variables were included such as gender (male, female, both) and age (mean). Sample size was recorded for each study for its impact on reliability estimates. Other possible moderator data collected was disease status (diseased, non-diseased), reliability type (internal consistency, test-retest) and study design (randomized controlled trial [RCT], other). Also, the SF-36 form used for the HRQOL measure was recorded (SF-36, SF-12, SF-8).

Data analysis

A separate meta-analysis was performed for each HRQOL domain (physical and mental). Since internal consistency reliability coefficients represent a proportion of variance not accounted for by error, the square root of each coefficient was first taken. Fisher's r to Z transformation was then performed as recommended for correlation coefficients (Lipsey & Wilson, 2001). Meta-analytic mean effect sizes and 95%

confidence intervals were computed for each reliability study using a random effects model. A mean effect size of .70 or greater was considered acceptable evidence of reliability (Gliner, Morgan, & Harmon, 2001). A moderator analysis was performed to account for the extra variance by examining different study characteristics. The *Q* test of homogeneity was used to support the moderator analysis in determining if the variance in reliability coefficients was significantly different from zero. The Comprehensive Meta-Analysis Version 2.0 software was used for all meta-analyses (Comprehensive Meta-Analysis, 2006).

Results

Search results

A total of 1,358 articles were retrieved using the search strategy. After reviewing titles and/or abstracts, 1,080 were eliminated using exclusion criteria. After full review of the remaining 278 articles, 20 physical domain and 21 mental domain reliability coefficients were retrieved. After contacting study authors from the remaining articles, an additional 24 and 22 coefficients were retrieved for the physical and mental scales, respectively. Table 1 contains 87 effect sizes used in this study, 44 for the physical health domain and 43 for the mental health domain.

Physical health domain

Figure 1 shows the effect sizes and 95% confidence intervals for the physical health domain across each study. The weighted mean effect size from the physical health domain was computed using a random effects model. The effect size was strong and significantly different from zero, ES = .90 (95% CI: .88, .92), $p < .001$. Therefore, the physical health scales of the SF-36 assessment are reliable across a wide variety of physical activity studies. The test of homogeneity was significant, $Q = 2057.21$, $df = 43$, $p < .001$, $I^2 = 97.91$. This indicates that some effect sizes may come from different populations and a moderator analysis should be performed. Table 2 shows results of the moderator analysis for the physical health studies. Results showed that studies of various gender (female or both sexes) did not account for significant variance in effect sizes, $Q = 1.77$, $df = 1$, $p = .184$. Studies using different designs (RCT or other types) did not account for significant variance in effect sizes, $Q = 0.02$, $df = 1$, $p = .884$. Also, studies using participants of different health status (diseased or non-diseased) did not account for significant variance in effect sizes, $Q = 0.77$, $df = 1$, $p = .379$. However, studies using different forms of the SF-36 (SF-36, SF-12 or SF-8) did account for significant variance in effect sizes, $Q = 12.82$, $df = 2$, $p = .002$. As well, studies publishing different reliability coefficients (alpha or retest) accounted for significant variance in effect sizes, $Q = 8.00$, $df = 1$, $p = .005$. Finally, mean age of study participants was a significant predictor ($b = -.003$) of effect size, $Q = 9.59$, $df = 1$, $p = .002$. Figure 3 displays the linear relationship between age and the transformed effect sizes.

Table 1

Characteristics of the physical activity and SF-36 studies

Study	Np	Form	Type	Design	Disease	Age	Gender	ES_{phys}	ES_{ment}
Fisher 2004	582	SF-12	alpha	RCT	no	74	both	.911	.900
Li 2003	40	SF-12	alpha	RCT	no	72.6	both	.889	.883
Barnason 2009	55	SF-36	alpha	RCT	yes	71.6	both	.949	.949
Basen-Engquist 2006	60	SF-36	alpha	RCT	yes	55	female	.927	.927
Blacklock 2007	341	SF-36	alpha	other	no	55	both	NA	.917
Ciairano 2010	22	SF-36	alpha	RCT	no	80.6	both	.775	.812
Conroy 2007	497	SF-36	alpha	other	no	56.9	female	.943	.933
Cook 2011	539	SF-36	alpha	other	no	19.8	both	.900	.943
Coups 2009	175	SF-36	alpha	other	yes	68.7	both	.922	.922
Griffith 2009	126	SF-36	alpha	RCT	yes	60.2	both	.920	NA
Isaacs 2007	943	SF-36	retest	RCT	yes	57	both	.680	.800
Johansen 2001	38	SF-36	alpha	other	yes	52	both	.794	NA
Li 2009	599	SF-36	alpha	other	no	NG	both	.837	.837
Li 2010	187	SF-36	alpha	other	yes	59	both	.910	.910
McGrath 2011	143	SF-36	alpha	other	no	NG	both	.854	.854
Smith 2009	736	SF-36	alpha	other	yes	57	female	NA	.866
Tessier 2007	3891	SF-36	alpha	other	no	51.8	both	.906	.906
Tung 2010	70	SF-36	alpha	other	yes	69.6	both	.926	.896
Turner 2009	2995	SF-36	alpha	other	yes	55.3	both	.975	.927
Volkmann 2010	242	SF-36	alpha	other	yes	43	female	.959	NA
Yates 2003	64	SF-36	alpha	other	yes	NG	both	NA	.917
Zimmerman 2007	54	SF-36	alpha	RCT	yes	72.1	female	.967	.910
Tamari 2011	42	SF-36	alpha	other	no	75.7	both	.849	.900
Feldman 2009	50	SF-36	alpha	other	yes	51	both	.913	.931
Callaghan 2011	38	SF-12	retest	RCT	yes	53.5	female	.832	.832
Bennett 2007	56	SF-36	alpha	RCT	yes	58	both	.950	.961
Buessing 2009	388	SF-12	alpha	other	yes	60	female	.909	.922
Buys 2011	103	SF-36	alpha	other	yes	28.7	both	.905	.905
Vancampfort 2011	297	SF-12	alpha	other	yes	61.4	female	.831	.856
Rombaut 2010	64	SF-36	alpha	other	yes	38	female	.960	.920
Midtgaard 2006	55	SF-36	alpha	other	yes	42	both	.864	.879
Mueller 2009	57	SF-36	alpha	other	yes	NG	NG	.830	.869
Huisinga 2011	26	SF-36	alpha	other	yes	45.5	both	.922	.922
Aoyagi 2010	183	SF-36	alpha	other	no	73	both	.872	.922
Fassett 2009	120	SF-36	alpha	other	yes	60	both	.909	.884
Kerse 2008	193	SF-36	alpha	RCT	yes	NG	both	.949	.834
Logsdon 2009	37	SF-36	alpha	other	no	81.9	both	.854	.872
Ogilvie 2008	1322	SF-8	alpha	other	no	48	both	.930	.940
Van Uffelen 2007	152	SF-36	alpha	RCT	yes	75	both	.819	.793
Krousel-Wood 2008	76	SF-36	alpha	RCT	yes	56.6	both	.921	NA
Stroud 2009	121	SF-36	alpha	other	yes	50	both	.863	.871
Lund 2011	86	SF-36	alpha	RCT	yes	77	both	.906	.906
Lawton 2008	1089	SF-36	alpha	RCT	no	58.9	female	.869	.892
Poulin 2007	110	SF-36	alpha	other	yes	35.8	both	.900	.872
Park 2008	14	SF-36	alpha	other	no	72	both	.893	.815
Brandes 2011	53	SF-36	alpha	other	yes	65.8	both	.835	.870
MacMillan 2011	41	SF-36	alpha	RCT	no	NG	both	.815	.898

Note. NA indicates not applicable. NG indicates not given.

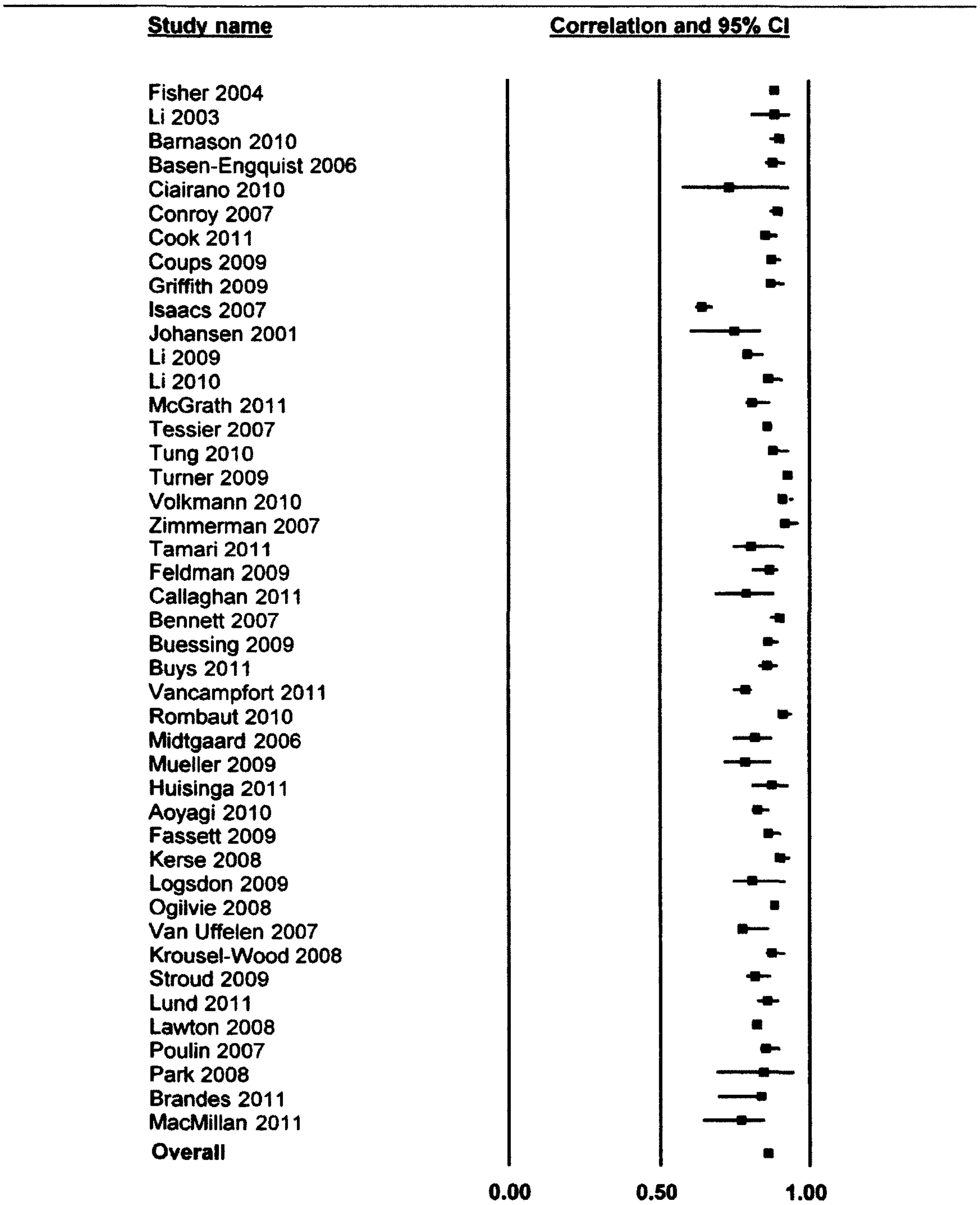


Figure 1

Forest plot of physical domain reliability effect sizes and 95% confidence intervals

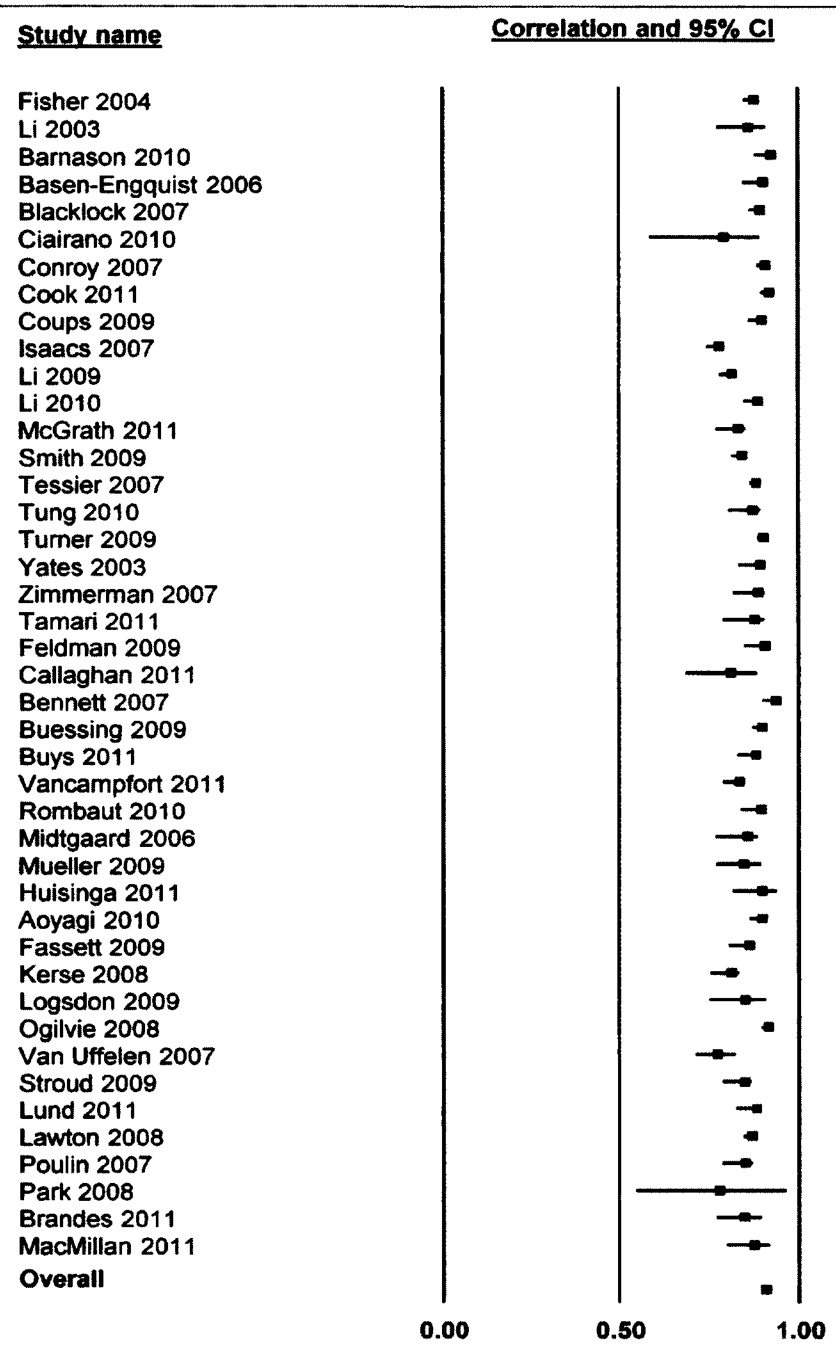


Figure 2

Forest plot of mental domain reliability effect sizes and 95% confidence intervals

