

MINIMAL DETECTABLE CHANGE AND PATIENT REPORTED OUTCOMES IN FALLS  
REHABILITATION

By

SERGIO ROMERO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

© 2008 Sergio Romero

To “Cu”

## TABLE OF CONTENTS

	<u>page</u>
LIST OF TABLES .....	6
LIST OF FIGURES .....	7
ABSTRACT .....	8
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW .....	10
Introduction.....	10
Theoretical Framework.....	12
Clinical Change .....	14
Clinical Important Differences .....	18
Calculating Clinically Meaningful Change.....	19
Anchor-Based Methods.....	20
Cross-sectional methods.....	20
Longitudinal methods.....	21
Distribution-Based Approaches.....	23
Paired t-Statistic .....	24
Growth curve analysis.....	24
Methods Based on Sample Variation .....	25
Effect size.....	25
Standardized Response Mean (SRM) .....	25
Methods Based on Measurement Precision.....	26
Standard Error of Mean (SEM).....	26
Reliable Change (RC) .....	26
Choosing an Appropriate Method .....	27
Patient Reported Outcomes (PRO).....	28
PRO Instruments .....	29
PROs and the International Classification of Function (ICF) .....	30
Advantages of Using PROs.....	32
Issues and Concerns about the Use of PRO .....	35
Summary.....	38
2 ESTABLISHING THE MINIMUM DETECTABLE CHANGE (MDC) FOR THE BERG BALANCE SCALE AND DYNAMIC GAIT INDEX .....	41
Introduction.....	41
Gait and Balance Assessment.....	41
The Berg Balance Scale (BBS) .....	41
The Dynamic Gait Index (DGI) .....	42
The Intraclass Correlation Coefficient (ICC), the Standard Error of Measurement (SEM) and the Minimal Detectable Change (MDC) .....	45

Methods .....	46
Subjects.....	46
Testing Procedure.....	47
Analysis .....	48
Results.....	50
Discussion.....	52
Conclusion.....	56
3 PATIENTS SUCCESS CRITERIA AND EXPECTATIONS IN FALLS REHABILITATION .....	61
Introduction.....	61
Methods .....	64
Subjects.....	64
Testing Procedure.....	66
Analysis .....	67
Results.....	69
Discussion.....	71
4 GENERAL SUMMARY AND CONCLUSIONS .....	79
Experiment I Summary.....	80
Experiment II Summary .....	81
General Conclusions and Future Directions.....	83
APPENDIX Patient’s Perspective Outcome Questionnaire (PPOQ) .....	85
LIST OF REFERENCES .....	88
BIOGRAPHICAL SKETCH .....	98

## LIST OF TABLES

<u>Table</u>		<u>page</u>
2-1	Absolute difference in BBS scores between initial and re-scored assessments.....	58
2-2	Absolute difference in DGI scores between initial and re-scored assessments .....	58
2-3	MDC values for the BBS and DGI .....	58
3-1	Demographics .....	77
3-2	Initial levels descriptive statistics .....	77
3-3	Success criteria (initial levels-success levels) descriptive statistics .....	77
3-4	Treatment expectations criteria (initial levels-expectations levels) descriptive statistics.....	78

## LIST OF FIGURES

<u>Figure</u>	<u>page</u>
1-1 International Classification of Functioning Disability and Health (ICF) model .....	40
2-1 Distribution of BBS difference scores (Initial BBS- re-scored BBS).....	57
2-2 Distribution of DGI difference scores (Initial DGI- re-scored DGI).....	57
2-3 BBS results for all participants. Mean value and difference between initial and re-scored values. ....	59
2-4 DGI results for all participants. Mean value and difference between initial and re-scored values. ....	59
2-2 Mean BBS score for different absolute differences in BBS between testing occasions....	60
2-3 Mean DGI score for different absolute differences in DGI between testing occasions.....	60

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

MINIMAL DETECTABLE CHANGE AND PATIENT REPORTED OUTCOMES IN FALLS  
REHABILITATION

By

Sergio Romero

August 2008

Chair: Craig A. Velozo  
Co-chair: Kathye E. Light  
Major: Rehabilitation Science

The overall aim of this project was to investigate the reliability of two instruments used for the assessment of balance and to explore patient's expectations and success criteria for the rehabilitation of falls.

The first experiment investigated minimal detectable change (MDC) for two common instruments used to assess gait and balance. The results of this study indicated that for the Berg Balance Scale and the Dynamic Gait Index, 6.6 and 3.1 points respectively were required to be 95% confident that "genuine" change had occurred. These results suggest that a significant amount of error is associated with these instruments. In addition, the results suggested that MDC values are not a constant feature of the instruments. MDC values for the high function group were 6.3 BBS points, as compared to 7.3 points for the low function group. That is, the values of MDC change based on the ability level of the persons assessed.

The second experiment investigated patient's success criteria and expectations with treatment. Participants reported considerable initial levels of impairment in energy and drive, mobility, and pain. Lower scores were seen in interaction with people and community and social

life. These findings suggest that domains with a strong social component were not as affected as domains with a strong physical component.

Participants in this study required significant improvement to consider their treatment successful. Domains such as mobility; and energy and drive, required significantly larger reductions than the community and social life; and interactions with people domains. This provides information about what is important to patients receiving this intervention. Participants expected mobility to change the most. However, similar finding was reported in the domain of energy and drive. An interesting finding was that, participant's expectation was that the treatment would *not* meet their success criteria, indicating that residual levels of impairment were expected.

Collectively, this series of studies promotes our understanding of significant change in patients receiving rehabilitation services related to falls. The results obtained indicate that current rehabilitation programs must consider the limitations of available instruments and take into consideration the needs and expectations of patients.

## CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW

### **Introduction**

The elderly population is the fastest growing group in the USA [1]. By the year 2030, it is expected that one in five Americans will be 65 or older. This represents an increase in the elderly population that will double that of the total population growth [2]. This epidemiological profile is particularly relevant to the national healthcare system, as the patient population will also experience similar growth. This growth in the number and proportion of older people will place increasing health and economic demands on the national healthcare system.

Although aging is a highly individual-dependent phenomenon, a progressive loss of biological function is expected with age. Sometimes, this can lead to functional decline, loss of independence, and, ultimately, a decrease in quality of life. In a number of elders, this biological decline can result in compromised mobility. This is one of the most disabling conditions elders can experience, since mobility limitations restrict the individual from fully participating in everyday activities and can contribute to further functional decline [3]. Individuals with mobility problems are prone to falls. The consequences of falls are far reaching and affect not only the individual, but also caregivers and the healthcare system in general.

Falls are a serious health problem in the older population. They are the leading cause of injury deaths among people 65 and older [4]. In 2002, nearly 13,000 people ages 65 and older died from fall-related injuries [4]. Falls affect more than one-third of adults aged 65 years and older each year [5]. As a result of the rapid growth in this segment of the population, fall-related medical care is imposing an enormous demand on the US healthcare systems. In 1994, the cost of fall-related injuries was \$27.3 billion. This number is expected to reach \$43.8 billion by the

year 2020 [6]. Preventive and rehabilitative strategies are needed to stop, or at least slow down, this trend.

Several important areas need to be addressed relating to falls in the elder population. First, there is a need to accurately identify the population at risk. Next, assessment tools used to identify individuals at risk and measure performance change must be reliable and convey information that is clinically relevant and easy to interpret. Finally, effective and feasible rehabilitation strategies must be developed to specifically target this population. To this end, rehabilitation goals must include the patients' criteria to ensure adherence to treatment and accurately determine treatment success. The overall goals of this dissertation are to investigate meaningful change in two assessment tools commonly used in falls rehabilitation programs, and to explore patients' expectations and success criteria in the rehabilitation of falls.

In the subsequent literature review I will develop the background to address the study's purpose. First, I will introduce the conceptual model that has guided this dissertation, the World Health Organization's International Classification of Function and Disability (ICF). The ICF provides a theoretical framework to encompass the multiple dimensions of the rehabilitation process and the factors that affect this process. Next, I will review the concept of clinical change and all of its variations (e.g. minimal detectable change, clinically important change, etc.). Rehabilitation decisions are based on assessing and reassessing patients to determine the effectiveness of treatment. A clear understanding of the properties of our assessment instruments and how to interpret the results obtained with these instruments is of utmost importance to clinicians and researchers. Third, I will review the literature pertaining to Patient Reported Outcomes (PRO), since understanding what the patient values in the rehabilitation process is

critical. Finally, I will provide a summary of this literature review and how it supports the overall purpose of this dissertation.

### **Theoretical Framework**

Several theoretical models have been developed and explored over the years. In the field of rehabilitation, there has been a shift from what was previously known as “disabling” models, to a more proactive theoretical framework termed “enablement models”. In rehabilitation sciences, an “enablement” model frequently used is the International Classification of Functioning, Disability, and Health (ICF), proposed by the World Health Organization [7]. This model was used to guide the research questions and analyses for this dissertation.

The ICF provides a theoretical framework for the analysis of health conditions, body structure and function, activity, and participation, and environmental and personal factors. There are two main components in the ICF model. The first is concerned with functioning and disability and includes the areas of body structure, body functions, activity, and participation. The second part includes the components of contextual factors, and includes environmental and personal factors (see figure 1.1).

In this dissertation, the health condition of interest includes a variety of disorders and diseases that influence gait and balance in the elder population and can lead to falls. The ICF model offers a holistic approach that considers the dynamic interaction of different aspects of a health condition from a biological, individual, and social perspective. In individuals with gait and balance disorders, a number of health conditions contribute to a decline in biological function that can lead to falls. In the ICF model, this is represented at the body function and structure side of the model (Figure 1.1). This can result in restriction and inability to perform certain activities and can also influence the individual’s social participation. This transition from body function and structure to activity and participation limitations can also work in the opposite direction. For

example, a decrease in participation and activities can lead to deterioration at the body function and structure level. In addition, personal and environmental factors are also taken into account by ICF model to draw a comprehensive picture of how an individual, within his or her limitations, and conditioned by the environment and his or her own personal characteristics, functions in society.

This dissertation focuses in two distinct, although directly connected, aspects of the rehabilitation of individuals with gait and balance disorders. First, two assessment instruments used in clinical practice to evaluate gait and balance will be studied. These instruments are used to test individual's ability to perform certain tasks related to gait and balance. Therefore, the first experiment in this dissertation will be investigating the activity component of the ICF model. Secondly, a patient reported outcome questionnaire will be used to investigate individual's expectations and success criteria when participating in a gait and balance intervention. Since the ICF is not only a theoretical model but also a classification system, using the coding system of the ICF, a comprehensive picture of an individual's experience when going through a gait and balance intervention can be drawn. The ICF categories facilitate the description and classification of all aspects of function and health in individuals, independent of a specific assessment instrument. The current ICF lists 1,424 categories referring to body functions and structure, activities and participation, and environmental factors [7]. The ICF model's classification system was used to produce a patient reported outcome questionnaire to identify specific areas of the model that are important to individuals with gait and balance problems including: mobility, self-care, interactions with people, community and social life, energy and drive, mental function, emotional distress, sensory function, and pain.

Figure 1.1 shows the areas of the model tested in the two experiments included in this dissertation. The ICF model emphasizes interconnectivity of all the areas of the model. This can lead to some overlapping, with some concepts being represented in more than one area of the model. Sensory function and pain, energy and drive, mental function, and emotional distress are included at the level of body function and structure. Some of these concepts are highly connected to personal factors and could also be represented in this area of the model. Balance, self care, and mobility are included under the area of activity. Again, one can argue that some of these concepts could be included under participation. Finally, interactions with people, and community and social life have been included under the participation area of the model. Although contextual factors are not specifically addressed in this dissertation, they must be considered when interpreting the results of this investigation. A combination of functioning and disability, together with contextual factors contributes to the ultimate goal of addressing quality of life and life satisfaction.

### **Clinical Change**

**Selecting outcome measures:** Health research is based on scientist's ability to measure the different dimensions of the health construct. In the past few decades, the medical community has shown an increased interest for the use of evidence-based clinical practice. This interest has been supported by governing agencies, insurance companies, and consumers. These groups recognize the need for interventions that are scientifically driven, produce measurable and meaningful results, and are based on theoretical models that cover the entire spectrum of health.

To this end, many measurement instruments have been developed to assess different aspects of health. The selection of a suitable instrument generally depends on the goals and characteristics of the population being assessed. In rehabilitation, most assessment tools aim at evaluating functional aspects related to health. In general, the relevance of measurement

instruments can be optimized by incorporating a theoretical framework of health and disability, establishing the purpose of the measurement, and assuring that these instruments are psychometrically sound.

A theoretical framework of health and disability provides the conceptual basis for developing and using an assessment instrument. For example, under the International Classification of Function Disability and Health (ICF) [7], a researcher can develop measurement instruments to assess the areas of body function and structure, activities, and participation, and the influence of personal and environmental factors [8]. Using a theoretical model helps researchers identify the domains of interest and provide a complete view of how the different domains interact with each other to affect the overall health status of the individuals under study.

Another essential factor in the development and selection of outcome measures is to establish the purpose of the measurement. In general, assessment instruments are used to discriminate among individuals, predict future outcomes, and evaluate interventions [9]. Discriminative instruments are used to differentiate between individuals based on specific criteria when no external gold standard exists for validating these measures [10]. These instruments are used for investigating between-subject differences where groups of individuals are assigned to separate treatment conditions. For example, a researcher interested in investigating differences in balance among two groups of elders with Parkinson's disease should use a discriminative assessment instrument. Predictive instruments are used to categorize individuals into predetermined category when a gold standard is available [10]. This gold standard is used to determine whether individuals have been classified correctly. For example, a shorter version of an instrument can be used to assess a particular condition. Later, the results

can be compared to the original or gold standard instrument. Finally, evaluative instruments are used to measure the magnitude of longitudinal change in an individual or group (within-subjects experimental design) [10]. In this type of investigation, two measurements are obtained from the same sample. Changes in performance within each participant across treatments are used to determine a treatment effect.

Measurement instruments used in rehabilitation must have sound psychometric properties. In classical test theory, these properties assure that the instruments we use can provide information that is meaningful, valid, and consistent with the construct we are measuring. The key traditional psychometric properties of an instrument are validity, reliability, responsiveness to change, and minimal clinically important difference (MCID) [12].

Validity refers to the ability of an instrument to measure what it is intended and presumed to measure [13]. A valid measure must be reliable (consistent). However, a reliable measure does not have to be valid. Validity can be investigated and defined in a number of ways. First, it is possible to correlate measures with a criterion measure known to be valid. This is considered the criterion validity. If the criterion measure is collected at the same time as the measure being validated, the concurrent validity is obtained. When the criterion is collected later, the validity obtained is the predictive validity. A different type of validity is based on the construct of the instrument been used. Construct validity refers to whether an instrument measures the construct it is supposed to measure. Finally, content validity, or face validity, is simply the extent to which a measure represents all facets of a given construct [14].

Instruments used in rehabilitation must also be reliable. A measurement is reliable when repeatedly testing a particular subject under the same conditions produces the same results [15]. A measurement can reliably measure the wrong attribute. Such a measurement will be reliable,

but not valid. There are a number of ways to determine reliability. First, inter-rater reliability is used to assess the degree to which different raters give consistent measurements of the same phenomenon. Second, test-retest reliability is used to assess the consistency of a measure from one time to another. And third, internal consistency reliability is used to assess the consistency of results across items within an instrument [16].

Responsiveness is also important to determine changes over time, which may be indicative of therapeutic effects. Instrument responsiveness is the ability of the instrument to precisely detect meaningful changes [17]. The measurements of an instrument used in rehabilitation must be able to identify clinically significant differences between and within patients over time [18]. In addition, the instrument should only be responsive to changes in the variable being assessed, and should not be influenced by changes in other variables [19]. Related to the responsiveness of an instrument are the concepts of sensitivity and specificity. Sensitivity is the ability of an instrument to detect changes in the variable under study, when they occur. It is a measure of the probability of correctly identifying a change. Specificity is the ability of an instrument to correctly identify when no changes in the variable under study occurs. It is a measure of the probability of correctly identifying no change [11].

Finally, related to the concept of responsiveness, is the ability of an instrument to detect minimal clinically important differences (MCID). The minimal important difference has been defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management” [21]. The concept of clinically important change introduces a new variable in measurement; the patient’s perspective. In addition, clinically important change can have consequences at the clinicians, researchers and societal level. Out of

all psychometric properties desirable in any assessment instrument, none has more practical implications than the ability of an instrument to detect MCID. At the end, the ultimate goal of an intervention is to produce results that are important to the persons receiving the treatment. An in depth analysis of MCID follows.

### **Clinical Important Differences**

Interpretation of clinical assessments is often difficult. This is especially true when the variables measured are based on some abstract construct. For example, interpreting a change in blood glucose levels of 20 mg/dL is easy, but interpreting a change of 4 points in a Quality of Life instrument can prove to be a difficult task. Therefore, establishing clinical importance is more difficult as the concept being assessed becomes more abstract. In addition, clinically important differences may be different across groups of patients defined by diseases, levels of severity, cultural background, socioeconomic status, and nationality [22]. Determining clinically meaningful differences is important because, sometimes, studies are based on small differences in mean scores between groups, which can lead to a statistically significant difference when the sample sizes are large. However, statistical significance is not equivalent to clinical significance [23].

**Different perspectives in clinically important differences:** There are several important issues to consider when looking at clinically important changes. First, from the point of view of the patient, a meaningful change may be one that results in a meaningful reduction in symptoms or improvement in functional status. On the other hand, clinicians can consider a meaningful change when the patient's improvement results in a change in treatment or disease prognosis. These two perspectives may not coincide all the time. In addition, societal and institutional perspectives for determining what constitutes a clinically important change can also differ from the clinicians' and patients' view [23]. From societies' perspective small changes can be

meaningful if the condition of interest affects a large number of people. Institutions may be more concerned with changes that influence health care policies [23].

Another important issue to consider is whether clinically meaningful changes are based on individual or group differences. When looking at group differences, it is important to take into consideration that mean changes do not provide information about individual scores. It is possible that even in groups with small mean differences a number of subjects could exhibit significant changes. On the other hand, these differences could be attributed to measurement error associated with the measurement instrument. Therefore, when investigating issues related to public health, where even small differences can have great impact, reporting group differences is appropriate [22]. Conversely, differences at the individual level are more relevant when individual decisions about a particular treatment must be made. Furthermore, the amount of change necessary to be considered clinically meaningful is also influenced by whether it is applied to an individual or a group. Relatively small improvements at the individual level may be considered clinically important when looking at the group level [22].

### **Calculating Clinically Meaningful Change**

To date, two broad strategies have been suggested for calculating MCID: 1) anchor-based measures and 2) distribution based approaches [24]. Anchor-based methods examine the relationship between an instrument's measure and an independent measure (or anchor) to explain the meaning of a particular degree of change. Therefore, anchor-based approaches need an independent standard or anchor that is itself interpretable and at least moderately correlated with the instrument being explored [25]. On the other hand, distribution based approaches rely on the statistical distribution of scores in a given instrument [25].

## **Anchor-Based Methods**

Anchor-based methods have been used to determine clinically meaningful change via cross-sectional and longitudinal approaches.

### **Cross-sectional methods**

A cross-sectional approach is used when comparing groups that are different in terms of some disease-related criterion [26]. The difference in mean values across groups is used to estimate the minimal clinically important difference. For example, in well known medical conditions where severity stages have been determined, such as in Parkinson's disease (Hoehn and Yahr scale), a difference equivalent of moving from one stage to the next can be used as MCID. Cross-sectional methods can also be used to compare individuals with and without a particular diagnosis. For example, Johnson et al. [27] investigated differences in SF-36 scores in patients with hypertension. They found that hypertensive patients scored on average 4.1 points lower on the SF-36 compared to those without hypertension. They determined that this difference (4.1 points) could be used to establish MCID. One disadvantage of this approach is that generalizing the results to other samples can be misleading because it is difficult to control for other variables that can cause the group differences. In addition, as with all cross-sectional designs, differences in mean scores may not accurately reflect true change [25].

MCID can also be inferred by linking the results to some external, non-disease related criteria. For example, Testa and Simonson [28] suggested that a 0.1 standard deviation decrease in the General Perceived Health scale was comparable in importance to the stress associated with experiencing the death of a close friend. These external criteria can provide a useful, easy to understand, anchor for comparison; however, interpretation of the results can be difficult in some cases. Again, as in other anchor-based approaches, there must be an assumption that all other variables remain stable (do not change).

Another cross-sectional approach consists of dichotomizing patients based on functioning level after treatment (functional vs. non-functional group). Based on the principle that a patient should be in the normal range of functioning after clinical intervention, Jacobson and Truax [29] propose three possibilities for identifying recovery status. First, a patient's score after clinical intervention is 2 standard deviations better compared to the dysfunctional group. Second, the patient's score after clinical intervention is within 2 standard deviations of the functional population. Finally, the patient's score after clinical intervention is closer to the mean of the functional population than the mean of the dysfunctional population. This method bases the identification of MCID on the ability of patient to move from one category to the next.

The use of cross-sectional comparisons has received two major criticisms. First, it is likely that when comparing two groups, more than one variable may be responsible for the differences between groups. Some researchers have suggested the use of regression models to control for other possible variables [30]. Second, some researchers argue that cross-sectional differences are not always equal to longitudinal changes using the same groups [31].

### **Longitudinal methods**

This approach is used when comparing group changes across time. One of the most commonly used anchor-based approaches for establishing clinically meaningful change in longitudinal studies is the use of global ratings of change.

Jaeschke et al. [32] used this approach to investigate MCID in patients with respiratory problems. They used the Chronic Respiratory Questionnaire and the Chronic Heart Failure Questionnaire to address change in dyspnea, fatigue, and emotion. After treatment, patients were asked about their global rating of change. Based on their responses, they were classified into four groups: "no change," "minimum," "moderate," and "largest," for each domain. After

investigating the mean change in the three domains, the authors concluded that a difference of half a point constituted an MCID.

Another longitudinal method for establishing clinically meaningful change involves the prognosis of future events. This method looks at individuals who experience a particular event such as mortality, use of medical care, cost of interventions or time to discharge [33]. Differences in individuals that experience and do not experience the event are used to determine MCID.

A final method for investigating MCID is the use of receiver operating characteristic (ROC) curve. The ROC curve method attempts to discriminate between patients who do and do not achieve clinically significant change using a single cutoff point [34]. Sensitivity, probability that a test result will be positive when the disease is present (true positive rate, expressed as a percentage), is plotted against specificity, probability that a test result will be negative when the disease is not present (true negative rate, expressed as a percentage). Each point on the curve represents a different cutoff. Usually, the point where sensitivity and specificity have the highest value is chosen as a MCIF cutoff point. Some studies choose the point where sensitivity equals specificity. Both of these cutoff choices are arbitrary because they do not consider the differences in importance between false positives, false negatives, and correct identification. MCID estimates based on arbitrary ROC curve cutoffs could be very different from estimates from cutoffs that compare correct and incorrect classifications [34].

In general, anchor-based techniques offer the advantage of linking changes in the variable of interest to outside meaningful anchors. This is particularly useful when investigating results for which the patient is the major source of information, such as in pain research and quality of life investigations. However, anchor-based approaches have several limitations. When using

global ratings, it is possible to obtain results that are affected by recall bias, especially when the time delay is long. Patients may simply forget or be affected by their current life situation. In addition, this method offers no information about the reliability and validity of the responses obtained. Another potential limitation is the generalizability of results obtained with different anchors. The use of different anchors may lead to different conclusions about the amount necessary to determine MCID. Some studies have found that conclusions obtained from anchor-based methods may vary depending on whether the anchors were obtained prospectively or retrospectively [35].

Most importantly, anchor-based methods do not take into consideration the precision of the instrument used. It is possible that MCID established by this method are within the range of error in the instrument. Any change within this range cannot be attributed to the treatment or intervention. Furthermore, interpretation of results may be difficult if there is not a linear relationship between the scores and the anchor chosen [35].

### **Distribution-Based Approaches**

Distribution-based approaches to determine MCID are based on the statistical characteristics of the obtained sample. Three categories of distribution-based measures have been proposed [23]. These are: methods based on statistical significance, methods based on sample variation, and methods based on measurement precision. Methods based on statistical significance evaluate change taking into consideration the probability that this change occurred by random variation. These methods are affected by sample size. Therefore, other things being equal, increasing the sample size may yield results that are statistically significant. Two approaches that use these methods include the paired t-statistic and growth curve analysis [36]. The second category includes methods based on sample variation. Different types of variation used include baseline variation of the sample (effect size), variation of change scores

(standardized response mean), and variation of change scores in a stable group (responsiveness statistic) [37]. These methods are independent of sample size, because variation is expressed as an average variation around a mean value. The last method is based on the measurement precision of the instrument. These methods evaluate change in relation to variation of the instrument instead of variation of the sample. They include the standard error of the mean (SEM) and the responsiveness statistic. These methods are also sample-independent.

### **Paired t-Statistic**

The t-statistic is used to test the hypothesis that there is no change in the average response on a measure over two time points. The paired t-statistic has been commonly used in a one-group repeated measures design [38]. It is calculated as the difference between pre-test and post-test scores divided by the standard error of measure change [39]. A concern with the use of this method to measure individual change is the fact that it only accounts for the statistical significance of the difference. This difference depends not only on the amount of change, but also on the sample size and the variability of the measure [39]. If used to establish a cutoff point, increasing the sample size will reduce the amount of difference necessary to reach this threshold. Statistical significance is not appropriate to establish the clinical importance of a change in score.

### **Growth curve analysis**

Individual growth curve coefficients can be estimated using hierarchical linear modeling [40]. Improvement rates are calculated by dividing the empirical Bayes estimated linear slope by the empirical Bayes estimated posterior standard error of the slope [40]. This method, like the t-statistic, is influenced by sample size. Speer and Greenbaum [41] claim that this method performs better than other distribution-based methods because it uses all data points to establish rates of change. They also report that a limitation of this method is the fact that it requires large samples to provide stable estimates of change [41]. Another limitation of this method is the

assumption of not having any missing data points or that missing data points are randomly missing. Violations of this assumption can result in biased conclusions [42].