Mental health domain

Figure 2 shows the effect sizes and 95% confidence intervals for the mental health domain across each study. The weighted mean effect size from the random effects model was strong and significantly different from zero, ES = .90 (95% CI: .89, .91), $p < .001$. Therefore, the mental health scales of the SF-36 assessment are also reliable across a various physical activity studies. The test of homogeneity was also significant for the mental health domain, $Q = 529.90$, $df = 42$, $p < .001$, $I^2 = 92.07$. This also shows that a moderator analysis should be performed. Table 3 shows results of the moderator analysis for the mental health studies. Results showed that studies of various gender (female or both sexes) did not account for significant variance in effect sizes, $Q = 0.03$, $df = 1$, $p = .864$. Studies using different designs (RCT or other types) did not account for significant variance in effect sizes, $Q = 1.03$, $df = 1$, $p = .310$. Also, studies using participants of different health status (diseased or non-diseased) did not account for significant variance in effect sizes, $Q = 0.34$, $df = 1$, $p = .557$. However, studies using different forms of the SF-36 (SF-36, SF-12 or SF-8) did account for significant variance in effect sizes, $Q = 46.89$, $df = 2$, $p < .001$. As well, studies publishing different reliability coefficients (alpha or retest) accounted for significant variance in effect sizes, $Q = 77.39$, $df = 1$, $p < .001$. Finally, mean age of study participants was also a significant predictor ($b = -.006$) of effect size, $Q = 58.04$, $df = 1$, $p < .001$. Figure 4 displays the relationship between age and the transformed effect sizes.

Table 2

Effect size by moderator for the SF-36 physical health studies

Moderator	*N*	*M ES*	*LL*	*UL*	*Q*	*df*	*p*
Gender					1.77	1	.184
Female	9	.92	.86	.98			
Both	34	.90	.86	.94			
Design					0.02	1	.884
RCT	16	.90	.86	.93			
Other	28	.90	.87	.93			
Disease					0.77	1	.379
No	14	.89	.87	.91			
Yes	29	.91	.87	.94			
Form					12.82	2	.002
SF-36	38	.90	.87	.93			
SF-12	5	.88	.84	.91			
SF-8	1	.93	.92	.94			
Reliability					8.00	1	.005
Alpha	41	.91	.88	.92			
Retest	2	.75	.55	.87			
Age	39	-.003	-.004	-.001	9.59	1	.002

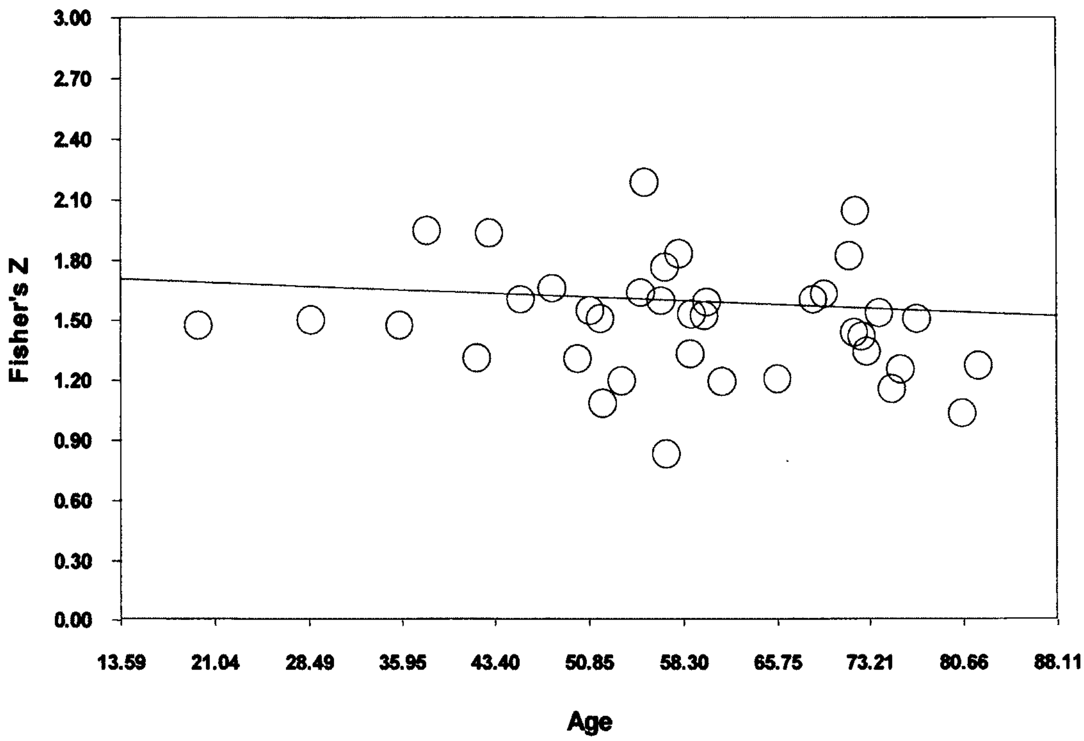
Note. M ES (95% CI) for age is a regression coefficient.

Table 3

Effect size by moderator for the SF-36 mental health studies

Moderator	*N*	*M ES*	*LL*	*UL*	*Q*	*df*	*p*
Gender					0.03	1	.864
Female	9	.90	.88	.92			
Both	33	.90	.88	.91			
Design					1.03	1	.310
RCT	14	.89	.86	.91			
Other	29	.90	.89	.91			
Disease					0.34	1	.557
No	15	.90	.89	.92			
Yes	27	.90	.88	.91			
Form					46.89	2	< .001
SF-36	37	.90	.88	.91			
SF-12	5	.89	.85	.92			
SF-8	1	.94	.93	.95			
Reliability					77.39	1	< .001
Alpha	41	.90	.89	.91			
Retest	2	.80	.78	.82			
Age	37	-.006	-.008	-.005	58.04	1	< .001

Note. M ES (95% CI) for age is a regression coefficient.



Note. Age is a significant (p = .002) moderator of physical domain effect sizes. Effect sizes were transformed to Fisher's Z.

Figure 3

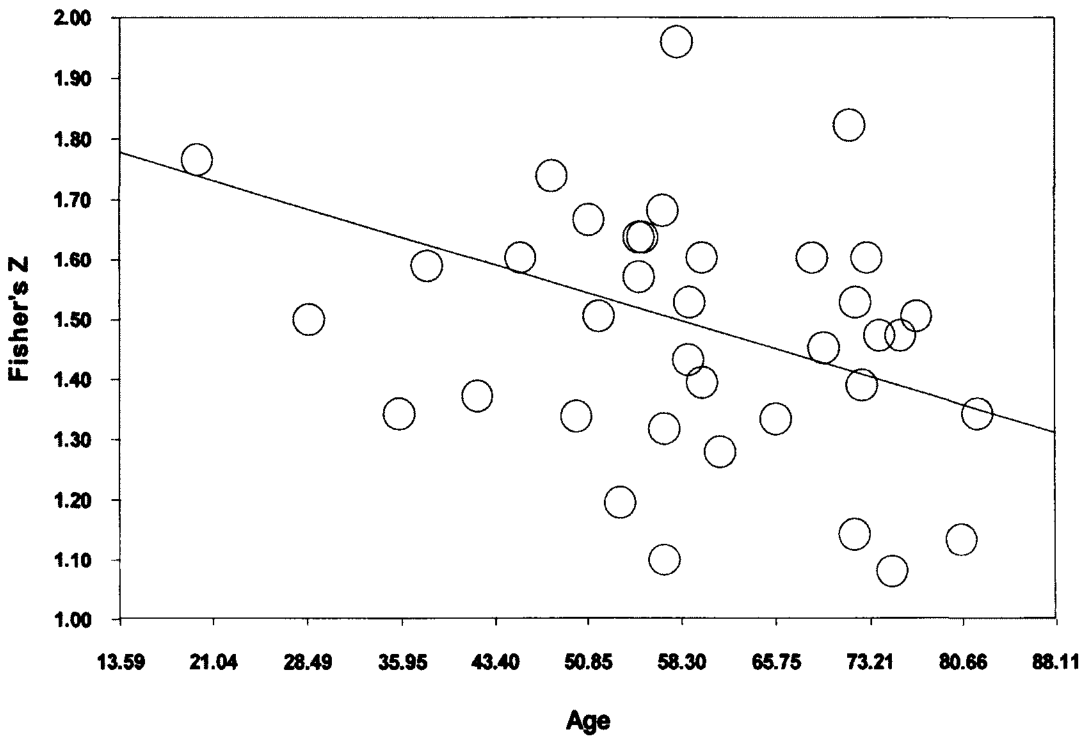
Weighted least squares regression of physical domain effect size regressed on age

Discussion

The primary purpose of this study was to perform a systematic review and meta-analysis of reliability coefficients from the physical and mental health scales of the SF-36 HRQOL assessment applied to physical activity research. The results indicated that both SF-36 HRQOL domains are highly reliable in physical activity research. The secondary purpose of this study was to examine potential moderators which may account for extra variance associated with the HRQOL measurements. Results of the moderator analysis showed that the different forms, the different types of reliability, and mean age significantly contributed to effect size variability.

The SF-36 was more reliable when used in its compact SF-8 form compared to its SF-12 or SF-36 form. This result is surprising since, generally, scales longer in length provide greater reliability coefficients (Allen & Yen, 1979). It is suggested that future research investigate the reliability comparison between the SF-8 and SF-12/SF-36 forms in physical activity research. The implications for finding the shorter form more reliable than the two longer forms are great when many physical activity studies are already burdened with many stages of intervention procedures. Caution should be taken; however, when interpreting the moderator analysis by form, since only one study was included using the SF-8 form.

More reliable scales were also seen when internal consistency reliability was the coefficient as compared to a test-retest coefficient. A lower reliability coefficient for test-retest situations compared to internal reliability seems appropriate since measures of stability are heavily affected by participant's carry-over effects such as memory, mood,



Note. Age is a significant ($p < .001$) moderator of mental domain effect sizes. Effect sizes were transformed to Fisher's Z.

Figure 4

Weighted least squares regression of mental domain effect size regressed on age or actual changes in the traits as well as length of time between administrations (Allen & Yen, 1979). Finally, study mean age was indirectly related to reliability for both scales, with the relationship stronger for the mental scales than for the physical scales. This observation could be explained by the fact that older research participants have more problems interpreting the SF-36 items as well as more problems responding and completing all items of the assessment (Sullivan, Karlsson, & Ware, 1995). The reasoning behind the stronger age effect in the mental health scales; however, is not as clear. It is possible that perceptions of mental health is not as clear to older individuals as is their perceptions of physical health.

Results of the moderator analyses showed that studies with females only did not have significantly different reliability of SF-36 scales from studies using both males and females. Results showed that studies utilizing RCTs did not have significantly different reliability from studies using other designs. Also, studies using the SF-36 to assess subjects with chronic disease did not have significantly different reliability from studies using non-diseased subjects. The non-significant moderators provide evidence of the robustness of the SF-36 reliability across gender, design, and disease.

This study has some limitations. First, the results of this study may have been affected by publication bias. The fact that the reliability coefficients were published in peer-reviewed journals makes it possible that studies reporting lower reliabilities for the SF-36 were not accepted by reviewers and therefore biased the results toward higher reliability estimates. It is also possible that researchers using the SF-36 in physical activity research, who computed low reliability measures, removed the SF-36 as an outcome variable from their study. Another limitation in this study is the fact that there were no male only studies used in the moderator analysis. Therefore, we are uncertain whether the reliability of the SF-36 is different for males compared to females or different from both male and female studies. The reason reliability coefficients from male only studies were not analyzed in this study was not because there were no physical activity articles using the SF-36 assessment on males only; but rather, there were no studies of this kind reporting the reliability. This bias should be recognized and future research is needed to assess the specific reliability generalization to the male population in physical activity research.

A strength of this study is that it was limited to the collection of reliability coefficients coming from publications on physical activity research. This type of focus allows for a more precise generalization of reliability. To know that the SF-36 assessment was on average reliable across a broad area of research topics would indeed be valuable; it is perhaps more valuable for the physical activity researchers to know that it performs reliably across a large span of physical activity research areas. Another strength of this study was its ability to separately analyze reliability in the physical and mental health domains. Many physical activity researchers use both physical and mental health components in their research (Li et al., 2003), while others use only the physical component (Griffith et al., 2009) or the mental component (Yates et al., 2003). From a measurement perspective, it is more valuable to research the reliability of a set of scores in the way in which they are used in research. Finding that the SF-36 on average produces scores that are reliable for both domains in physical activity research provides more psychometric information than if reliability was only generalized to the SF-36 as a whole.