## **Methods Based on Sample Variation**

### **Effect size**

Effect size is a broad name given to a number of indices that measure the magnitude of a treatment effect. Unlike previously presented significance tests, these indices are independent of sample size. Cohen [43] defined the standardized difference between two groups as the difference between the means, divided by the standard deviation of either group. Cohen concluded that the standard deviation of either group could be used when the variances of the two groups are homogeneous. He also established guidelines for the interpretation of effect size; for example, .20 for “small” effects, .50 for “moderate” effects, and .80 for “large” effects. Some researchers have investigated MCID based on effect size. Samsa and colleagues [22] propose an effect size of .20 as an appropriate definition of a MCID. Some limitations of using the effect size include the need for a homogeneous distribution; the size of the standard deviation, either at base line or after treatment, will have an inverse effect on effect size; and with large standard deviations producing smaller effect sizes [44].

### **Standardized Response Mean (SRM)**

SRM is defined as mean score change divided by the standard deviation of that score change [38]. A large SRM indicates that the change is large relative to the background variability in the measurements [45]. The SRM also uses cutoff points of .20, .50, and .80 to define small, moderate, and large effect sizes [46]. One limitation of the SRM is that comparable individual changes that have different SRM values depending on the variability of change in the sample [45].

## **Methods Based on Measurement Precision**

### **Standard Error of Mean (SEM)**

The SEM is a measure of the precision of an instrument. Therefore it is closely related to the concept of minimal detectable change. The SEM is the standard error in an observed score when the true score is not captured by the instrument used. In other words, it indicates how close a person's score is to their true score; the score that they would get if a test could be completely error-free [47]. It is calculated by using the sample standard deviation and the sample reliability coefficient. The exact formula estimates the SEM as the standard deviation of the instrument multiplied by the square root of one minus its reliability coefficient [47]. The SEM is considered to be an attribute of the measure and not a characteristic of the sample [47]. It is possible that the SEM of a particular measure would vary based on the method used to estimate the reliability coefficient and the presence of extreme scores. For MCID, SEM values of 1 SEM, 1.96 SEM, and 2.77 SEM have been suggested [48-49]. The SEM is expressed in the original metric of the measure it describes. This is important because it can help with interpretation of the results. Moreover, the SEM is a theoretically fixed parameter of a measure [49]. This means that for nearly all true scores, the deviation around the true score from repeated measurements is about the same [50].

### **Reliable Change (RC)**

A reliable change index is based on the amount of change that indicates the extent to which the observed change exceeds measurement error [50]. This index is referred to as the standard error of measurement difference (SEMD). The SEMD is directly related to the SEM, but produces smaller values. Therefore, this method is more conservative for a given cutoff value than the SEM approach, classifying fewer individuals as improved or deteriorated. A cutoff value of 1.96 has been suggested to determine whether an observed change in scores over time

should be categorized as unchanged, improved, or deteriorated [50]. A disadvantage of this method is that it assumes that measurement error is constant across the range of possible scores [50].

### **Choosing an Appropriate Method**

A number of methods have been proposed to investigate MCID. Anchor-based approaches have the advantage of linking changes to a meaningful external anchor. In addition, some of these methods include the most important measure of the significance of change; the patient's perspective. However, these methods do not consider the possible range of error associated with all instruments. In addition, interpretability of results is difficult when comparing investigations that use different external anchors. On the other hand, distribution based approaches provide a way to establish amount of change outside the limits of the instrument's error. In addition, these approaches provide a common metric that has equivalent meaning across measures, populations, and studies [50]. The distribution-based approaches that are better suited for establishing MCID are those based on the measurement precision of the instrument (SEM, RC). These measures establish the amount of error that is inherent to the instrument and the amount of random error that can be expected in repeated measures. In addition, they are not influenced by variability in the sample at baseline (as is the effect size), variability of the observed change (as is the Responsiveness Statistic), or the sample size (as are the t-statistic and growth curve analysis). Finally, these measures can be used to establish cutoff points based on a desired confidence level [50].

The first experiment in this dissertation uses the SEM method to investigate differences in balance scores (Berg Balance Scale and Dynamic Gait Index) that represent a minimal detectable change. The SEM is a measure of responsiveness and can also imply reliability of an instrument. The SEM expresses measurement error in the same units of the original tool and is

not influenced by variability among subjects. This technique is well suited for clinical practice. It provides an intuitive score, in the same unit of the original instrument, which can be used across measurements and populations. In addition, because the SEM looks at within individual variability, this technique is more appropriate than other responsiveness approaches when therapists need to make a decision about an individual [51].

### **Patient Reported Outcomes (PRO)**

Health professionals' approach to patients and their problems is greatly influenced by the conceptual models around which their knowledge is organized. Traditionally, the effects of a particular disease or medical condition have been assessed by using methods such as physiological exams, performance tests, and clinical observations. These methods have been employed, and still are, to fulfill the requirements of a conceptual model described as the "Biomedical Model". In this model, a disease or medical condition is the result of a pathophysiological event, intrinsic to the individual, resulting in a reduction of the individual's quality of life. As a result, curing or managing a disease or medical condition revolves around identifying the disease, understanding it, and learning to control and alter its course [52]. "Biopsychosocial Models", such as the one used to guide this dissertation, the ICF, expands the concept of disease to include not only the pathophysiological event, but also the psychological and societal consequences of the disease process [53]. With the acceptance of this new conceptual model, comes a need to develop tools that can effectively measure individual's social and personal factors that affects the disease process and can play an important role in the diagnostic and rehabilitation process.

Measuring the impact of social and psychological factors in the disease process requires participation of the patient. For some medical conditions the patient is the only source of information. For example, depression is a condition that, often, has no observable or measurable

physical symptoms. Clinicians must rely on information provided by the patient to correctly diagnose and treat patients with this medical condition. In addition, to establish endpoints in the rehabilitation of these patients, clinicians must use tools that take into consideration the patient's perspective. Patient reported outcomes (PRO) instruments become indispensable in these situations.

### **PRO Instruments**

PRO has been defined by the Federal Drug Administration as: "Any report coming directly from patients (i.e., study subjects) about a health condition and its treatment" [54]. PRO instruments are used to measure treatment benefits by capturing concepts related to how a patient feels or functions with respect to his or her health or condition. The ideas, activities, behaviors, or feelings measured by PRO instruments can be either verifiable in nature, such as walking, or can be non-observable, known only to the patient, such as pain, depression etc. Although these symptoms are highly dependent on the patient perception, historically, these assessments were often made by clinicians who observed and interacted with patients. Recently, these kinds of assessments are increasingly performed with PRO instruments.

The idea of asking patients about their feelings and symptoms is not new. In fact, doctors have used this technique throughout the history of the medical profession [55]. What makes PRO instruments different is the fact that information about symptoms and performance is being obtained directly from patients. This is done without interpretation from clinicians, using structured questionnaires that are shown to give reproducible, meaningful, and quantitative assessments of how patients feel and how they function [56]. Therefore, the adequacy of a PRO instrument is based on its ability to capture the patient's evaluation of the impact of disease on their functioning and well-being [57-58]. For this reason, PRO instruments can be categorized

under a new set of instruments that uses a conceptual framework where the patient is the focus of the assessment.

### **PROs and the International Classification of Function (ICF)**

The ICF is a thorough framework that can be used as a reference to evaluate, compare and classify instruments used in the rehabilitation field. In addition, the ICF is a classification system and can be used to develop new instruments, using the coding system it provides.

Because the ICF was developed to include social, personal and environmental aspects of health and disability, it is an ideal framework for the evaluation of PRO instruments.

PRO instruments provide information about how patients feel or function with respect to their health or condition. Information generated by a PRO instrument can serve to measure treatment benefit, from the patient perspective. To arrive at this conclusion, there must be evidence that the PRO instrument is based on a valid theoretical construct. In fact, one of the most important psychometric properties of a tool is its construct validity. Construct validity refers to the degree to which inferences can legitimately be made from the measures in a study to the theoretical constructs on which those measures were based [59]. PRO instruments can be used to measure simple constructs such as the ability to perform activities of daily living or more complex constructs such as quality of life, which includes physical, psychological and social components.

The ICF model can be used as a reference to investigate what areas of the rehabilitation spectrum are covered by PRO instruments. Activities and participation (right side of the model) are the areas of the ICF model measured by PRO instruments. For example, one of the most widely used PRO instruments is the SF-36 [60]. This instrument measures generic health status in the general population. Patients are asked to evaluate their general health and limitations in activities as a result of their physical health or emotional problems. Although this instrument was

not developed using the ICF as a reference, it covers many areas of the model. Questions such as: “Does your health now limit you in these activities? If so, how much?” or “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?” assess the ICF areas of activity and participation. Although personal factors are not directly assessed with this instrument, this must be considered when interpreting the results. Individuals’ perception of their mental and physical function is heavily influenced by personal factors such as their personality, cultural background, and societal roles. Except for measuring the influence of the environment, the SF-36 covers important psycho-social areas of the ICF model.

Other PRO instruments are specifically designed using the ICF as a theoretical model and the code it provides as the basis for classification. For example, the PAR-PRO, a measure of home and community participation, was developed as a broad measure of home and community involvement for persons with disabilities [60]. This instrument is based on the domains of activities and participation in the ICF, including learning and applying knowledge, communication, mobility, self-care, domestic life, interpersonal interactions and relationships, major life areas, and community social and civic life. A particular challenge of the ICF is to be able to distinguish between activities and participation when using the coding in the activities and participation domains. Often, the only possible indicator of participation is coding through performance, which might also have to be coded as frequency of participation in a particular activity [60]. However, instruments such as the PAR-PRO demonstrate the potential for using the extensive classification system of the ICF for the development of PRO instruments. In this dissertation, a questionnaire to assess the patients’ perspective was developed using the ICF

classification system to assess areas of the model of particular importance to individuals undergoing gait and balance rehabilitation related to falls.

### **Advantages of Using PROs**

Many chronic conditions have a debilitating effect that progressively deteriorates the patients' quality of life. Chronic diseases have social, personal and mental implications that can lead to fatigue, depression, pain, and isolation. Traditional measurement tools (e.g. blood pressure instruments, or blood sugar level counts), are very accurate at measuring physiological function. However, a new set of sensitive and well-validated tools are needed to improve and standardize measurements of symptoms related to social, personal, and environmental factors associated with health. In general, the use of the patient's perspective becomes more important as the variable being measured becomes more abstract. Concepts that do not have a well define physiological component such as pain, quality of life, life satisfaction, or self-efficacy, can be better understood by considering the patients' perspective.

PRO instruments are ideal and, sometimes, the only way to measure ICF concepts of activity, participation and the influence of personal and environmental factors in health and disability. For example, when measuring quality of life (QOL), a concept that is highly individual and context dependent, we have no choice but to use PRO instruments. In fact, some researchers and clinicians use these two terms indistinctively [61]. QOL or health related quality of life (HQRL) questionnaires are PRO instruments that explicitly include the patient's perception of the broad impact of disease on their functioning and overall wellbeing [62]. Ultimately, the goal of a therapeutic intervention is to impact health and increase function and quality of life. Consequently, the use of PRO instruments in clinical practice to determine end points and evaluate treatment effectiveness is critical. This has been recognized by many national and international agencies (FDA [54], NIH [63], European Agency for the Evaluation of Medical

Products [10], WHO [65]) that identify the need to use PRO instruments to improve health and quality of health care.

There is enough evidence that support the use of PRO instruments, especially those that assess QOL, as a treatment outcome. In fact, QOL has been a strong prognostic variable for survival in several cancer related studies [66-67]. QOL data can be especially important when two treatment options with similar survival outcomes are available. In these cases, QOL outcomes can be the deciding factor for choosing a particular treatment. For example, a woman with breast cancer might face a decision of whether to opt for a mastectomy or conservation of the breast. Both treatment options produce similar survival outcomes, but each has implications for QOL that will determine the woman's ultimate decision. For some women avoiding the radiation therapy required to conserve the breast is important. For others, conservation of the breast is of most importance. Even when survival outcomes are different, some patients may select a less effective treatment because of the effect the treatment may have on their QOL [68]. It is clear that the availability of QOL data is essential for making a balanced and informed decision about treatment options.

PRO instrument can also play a role in the diagnose and treatment effect of health issues related to the ICF domain of body function. Some symptoms and treatment effects are not measurable and only known to the patient. Concepts such as pain intensity or pain relieve are not observable and have no direct physical manifestation. Again, PRO instruments are needed to assess these areas. This is important because, sometimes, improvements in a particular clinical measurement may not correlate with how the patient functions or feels. For instance, a patient can demonstrate an improvement in a test of muscle strength, but this may not correlate with improvements in walking or impact the patient's ability to perform daily activities.

The pain literature provides valuable information on the importance of patient's perceptions and perspectives. Hodgkins and Daltroy [69] investigated the assessment of pain by physicians and patients. They found that physicians' rating of pain is generally lower than that of the patients. In addition, male physicians tended to rate female patients' pain lower than that of male patients. The patients' perception of their pain can also be variable. When distracted, a patient may provide a lower rating of their pain than when they focus on the pain sensation. Inconsistency in how the patients evaluate a particular condition does not change the value of the information obtained when using PRO instruments. The patients' experience is what the patient says it is at any given time [69]. This information is valuable for diagnosing and treating the condition. It can lead to a better understanding of the nature of the experience and how patients' personal factors can affect their perception.

PRO instruments are often developed to measure what patients want and expect from their treatment and what is most important to them [54]. The patient's perspective is critical to evaluate treatment effect and patient satisfaction with treatment. Ideally, a treatment intervention or health strategy should be aimed at addressing all aspects of the construct of health. PRO instruments allow clinicians and researchers to assess an area of the model that was previously not well understood because of its abstract nature. Adding these instruments to physiological clinical measures will help to obtain a more complete picture of the patient's health and how it is affected by their personal and societal circumstances.

As previously mentioned, seeking information from the patients about their health condition and how it affects their function and participation is not new. However, PRO instruments provide a formal assessment that may be more reliable than the traditionally used informal patient interview. PRO instruments use a predetermined format to minimize

measurement error and ensure consistency [54]. Instruments can be self-reported or clinician administered. In the first case, PRO instruments avoid possible clinicians' bias and offer an unfiltered response that reflects more closely how the patients rate their health. Well-developed and adequately validated PRO instruments have been shown to provide information that matches the results obtained by experts in the particular field of interest. In fact, often, this is the method used to study the validity of PRO instruments [54].

### **Issues and Concerns about the Use of PRO**

The use of PRO instruments to evaluate health and disability is widely accepted in the medical and research community [70]. They add value to traditional clinical assessments and offer a unique perspective of the patient experience. There are few disadvantages for the use of PRO instruments; however several methodological, theoretical, and practical considerations must be critically reviewed to ensure the information obtained is accurate and useful.

One major concern with the use of PRO is that of definition. PRO instruments are commonly used to assess concepts of activity, participation, and social and personal interactions. These concepts are abstract in nature; therefore generalization of findings obtained with the use of PRO instruments must be done with caution. For example, the concept of quality of life (QOL) is receiving increased attention in the clinical and research arena. QOL is an abstract concept for which it is impossible to create a single instrument that assesses it. Researchers have attempted to solve this issue by using a narrower definition of QOL. The idea of health related quality of life comes from these efforts of having a more precise definition of the concept being measured. Another approach to this issue is to create specific instruments that relate to a particular health condition. This serves a double purpose because it also helps to reduce the ceiling and floor effect experienced when instruments are too general.

In the past few decades researchers have adopted this strategy and developed a number of instruments aiming at particular health conditions. The list of instruments is so extensive that some organizations, realizing the difficulty of having access to these instruments, have created databases where clinicians and researchers can search using criteria such as disease, patient population, or type of instruments. For example, the ProQolid database, developed by Mapi Research Institute, contains over 500 instruments and aims at identifying and describing PRO and QOL instruments to help researchers and clinicians choose appropriate instruments and facilitate access to them.

Another concern relates to the administration of PRO instruments. The use of self or clinician administered questionnaires is difficult in certain patients with communication and cognitive impairments. For example, patients with stroke often have speech or cognitive deficits that make the use of PRO instruments difficult. Similar difficulties are encountered when using these instruments with those who have a low education level, do not speak the language, or come from a different cultural background. Of particular interest is the issue of cultural relevance, especially when people from different cultural backgrounds are compared using the same instrument. PRO instruments require an internal evaluation of several aspects of one's life and how these aspects are influenced by the health condition of interest. These values are influenced by the patient's culture and previous experiences. For example, in the pain literature there is enough evidence to support differences in the pain experience based on ethnic, social, gender, and geographical factors [71].

Another issue to consider is the psychometric characteristics of some PRO measures. There are concerns in the research literature about some PRO measures being inadequately conceptualized, lacking psychometric rigor, and having inconsistently applied psychometric

methods [72]. To solve this issue, some national and international organizations have released guidelines for the creation and evaluation of PRO instruments. Recently, the Scientific Advisory Committee (SAC) of the Medical Outcomes Trust created a document to guide in the evaluation of PRO instruments. In this guide the SAC states that PROs should be evaluated on the following seven dimensions; 1) the use of pre-specified conceptual and measurement models; 2) the strength of empirical support for the reliability and validity of the scale(s); 3) the responsiveness of PRO to clinical change; 4) the method(s) for interpreting scores; 5) the level of respondent and administrative burden; 6) the equivalence of alternative forms of administration; and 7) the rigor with which translations are adapted for use in specific cultural contexts [73]. This comprehensive list is useful for researchers interested in evaluating PRO instruments, but may have little use for clinicians. There is a need to create a way of translating these guidelines efforts into clinical practice.

Patient reported outcomes (PROs) are a necessary and valid way of including the patients' perspective into research and clinical practice. Using the International Classification of Functioning, Disability, and Health (ICF) model clinicians and researchers can evaluate existing PRO instruments, and propose new instruments to assess particular areas of the model. PRO instruments must be carefully selected to meet the needs of the specific population of interest. No single instrument has universal application in health assessment.

There are some concerns about the psychometric properties of PRO instruments. These instruments are often used to explore abstract concepts that are difficult to interpret and conceptualize. Assigning a single score value to broad concepts such as quality of life may be an oversimplification of a very complex component of life. However, researchers have found strategies to minimize this problem. Using instruments that are disease and population specific,

based on sound theoretical concepts, reliable, responsive to clinically important change, and culturally sensitive may ameliorate some of the concerns expressed by members of the research community.

PRO measurements, such as participation, satisfaction with care, and quality of life are increasingly being required by accreditation agencies, patients, policy-makers, and medical insurers as quality indicators of best practice [73]. These groups recognize the need to empower patients by giving them the opportunity to include their perspective in the selection of treatments and evaluation of treatment outcomes. To my knowledge, the second study in this dissertation is an attempt at using an ICF-based PRO instrument to assess expectations and patient satisfaction in the rehabilitation of falls.

### **Summary**

The number of people 65 and older is likely to continue rising in the future. Falls are already a major health concern in this population and will likely continue to be a burden to elders and the health system. To offer the best possible, most efficient medical care to the increasing number of elders, clinicians should use the best assessment instruments available and include evidence base information in their clinical practice. In addition, to provide patients the treatment they deserve, clinicians need to understand the patient's expectations and goals. Patient satisfaction is the ultimate goal in any service provider, including medical services.

In research, most rehabilitation interventions are evaluated by using some test of statistical significance. Assessment instruments used during these interventions provide a measure of performance that can be used to assess disease severity or monitor improvement. Tests of statistical significance offer important information about groups of patients, but fail to capture differences at the individual level. In recent times, a number of investigators are looking at other statistical methods that can provide information about clinically relevant change [29, 41,

47, 50]. This dissertation is a first attempt at investigating clinically meaningful changes in measures of gait and balance used to evaluate geriatric patients who have fallen or are at risk of falling.

In addition, a new set of instruments are also receiving a great deal of attention from the research community. Recognizing the need to empower patients and understand their personal needs, patient reported outcomes are been included in clinical trials to assess the relevance of the intervention, from the patient's perspective. In fact, during the inaugural ceremony of the new NIH PRO initiative Director Elias A. Zerhouni said 'There is a pressing need to better quantify clinically important symptoms and outcomes that are now difficult to measure. Clinical measures of outcome such as x rays and lab tests have minimal relevance to the day-to-day functioning of patients with such chronic diseases as arthritis, multiple sclerosis, and asthma, as well as chronic pain conditions' (NIH, 2005). This dissertation is a first attempt at using a PRO instrument to assess expectations and success criteria, from the patient's perspective, of a rehabilitation program for patients with gait and balance problems who have fallen or are at risk of falling.

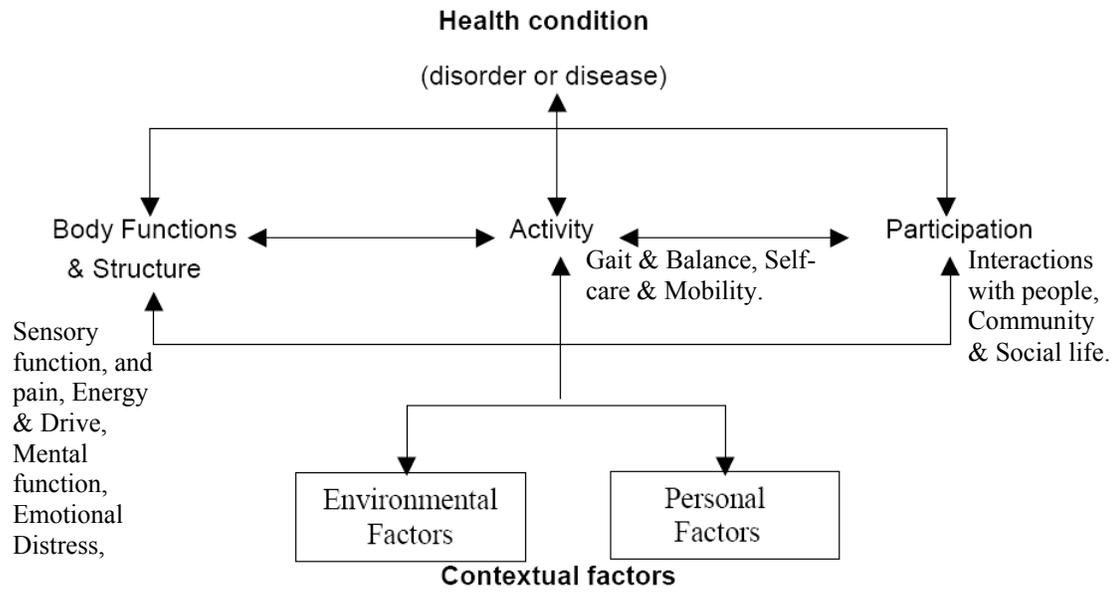


Figure 1-1. International Classification of Functioning Disability and Health (ICF) model

## CHAPTER 2 ESTABLISHING THE MINIMUM DETECTABLE CHANGE (MDC) FOR THE BERG BALANCE SCALE AND DYNAMIC GAIT INDEX

### **Introduction**

Most rehabilitation efforts to treat falls include a component of physical rehabilitation. These treatment plans are based on assessments performed with instruments specifically developed to obtain information related to the condition of interest. Therapists assess and reassess patients to identify specific problem areas and establish improvement criteria. Two assessment tools commonly used in physical therapy, and specifically in the evaluation of elder individuals who have fallen or are at risk of falling, are the Berg Balance Scale (BBS) and the Dynamic Gait index (DGI) [74].

### **Gait and Balance Assessment**

#### **The Berg Balance Scale (BBS)**

The BBS is a frequently used performance-based scale that assesses postural balance [75]. The test consists of 14 commonly used tasks: Sitting to standing, standing unsupported, sitting unsupported, standing to sitting, transfers, standing with eyes closed, standing with feet together, reaching forward with outstretched arm, retrieving an object from floor, turning to look behind, turning 360°, placing alternate foot on stool, standing with 1 foot in front, and standing on 1 foot. The scoring method is based on a 5-point ordinal scale of 0 (indicates the lowest level of function) to 4 (indicates the highest level of function), with the total score ranging from 0 to 56. The BBS was specifically designed to be used at the clinic. It requires minimal equipment (stopwatch, chair, stool and ruler) and can be applied in under 15 minutes.

A considerable amount of evidence suggests that the BBS is a valid measure of standing balance. Initially, Berg et al. [75] correlated BBS scores with a general rating of balance made by therapist (Pearson  $r=.81$ ). Other studies by the same author have also demonstrated high

correlation values between the BBS and other measures of balance. For instance, the Pearson  $r$  correlations between the BBS and the balance subscale of the Tinetti Performance-Oriented Mobility Assessment and the Barthel Index mobility subscale were .91 and .67 respectively [76]. Other researchers have also found high correlations between BBS scores and other motor and functional measurements: Fugl-Meyer Test motor and balance subscales (Pearson  $r$  = .62–.94), Timed Up & Go Test (TUG) scores (Pearson  $r$  = .76), Emory Functional Ambulation Profile (Pearson  $r$  = .60), and Gait Speed (Pearson  $r$  = .81) [77- 78]. The BBS scores also correlated moderately with data obtained for the Dynamic Gait Index (Spearman coefficient = .67), and center-of-pressure measures (–.40 to –.67 [Kendall coefficient of variance]) [78].

Several studies have also reported high intra- and inter-rater reliability for the BBS. Berg et al. [79] used videotaped evaluations of the BBS to obtain inter-rater reliability (ICC = .98 for total BBS scores). The same researchers replicated these results in a test-retest format, producing a within rater ICC = .97 and between rater ICC = .98. The large majority of studies investigating reliability of the BBS have used some form of correlation coefficient such as the Pearson's Product-Moment Correlation Coefficient ( $r$ ) and the intra-class correlation coefficient [ICC], with the latest becoming more popular in recent times. A fundamental problem of ratio indexes such as the ICC is that the error of measurement and true variability are expressed in relative terms. An ICC score is a ratio of within subject and between subject variability. Thus, the range of genuine differences in any attribute is sample dependent. Therefore, previously reported high ICC values for the BBS must be considered with caution.

### **The Dynamic Gait Index (DGI)**

The DGI was developed by Shumway-Cook and Woollacott [80] to assess balance in the older adult at risk for falling. This functional gait scale consists of 8 common gait tasks: walking at different speeds on a level surface, walking with horizontal and vertical head turns, ambulating

over and around obstacles, ascending and descending stairs, and making quick turns. Each item is scored on a 4-level ordinal scale, where 3= “normal”, 2= minimal impairment, 1= moderate impairment and 0 = severe impairment. The maximum possible score is 24 points. The DGI can be administered in 10 minutes and requires minimal equipment.