In conclusion, the evidence from this study overwhelmingly supports the use of the SF-36 HRQOL assessment in physical activity research. Both physical and mental health domains maintain very strong reliability across studies of different gender, different research design, and different diseases states. The SF-8 may provide slightly greater reliability compared to its longer counterparts. Measures of internal consistency provide greater reliability coefficients as compared to measures of stability. Finally, reliability may slightly decrease among the aging population.

Chapter III References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, Calif.: Brooks/Cole Pub. Co.

Aoyagi, Y., Park, H., Park, S., & Shephard, R. J. (2010). Habitual physical activity and health-related quality of life in older adults: interactions between the amount and intensity of activity (the Nakanojo Study). *Quality of Life Research, 19*(3), 333-338.

Barnason, S., Zimmerman, L., Schulz, P., & Tu, C. (2009). Influence of an early recovery telehealth intervention on physical activity and functioning after coronary artery bypass surgery among older adults with high disease burden. *Heart & Lung, 38*(6), 459-468.

Basen-Engquist, K., Taylor, C. L. C., Rosenblum, C., Smith, M. A., Shinn, E. H., Greisinger, A., Rivera, E. (2006). Randomized pilot test of a lifestyle physical activity intervention for breast cancer survivors. *Patient Education and Counseling, 64*(1-3), 225-234.

Bennett, J. A., Lyons, K. S., Winters-Stone, K., Nail, L. M., & Scherer, J. (2007). Motivational interviewing to increase physical activity in long-term cancer survivors - A randomized controlled trial. *Nursing Research, 56*(1), 18-27.

Blacklock, R. E., Rhodes, R. E., & Brown, S. G. (2007). Relationship between regular walking, physical activity, and health-related quality of life. *Journal of physical activity & health, 4*(2), 138-152.

Brandes, M., Ringling, M., Winter, C., Hillmann, A., & Rosenbaum, D. (2011). Changes in Physical Activity and Health-Related Quality of Life During the First Year After Total Knee Arthroplasty. *Arthritis Care & Research, 63*(3), 328-334.

Buessing, A., Fischer, J., Fouladbakhsh, J. M., Ostermann, T., & Matthiessen, P. F. (2009). Perceived and expected effects of physical activity in female cancer survivors. *European Journal of Integrative Medicine, 1*(1), 9-17.

Buys, R., Budts, W., Delecluse, C., & Vanhees, L. (2011). Exercise Capacity, Physical Activity, and Obesity in Adults With Repaired Aortic Coarctation. *Journal of Cardiovascular Nursing*. In Print.

Callaghan, P., Khalil, E., Morres, I., & Carter, T. (2011). Pragmatic randomised controlled trial of preferred intensity exercise in women living with depression. *BMC Public Health, 12*(11), 465.

Ciairano, S., Liubicich, M. E., & Rabaglietti, E. (2010). The effects of a physical activity programme on the psychological wellbeing of older people in a residential care facility: an experimental study. *Ageing & Society, 30*, 609-626.

Conroy, M. B., Simkin-Silverman, L. R., Pettee, K. K., Hess, R., Kuller, L. H., & Kriska, A. M. (2007). Lapses and psychosocial factors related to physical activity in early postmenopause. *Medicine and Science in Sports and Exercise, 39*(10), 1858-1866.

Cook, B., Hausenblas, H., Tuccitto, D., & Giacobbi, P. R., Jr. (2011). Eating Disorders and Exercise: A Structural Equation Modelling Analysis of a Conceptual Model. *European Eating Disorders Review, 19*(3), 216-225.

Coups, E. J., Park, B. J., Feinstein, M. B., Steingart, R. M., Egleston, B. L., Wilson, D. J., & Ostroff, J. S. (2009). Physical activity among lung cancer survivors: changes across the cancer trajectory and associations with quality of life. *Cancer Epidemiology Biomarkers & Prevention, 18*(2), 664-672.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Dominick, K. L., Ahern, F. M., Gold, C. H., & Heller, D. A. (2004). Health-related quality of life among older adults with arthritis. *Health and Quality of Life Outcomes,13*(2), 2-5

Fassett, R. G., Robertson, I. K., Geraghty, D. P., Ball, M. J., Burton, N. W., & Coombes, J. S. (2009). Physical Activity Levels in Patients with Chronic Kidney Disease Entering the LORD Trial. *Medicine and Science in Sports and Exercise, 41*(5), 985-991.

Feldman, L. S., Kaneva, P., Demyttenaere, S., Carli, F., Fried, G. M., & Mayo, N. E. (2009). Validation of a physical activity questionnaire (CHAMPS) as an indicator of postoperative recovery after laparoscopic cholecystectomy. *Surgery, 146*(1), 31-39.

Fisher, K. J., & Li, F. Z. (2004). A community-based walking trial to improve neighborhood quality of life in older adults: A multilevel analysis. *Annals of Behavioral Medicine, 28*(3), 186-194.

Gliner, J. A., Morgan, G. A., & Harmon, R. J. (2001). Measurement reliability. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(4), 486-488.

Griffith, K., Wenzel, J., Shang, J., Thompson, C., Stewart, K., & Mock, V. (2009). Impact of a Walking Intervention on Cardiorespiratory Fitness, Self-Reported Physical Function, and Pain in Patients Undergoing Treatment for Solid Tumors. *Cancer, 115*(20), 4874-4884.

Hennessy, C. H., Moriarty, D. G., Zack, M. M., Scherr, P. A., & Brackbill, R. (1994). Measuring health-related quality of life for public health surveillance. *Public Health Reports, 109*(5), 665-672.

Huisinga, J. M., Filipi, M. L., & Stergiou, N. (2011). Elliptical exercise improves fatigue ratings and quality of life in patients with multiple sclerosis. *Journal of Rehabilitation Research and Development, 48*(7), 881-889.

Isaacs, A. J., Critchley, J. A., Tai, S. S., Buckingham, K., Westley, D., Harridge, S. D. R., . . . Gottlieb, J. M. (2007). Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only. *Health Technology Assessment, 11*(10), 1-165.

Johansen, K. L., Painter, P., Kent-Braun, J. A., Ng, A. V., Carey, S., Da Silva, M., & Chertow, G. M. (2001). Validation of questionnaires to estimate physical activity and functioning in end-stage renal disease. *Kidney International, 59*(3), 1121-1127.

Kerse, N., Falloon, K., Moyes, S. A., Hayman, K. J., Dowell, T., Kolt, G. S., & Arroll, B. (2008). DeLLITE depression in late life: an intervention trial of exercise. Design and recruitment of a randomised controlled trial. *BMC Geriatrics, 24*(8), 12-18.

Kline, T. (2005). *Psychological testing : a practical approach to design and evaluation*. Thousand Oaks, Calif.: Sage Publications.

Krousel-Wood, M. A., Berger, L., Jiang, X., Blonde, L., Myers, L., & Webber, L. (2008). Does home-based exercise improve body mass index in patients with type 2 diabetes? Results of a feasibility trial. *Diabetes Research and Clinical Practice, 79*(2), 230-236.

Lawton, B. A., Rose, S. B., Elley, C. R., Dowell, A. C., Fenton, A., & Moyes, S. A. (2008). Exercise on prescription for women aged 40-74 recruited through primary care: two year randomised controlled trial. *British Medical Journal, 337*.

Li, F. Z., Harmer, P., Wilson, N. L., & Fisher, K. J. (2003). Health benefits of cobblestone-mat walking: Preliminary findings. *Journal of Aging and Physical Activity, 11*(4), 487-501.

Li, G. S. F., Lu, F. J. H., & Wang, A. H. H. (2009). Exploring the relationships of physical activity, emotional intelligence and health in Taiwan college students. *Journal of Exercise Science & Fitness, 7*(1), 55-63.

Li, M., Li, L., & Fan, X. (2010). Patients having haemodialysis: physical activity and associated factors. *Journal of Advanced Nursing, 66*(6), 1338-1345.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, Calif.: Sage Publications.

Logsdon, R. G., McCurry, S. M., Pike, K. C., & Teri, L. (2009). Making Physical Activity Accessible to Older Adults With Memory Loss: A Feasibility Study. *Gerontologist, 49*. doi: 10.1093/geront/gnp082

Lund, A., Michelet, M., Sandvik, L., Wyller, T. B., & Sveen, U. (2011). A lifestyle intervention as supplement to a physical activity programme in rehabilitation after stroke: a randomized controlled trial. *Clinical Rehabilitation*. doi: 10.1177/0269215511429473

MacMillan, F., Fitzsimons, C., Black, K., Granat, M. H., Grant, M. P., Grealy, M., . . . Mutrie, N. (2011). West End Walkers 65+: A randomised controlled trial of a primary care-based walking intervention for older adults: Study rationale and design. *BMC Public Health, 11*. doi: 120 10.1186/1471-2458-11-120

McGrath, J. A., O'Malley, M., & Hendrix, T. J. (2011). Group exercise mode and health-related quality of life among healthy adults. *Journal of Advanced Nursing, 67*(3), 491-500.

Midtgaard, J., Rorth, M., Stelter, R., & Adamsen, L. (2006). The group matters: an explorative study of group cohesion and quality of life in cancer patients participating in physical exercise intervention during treatment. *European Journal of Cancer Care, 15*(1), 25-33.

Moriarty, D. G., Zack, M. M., & Kobau, R. (2003). The Centers for Disease Control and Prevention's Healthy Days Measures - population tracking of perceived physical and mental health over time. *Health and Quality of Life Outcomes, 1*, 1-37

Mueller, J., Christov, F., Schreiber, C., Hess, J., & Hager, A. (2009). Exercise capacity, quality of life, and daily activity in the long-term follow-up of patients with univentricular heart and total cavopulmonary connection. *European Heart Journal, 30*(23), 2915-2920.

Ogilvie, D., Mitchell, R., Mutrie, N., Petticrew, M., & Platt, S. (2008). Personal and environmental correlates of active travel and physical activity in a deprived urban population. *International Journal of Behavioral Nutrition and Physical Activity, 5*. doi: 43 10.1186/1479-5868-5-43

Park, S. A., Shoemaker, C., & Haub, M. (2008). Can Older Gardeners Meet the Physical Activity Recommendation through Gardening? *HortTechnology, 18*(4), 639-643.

Poulin, M.-J., Chaput, J.-P., Simard, V., Vincent, P., Bernier, J., Gauthier, Y., Tremblay, A. (2007). Management of anti psychotic-induced weight gain: a prospective naturalistic study of the effectiveness of a supervised exercise program. *Australian and New Zealand Journal of Psychiatry, 41*(12), 980-989.

QualityMetric. (2011). *SF Health Surveys*. Retrieved August 15, 2011, from http://www.qualitymetric.com/WhatWeDo/GenericHealthSurveys/tabid/184/Default.aspx

Ragan, B. G., & Kang, M. (2005). Reliability: Current Issues and Concerns.. *Athletic Therapy Today, 10*(6), 30-33.

Rombaut, L., Malfait, F., Cools, A., De Paepe, A., & Calders, P. (2010). Musculoskeletal complaints, physical activity and health-related quality of life among patients with the Ehlers-Danlos syndrome hypermobility type. *Disability and Rehabilitation, 32*(16), 1339-1345.

United States Department of Health & Human Services. (2011). *Healthy people 2020*, from http://www.healthypeople.gov/2020/topicsobjectives2020/default.aspx

Smith, A. W., Alfano, C. M., Reeve, B. B., Irwin, M. L., Bernstein, L., Baumgartner, K., Ballard-Barbash, R. (2009). Race/ethnicity, physical activity, and quality of life in breast cancer survivors. *Cancer Epidemiology & Biomarkers Prevention, 18*(2), 656-663.

Stroud, N. M., & Minahan, C. L. (2009). The impact of regular physical activity on fatigue, depression and quality of life in persons with multiple sclerosis. *Health and Quality of Life Outcomes, 7*. doi: 68 10.1186/1477-7525-7-68.

Sullivan, M., Karlsson, J., & Ware, J. E., Jr. (1995). The Swedish SF-36 Health Survey--I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Social Science Medicine, 41*(10), 1349-1358.

Tamari, K. (2011). Self-reported Home Exercise and Younger Age Predict Improved Health-related Quality of Life among Community-Dwelling Elderly Participants in a Three-month Group Exercise Class. *Journal of Physiological Anthropology, 30*(2), 77-85.

Tessier, S., Vuillemin, A., Bertrais, S., Boini, S., Le Bihan, E., Oppert, J. M., . . . Briancon, S. (2007). Association between leisure-time physical activity and health-related quality of life changes over time. *Preventive Medicine, 44*(3), 202-208.

Tung, H.-H., Chen, Y.-C., Wei, J., Liu, C.-Y., Chang, C.-Y., & Wang, T.-J. (2010). Leisure physical activity and quality of life after coronary artery bypass graft surgery for patients with metabolic syndrome in Taiwan. *Heart & Lung, 39*(5), 410-420.

Turner, A. P., Kivlahan, D. R., & Haselkorn, J. K. (2009). Exercise and Quality of Life Among People With Multiple Sclerosis: Looking Beyond Physical Functioning to Mental Health and Participation in Life. *Archives of Physical Medicine and Rehabilitation, 90*(3), 420-428.