The psychometric properties of the DGI have not been extensively investigated. Validity of the scale has been supported by moderate correlation with the BBS (Spearman rank order correlation,  $r = 0.71$ ), [81]. Sensitivity and specificity to identify individuals with a history of falls has been established at 59% and 64% respectively [82]. The test developers investigated the inter-rater and test-retest reliability of the scale using a small sample of 5 older adults and 5 raters. They found ICC values of .96 (inter-rater) and .98 when subjects were re-tested a week later by 2 therapists. Intra-rater reliability has not been reported in the literature. Despite being widely used in the clinic, the psychometric properties of the DGI have not been investigated sufficiently.

The BBS and DGI are used to assess different dimensions of balance (i.e. static and dynamic balance). Since the assessment of balance control is a crucial step to identify individuals who are at risk of falling, these two instruments are often the central components of the physical therapy evaluation. Scores from these instruments are regularly used to determine a particular treatment and to monitor improvement. Moreover, low scores in these instruments have been associated with an increased risk of falling in the elder population (i.e., a BBS score of <45 and DGI of <19) [76, 82-84]. Therapists use these cut-off points as a reference to guide treatment and monitor progress. Cut-off scores are taken as absolute values to determine the success of a particular intervention and are often used to report patient progress. For example, a patient who improves from an initial BBS score of 40 to a final score of 46 can be reported as being outside

of the range associated with a high risk of falling. This anchor-based method must be used with caution, because assessment instruments are not always responsive enough to detect small changes in performance, and these changes can be masked by the multiple sources of error associated with the instrument.

Unfortunately, when results differ from one assessment to the next, it cannot be assumed that true change has occurred; some or all of the change could be attributed to measurement error. Error can be inherent to the test used or represent the naturally existing fluctuation in patients' performance. The amount of error across measurements of the same test is related to the reliability of the test. Reliability is the ability of a particular test to consistently provide the same value when no change has occurred [85]. It is also a measurement of the objectivity of the test. There are several statistical methods that have been used to measure reliability. They can be divided in two major groups: measures of relative and absolute reliability. Relative reliability refers to the degree of association between repeated measurements. In other words, relative reliability measures the strength of the correlation between repeated measures. It takes into account the total group variability (between subject/measurement) and the individual measurements variability (within subject/measurement) to obtain a correlation coefficient, for example, the Pearson correlation coefficient or the Intra-class correlation coefficient (ICC) [86]. Absolute reliability refers to the variability of the scores from measurement to measurement (within subject/measurement). This approach does not take into account the range of individual scores and is not sample-dependent [87]. Some of the tools used to calculate absolute reliability include: the coefficient of variation (CV) and the standard error of the measurement (SEM), [88-91].

While correlation methods used to calculate relative reliability are excellent sources of information to compare groups of patients, the SEM is more appropriate for clinical practice, that is, when making decisions about individual patients [90]. However, to date, only one published study has used a measurement of absolute reliability to look at the psychometric properties of the BBS and no published study has addressed this issue with the DGI. Stevenson [91] used the SEM to investigate error associated with the use of the BBS in stroke patients. He found a SEM (in BBS units) of 2.49 in patients with stroke receiving inpatient rehabilitation. In addition, he calculated a confidence interval around the SEM and found that a change of 6 BBS points was needed to be 90% confident of genuine change. This finding is somewhat surprising and questions previously reported high BBS reliability scores. Further investigation in different populations is needed.

#### **The Intraclass Correlation Coefficient (ICC), the Standard Error of Measurement (SEM) and the Minimal Detectable Change (MDC)**

The ICC, or intraclass correlation coefficient, is the most commonly reported reliability measure in the literature. The ICC provides information about the measure's ability to differentiate among subjects. The ICC, as with other correlation coefficients, provides values between 1.0 and -1.0, with high absolute values indicating less variability between scores. The ICC incorporates total variability (between subject/measurement), and error associated with it, and the individual variability (within subject/measurement) to obtain a ratio. This technique is most appropriate for investigating differences between groups of patients.

A less frequently used reliability index is the standard error of measure (SEM). While the ICC expresses the proportion of variance of an observation due to between-subject variability in the true scores, the SEM is a measure of within subject variability. The SEM expresses measurement error in the same unit of the original tool and is not influenced by variability

among subjects. The SEM is closely related to the concept of minimal detectable change (MDC) expressed by Stratford et al [92]. The MDC is the amount of change in a given measure that must be obtained for a clinician to determine that true change has occurred. The MDC is expressed as a confidence interval around the SEM, indicating the values that are within the range of error attributable to the measuring instrument. The MDC provides the clinician useful and easy to understand criterion for change in patients' performance.

Therefore, investigations are needed to determine the amount of change in the Berg Balance Scale and Dynamic Gait Index necessary for a therapist to conclude that "true" change has occurred (MDC). In addition, understanding how MDC values change at different score levels of the BBS and DGI (cut-off points, <45 and <19 respectively) may prove to be valuable for clinicians that are faced with making treatment decisions based on these values.

Thus, the purpose of this study was to use the standard error of measurement (SEM) to investigate the Minimal Detectable Change associated with the Berg Balance Scale and Dynamic Gait Index. In this study, I attempted to improve on previous reliability investigations by providing clinicians with estimates of measurement error that are easy to interpret and can be used to make clinical decisions.

## **Methods**

### **Subjects**

The sample consisted of 42 subjects (26 Males and 16 Females, age 55 and older) participating in a larger, funded, research study looking at the link between smoking and recovery from frailty in older Floridians. This study was supported by a grant from the Florida Department of Health and received approval of the Institutional Review Board for the University of Florida and the Research and Development Committee at the North Florida/South Georgia VA Medical Center. Inclusion criteria included: community dwellers with a history of falling

twice in the past 12 months, the ability to walk 20 feet (with or without an assistive device), and a score of 24 or higher on the Mini-Mental State Exam [97].

Participants were recruited in two different ways. First, patients from a Gait and Balance disorders clinic at the North Florida/South Georgia VA Medical Center were approached by their therapists and asked if they were interested in participating in a research study. If the patients agreed, a research coordinator explained the study and, if still interested, consented the patients. Secondly, letters were sent to doctor's offices explaining the study and flyers were distributed throughout the community. Participants responded to the advertisement and, if they met the study's inclusion criteria, were enrolled in the study.

### **Testing Procedure**

Enrolled participants were tested with an extensive battery of tools including: Berg Balance Scale (BBS), Dynamic Gait Index (DGI), isometric lower extremity strength, physical function domain of the MOS36, Falls Efficacy Scale, Geriatric Depression Score, pain experience VAS, timed tests of gait, Frenchay IADL scale, Pleasant Event Schedule, and spontaneous self-selected gross motor activity as measured by accelerometry. After the initial assessment, a home exercise program was prescribed. Participants were instructed to keep exercise logs to record adherence to the program. Participants were monitored for 3 months. During this time, a total of 3 evaluations were performed at 4, 8 and 12 weeks. In each of these evaluations, participants received the previously mentioned battery of assessment tools. In addition, exercise logs were collected.

Evaluations were conducted by two experienced physical therapists specialized in gait and balance disorders in the elder population (>7 years experience in geriatric Physical Therapy). From each evaluation, the BBS and DGI tests were videotaped by a research assistant with a Sony DCR-VX2100 digital camcorder. Videotapes were converted to DVD format for later

view. Recorded sessions were re-scored at a later time (time between initial and re-scores > two weeks) by the same therapists. Therapists used a TV screen or computer monitor to view the recorded evaluations. During the re-scoring of the videotaped evaluations, therapists were allowed to pause, play in slow motion, and/or replay any portion of the evaluation they were unsure about. Therapists were blinded to previous score and whether the recordings were from an initial, 4 weeks, 8 weeks, or 12 weeks evaluation. For the purpose of this study, only initial evaluations were used. All participants were assessed and re-assessed by the same therapist. Data was recorded by the physical therapists and research assistant. All data was later entered into a central database.

### **Analysis**

All statistical analysis and graphical representations were performed with SPSS 13.0 software for Windows (SPSS Inc., Chicago, IL, USA) and Microsoft Office Excel software for Windows (Microsoft Corporation, Redmond, Washington, USA).

Box plots were used to investigate the presence of outliers in the data (figures 2-1, 2-2). The distribution of the absolute differences between tests (initial BBS and DGI and re-scored BBS and DGI) was plotted. Cases with values between 1.5 and 3 box lengths (interquartile range) from the upper or lower edge of the box were considered mild outliers. Cases with values more than 3 box lengths from the upper or lower edge of the box were considered extreme outliers. For the BBS data, 3 outliers were identified. The DGI data did not present any outliers. Since no extreme outliers were observed in either data-set, all scores were considered valid for subsequent analysis.

The procedure suggested by Stratford [86, 92] was used to calculate the standard error of measurement (SEM), also referred as the absolute reliability. The equation for the SEM is:

$SEM = SD \sqrt{1 - ICC}$ , where SD= sample standard deviation, and ICC= intraclass correlation coefficient. However, Stratford stated that SEM can also be calculated from the square root of the mean square error term in a repeated measure ANOVA. In addition, the SEM was used to calculate the Minimal Detectable Change (MDC). The MDC is the product of the SEM, the tabled z-score for a desired confidence interval and the  $\sqrt{2}$ . The  $\sqrt{2}$  term acknowledges two measurements are being compared. For a 95% confidence interval the  $MDC = SEM * 1.96 * \sqrt{2}$  (1.96= Z-value associated with a two-sided 95% confidence interval). Confidence intervals were also calculated with a 90% and 80% ( $MDC(90\%) = SEM * 1.645 * \sqrt{2}$ , and  $MDC(80\%) = SEM * 1.28 * \sqrt{2}$ ). The use of a parametric test (ANOVA) requires that the data meet the normality assumption. Normality was visually explored with Normal Q-Q plots and tested with the Kolmogorov-Smirnov normality test. In addition, the use of the SEM, because it assumes a normal distribution of error, requires that the measurement error is not related to the magnitude of the measured variable. This is referred as heteroscedasticity. Heteroscedastic data shows a relationship between the amount of measurement error and the magnitude of the measurement. Heteroscedasticity was formally examined by plotting the absolute differences between initial value and re-scored value, against the mean score. Additionally, Spearman's rho correlation was used to rule out a relationship between each individual's absolute score difference and his or her mean.

The SEM and MDC procedures described above were also used to investigate the amount of error associated with individuals at different levels of the BBS and DGI rating scale. Because the "true" score in these two assessments is unknown, the mean value between the initial and re-scored values of the BBS and DGI was used to dichotomize the participants in two groups.

Commonly used cut-off points (<45 for the BBS and <19 for the DGI) were used to form the groups. The SEM and MDC were calculated for the four resulting groups.

Finally, a correlation analysis was performed to investigate a possible relationship between individual's difference (BBS initial – BBS re-scored and DGI initial –DGI re-scored) and absolute difference in the BBS and the DGI. To acknowledge the ordinal nature of the BBS and DGI, a non-parametric Spearman's rho correlation analysis was selected.

## **Results**

A total of 42 participants were assessed with the BBS and the DGI. The average age was 75.6 years (range 59 to 88 years). The ratio of males to females was 26 males (62 %) and 16 females (38%). The participant's mean initial BBS score was 40.7 points (SD=7.3, range 18-53). The re-scored mean value was 41.8 points (SD=7.5, range 24-55). For the DGI, the mean initial value was 13.4 (SD=4.2, range 3-21), and the re-scored mean was 13.1 (SD=4.3, range 4-22). A distribution of the absolute difference between initial and re-scored values is found in table 2.1 for the BBS and table 2.2 for the DGI. The mean absolute difference was 2.57 (SD=2.4, range 0-11) for the BBS, and 1.29 (SD=.99, range 0-3).

The distribution of absolute values of the difference between initial BBS and re-scored BBS was investigated with a box plot (Figure 2.1). For the BBS, three participants' scores were identified as outliers. Their absolute values were 8, 11, and 6 respectively. These scores were considered mild outliers because their values laid between 1.5 times and 3.0 times the interquartile range below the first quartile or above the third quartile. Therefore these 3 scores were included in all subsequent analysis. The distribution of absolute values of the difference between initial DGI and re-scored scores for the DGI was also explored with a Box plot (Figure 2.2). No outliers were identified in this case.

The absolute differences in the BBS ranged from 0 to 11 points (Table 2.1). Fifty seven percent of the participants had a BBS absolute difference of 2 BBS points or less. The mean absolute difference was 2.45 BBS points. A graphical representation of the distribution of score differences between initial and re-scored BBS is presented in figure 2.3. For the DGI, the absolute differences in scores ranged from 0 to 3. Seventy four percent of participants had a difference in score of 1 or less DGI points. The mean absolute difference was 1.13 DGI points. Figure 2.4 shows a graphical representation of the differences in DGI score between the 2 testing scenarios.

The distribution of scores from the BBS and DGI were visually inspected for normality using normal Q-Q plots. Both distributions showed no significant departure of the data from normality. Formal analysis of the data with a Kolmogorov-Smirnov test showed no significant departures from normality ( $D_{42} = .143, p > .05$ ,  $D_{42} = .09, p > .05$ , for the BBS and DGI respectively).

Visual inspection of the distribution of absolute values of the difference between the two test conditions plotted against their mean, show that the data was reasonably homoscedastic for both the BBS (figure 2-3) and DGI (figure 2-4). Spearman's rho correlations (BBS,  $r_s = -.09, p > .05$ , and DGI,  $r_s = .02, p > .05$ ) confirmed the lack of relationship between the mean score and the difference in scores (initial minus re-scored values).

Repeated measures ANOVA were performed to calculate the mean square error term, as described by Stratford [86, 92]. The BBS provided a mean square error of 5.2 and the DGI 1.1. For a 95% confidence interval, the minimal detectable change (MDC) is  $\sqrt{5.2} * 2.77 = 6.4$  for the BBS and  $\sqrt{1.1} * 2.77 = 2.95$  for the DGI. Therefore, a change greater than 6.4 BBS points and 2.9 DGI points is necessary to reveal a change that exceeds the measurement error associated with

this instruments and show “genuine” change. Additional confidence intervals of 90% and 80% were performed for the BBS and DGI (table 2.3). This table also includes BBS MDC results when dividing the participants in functional groups (above and below 45 BBS points). This grouping resulted in 30 cases classified as low function (>45) and 12 cases classified as high function (<45). The  $MDC_{95\%}$  was 7.3 BBS points for the low function group, and 6.3 BBS points for the high function group.

Additional correlation analysis was performed to investigate a possible relationship between absolute change in the BBS and DGI. A Spearman’s rho correlation value of  $r_s = .15$   $p > .05$  suggest that there was no relationship between participants’ score differences in both instruments. Therefore, differences in scores in the BBS (initial BBS minus re-scored BBS) were not correlated with differences in the DGI (initial DGI minus re-scored DGI).

### **Discussion**

The Berg Balance Scale and the Dynamic Gait index are two instruments widely used in clinical practice to measure individual’s gait and balance ability, and monitor improvement in these areas. High reliability values have been reported for both instruments [79, 82]. However, previous investigations have used a form of correlation, such as the Pearson's product-moment correlation or the Intraclass Correlation Coefficient (ICC), to investigate the reliability of these instruments. While correlational investigations are suitable for investigating the degree of agreement between groups of subjects in repeated measures, they offer little information about the amount of change an individual needs to achieve “genuine” change. That is, the amount of change beyond the error associated with the instrument.

Absolute reliability is a more appropriate way of investigating the reliability of an instrument intended for use in a clinical setting, where clinicians are more concerned about individual change. In this investigation, the results of the absolute reliability of the Berg Balance

Scale and the Dynamic Gait Index indicate that 6.4 and 2.9 points respectively are required to be 95% confident genuine change has occur between 2 testing occasions. This information is valuable for clinicians that can apply these numbers, at the individual level, to assess improvement in function over time.

Although the 95% confident level is widely accepted in the research community, one could argue that, in clinical practice, a lower confidence level could be of practical use to make appropriate clinical decisions. In this investigation, confidence levels of 90% and 80% were calculated. The BBS showed  $MDC_{90\%}$  of 5.4 and  $MDC_{80\%}$  of 4.2 points. The DGI presented  $MDC_{90\%}$  of 2.5 and  $MDC_{80\%}$  of 1.9 points. Even at the lowest confidence level both instruments demonstrated an estimated amount of error that should be considered when making clinical judgments. It is worth noting that, in this investigation, the DGI demonstrated half the amount of estimated error as compared to the BBS. However, this comparison does not take into account the range of values of both instruments. That is, the MDC of 6.4 points in the BBS is equal to 11.5% of the total possible score of the BBS (56 points) while the MDC of 2.9 DGI points is equal to 12% of the total possible score for the DGI (24 points). Therefore, based on these results, both instruments present similar amounts of variability between the two testing occasions.

This investigation is not without limitations. The main disadvantage of using a distribution-based approach to assess change is that the results offer no indication of the importance of this change. That is, minimal detectable change (MDC) is not equal to minimally important change (MIC). In fact, is possible that MIC change is smaller than MID. In this case, the instrument would not be able to detect the desired change, since the MIC would be smaller then the measurement error of the instrument. Identifying when genuine change (beyond

measurement error) occurs is a necessary but incomplete step in the process of judging the importance of an observed clinical change.

Distribution based approaches such as the standard error of measurement (SEM) used in this investigation, assumes that the measurement error is constant across the range of possible scores. In this investigation, individuals were dichotomized into two functional groups to investigate the possible fluctuation of the SEM at two different levels of the scale. For the Berg Balance Scale, individuals with lower performance ( $BBS \leq 45$ ) demonstrated higher SEM values. Their  $MDC_{95\%}$  was 7.3 BBS points. In contrast, participants with higher performance scores ( $BBS \geq 45$ ) showed lower SEM values. Their  $MDC_{95\%}$  was 6.3 BBS points. However, dichotomizing the initial pool of participants reduced the sample size of the groups. Therefore, these results should be interpreted with caution. Further investigation is needed to substantiate the possibility of different MDC levels based on performance. The same approach was not used with scores from the DGI, because dichotomizing this group resulted in a sample size for the high level group of only 10 participants.

A methodological issue worth considering is the use of videotaped evaluations to establish the reliability of an instrument, especially when the scores of the initial live evaluation are compared with scores obtained by evaluating videotaped performances. This method has been widely used and published. In fact, the initial reliability study conducted by Berg et al [79] used videotaped assessments to investigate the intra-rater reliability of the BBS. A clear disadvantage of this design is that it does not take into account the natural fluctuation of the participant's performance when tested in two separate occasions. Therefore, clinical decisions based on this must be made with caution, since not all sources of error are considered. However, in a recent publication Stevenson [91] found a  $MDC_{95\%}$  value of 6.9 BBS points when assessing stroke

patients in a test re-test design. The fact that Stevenson's results are comparable to what was found in the present experiment, suggest that the variation seen in both experiments is mostly due to the instrument's reliability and not within patient reliability. Interestingly, Stevenson's test re-test experiment used the best performance of 3 trials as their value for each item. With this approach, it seems plausible to conclude that the "true" score is more easily captured, and the within subject variability decreased. In addition, Stevenson used the data reported by Berg and colleagues [93] in their reliability study to calculate the  $MDC_{95\%}$ . He found a  $MDC_{95\%}$  of 6.2. Again, the investigation by Berg employed a test re-test design with stroke subjects and produced similar results to the present study. To address the possibility of increased variability when using two distinct methods of evaluation (live performance vs. videotaped evaluation) future research should consider using only one of the methods to evaluate MDC in these instruments.

The therapists participating in this experiment offered feedback about the appropriateness of using videotaped sessions to investigate MDC. In general, they agreed there were some limitations in terms of having the right camera perspective to accurately assess a particular task. For example, in some static items of the BBS where body sway is an important factor, the therapists found the camera was not stable enough to judge the participant's sway. However, the therapists also indicated the use of the video allowed them to pause, slow down, or rewind and play again the tape, if they felt unsure about a particular performance. In addition, an advantage of using videotaped sessions is that it eliminates the possibility of a learning effect when assessing participants at two separate occasions. When testing and re-testing subjects within a short period of time, it is plausible to assume that subjects could perform better after being familiar with the test and the testing environment. From this experience, it is clear that a more

standardized way of filming the assessments, possibly using multiple fixed cameras, would result in a more accurate assessment of participants' performance.

While often time researchers focus on significant group mean changes in the variable of interest to draw conclusions about the effectiveness of a particular intervention, clinicians face the need to assess individual patients to judge a particular condition or monitor improvement. In this study, the BBS and DGI demonstrated mean values between test occasions of less than one point, suggesting that, as a group, both testing occasions provided almost indistinguishable results. However, at the individual level, these two instruments demonstrate an important amount of variability. Therefore, clinicians must be aware of this issue and consider the minimal detectable change values when making individual decisions based on these instruments.

### **Conclusion**

The procedure outlined by Stratford [86, 92] take reliability assessment to a more sophisticated, yet much more user friendly level. This statistical approach allows clinicians to simply determine when genuine change occurs among two testing occasions.

This experiment is a first attempt at investigating the Minimal Detectable Change (MDC) of the Berg Balance Scale (BBS) and the Dynamic Gait Index (DGI) in elder community dwellers participating in a rehabilitation program. The results from this investigation demonstrate that a change of 6.4 point in the BBS and 2.9 points in the DGI is necessary to be 95% confident genuine change in function has occurred between 2 assessments. These guidelines are important assessing individuals' performance to monitor progress and guide treatment in clinical practice. Future investigations are needed to explore MDC at different functional levels.

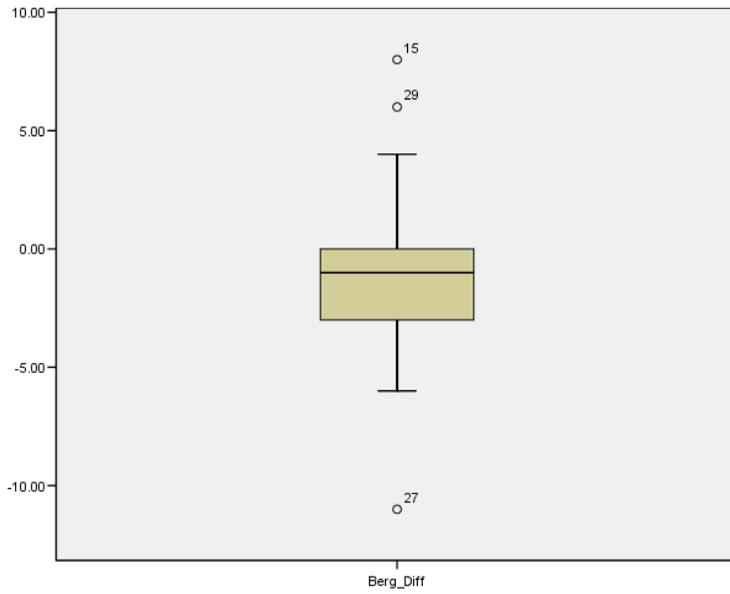


Figure 2-1. Distribution of BBS difference scores (Initial BBS- re-scored BBS)



Figure 2-2. Distribution of DGI difference scores (Initial DGI- re-scored DGI)

Table 2-1. Absolute difference in BBS scores between initial and re-scored assessments

Difference in BBS points	Number of subjects	Percent	Cumulative percent
0	9	21.4	21.4
1	8	19.0	40.5
2	7	16.7	57.1
3	8	19.0	76.2
4	5	11.9	88.1
5	1	2.4	90.5
6	2	4.8	95.2
8	1	2.4	97.6
11	1	2.4	100.0
Total	42	100.0	

Table 2-2. Absolute difference in DGI scores between initial and re-scored assessments

Difference in DGI points	Number of subjects	Percent	Cumulative percent
0	13	31.0	31.0
1	18	42.9	73.8
2	5	11.9	85.7
3	6	14.3	100.0
Total	42	100.0	

Table 2-3. MDC values for the BBS and DGI

	MDC values for the BBS and DGI			
	BBS	Low function BBS (n=30)	High function BBS (n=12)	DGI
MDC <sub>95%</sub>	6.4	7.3	6.3	2.9
MDC <sub>90%</sub>	5.4	5.8	4.2	2.5
MDC <sub>80%</sub>	4.2	4.5	3.2	1.9

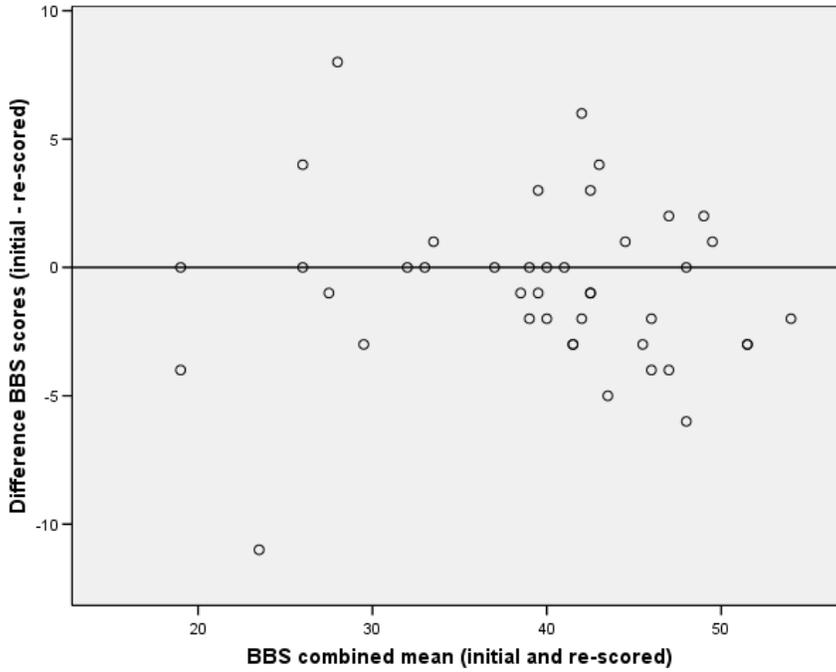


Figure 2-3. BBS results for all participants. Mean value and difference between initial and re-scored values.