United States Deptartment of Health and Human Services. Physical Activity guidelines Advisory Committee., & United States Deptartment of Health and Human Services. (2008). Physical Activity Guidelines Advisory Committee report, 2008 to the Secretary of Health and Human Services Retrieved from http://www.health.gov/paguidelines/committeereport.aspx

United States Public Health Service., Office of the Surgeon General., National Center for Chronic Disease Prevention and Health Promotion (U.S.), & President's Council on Physical Fitness and Sports (U.S.). (1996). *Physical activity and health : a report of the Surgeon General*. Atlanta, Ga. Washington, D.C. Pittsburgh, PA: U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention President's Council on Physical Fitness and Sports.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*(1), 6-20.

van Uffelen, J. G. Z., Paw, M. J. M. C. A., Hopman-Rock, M., & van Mechelen, W. (2007). The effect of walking and vitamin B supplementation on quality of life in community-dwelling adults with mild cognitive impairment: a randomized, controlled trial. *Quality of Life Research, 16*(7), 1137-1146.

Vancampfort, D., Probst, M., Scheewe, T., Maurissen, K., Sweers, K., Knapen, J., & De Heft, M. (2011). Lack of physical activity during leisure time contributes to an impaired health related quality of life in patients with schizophrenia. *Schizophrenia Research, 129*(2-3), 122-127.

Volkmann, E. R., Grossman, J. M., Sahakian, L. J., Skaggs, B. J., FitzGerald, J., Ragavendra, N., . . . McMahon, M. (2010). Low Physical Activity Is Associated With Proinflammatory High-Density Lipoprotein and Increased Subclinical Atherosclerosis in Women With Systemic Lupus Erythematosus. *Arthritis Care & Research, 62*(2), 258-265.

Ware, J.E. (2004). SF-36 Health Survey Update. Retrieved August 15, 2011, from http://www.sf-36.org/announcements/Updated_SF36_bookChapter_Sept04.pdf

Ware, J.E., Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, 30(6), 473.

Yates, B. C., Price-Fowlkes, T., & Agrawal, S. (2003). Barriers and facilitators of self-reported physical activity in cardiac patients. *Research in Nursing & Health, 26*(6), 459-469.

Zimmerman, L., Barnason, S., Schulz, P., Nieveen, J., Miller, C., Hertzog, M., . . . Tu, C. (2007). The effects of a symptom management intervention on symptom evaluation, physical functioning, and physical activity for women after coronary artery bypass surgery. *Journal of Cardiovascular Nursing, 22*(6), 493-500.

CHAPTER IV

Evaluation of the Short-Form Health Survey (SF-36) Using the Rasch model

Introduction

Health-related quality of life (HRQOL) is an outcome measure that has seen considerable attention in physical activity and public health research (Heath & Brown, 2009). HRQOL is a broad concept that generally includes dimensions of physical, mental, and social well being. Wilson and Cleary (1995) expand on the complexity of HRQOL by stating that HRQOL is a function of biological and physiological variables, symptom status, functional status, and general health perceptions. Because HRQOL is such an all-encompassing health measure, it has become a standard outcome variable in public health research (Centers for Disease Control and Prevention [CDC], 2000). HRQOL has also shown to be a valuable predictor of health status, predicting the number of physician visits, hospitalization events, and mortality among adults (Dominick, Ahern, Gold, & Heller, 2002).

Item response theory (IRT) is a modern approach to measurement theory. IRT works differently from classical test theory (CTT) in that it focuses on each item by examining the response of an individual at a specific ability level and the characteristics of that item (Embretson & Reise, 2000). An IRT model that is only concerned with an item's difficulty level (*b*-parameter) and the individuals' ability (θ), is considered a 1-parameter model, and commonly called a Rasch measurement model (Rasch, 1960).

Rasch analysis can be applied to health and behavioral assessments containing dichotomous response (yes/no) items, polytomous response (Likert-type) items, or a mix of both (Bond & Fox, 2007).

Given the overwhelming interest in HRQOL as a measure in physical activity research, there is a strong need for a better understanding of the measurement properties of HRQOL assessments commonly used in physical activity research. The Short Form-36 Health Survey (SF-36) is the leading HRQOL assessment used in physical activity research. The majority of physical activity researchers use either one or both of the SF-36 domain component scores (physical and mental). There are currently no studies that assess the measurement properties of these two commonly used domains using the Rasch measurement model. Therefore, the purpose of this study was to evaluate the measurement properties of each SF-36 domain using the Rasch model. The results of this study will serve as a critical evaluation of the SF-36 and possibly find needed modifications due to poor measurement properties or validate its continued use.

Methods

Participants. Data for this study came from a survey administered to adults in and around a large southeastern U.S. university community. Participants were recruited via public advertisement and announcements to local social group networks. Participants were allowed to complete the survey if they were 18 years of age or older. Human subject clearance was obtained before conducting research from the campus Institutional Review Board. Each HRQOL assessment was converted to electronic form for webbased administration and the ordering of HRQOL assessments was counterbalanced. The online survey took approximately 15 minutes to complete. Participants completed the survey during the months of January-February, 2012.

A total of 634 participants completed the SF-36 HRQOL assessment of which 72.2% were female (see Table 2). For age, 54.3% were between 18 and 24 years, 32.0% between 25 to 49 years, and 13.4% were between 50 and 78 years. For race, majority (83.4%) of the participants were White followed by Black (9.1%). Of the participants, 3.2% reported having only a high school education or less, 50.0% reported having some college education and 46.5% reported having completed a college degree. Finally, 60.3% of participants reported being single, 18.3% reported being married, 13.0% reported being either separated or divorced, 6.7% reported living with a partner, and 1.4% reported being widowed.

HRQOL assessments. The *Short-Form Health Survey (SF-36)* is one of the most widely used HRQOL instruments in physical activity research. The SF-36 was developed from the Medical Outcomes Study (MOS) conducted by RAND (Ware & Sherbourne, 1992). The SF-36 is a multi-dimensional scale consisting of 36 items, 8 health-related dimensions, and two domains (see Table 1). The dimensions include: 1) vitality, 2) physical functioning, 3) bodily pain, 4) general health, 5) physical role functioning, 6) emotional role functioning, 7) social role functioning, and 8) mental health. The physical domain consists of the physical functioning, bodily pain, general health, and physical role

Table 1

Characteristics of participants completing the SF-36 HRQOL assessment (N = 634)

Characteristic	*N*	%
Gender		
Male	175	27.6
Female	458	72.2
Age Group		
18-24	344	54.3
25-49	203	32.0
50-78	85	13.4
Race		
White	529	83.4
Black	58	9.1
Hispanic	8	1.3
Asian	12	1.9
Other	24	3.8
Education		
High School or less	20	3.2
Some College	317	50.0
College Degree	295	46.5
Marital Status		
Single	382	60.3
Married	116	18.3
Separated/Divorced	82	13.0
Widowed	9	1.4
Living w/Partner	42	6.7

functioning dimensions and the mental domain consists of the vitality, emotional role functioning, social role functioning, and mental health dimensions (Ware, 2004).

The SF-36 is intended to measure HRQOL in adults and can be self-administered, administered via computer, with aid of an interviewer, or by telephone. The instrument can be modified to include either a (standard) 4-week recall or a 1-week recall and has been incorporated into both observational as well as intervention-type studies. The *SF-12* is a shorter version of the original that maintains the measurement of all 8 dimensions as well as the two domain-specific summary scores (QualityMetric, 2011).

Table 2

Number of Items by Domain of the SF-36 and CDC HRQOL-4 assessment tools

SF-36		CDC HRQOL-4	
Domain	Items	Domain	Items
Physical Health	21	Physical Health	3
Mental Health	14	Mental Health	1
		Healthy Days Index	1
Total Items	35		5

Note. Healthy Days Index is a composite variable from items 2 and 3 of the HRQOL-4 core.

The *CDC Healthy Days* module (HRQOL-9) is a widely used module in national surveillance systems such as the National Health and Nutrition Examination Survey (NHANES) and the Behavioral Risk Factor Surveillance System (BRFSS). The HRQOL-4 is a simple 4-item tool for assessing HRQOL in large scale studies and is considered the core (see Table 1). The first item assesses perceived general health and asks respondents to rate their health in general on a 5-point categorical scale ranging from *excellent* to *poor*. The last three items ask for the number of days out of the previous 30 in which (1) your physical health was poor, (2) your mental health was poor, and (3) you were unable to engage in usual activities due to poor health. A fifth summary measure of healthy days (or unhealthy days) can be computed by summing the physical and mental items and creating a ceiling at 30 days (CDC, 2000). For the current study, the Healthy Days Index was used for a validation of the SF-36 ability scores. The Healthy Days Index contains two items which combined represent both domains of HRQOL (Mielenz, Jackson, Currey, DeVellis, & Callahan, 2006).

Rasch model. The Rasch model is a probability model which includes a persons' ability and an item's difficulty as parameters. The Rasch model converts responses from a rating scale to a new scale with interval level measurement properties (Bond & Fox, 2007). The new scale values are called logits (log odds) and are so for a persons' ability (θ) as well as an item's difficulty (b). Logits take the same presence as Z-scores, with a mean of zero. A person with a positive logit generally has a greater "ability" concerning the trait being measured (i.e., has a higher overall HRQOL) and a person with a negative logit generally has a lower ability concerning the trait. An item with a positive logit generally indicates higher item "difficulty" and an item with a negative logit generally indicates lower item difficulty (Bond & Fox, 2007). A larger item difficulty indicates that individuals are less likely to endorse that item.

The primary assumption of the Rasch model is that the measurement scale should be unidimensional. For this study, this means that each scale should measure its respective HRQOL domain and nothing more. This assumption can be examined by examining item fit statistics. Once data are fit to the Rasch model and the assumption of unidimensionality is met, a researcher can proceed in inspecting several of the Rasch model statistics. Person reliability estimates and item reliability estimates are reported from a Rasch analysis and provide analogous information as that of Cronbach's alpha, with a range of 0 to 1.00. Person separation and item separation indices are standard error units representing the spread or separation of persons (or items) on the ability scale, where a larger value indicates the scale's ability to better separate persons (or items). The Infit and Outfit statistics from a Rasch analysis are mean square statistics with expected values of one and an acceptable range of 0.50 to 1.50 (Wright & Linacre, 1994). Item Infit and Outfit statistics provide evidence of construct validity. Person Infit and outfit statistics represent whether individuals respond in an expected way given their response pattern and item difficulty (Bond & Fox, 2007).

Proper category functioning can also be examined by the Rasch model. Item-person map (Wright map) distributions can be examined from a Rasch analysis. The item-person map is a single dimensional graph linking item difficulty and person ability estimates on the same common scale (logits). The item-person map shows both distributions as well as the relative position of an individual's trait (i.e., HRQOL) for the items.

Data analysis. The plan was to run two separate analyses on the two HRQOL domains (physical & mental) of the SF-36 assessment. A 7-step procedure was followed to evaluate each SF-36 domain by Rasch analysis. The first step included evaluating each item for proper category functioning. The evaluation criteria included (1) regular frequency distributions, (2) average logit score measures increasing as categories increase, (3) Infit and Outfit mean square residuals are appropriate for each category, and (4) category thresholds arranged in order (Linacre, 1999; 2002b). The second step included an evaluation of model-data fit. The model-data fit criteria included inspection of the Infit and Outfit statistics for each item. If these fit statistics were greater than 1.5 or less than 0.5, the item was considered misfit (Lunz, 1990) and were subsequently discarded. The third step included an inspection of the item-person map. This step evaluates how evenly spread the items are relative to the participants in terms of the HRQOL trait. The fourth step included the evaluation of each item in terms of item difficulty parameters, item separation, and item separation reliability. The fifth step involved the evaluation of individuals fitting the Rasch model in terms of person ability (θ) fit, person separation index, and person separation reliability. The sixth step included convergent validity evidence for the SF-36 domains by computation of bivariate correlations between each of the SF-36 ability (θ) scores and the CDC Healthy Days

Index from the HRQOL-4 core. The seventh and final step included construct validity evidence for the SF-36 ability (θ) scores by showing differences in the scores between groups of participants with known theoretical differences in HRQOL. The grouping variables were all dichotomized (yes/no) and included obesity, smoking status, chronic illness, vigorous activity participation, moderate activity participation, strength training participation, hypertension, high cholesterol, and diabetes. All analyses were carried out using SAS version 9.3 and Winsteps v3.65 (Linacre, 2006).

Results

Table 3 displays distribution information for the SF-36 physical domain items. Twenty one items had responses across all categories. Relative frequencies per category ranged from .005 to .959 across all items in the physical domain. Ten items have categorical rating scales consisting of 3 points, another ten items have a 5-point scale, and one item is a 6-point scale. The overall average response across all 21 items was 3.55, ranging from 2.45 to 4.74. All items were coded to reflect greater HRQOL with higher scores. Table 4 displays item distribution information for the SF-36 mental domain. Relative frequencies per category ranged from .008 to .549 across all items in the mental domain. All 14 items were on a 5-point scale and each had responses across all categories. The overall average response across all 14 items was 3.76, ranging from 2.91 to 4.28. Each item in the mental domain was also coded to reflect greater HRQOL with higher scores.

Table 3

Item category distributions (%) and item means of the physical health domain of the SF-36 HRQOL assessment (N = 634)

		Item Categories					
Item	*M*	1	2	3	4	5	6
SF1	3.56	0.5	9.1	36.4	41.3	12.6	*
SF3a	2.45	10.7	33.4	55.8	*	*	*
SF3b	2.87	2.2	8.2	89.6	*	*	*
SF3c	2.91	1.4	6.2	92.4	*	*	*
SF3d	2.71	4.6	19.9	75.6	*	*	*
SF3e	2.90	1.6	7.3	91.2	*	*	*
SF3f	2.78	3.5	14.7	81.9	*	*	*
SF3g	2.83	3.6	10.1	86.3	*	*	*
SF3h	2.89	2.5	6.2	91.3	*	*	*
SF3i	2.91	2.7	4.1	93.2	*	*	*
SF3j	2.93	2.7	1.4	95.9	*	*	*
SF4a	4.60	0.8	1.7	6.8	18.5	72.2	*
SF4b	4.28	1.6	6.9	10.7	23.7	57.1	*
SF4c	4.60	0.9	2.5	6.2	16.1	74.3	*
SF4d	4.55	0.9	2.4	8.0	17.8	70.8	*
SF7	4.74	0.5	3.3	11.2	18.9	39.1	27.0
SF8	4.52	0.5	3.2	6.6	23.8	65.9	*
SF11a	4.12	3.5	9.6	7.7	30.3	48.9	*
SF11b	3.75	6.0	10.1	15.5	40.1	28.4	*
SF11c	4.06	1.6	7.3	19.4	26.8	45.0	*
SF11d	3.67	5.4	12.0	12.6	50.2	19.9	*

Note. Categories reflect reverse coding with higher categories representing higher HRQOL. *Represents a category which is not present for the item.

Table 4

Item category distributions (%) and item means of the mental health domain of the SF-36 HRQOL assessment (N = 634)

		Item Categories				
Item	*M*	1	2	3	4	5
SF5a	4.22	1.4	3.3	18.3	25.7	51.3
SF5b	3.96	2.1	10.3	17.7	30.0	40.1
SF5c	4.26	0.8	5.8	14.4	24.1	54.9
SF6	4.15	1.7	5.8	13.9	32.8	45.7
SF9a	3.49	2.1	13.1	28.7	46.5	9.6
SF9b	3.69	2.2	11.4	24.4	39.4	22.6
SF9c	4.28	1.1	4.6	14.4	25.6	54.4
SF9d	3.30	3.6	17.2	28.9	45.7	4.6
SF9e	3.22	4.4	17.2	36.3	36.4	5.7
SF9f	3.95	2.2	6.5	18.6	39.4	33.3
SF9g	3.22	7.3	17.0	31.4	34.9	9.5
SF9h	3.68	0.8	8.4	24.4	55.2	11.2
SF9i	2.91	10.6	22.1	36.4	27.3	3.6
SF10	4.27	0.8	4.1	15.9	25.9	53.3

Note. Categories reflect reverse coding with higher categories representing higher HRQOL.

Table 5 displays the criteria for the SF-36 physical health domain. Overall, the item categories for the physical health domain functioned well, meeting 87.0% of the total criteria used to evaluate proper functioning. A total of ten items were flagged for negative validity. Two items (4a, 8) had Outfit mean square values greater than 2.0. Seven items (3b, 3c, 3g, 3h, 3i, 11a, 11d) lacked ordered thresholds. And one item (3j) had both an Outfit mean square greater than 2.0 and disordered thresholds. Figure 1 displays the category probability curves for the SF-36 physical health domain items. The graphs depict unordered thresholds for the eight items mentioned. For the mental health domain, the item categories functioned very well, meeting 100% of the total criteria used to evaluate proper functioning. Table 6 displays the criteria for the SF-36 mental health domain and Figure 2 displays proper ordering of the category thresholds.

In terms of model-data fit, the physical health domain data did not initially fit the Rasch model well. Table 7 displays the individual fit statistics for each item (mis-fit items not shown). Seven (3a, 3b, 3c, 3e, 3j, 3i, & 11c) out of the 21 items had fit statistics out of the acceptable range (i.e., 0.5 to 1.50). After the misfit items were discarded, a 14-item physical health domain fit the Rasch model well. The mean (SD) of the Infit and Outfit statistics were 1.00 (0.20) and 0.98 (0.30), respectively. The mental health domain did initially fit the Rasch model well. The mean (SD) of the Infit and Outfit statistics were 1.00 (0.16) and 0.96 (0.17), respectively. Table 8 displays the individual fit statistics for each item in the mental health domain.

The item-person map for the adjusted physical health domain is shown in Figure 3. The leftmost vertical axis represents the logit scale where larger values signify better

HRQOL. The pound signs (#) represent the distribution of person-level HRQOL relative to the logit scale. The rightmost side of the graph represents each item relative to its difficulty. The map shows that the distribution of items, with mean (SD) of 0.0 (0.67) was not well matched to the persons' HRQOL, with mean of 2.21 (1.61). The item locations indicate that the items are not targeting people of high HRQOL (> 1.5 logits) or low HRQOL (< -1.0 logits).

The item-person map for the mental health domain is shown in Figure 4. The map shows that the distribution of items with mean (SD) of 0.0 (0.95) was matched better to the persons' HRQOL with mean of 1.46 (1.77), than the physical health domain. The item locations also indicate that the items have better coverage across persons than the physical health domain, with coverage between -1.25 to 2.00 logits.

Item difficulty values resulting from the Rasch calibration are displayed in Tables 7 and 8 for the physical and mental health domains, respectively. The larger an item's logit value is the higher the trait (HRQOL) must be for a person to endorse the item. Physical domain item difficulty ranged from -0.89 to 1.24 logits. The most difficult item was item 11d (How true is the following statement: My health is excellent.). The least difficult item was item 3h (Does your health now limit you in: Walking several hundred yards?). Mental domain item difficulty ranged from -1.13 to 1.88 logits. The most difficult item was item 9i (How much of the time during the past 4 weeks: Did you feel tired?). The least difficult item was item 10 (During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities [like visiting friends, relatives, etc.]?).

Table 5

Rating scale properties and decisions resulting from Rasch analysis of the SF-36 physical health domain (Items = 21)

Item	Response Scale	Regular Frequency Distribution	Average Advances w/Categories	Outfit MS < 2.0	Thresholds Ordered	Valid Criteria
SF1	5	Yes	Yes	Yes	Yes	4
SF3a	3	Yes	Yes	Yes	Yes	4
SF3b	3	Yes	Yes	Yes	No	3
SF3c	3	Yes	Yes	Yes	No	3
SF3d	3	Yes	Yes	Yes	Yes	4
SF3e	3	Yes	Yes	Yes	Yes	4
SF3f	3	Yes	Yes	Yes	Yes	4
SF3g	3	Yes	Yes	Yes	No	3
SF3h	3	Yes	Yes	Yes	No	3
SF3i	3	Yes	Yes	Yes	No	3
SF3j	3	Yes	Yes	No	No	2
SF4a	5	Yes	Yes	No	Yes	3
SF4b	5	Yes	Yes	Yes	Yes	4
SF4c	5	Yes	Yes	Yes	Yes	4
SF4d	5	Yes	Yes	Yes	Yes	4
SF7	6	Yes	Yes	Yes	Yes	4
SF8	5	Yes	Yes	No	Yes	3
SF11a	5	Yes	Yes	Yes	No	3
SF11b	5	Yes	Yes	Yes	Yes	4
SF11c	5	Yes	Yes	Yes	Yes	4
SF11d	5	Yes	Yes	Yes	No	3

Note. Seventy-three out of 84 (87.0%) total criteria met. Fifty out of 56 (89.3%) total criteria met after misfit items discarded.

Table 6

Rating scale properties and decisions resulting from Rasch analysis of the SF-36 mental health domain (Items = 14)

Item	Response Scale	Regular Frequency Distribution	Average Advances w/Categories	Outfit MS < 2.0	Thresholds Ordered	Valid Criteria
SF5a	5	Yes	Yes	Yes	Yes	4
SF5b	5	Yes	Yes	Yes	Yes	4
SF5c	5	Yes	Yes	Yes	Yes	4
SF6	5	Yes	Yes	Yes	Yes	4
SF9a	5	Yes	Yes	Yes	Yes	4
SF9b	5	Yes	Yes	Yes	Yes	4
SF9c	5	Yes	Yes	Yes	Yes	4
SF9d	5	Yes	Yes	Yes	Yes	4
SF9e	5	Yes	Yes	Yes	Yes	4
SF9f	5	Yes	Yes	Yes	Yes	4
SF9g	5	Yes	Yes	Yes	Yes	4
SF9h	5	Yes	Yes	Yes	Yes	4
SF9i	5	Yes	Yes	Yes	Yes	4
SF10	5	Yes	Yes	Yes	Yes	4

Note. Fifty-six out of 56 (100%) total criteria met.

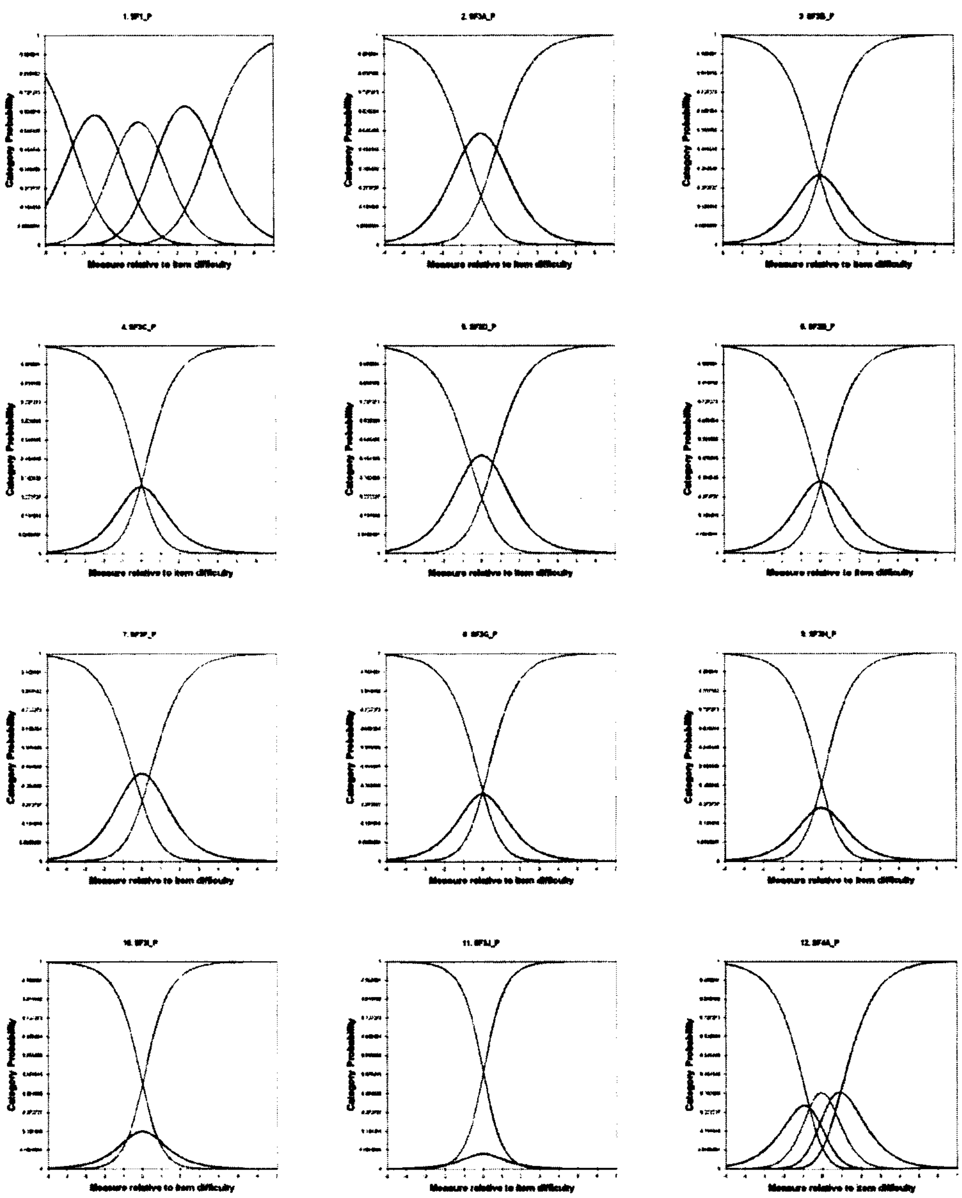


Figure 1

Category probability curves of the SF-36 HRQOL physical domain items

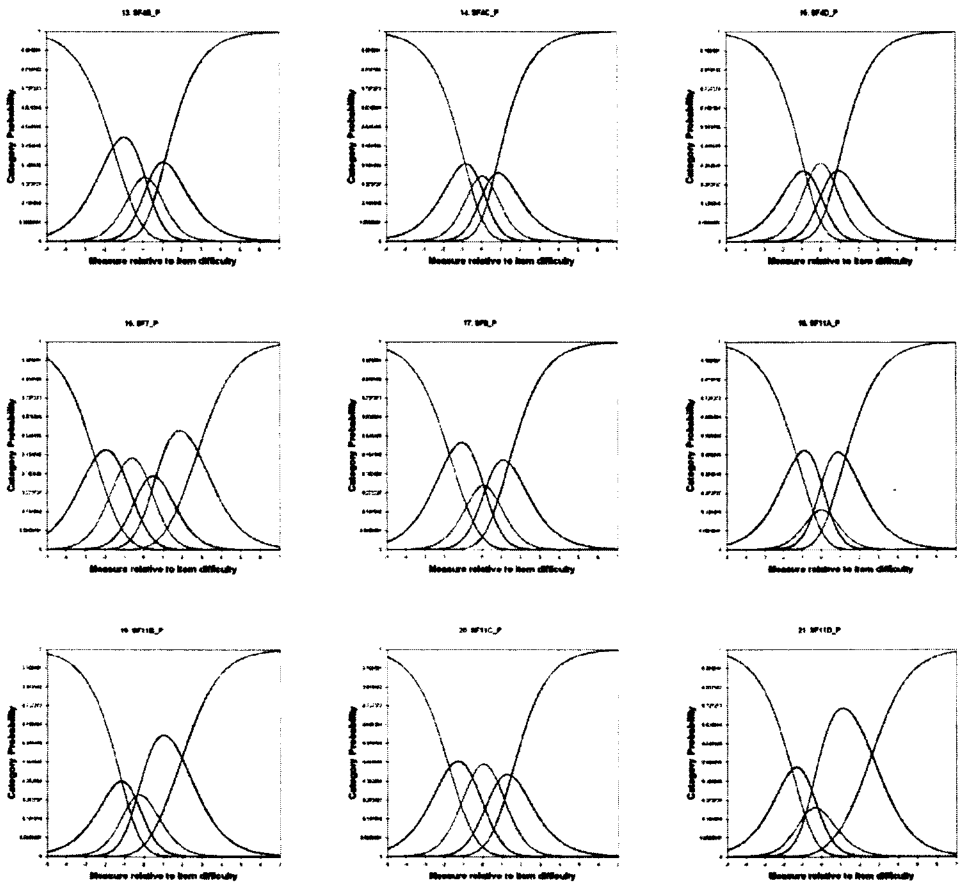


Figure 1 (continued)