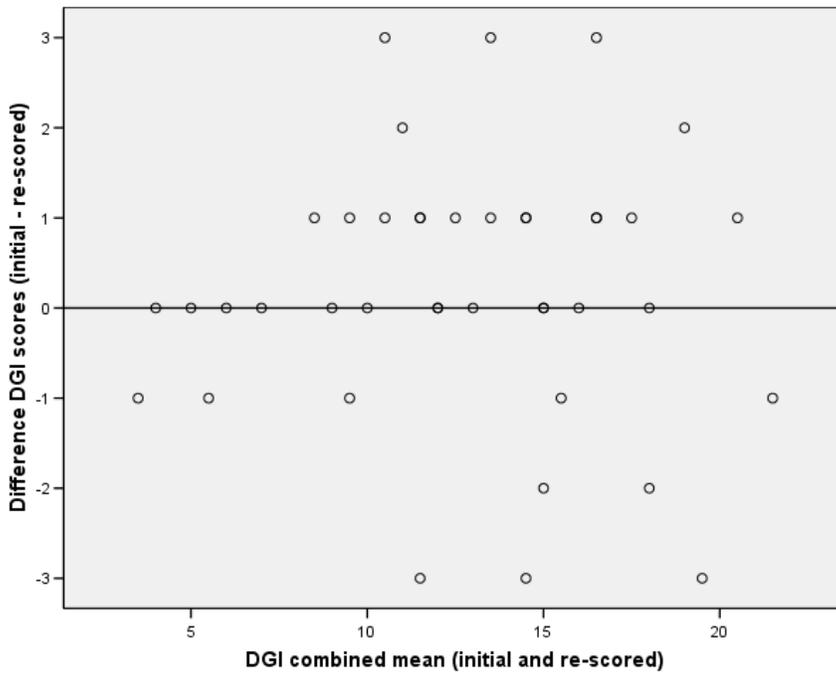


Figure 2-4. DGI results for all participants. Mean value and difference between initial and re-scored values.

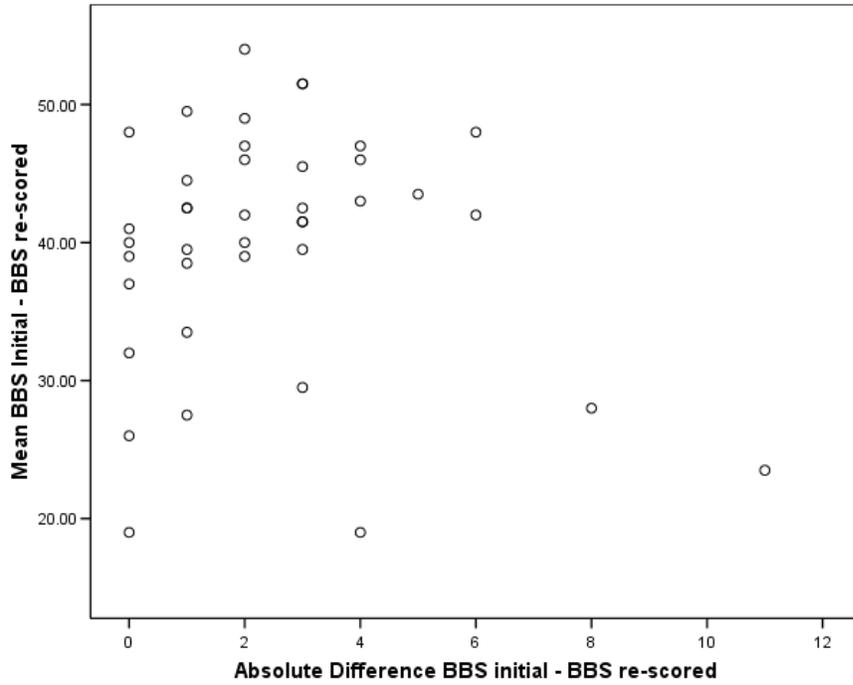


Figure 2-2. Mean BBS score for different absolute differences in BBS between testing occasions

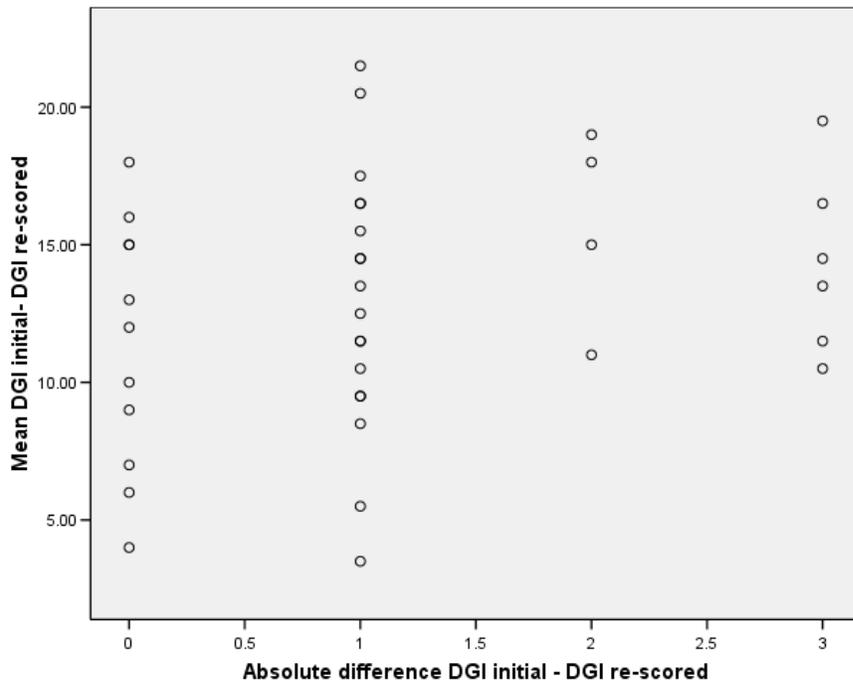


Figure 2-3. Mean DGI score for different absolute differences in DGI between testing occasions

## CHAPTER 3 PATIENTS SUCCESS CRITERIA AND EXPECTATIONS IN FALLS REHABILITATION

### **Introduction**

Despite the efforts of researchers and funding agencies to understand the mobility problems many older adults endure, falls and its consequences are taking epidemic proportions in the western world. In the US, falls are the leading cause of injury deaths among persons over 65. Falls among elderly persons account for approximately 16,000 deaths and 1.8 million emergency room visits annually, and fall-related injuries for people 65 and older cost a total of \$27.3 billion in 2003 alone [105]. These alarming statistics have prompted the US senate to recently pass a new legislation to reduce and prevent elder falls through public education campaigns and research [117]. This new effort acknowledges the need for the research community to continue working towards finding solutions to prevent the incidence of falls, improve the treatments available and ameliorate the consequences of falling.

The causes of falls have been heavily studied and multiple factors have been identified. Many of these studies indicate that fall prevention programs that include a multidisciplinary approach with a component of physical rehabilitation can significantly reduce the incidence of falls [99-101]. However, although these programs may offer significant improvements, reducing the rates of falling may be dependent on adherence to the programs. Adherence rates for participation in fall prevention programs have been historically low [102-103]. To maximize acceptability and adherence among older people, we need to have a better understanding of the factors that contribute or detract from patients adhering to fall prevention programs. Therefore, it seems intuitive to turn to the patient to identify factors that are important to them, and determine the type of treatment that will most satisfy their needs. In addition, to evaluate treatment,

therapists should be sensitive to the amount of improvement patients need to experience to consider their treatment successful.

Most rehabilitation strategies include end-points based on clinical assessment tools and conclusions drawn from standard statistical methods of significance. Statistical significance is important, but not sufficient to establish clinically relevant conclusions [104]. A better approach is to consider the number of patients that reach a clinically important end-point. To date, few investigations have used the patient's perspective to arrive at these end-points. In the falls literature, this issue has not been explored. Determining endpoints in balance rehabilitation interventions that reflect the patient's view and contribute to their satisfaction would appear to be a valuable empirical endeavor.

There is an increased interest in the research community about establishing what constitutes a clinically important change for healthcare interventions [106-108]. The previous study in this dissertation demonstrated that available assessment instruments are not always accurate at detecting change. Regardless of whether a change can be measured or observed, the interpretation of this change will depend on whose perspective we consider. From the clinician's perspective, a change that results in a modification of treatment or patient's prognosis is certainly considered a significant change. Researchers consider a change that achieves statistical significance as a relevant change. In addition, most researchers are concerned with differences in group of patients. However, group differences could reflect large change in some patients and modest change in others, or modest change in many patients. Therefore, group changes are difficult to interpret and apply, since there is no indication of the likelihood of a positive change in a single patient. A better approach, but seldom reported in the literature, is to turn to the patient to identify relevant change. Patient reported outcomes are crucial to identify relevant

domains that impact the quality of life of the patient. It is the patient who experiences their quality of life, and only they are in a position to ultimately judge whether a change is important [109].

Rehabilitation strategies need to focus on improvements that are perceived by the patient as being beneficial. This will ensure higher compliance with treatments and a more positive rehabilitation experience. Empowering the patient can lead to positive outcomes and influence patient's satisfaction with treatment. Most of the patient satisfaction literature has concentrated on treatment of non-chronic conditions. In these cases, patients who experience amelioration of their symptoms report high levels of satisfaction. Inversely, patients with chronic conditions are thought to experience less satisfaction with treatment. There is some challenge to this expectation in the chronic pain literature. Chronic pain patients have reported moderate to high levels of satisfaction even with small reductions of pain [110-111]. These patients report that, while pain was still present, their therapy had helped them reduce some of the collateral consequences of chronic pain, such as mood disturbances, sleeping problems, etc.

Most elder individuals who have fallen or are at risk of falling, can be classified as chronic patients. In fact, Calkins and colleagues [112] reported that almost 75 percent of the elderly (age 65 and over) have at least one chronic illness. Furthermore, having one or more chronic condition has been associated with an increased risk of falling by several investigators [113-115]. Therefore, it is possible that individuals who fall will share some of characteristics of patients with other chronic illnesses, and base their satisfaction with treatment due to collateral consequences. Other domains important to this group, such as energy and drive, emotional distress or social interactions, are areas that may be impacted by fall treatment interventions.

Surprisingly, to date, no published study has examined patient satisfaction in falls rehabilitation programs.

After identifying relevant domains, investigators need to assess how much improvement in each domain represents treatment success. Traditionally, this has been accomplished by using clinical assessment tools that mimic “real life” situations. However, these tools often fail to cover the domains that are important to patients. Thus, a comprehensive approach that takes into account the multiple areas of concern to patients is warranted. After establishing the domains of relevance to this population, these criteria can be used to compare satisfaction across varied studies and treatment options.

Therefore, the primary aim of this study is to investigate the patient’s success criteria across several domains including: mobility, self-care, interactions with people, community and social life, energy and drive, mental function, emotional distress, sensory function, and pain. In addition, a secondary aim is to investigate patient expectations for treatment across above mentioned domains.

## **Methods**

### **Subjects**

A total of 50 participants (age 55 and older) were enrolled in this study. Twenty of these participants were also part of a larger, funded, research study looking at the link between smoking and recovery from frailty in older Floridians (DOH- 04NIR-15). This larger study was supported by a grant from the Florida Department of Health. The remaining 30 participants were only participating in the present study. Both studies were individually approved by the Institutional Review Board for the University of Florida and the Research and Development Committee at the North Florida/South Georgia VA Medical Center. Inclusion criteria for both studies included: community dwellers with a history of falling, the ability to walk 20ft (with or

without an assistive device), and a score of 24 or higher on the Mini-Mental State Exam [97]. Participants were not screened for any particular condition. Therefore, the participants' pool consisted of individual with an extensive number of medical conditions including: diabetes, hypertension, neuropathies, orthopedic problems, general dizziness, history of stroke, general frailty, etc.

Participants in the DOH- 04NIR-15 study were recruited in two different ways. First, patients from a Gait and Balance disorders clinic, at the North Florida/South Georgia VA Medical Center, who were receiving outpatient physical therapy services related to their history of falling or being at risk of falling, were approached by their therapists and asked if they were interested in participating in a research study. If the patients agreed, a research coordinator explained the study and, if still interested, consented the patients. Secondly, letters were sent to doctor's offices explaining the study and flyers were distributed throughout the community. Participants responded to the advertisement and, if they met the study's inclusion criteria, were enrolled in the study. The remaining participants, only enrolled in the present study, were also recruited at the North Florida/South Georgia VA Medical Center Gait and Balance disorders clinic during their scheduled appointments at the clinic. These participants were also receiving outpatient physical therapy services related to their history of falling or being at risk of falling.

All participants received identical initial evaluations, consisting of a battery of assessment instruments including: Patient's Perspective Outcome Questionnaire (PPOQ), Berg Balance Scale (BBS), Dynamic Gait Index (DGI), isometric lower extremity strength, physical function domain of the MOS36, Falls Efficacy Scale, Geriatric Depression Score, pain experience VAS, timed tests of gait, Frenchay IADL scale, Pleasant Event Schedule, and spontaneous self-selected gross motor activity as measured by accelerometry. All physical assessment instruments were

administered by licensed Physical Therapist specialized in the rehabilitation of falls.

Questionnaires were administered by a research assistant. A home exercise program, based on the initial assessment, was then prescribed by the therapist. Subsequent re-evaluations at 4 weeks, 8 weeks and 12 weeks, were completed to record progress and compliance with the program.

### **Testing Procedure**

The same procedure was used for both groups of participants. In brief, during the initial evaluation, participants were administered the Patient's Perspective Outcomes Questionnaire (PPOQ, Appendix 1). A research assistant read the questionnaire out loud and recorded the answers in paper forms. Participants were allowed to ask questions and the research assistant offered clarification for any questionnaire items that participants felt unsure about.