Category probability curves of the of the SF-36 HRQOL physical domain items

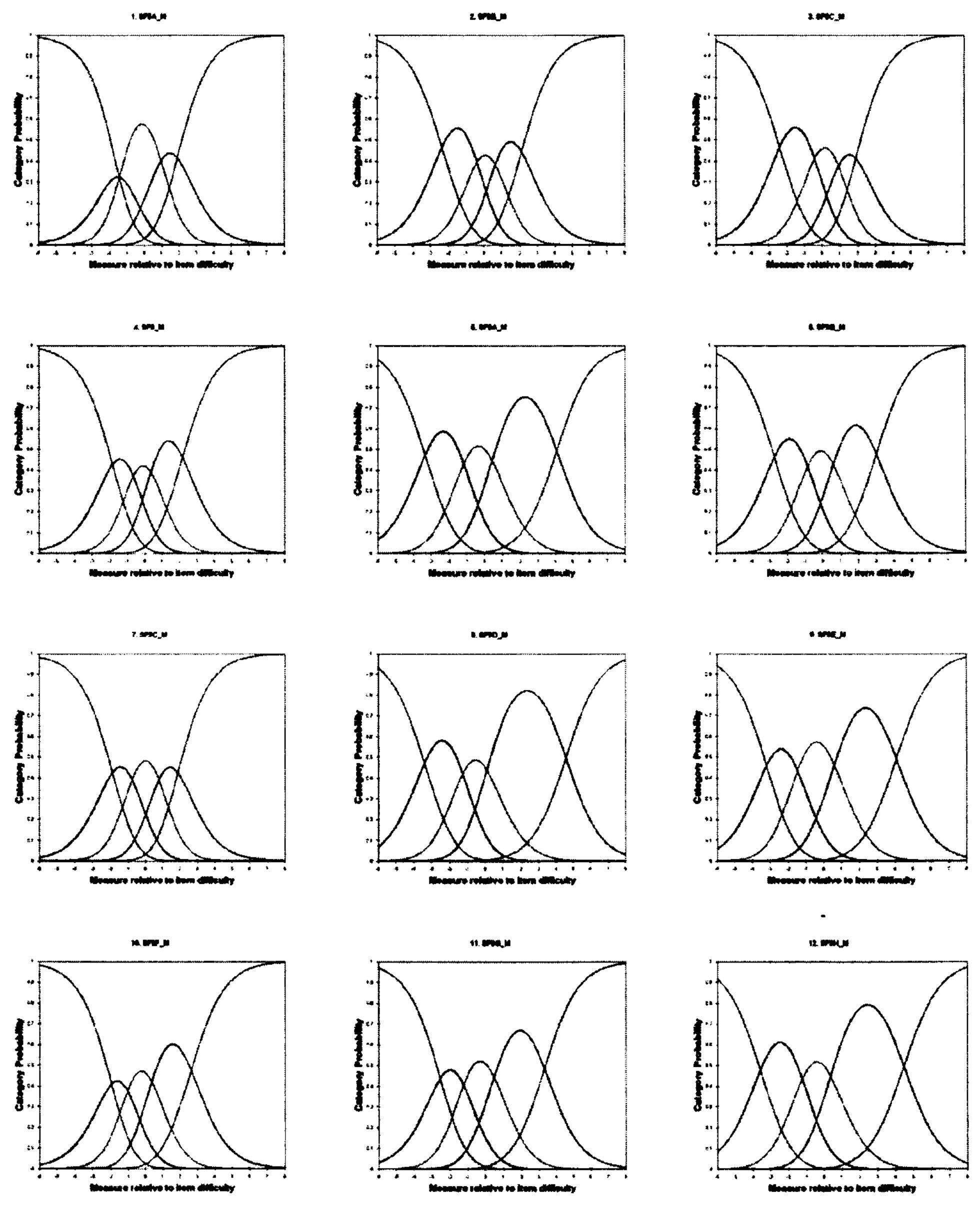


Figure 2

Category probability curves of the of the SF-36 HRQOL mental domain items

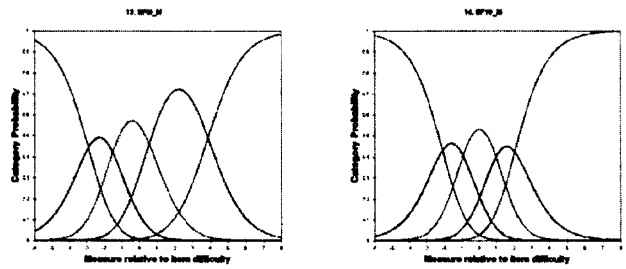


Figure 2 (continued)

Category probability curves of the of the SF-36 HRQOL Mental domain items

Table 7

Summary of Rasch calibration of the SF-36 physical health domain

Item	Response Scale	Calibration logits	SE logits	Infit MnSq	Outfit MnSq
SF1	5	0.82	0.07	0.98	0.97
SF3d	3	-0.21	0.09	0.95	0.85
SF3f	3	-0.5	0.1	1.05	1.1
SF3g	3	-0.56	0.1	0.91	0.75
SF3h	3	-0.89	0.12	1.04	1.24
SF4a	5	-0.52	0.07	0.93	0.9
SF4b	5	0.1	0.06	0.88	0.93
SF4c	5	-0.45	0.07	0.72	0.55
SF4d	5	-0.39	0.07	0.74	0.73
SF7	6	0.31	0.05	1.28	1.41
SF8	5	-0.54	0.07	0.88	0.82
SF11a	5	0.5	0.05	1.49	1.43
SF11b	5	1.1	0.05	1.18	1.15
SF11d	5	1.24	0.05	0.9	0.8

Note. Items 3a, 3b, 3c, 3e, 3j, 3i, & 11c were discarded due to misfit criteria.

Table 8

Summary of Rasch Calibration of the SF-36 mental health domain

Item	Response Scale	Calibration logits	SE logits	Infit MnSq	Outfit MnSq
SF5A	5	-0.9	0.06	1	0.91
SF5B	5	-0.39	0.06	0.91	0.82
SF5C	5	-1.1	0.06	1.07	0.88
SF6	5	-0.72	0.06	0.89	0.82
SF9A	5	0.47	0.06	0.86	0.85
SF9B	5	0.04	0.06	1.37	1.37
SF9C	5	-1.04	0.06	0.85	0.79
SF9D	5	1.06	0.06	0.92	0.95
SF9E	5	1.15	0.06	1.07	1.08
SF9F	5	-0.36	0.06	0.82	0.81
SF9G	5	1.13	0.06	1.25	1.24
SF9H	5	-0.09	0.07	0.85	0.88
SF9I	5	1.88	0.06	1.1	1.08
SF10	5	-1.13	0.06	1.03	0.90

Note. No items were discarded for mis-fitting.

```
   6         .####  +
                    |
                    |
                    |
           #######  |
                    |
   5               T+
                    |
                    |
                    |
           .######  |
                    |
   4                +
                    |
         .########  |
                   S|
                    |
         .########  |
   3                +
                    |
      .###########  |
                    |
         #########  |
        ########## M|
   2       .######  +
                    |
          .#######  |
             .####  |
        .#########  |T SF11D
              .###  |  SF11B
   1      .#######  +
              .###  |  SF1
             .#### S|S
             .####  |  SF11A
                ##  |  SF7
                ##  |  SF4B
   0            .#  +M
                .#  |  SF3D
                .#  |  SF4C   SF4D
                 .  |  SF3F   SF3G   SF4A   SF8
                 . T|S
                 .  |  SF3H
  -1                +
                 .  |
                 .  |T
                 .  |
                    |
                    |
  -2                +
              <less>|<frequ>
```

Note. Each # represents 5 participants. M = mean. S = 1 SD. T = 2 SD.

Figure 3

Person-item map of the SF-36 physical HRQOL domain

```
 7             .  +
               .  |
                  |
                  |
                  |
 6                +
              .#  |
                  |
                  |
             .##  |
 5                +
             .## T|
                  |
                  |
            .###  |
 4                +
         .######  |
                  |
        ########  |
                 S|
 3        ######  +
         #######  |
                  |
       .########  |
        ########  |
 2        .#####  +T
        .#######  |  SF9I
        .#######  |
          .##### M|
      ##########  |  SF9E  SF9G
 1          .###  +S SF9D
    .###########  |
           .####  |
        .#######  |  SF9A
           .####  |
 0         .####  +M SF9B  SF9H
            .### S|
          .#####  |  SF5B  SF9F
             .##  |
         .######  |  SF5A  SF6
-1          .###  +S SF9C
              .#  |  SF10  SF5C
              .#  |
              ##  |
               #  |
-2             . T+T
               .  |
               .  |
               .  |
               .  |
-3                +
                  |
               .  |
                  |
                  |
-4                +
           <less>|<frequ>
```

Note. Each # represents 5 participants. M = mean. S = 1 SD. T = 2 SD.

Figure 4

Person-item map of the SF-36 mental HRQOL domain

Table 9

Model-data fit statistics for each stage of the Rasch analysis

						Person		Item			
Analysis	Domain	Items	%PVE	%IF	%PF	Separation	Reliability	Separation	Reliability	r_{RM}	Alpha
1	Physical	21	87.0	85.7	85.0	2.42	.85	8.03	.98	.87	.90
	Mental	14	100	100	84.0	3.55	.93	15.26	1.00	.96	.94
2	Physical	14	89.3	100	86.0	2.27	.84	8.81	.99	.88	.88

Note. Analysis #1 is the initial analysis with all items. Analysis #2 is after dropping misfit items. Alpha is Cronbach alpha and r_{RM} is the correlation between raw scores and person abilities (θ). %PVE is percent of validity criteria met for Rasch optimal category analysis. %IF is percent of items fitting the Rasch model. %PF is percent of person abilities fitting the Rasch model.

Item separation and item separation reliability were examined next (see Table 9). The item separation index indicates how well the scale separates the items along the ability continuum. The item separation index was 8.81 and 15.26 for the final physical and mental health Rasch models, respectively. A high item separation index (2.0 or greater) indicates adequate discrimination by the items. The item separation reliability indicates the capability to replicate item placement within measurement error for another sample. The item separation reliability was.99 and 1.00 for the final physical and mental health Rasch models, respectively. An item separation reliability close to 1.00 indicates a high degree of confidence for items (Bond & Fox, 2007).

The persons' HRQOL was estimated for each domain during the Rasch calibration process in logits, where a higher logit value indicated a greater (positive) level of HRQOL. Table 10 displays descriptive statistics for the person-level HRQOL trait (θ). The mean (SD) level of persons' physical HRQOL was 2.21 (1.61). The range of persons' physical HRQOL was from -1.72 to 5.21, indicating a large spread of physical HRQOL. Participant HRQOL was consistent across gender and age. The overall person fit was examined by evaluating the percentage of persons with acceptable fit criteria. Of the total sample, 545 (86%) ability estimates were well fit. Person separation was 2.27, which indicates that people were well spread along the physical HRQOL continuum. The person separation reliability was .84, which indicates an acceptable degree of confidence in replicating person placement within measurement error. The mean (SD) level of persons' mental HRQOL was 1.46 (1.77). The range of persons' mental HRQOL was

Table 10

Descriptive statistics for person HRQOL trait (θ) from Rasch analyses

	Mental HRQOL		Physical HRQOL	
	M	*SD*	*M*	*SD*
Overall	1.46	1.77	2.21	1.61
Gender				
Males	2.03	1.80	2.71	1.68
Females	1.24	1.69	2.00	1.53
Age (years)				
18-24	1.20	1.59	1.97	1.43
25-49	1.46	1.90	2.43	1.81
50-78	2.47	1.72	2.55	1.63

Note. HRQOL trait is measured in logits.

Table 11

Bivariate Spearman correlations between person HRQOL trait (θ) from Rasch analyses and CDC Healthy Days Index

		Mental HRQOL	Physical HRQOL
	N	r_s	r_s
Overall	634	-.729	-.532
Gender			
Males	175	-.681	-.440
Females	458	-.724	-.527
Age (years)			
18-24	344	-.658	-.458
25-49	203	-.777	-.588
50-78	85	-.765	-.449

Note. All correlations were significant (p's < .001). CDC Healthy Days Index represents number of unhealthy days.

from -3.50 to 6.78, indicating a large spread of mental HRQOL. Of the total sample, 533 (84%) ability estimates were well fit. Person separation was 2.27, which indicates acceptable spread along the mental HRQOL continuum. The person separation reliability was .84, which was acceptable.

Table 11 contains results of the convergent validity evidence for the physical and mental HRQOL person scores resulting from the Rasch analyses. Overall, the physical health scores were moderately correlated ($r_s = -.53$) with the CDC's Healthy Days Index. Analysis by gender and age for the physical health scores showed similar results with correlations ranging from -.44 to -.59. For mental health, person scores were strongly correlated ($r_s = -.73$) with the CDC's Healthy Days Index. Analysis by gender and age for the mental health scores showed similar results with correlations ranging from -.66 to -.78.

Table 12 contains results of the construct validity evidence for the physical and mental HRQOL person scores. Dichotomous groups were compared with known differences in HRQOL. For physical health, HRQOL person scores were significantly greater for those participants who were not obese, non-smokers, did not have an illness, did engage in vigorous activity, did engage in moderate activity, did engage in strength training exercises, did not have hypertension, did not have high blood cholesterol, and did not have diabetes (all p's < .01). For mental health, HRQOL person scores were also significantly greater for those participants who were not obese, non-smokers, did not have an illness, did engage in vigorous activity, did engage in moderate activity, did

Table 12

Mean differences between known groups in person HRQOL trait (θ) from Rasch analyses

	Mental HRQOL			Physical HRQOL		
Health Status	*M*	*SD*	*p*	*M*	*SD*	*p*
Obesity						
Yes	1.16	1.89	.008	1.56	1.51	< .001
No	1.52	1.73		2.33	1.60	
Current smoker						
Yes	0.69	1.54	.002	1.46	1.34	.001
No	1.51	1.76		2.26	1.61	
Has an illness						
Yes	.43	1.48	< .001	0.67	1.07	< .001
No	1.52	1.76		2.32	1.58	
Vigorously active						
Yes	1.88	1.69	< .001	2.69	1.72	< .001
No	0.96	1.71		1.61	1.24	
Moderately active						
Yes	1.68	1.73	< .001	2.38	1.61	< .001
No	.49	1.56		1.40	1.33	
Strength trains						
Yes	1.71	1.73	< .001	2.49	1.68	< .001
No	1.06	1.75		1.76	1.39	
Hypertension						
Yes	1.15	2.05	< .001	1.96	1.48	.009
No	1.50	1.70		2.24	1.62	
High cholesterol						
Yes	1.32	1.92	< .001	1.92	1.50	< .001
No	1.46	1.73		2.23	1.62	
Diabetes						
Yes	0.32	1.45	.001	0.99	1.13	< .001
No	1.49	1.76		2.24	1.61	

Note. p-values are from age-adjusted analysis of covariance (ANCOVA).

engage in strength training exercises, did not have hypertension, did not have high blood cholesterol, and did not have diabetes (all p's $< .01$).

Discussion

The purpose of this study was to separately evaluate the two HRQOL domains (physical and mental) of the SF-36 assessment using the Rasch model. The initial stages of the analysis evaluated the category functioning of each item. Using four criteria per item, it was found that majority of the total criteria were met for the physical health domain and all of the total criteria were met for the mental health domain. Despite the high percentage of validity evidence in the physical domain, eight items were flagged for disordered thresholds. The issue of ordered thresholds is an important item category characteristic. Order in the thresholds indicates that persons responding to higher levels (or lower levels) of a categorical scale in fact posses higher levels (or lower levels) of the trait being assessed. When thresholds are disordered, it is possible that some categories in the scale are unnecessary and/or redundant (Linacre, 2002b).

Six of the 8 disordered items came from the physical functioning section of the physical health domain. These items were 3b, 3c, 3g, 3h, 3i, and 3j. All physical functioning items share the same stem (Does your health now limit you in these activities?) and the same categorical scale: 1) Yes, limited a lot, 2) Yes, limited a little, and 3) No, not limited at all. One solution in this case may be to collapse the two "Yes" categories (i.e., 112) for each of these items. This would form a dichotomous item of 1) Yes, limited at least a little and 2) No, not limited at all.

The other two items with disordered thresholds came from the general health section of the physical health domain. These items were 11a and 11d. All general health items share the same stem (How true or false is each of the following statements for you?) and the same categorical scale: 1) Definitely true, 2) Mostly true, 3) Don't know, 4) Mostly false, and 5) Definitely false. One solution in this case may be to remove the "Don't know" category completely from the scale. This option could be explored by combining the "Don't know" category with the "Mostly true" category (i.e., 12234) or combining the "Don't know" category with the "Mostly false" category (i.e., 12334). The exploration of collapsing categories and re-running the Rasch model is a process that should be backed by a confirmatory stage (Linacre, 2002b) and is beyond the scope of this paper. This exploratory and confirmatory procedure is, however, needed and suggested for future research on the SF-36 HRQOL assessment.

Model-data fit was evaluated next and found that the mental health domain items adequately fit the Rasch model. This provides evidence that the SF-36 assesses a unidimensional mental HRQOL domain. The physical health domain data, however, did not initially fit the Rasch model well. Seven (3a, 3b, 3c, 3e, 3j, 3i, & 11c) out of the 21 items had fit statistics out of the acceptable range. After the misfit items were discarded, a 14-item physical health domain fit the Rasch model well and provided evidence for a unidimensional physical HRQOL domain. Six of the 7 items deleted were from the physical functioning scale (3a, 3b, 3c, 3e, 3j, and 3i). These items assessed participant's limitations in vigorous activity, moderate activity, lifting or carrying groceries, climbing one flight of stairs, walking one hundred yards, and bathing or dressing yourself, respectively.

One of two factors might be the underlying cause of these mis-fitted items. One factor is the 3-point scale previously mentioned regarding the physical functioning section of the SF-36. Four of the 6 mis-fitted physical functioning items were also flagged for having disordered thresholds. This type of category dysfunction is likely to explain the item's mis-fitting the Rasch model (Bond & Fox, 2007). The other factor is the possibility that the mis-fitted items are not of the same unidimensional construct as the other items. However, since these items concerning limitations in movement-related activities are of similar nature to other well-fitted items (i.e., Climbing several flights of stairs, Walking more than a mile, etc.), it is more likely they are mis-fitting due to improper category functioning.

The item-person maps for both the physical and mental health domains showed that the item locations were not well matched to persons of very high (very good) HRQOL or very low (very poor) HRQOL. The items were targeted well to persons of moderately poor to moderately good HRQOL. In other words, the items were too easy for the many of the participants in both domains. The most difficult physical domain item was item 11d (How true is the following statement: My health is excellent.) followed by item 11b (How true is the following statement: I am as healthy as anybody I know.). The least difficult physical domain item was item 3h (Does your health now limit you in: Walking several hundred yards?) followed by item 8 (During the past 4 weeks, how much did pain interfere with your normal work [including both work outside the home and housework]?). Discrimination among persons of better physical HRQOL may be increased by the addition of more difficult items (Bond & Fax, 2007).

The most difficult mental domain item was item 9i (How much of the time during the past 4 weeks: Did you feel tired?) followed by item 9e (How much of the time during the past 4 weeks: Did you have a lot of energy?). The least difficult mental domain item was item 10 (During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities [like visiting friends, relatives, etc.]?) followed by item 5c (how much of the time have you had any of the following problems… as a result of any emotional problems: Did work or activities less carefully than usual.). As well, discrimination among persons of better mental HRQOL may be increased by the addition of more difficult items. This should be explored in future research on the SF-36 HRQOL assessment.

A final stage of the present study was to show validity evidence for the persons' HRQOL scores resulting from the Rasch analyses. Results showed that the physical health Rasch scores and the CDC's Healthy Days Index were moderately correlated with each other. This provides convergent validity evidence in that both measures theoretically attempt to assess the same construct (Hennessy, Moriarty, Zack, Scherr, & Brackbill, 1994; Ware & Sherbourne, 1992). For the mental health domain, the Rasch scores and the CDC's Healthy Days Index were moderately to strongly correlated with each other. Overall, the convergent validity evidence provided substantial confirmation for the use of the Rasch person scores as a measure of HRQOL.

Construct validity evidence was also tested in this study by showing differences in Rasch HRQOL person scores between groups of participants with known differences. Results showed significantly greater physical and mental HRQOL scores for participants who were not obese compared to participants who were obese. This relationship has been confirmed before in large scale studies showing decreased physical HRQOL as well as mental HRQOL among obese adults compared to normal weight adults (Jia & Lubetkin, 2005). The same relationship was also found in a large population-level study using the CDC's Healthy Days HRQOL core (Hassan, Joshi, Madhavan, & Amonkar, 2003). Significantly greater physical and mental HRQOL scores were also seen for participants who did not smoke compared to participants who did smoke. This relationship has also been shown in national data with smokers who made no attempts to quit having significantly lower mental and physical HRQOL (McClave, Dube, Strine, & Mokdad, 2009).

Our finding of greater physical and mental HRQOL among participants with some form of illness has also been confirmed before by others showing that adults with some form of chronic illness were significantly more likely to report lower levels of HRQOL (Strine, Chapman, Balluz, Moriarty, & Mokdad, 2008). Significantly greater physical and mental HRQOL scores were also seen for participants who engaged in various physical activities as compared to participants who did not engage in those activities. This relationship has also been shown in quality of life research where adults who engaged in physical activity were less likely to report poor HRQOL than their non-active counterparts (Brown et al., 2003). Finally, this study showed that participants who reported having hypertension, high blood cholesterol, or diabetes had significantly lower physical and mental HRQOL compared to those participants who did not report those health problems. These findings have also been confirmed before (Hayes, Denny, Keenan, Croft, & Greenlund, 2008).

This study has many strengths worth mentioning. The large sample size was useful and necessary for proper Rasch parameter estimates and fit statistics. Samples of size 200 and greater are suggested for proper estimation (Kline, 2005). Another strength of this study is the use of the partial credit Rasch model to allow for the evaluation of proper category functioning per item (Bond & Fox, 2007). This was essential for the SF-36 HRQOL assessment because the instrument contains 35 items (21 for physical health and 14 for mental health) with three different categorical scales (3-point, 5-point, and 6-point). A final strength in this study was the administration of the CDC's Healthy Days HRQOL core to the same sample of participants for its use in validating the Rasch person HRQOL scores.

A limitation of this study was the use of the general population as a sampling frame. The SF-36, like many HRQOL assessments, is often used to differentiate perceived health among people suffering from disease states (Ware, 2000). It was found in the current study that, for a general sample of adults, the SF-36 items were too easy (ceiling effect). However, it is useful for researchers to know that when administering the SF-36 to a general sample of adults, the assessment may not be useful in effectively separating those individuals in terms of HRQOL.

In conclusion, a Rasch partial credit model was used to analyze the two dominant HRQOL domains (physical and mental) of the SF-36 assessment. The majority of the total criteria used for optimal category functioning were met for the physical health domain and all of the total criteria were met for the mental health domain. Eight items were flagged for disordered thresholds, of which 6 items came from the physical functioning subscale. Seven physical health items had fit statistics out of the acceptable range and were dropped from the final Rasch analysis. It is suggested that exploratory and confirmatory re-categorization of the 8 identified items be investigated. Also, adding more difficult items to the SF-36 should be investigated to help target healthier individuals. Finally, both convergent and construct validity evidence provided substantial confirmation for the use of the Rasch physical and mental health person scores as measures of HRQOL.

Chapter IV References

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model : fundamental measurement in the human sciences* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Brown, D. W., Balluz, L. S., Heath, G. W., Moriarty, D. G., Ford, E. S., Giles, W. H., & Mokdad, A. H. (2003). Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. *Preventive Medicine, 37*(5), 520-528.

Centers for Disease Control and Prevention. *Measuring healthy days: Population assessment of health-related quality of life*. Centers for Disease Control and Prevention, Atlanta, Georgia 2000.

Dominick, K. L., Ahern, F. M., Gold, C. H., & Heller, D. A. (2002). Relationship of health-related quality to health care utilization and mortality among older adults. *Aging Clinical and Experimental Research, 14*(6), 499-508.

Hassan, M. K., Joshi, A. V., Madhavan, S. S., & Amonkar, M. M. (2003). Obesity and health-related quality of life: a cross-sectional analysis of the US population. *International Journal of Obesity and Related Metabolic Disorders, 27*(10), 1227-1232.

Hayes, D. K., Denny, C. H., Keenan, N. L., Croft, J. B., & Greenlund, K. J. (2008). Health-related quality of life and hypertension status, awareness, treatment, and control: National Health and Nutrition Examination Survey, 2001--2004. *Journal of Hypertension, 26*(4), 641-647.

Heath, G. W., & Brown, D. W. (2009). Recommended levels of physical activity and health-related quality of life among overweight and obese adults in the United States, 2005. *Journal of Physical Activity and Health, 6*(4), 403-411.

Hennessy, C. H., Moriarty, D. G., Zack, M. M., Scherr, P. A., & Brackbill, R. (1994). Measuring health-related quality of life for public health surveillance. *Public Health Reports, 109*(5), 665-672.

Jia, H., & Lubetkin, E. I. (2005). The impact of obesity on health-related quality-of-life in the general adult US population. *Journal of Public Health (Oxf), 27*(2), 156-164.

Kline, T. (2005). *Psychological testing : a practical approach to design and evaluation.* Thousand Oaks, Calif.: Sage Publications.

Linacre, J. (2006). *A user's guide to Winsteps: Rasch model computer programs.* Chicago: Winsteps.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3,* 103–122.

Linacre, J. M. (2002a). What do Infit and Outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16,* 878.

Linacre, J. M. (2002b). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85–106.

Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity of examination scores. *Applied Measurement in Education, 3*(4), 331–345.

McClave, A. K., Dube, S. R., Strine, T. W., & Mokdad, A. H. (2009). Associations between health-related quality of life and smoking status among a large sample of U.S. adults. *Preventive Medicine, 48*(2), 173-179.

Mielenz, T., Jackson, E., Currey, S., DeVellis, R., & Callahan, L. F. (2006). Psychometric properties of the Centers for Disease Control and Prevention Health-Related Quality of Life (CDC HRQOL) items in adults with arthritis. *Health and Quality of Life Outcomes, 24*(4), 66.

QualityMetric. (2011). SF Health Surveys. Retrieved August 15, 2011, from http://www.qualitymetric.com/WhatWeDo/GenericHealthSurveys/tabid/184/Default.aspx

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). *Probabilistic models for some intelligence and achievement tests.* (Expanded ed.). Chicago: University of Chicago Press.

Strine, T. W., Chapman, D. P., Balluz, L. S., Moriarty, D. G., & Mokdad, A. H. (2008). The associations between life satisfaction and health-related quality of life, chronic illness, and health behaviors among U.S. community-dwelling adults. *Journal of Community Health, 33*(1), 40-50. doi: 10.1007/s10900-007-9066-4

Ware, J.E. (2004). SF-36 Health Survey Update. Retrieved August 15, 2011, from http://www.sf-36.org/announcements/Updated_SF36_bookChapter_Sept04.pdf

Ware, J.E., Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, *30*(6), 473.

Ware, J. E., Jr. (2000). SF-36 health survey update. *Spine (Phila Pa 1976), 25*(24), 3130-3139.

Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Journal of the American Medical Association, 273*(1), 59-65.

CHAPTER V

Conclusion

Since the publication of *Physical Activity and Health: A Report of the Surgeon General*, many studies have shown the positive effects of regular physical activity on specific health outcomes. Health-related quality of life (HRQOL) is one such health outcome that has seen growing interest in physical activity research. HRQOL is a broad construct that includes both subjective and objective indicators of people's lives that affect their physical and/or mental health status (CDC, 2000). Physical activity has been shown to be directly associated with HRQOL (Heath & Brown, 2009). Specifically, meeting recommended levels of physical activity has shown to be related to superior levels of HRQOL (Brown et al., 2003). HRQOL is now more than ever being included in physical activity research studies alongside the more conventional measures of health status (Dominick, Ahern, Gold, & Heller, 2004). Ultimately, increasing HRQOL has been described as the most important goal in physical activity interventions (Bertheussen et al., 2011). Given the overwhelming interest in HRQOL as an outcome measure in physical activity research, there is a strong need for a better understanding of the measurement properties of HRQOL assessments commonly used in physical activity research.

The first study, "*A Systematic Review of Health-Related Quality of Life Assessments in Physical Activity Research*," examined 10 HRQOL assessments used in physical activity research. The identified assessments were the Short Form Health

Survey (SF-36), Sickness Impact Profile (SIP), Euroqol (EQ-5D), Nottingham Health Profile (NHP), WHO Quality of Life (WHOQOL-BREF), Quality of Well-Being Scale (QWB), Health Utilities Index 3 (HUI3), CDC Healthy Days Core (HRQOL-4), Assessment of Quality of Life (AQoL), and the Duke Health Profile (DHP). The SF-36 was by far the most commonly used assessment tool. As well, the SF-36 had the most published evidence supporting its psychometric properties. Other HRQOL assessments with good potential include AQoL, DHP, and WHOQOL-BREF. These 3 assessments provide a lot of HRQOL information given the number of items they contain. If time is the most important factor, the EQ-5D and CDC HRQOL-4 are useful and valid scales.

The second study, "*Reliability of the Short-Form Health Survey (SF-36) in Physical Activity Research Using Meta-Analysis*," examined the reliability generalization of the SF-36 applied to physical activity research using meta-analysis. Results showed that the reliability of the SF-36 generalizes very well across various physical activity studies. The reliability coefficients were strong and significant for both physical and mental health domains of the assessment. Reliability was maintained across studies of different gender, different research design, and different diseases status. However, reliability was significantly higher if the internal consistency coefficient was reported (opposed to test-retest), the SF-8 was used (opposed to SF-36 or SF-12), or if the mean age of participants was lower (compared to older).

The third and final study, "*Evaluation of the Short-Form Health Survey (SF-36) Using the Rasch model*," examined the two HRQOL domains (physical and mental) of the SF-36 assessment using Rasch analysis. Results showed that the majority of the total criteria used for optimal category functioning were met for the physical health domain and all of the total criteria were met for the mental health domain. Eight items were flagged for disordered thresholds. It is suggested that exploratory and confirmatory categorization of the 8 identified items be investigated. Seven physical health items also had fit statistics out of the acceptable range and were dropped from the final Rasch analysis. Furthermore, adding more difficult items to the SF-36 should be investigated to better target healthier individuals. Finally, both convergent and construct validity evidence provided substantial confirmation for the use of the Rasch physical and mental health person scores as measures of HRQOL.

In conclusion, the three studies contained in this dissertation address the measurement issues of HRQOL assessment in physical activity research. It was found that the SF-36 was the dominant assessment tool used in physical activity research. It was also found, through meta-analytical procedures, that the SF-36 provides strong reliability across a wide range of physical activity research. The SF-36 also met stringent modern measurement criteria using the Rasch model. Future research recommendations include 1) the examination of potentially useful HRQOL assessments in physical activity research such as the AQoL, DHP, WHOQOL-BREF for their large information to item ratios and the EQ-5D and CDC HRQOL-4 for their ability to be placed in large surveys; 2) exploratory and confirmatory categorization of several items in the SF-36 assessment so as to provide a more optimal functioning category scale; and 3) exploring the addition of more difficult items to the SF-36 to better target healthier individuals.

Chapter V References

Bertheussen, G. F., Romundstad, P. R., Landmark, T., Kaasa, S., Dale, O., & Helbostad, J. L. (2011). Associations between physical activity and physical and mental health--a HUNT 3 study. *Medicine and Science in Sports and Exercise*, *43*(7), 1220-1228.

Brown, D. W., Balluz, L. S., Heath, G. W., Moriarty, D. G., Ford, E. S., Giles, W. H., & Mokdad, A. H. (2003). Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. *Preventive Medicine*, *37*(5), 520-528.

Centers for Disease Control and Prevention. Measuring healthy days: Population assessment of health-related quality of life. Centers for Disease Control and Prevention, Atlanta, Georgia 2000.

Dominick, K. L., Ahern, F. M., Gold, C. H., & Heller, D. A. (2004). Health-related quality of life among older adults with arthritis. *Health and Quality of Life Outcomes*, *13*(2), 5-12.

Heath, G. W., & Brown, D. W. (2009). Recommended levels of physical activity and health-related quality of life among overweight and obese adults in the United States, 2005. *Journal of Physical Activity and Health*, *6*(4), 403-411.

APPENDICES

APPENDIX A

IRB Approval Letter

December 14, 2011

Peter D. Hart
Department of Health and Human Performance
pdh2k@mtmail.mtsu.edu , minsoo.kang@mtsu.edu

Protocol Title: "Measurement Issues in Health-Related Quality of Life Assessments in Physical Activity Research"
Protocol Number: **12-146**

Dear Investigator(s),

The MTSU Institutional Review Board, or a representative of the IRB, has reviewed the research proposal identified above. The MTSU IRB or its representative has determined that the study poses minimal risk to participants and qualifies for an expedited review under the 45 CFR 46.110 Category7.

Approval is granted for one (1) year from the date of this letter for 500participants.

According to MTSU Policy, a researcher is defined as anyone who works with data or has contact with participants. Anyone meeting this definition needs to be listed on the protocol and needs to provide a certificate of training to the Office of Compliance. **If you add researchers to an approved project, please forward an updated list of researchers and their certificates of training to the Office of Compliance (c/o Emily Born, Box 134) before they begin to work on the project.** Any change to the protocol must be submitted to the IRB before implementing this change.

Please note that any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918.

You will need to submit an end-of-project form to the Office of Compliance upon completion of your research located on the IRB website. Complete research means that you have finished collecting and analyzing data. **Should you not finish your research within the one (1) year period, you must submit a Progress Report and request a continuation prior to the expiration date.** Please allow time for review and requested revisions. Your study expires December 14, 2012.

Also, all research materials must be retained by the PI or faculty advisor (if the PI is a student) for at least three (3) years after study completion. Should you have any questions or need additional information, please do not hesitate to contact me.

Sincerely,

Emily Born

Emily Born
Research Compliance Officer
Middle Tennessee State University
eborn@mtsu.edu

APPENDIX B

Healthy Days Core Module (CDC HRQOL– 9)

1. Would you say that in general your health is:

Excellent	1
Very good	2
Good	3
Fair	4
Poor	5

2. Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good? ____

3. Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good? ____

4. During the past 30 days, for about how many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation? ____

5. During the past 30 days, for about how many days did PAIN make it hard for you to do your usual activities, such as self-care, work, or recreation? ____

6. During the past 30 days, for about how many days have you felt SAD, BLUE, or DEPRESSED? ____

7. During the past 30 days, for about how many days have you felt WORRIED, TENSE, or ANXIOUS? ____

8. During the past 30 days, for about how many days have you felt you did NOT get ENOUGH REST or SLEEP? ____

9. During the past 30 days, for about how many days have you felt VERY HEALTHY AND FULL OF ENERGY? ____

APPENDIX C

Short-Form Health Survey (SF-36)

1. In general, would you say your health is:

Excellent	Very good	Good	Fair	Poor
☐	☐	☐	☐	☐

2. <u>Compared to one year ago</u>, how would you rate your health in general <u>now</u>?

Much better now than one year ago	Somewhat better now than one year ago	About the same as one year ago	Somewhat worse now than one year ago	Much worse now than one year ago
☐	☐	☐	☐	☐

3. The following questions are about activities you might do during a typical day. Does <u>your health now limit you</u> in these activities? If so, how much?

	Yes, limited a lot	Yes, limited a little	No, not limited at all
a <u>Vigorous activities</u>, such as running, lifting heavy objects, participating in strenuous sports	☐	☐	☐
b <u>Moderate activities</u>, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	☐	☐	☐
c Lifting or carrying groceries	☐	☐	☐
d Climbing <u>several</u> flights of stairs	☐	☐	☐
e Climbing <u>one</u> flight of stairs	☐	☐	☐
f Bending, kneeling, or stooping	☐	☐	☐
g Walking <u>more than a mile</u>	☐	☐	☐
h Walking <u>several hundred yards</u>	☐	☐	☐
i Walking <u>one hundred yards</u>	☐	☐	☐
j Bathing or dressing yourself	☐	☐	☐

4. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a Cut down on the amount of time you spent on work or other activities	☐	☐	☐	☐	☐
b Accomplished less than you would like	☐	☐	☐	☐	☐
c Were limited in the kind of work or other activities	☐	☐	☐	☐	☐
d Had difficulty performing the work or other activities (for example, it took extra effort)	☐	☐	☐	☐	☐

5. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a Cut down on the amount of time you spent on work or other activities	☐	☐	☐	☐	☐
b Accomplished less than you would like	☐	☐	☐	☐	☐
c Did work or activities less carefully than usual	☐	☐	☐	☐	☐

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐	☐	☐	☐	☐

7. How much bodily pain have you had during the past 4 weeks?

None	Very mild	Mild	Moderate	Severe	Very severe
☐	☐	☐	☐	☐	☐

8. **During the <u>past 4 weeks</u>, how much did <u>pain</u> interfere with your normal work (including both work outside the home and housework)?**

Not at all	A little bit	Moderately	Quite a bit	Extremely
☐	☐	☐	☐	☐

9. **These questions are about how you feel and how things have been with you <u>during the past 4 weeks</u>. For each question, please give the one answer that comes closest to the way you have been feeling.**

How much of the time during the <u>past 4 weeks</u>...

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a Did you feel full of life?	☐	☐	☐	☐	☐
b Have you been very nervous?	☐	☐	☐	☐	☐
c Have you felt so down in the dumps that nothing could cheer you up?	☐	☐	☐	☐	☐
d Have you felt calm and peaceful?	☐	☐	☐	☐	☐
e Did you have a lot of energy?	☐	☐	☐	☐	☐
f Have you felt downhearted and depressed?	☐	☐	☐	☐	☐
g Did you feel worn out?	☐	☐	☐	☐	☐
h Have you been happy?	☐	☐	☐	☐	☐
i Did you feel tired?	☐	☐	☐	☐	☐

10. **During the <u>past 4 weeks</u>, how much of the time has your <u>physical health or emotional problems</u> interfered with your social activities (like visiting friends, relatives, etc.)?**

All of the time	Most of the time	Some of the time	A little of the time	None of the time
☐	☐	☐	☐	☐

11. How TRUE or FALSE is <u>each</u> of the following statements for you?

		Definitely true	Mostly true	Don't know	Mostly false	Definitely false
A	I seem to get sick a little easier than other people	☐	☐	☐	☐	☐
B	I am as healthy as anybody I know	☐	☐	☐	☐	☐
C	I expect my health to get worse	☐	☐	☐	☐	☐
D	My health is excellent	☐	☐	☐	☐	☐

APPENDIX D

Health and Demographic Survey

1. What is your gender?

    a. male
    b. female

2. What is your age? _______

3. What is your race/ethnicity?

    a. Caucasion or white
    b. Black
    c. Hispanic
    d. Asian
    e. Bi- or Mixed-race
    f. Other

4. What is the highest education you received?

    a. None
    b. Elementary School
    c. Junior High School
    d. Senior High School
    e. Some College
    f. College Degree

5. What is your marital status?

    a. Single
    b. Married
    c. Separated
    d. Divorced
    e. Widowed
    f. Living with Partner

6. What is your height in inches? ________

7. What is your weight in pounds? ________

8. Do you smoke cigarettes?

   a. yes
   b. no

9. Do you have a major illness?

   a. yes
   b. no

10. Do you do any vigorous-intensity sports, fitness, or recreational activities that cause large increases in breathing or heart rate like running or basketball for at least 10 minutes continuously?

    a. yes
    b. no

11. Do you do any moderate-intensity sports, fitness, or recreational activities that cause a small increase in breathing or heart rate such as brisk walking, bicycling, swimming, or golf for at least 10 minutes continuously?

    a. yes
    b. no

12. Do you do any physical activities specifically designed to strengthen your muscles such as lifting weights, push-ups or sit-ups?

    a. yes
    b. no

13. Have you ever been told by a health professional that you have high blood pressure?

    a. yes
    b. no

14. Have you ever been told by a health professional that you have high cholesterol?

    a. yes
    b. no

15. Have you ever been told by a health professional that you have high blood sugar or diabetes?

    a. yes
    b. no