**The Patient's Perspective Outcome Questionnaire (PPOQ):** Currently, there are no instruments in the literature to assess success criteria and expectations for treatment of falls. The PPOQ is an adaptation of the Patient Centered Outcomes Questionnaire developed by Robinson and colleagues [104]. These researchers used this questionnaire to assess success criteria and expectations for treatment of chronic pain. Dr. Robinson participated in the development of the PPOQ.

The PPOQ is identical to the Patient Centered Outcomes Questionnaire except for the domains it measures. It consists of 4 questions that address 9 different health domains. These four questions include: 1) current levels of involvement for each domain, 2) changes in each domain that will represent a successful treatment, 3) treatment outcome expectations, and 4) importance of each of the domains. These domains are based on common problem-areas associated with elder individuals who have fallen or are at risk of falling. The PPOQ uses the language of the World Health Organization (WHO) International Classification of Functioning

Disability and Health (ICF) classification system [116]. Domains used in this questionnaire refer to specific domains within the ICF classification system and include: mobility; self-care; interactions with people; community and social life; energy and drive; mental function; emotional distress, sensory function; and pain. Participants rate their perception on a scale of 0 (none/not affected/not important) to 100 (worst imaginable/most affected/most important).

Clarification about each of the domains is included in the PPOQ. To ensure uniformity and standardization of the instrument, the ICF definition for each of the domains in the PPOQ is included. For instance, the ICF defines mobility as: “this term refers to the ability to change location or transfer from one place to another. It also includes actions such as carrying, moving or manipulating objects and capacity to walk, run or climb. Lastly, mobility also refers to the ability to use various forms of transportation” [116]. This exact definition of mobility is included in the PPOQ and was used by the research assistant to explain the different domains. For example, when asking the question: On a scale of 0 (not at all important) to 100 (most important), please indicate how important it is for you to see improvement in you mobility”, the research assistant provides the above mentioned definition of mobility. The same procedure was used to explain all domain definitions in the questionnaire.

## **Analysis**

All statistical analysis and graphical representations were performed with SPSS 13.0 software for Windows (SPSS Inc., Chicago, IL, USA) and Microsoft Office Excel software for windows (Microsoft Corporation, Redmond, Washington, USA).

First, descriptive statistics were generated for each of the domains. A repeated measures ANOVA was performed to determine whether differences existed across domains in the usual levels of involvement. Follow up paired t-tests, corrected for multiple comparisons (Bonferroni correction) were performed to investigate differences between mobility and all other domains.

Then, the success criteria were determined by subtracting the usual level across domains from their expected level. For example, an individual with usual levels of mobility of 60/100 and success criteria of 40/100 requires 20 point change in mobility to consider their treatment successful. Success criteria were transformed to percentage change. Therefore, in the previous example, 20 point change represents 33.3% of the initial 60/100. Repeated measures ANOVA were performed on the percentage change success criteria scores to determine whether differences existed across domains in the amount of change necessary for participants to consider their treatment successful. Then, paired-t-tests, corrected for multiple comparisons (Bonferroni corrections) were performed to investigate possible differences between domains.

Next, usual levels across domains were subtracted from their expected levels to obtain treatment expectations criteria. Again, these scores were transformed to percentages to represent the percentage amount of change participants expected after treatment. A repeated measures ANOVA was performed to determine whether differences existed across domains in the percentage amount of change participants expected after treatment. Comparisons between mobility and all other domains were performed with paired t-tests corrected for multiple comparisons (Bonferroni correction).

Next, participants were dichotomized to form two groups (compliant vs. non-compliant). Compliance was defined as a participant who completed the standard 12 weeks program. Participants were re-evaluated 3 times after the initial evaluation, at 4, 8, and 12 weeks. Participants who attended the last re-evaluation (12 weeks) were considered compliant. Participants who missed some of the intermediate evaluations (4 or 8 weeks), but attended the last re-evaluation (12 weeks) were still considered compliant. A multivariate analysis of variance

(MANOVA) was used to explore possible differences between compliant and non-compliant groups in treatment expectations across domains.

## Results

Demographic characteristics of the sample are presented in table 3-1. Table 3-2 contains descriptive statistics from the PPOQ initial levels. Participants reported low to moderate initial levels of restriction in mobility, self-care, interactions with people, community and social life, energy and drive, mental function, emotional distress, sensory function, and pain associated with their conditions (higher scores, worse the condition) . Energy and drive; mobility; and pain received the highest scores (53, 47, and 44 respectively), while interactions with people and self-care received the lowest (21 and 24 respectively). Differences of initial levels across domains were explored with repeated measures ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated ( $\chi^2(35)= 52.02, p<.05$ ); therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ( $\epsilon = .79$ ). Results indicated the existence of significant differences of initial levels among domains  $F= (6.35, 310.98)= 8.6, p<.05$ . Paired t-tests, adjusted for multiple comparisons (Bonferroni correction), were used for posttests. The main analysis comparing mobility to other domains resulted in 8 comparisons. Therefore, for 8 domain comparisons a .006 level of significance was selected. Participants reported higher levels of impairment in the mobility domain, compared to self-care and interactions with people ( $P<.006$ ). The community and social life domain showed similar trend but the analysis failed to reach statistical significance  $t(49)= 2.79, P= .008$ .

To determine the amount of change necessary for participants to deem their treatment successful, their initial levels across domains were subtracted from their success criteria (table 3-3). Participants considered a mean reduction of 52% in mobility as a successful outcome. Lower

results were found for the domains of sensory function (34%), self-care (35%), interactions with people (22%), and, community and social life. Energy and drive required the largest reduction (59%). A repeated measures ANOVA was used to investigate whether differences existed in the amount of change necessary for participants to consider their treatment successful. Mauchly's test indicated that the assumption of sphericity had not been violated ( $\chi^2(35)= 46.9, p>.05$ ); therefore degrees of freedom were not corrected. The amount of change required for participants to consider their treatment successful was significantly different among domains  $F=(8, 368)= 7.04, p<.05$ . Paired t-tests were used for posttests. A Bonferroni correction was applied and so all effects are reported at .006 level of significance. It appeared that the reduction in mobility necessary for successful treatment was significantly greater than the reductions necessary for successful treatment of self-care, interactions with people and community and social life ( $P<.006$ ). Therefore, in this group of participants, treatment success criteria was significantly different across domains, suggesting participants require different amounts of change across domains to consider their treatment successful.

Next, initial levels across domains were subtracted from their expected levels to obtain treatment expectations criteria. Subjects expected similar levels of mobility and energy and drive (both 42%). Interactions with people received the lowest score (18%) (Table 3-4). A Repeated measures ANOVA was used to investigate whether differences existed in the amount of change participants expected after treatment. Sphericity assumption was tested with Mauchly's test. Results indicated that the assumption had been violated ( $\chi^2(35)= 76.02, p<.05$ ); therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ( $\epsilon = .74$ ). Results indicated the existence of significant differences among domains in the percentage amount of change participants expected after treatment  $F=(5.91, 277.78)= 3.94, p<.05$ . Paired t-

tests, adjusted for multiple comparisons (Bonferroni correction), were used for posttests. For this group, the percentage change participants expected for the mobility domain was significantly greater than the percentage change expected for the domains of self-care, interactions with people and pain ( $P < .006$ ). Lastly, the results from multivariate analysis of variance (MANOVA) indicated there were no significant differences between compliant and non-compliant groups in treatment expectations across domains. The Wilks Lambda multivariate test of overall differences among groups was not significant ( $p = 0.934$ ). An exploratory descriptive investigation of groups of participants, based on compliance, revealed that VA participants were 40% compliant, while community participants presented 80% compliance.

### **Discussion**

Consistent with the theoretical model that guided this research, participants with mobility problems leading to falls demonstrated significant levels of interference across several health domains, representing activity and participation, and body function. In this group, participants reported considerable initial levels of interference in domains such as energy and drive (54/100), and pain (45/100). Not surprisingly, the mobility domain also received high scores (48/100). Lower scores were seen in domains such as interaction with people (21/100), community and social life (33/100), and self-care (24/100). These findings suggest that rehabilitation strategies should take into consideration the complex and multidimensional nature of falls and provide interventions that target the different domains that affect this population. A number of publications have investigated this issue. For example, Gillespie and colleagues [117] conducted a systematic literature review of randomized controlled trial programs designed to reduce the number of falls in community-dwelling, institutionalized, or hospitalized elderly people. The authors concluded that, interventions targeting only the physical aspects related to falls produced

lower outcomes when compared to multidimensional interventions that took into account the intrinsic and environmental risk factors of patients.

The present study is a first attempt at using patient reported outcomes to investigate several health domains related to individuals receiving rehabilitation services related to falling. More specifically, this investigation focused on three fundamental aspects of rehabilitation: how participants perceived their levels of impairment across several health domains, how much change was necessary across domains to consider their treatment successful, and what were their treatment expectations.

As mentioned before, significant levels of impairment were reported across domains. Interestingly, several differences were found between domains. The energy and drive domain received the highest score, suggesting that, for this group, issues such as feelings of fatigue, motivation and energy level are commonly present. There is support in the literature for this finding. Fatigue has been associated with a number of conditions in the elderly population, including diabetes, heart failure, Parkinson's disease, cancer, sleep disorders, and hormonal changes [118-121]. In addition, participants reported considerable levels of pain. There is supporting evidence that suggest pain is a risk factor for falls and also that pain can lead to activity avoidance [122-123]. The findings from the present investigation suggest that pain and fatigue should be considered when assessing this population.

Patient reported outcomes provide a unique opportunity to evaluate clinical practice, from the patient's perspective. In the present study, participants required significant changes in a number of health domains to consider their treatment successful. The mobility domain required significantly larger reductions than the community and social life; and interactions with people domains. The clinical implication of this finding is that success criteria differ across domains,

suggesting that rehabilitation intervention should be guided towards areas of greatest concern to the patient.

This study also explored participants' expectations with treatment outcomes. Not surprisingly, participants expected mobility to change more than in other domains (42%). An unanticipated finding was that participants expected the same amount of change in the domain of energy and drive. This finding, suggest that perhaps participants view this domain as an extension of the mobility domain. That is, participants may connect improvements in mobility with increased energy and drive and, therefore, expect that energy and drive will increase as mobility improves.

In this study, participants expected different amounts of change across domains. Again, this finding must be interpreted with caution, since percentage change are heavily influenced by initial scores. For example, a participant reporting a 10 point change from an initial level of 90/100 and an expected level of 80/100 produces a percentage change of 11%, while a participant reporting a 10 point change from an initial level of 30/100 and an expected level of 20/100 produces a percentage change of 33%. In this group, participants reported low initial levels of interference with self-care and their expected level was also low. A similar trend was apparent in other domains, where high initial levels also resulted in high expectations, while low initial levels resulted in low expectations. It is plausible to speculate that participants were influenced by their initial scores and based their subsequent answers proportionally.

In the present study, participants had reasonable treatment expectations. Their expectations were lowest in domains related to participation, such as community and social life, and interactions with people. Perhaps, participants found it difficult to see the connection between improvement in physical function and improvement in social roles. Across all domains,

participants reported lower expected scores compared to their success criteria. This indicates that, in this population, participants did not expect their treatment to meet their success criteria. That is, participants expected residual levels across domains following treatment.

Results of the multivariate analysis of variance indicated no differences in expectations existed between groups of participants based on compliance with the treatment. However, these results must be interpreted with caution, due to the lack of statistical power resulting from dividing the participant pool into two groups. In addition, for a number of participants compliance was unknown, resulting in an even smaller sample size (compliant group N= 14, non-compliant N= 7). Further investigation is warranted in this area. A larger sample size that represents the general geriatric population should be used to investigate compliance. The veteran population presents certain characteristics that can have an effect on compliance. In the present study, although the sample size was too small to draw statistical conclusions, there were marked differences between VA participants and community participants. VA participants were 40% compliant, while community participants presented 80% compliance. In the larger DOH-04NIR-15 study, VA compliance was 33%. Compliance with rehabilitation interventions is essential for individuals who fall, and should be further investigated. For researchers, lack of compliance represents an additional problem, because of the bias sample selection. Drawing conclusions about the effectiveness of a particular rehabilitation strategy when samples are only representative of a small group of compliant participants is highly suspect. There could be confounding factors that make this group different and affect how this group reacts to treatment.

This study presents a number of limitations. The results obtained are based on a questionnaire specifically design for this investigation. Although this questionnaire is based on a previously published instrument [104], the psychometric properties of the former and the current

questionnaire have not been investigated. Further work is required to ascertain the reliability, validity, and generalizability of findings obtained with this questionnaire. A methodological strength of this questionnaire is that it uses the language of the ICF. This allows for comparison with other instruments that use the same language. The ICF classification system has been successfully used to develop a number of instruments [124]. Using the ICF classification system, a standardized set of relevant categories can be investigated for specific populations.

Another limitation of this study concerns the limited information about participants' characteristics. More specifically, the medical diagnosis of the pool of participants was unknown. In addition, two distinct populations were included in this study, veterans and community participants. Although both groups were community dwellers, there are certain characteristics of the veteran population that could influence their perception, expectation, and success criteria. Veterans participating in this study were recruited during their initial visit to physical therapy services. In contrast, the community dwellers participating in this study responded to advertisements distributed by the research team. The difference in recruitment approaches could produce bias and have an effect on participants' expectations.

The present study is a preliminary attempt at exploring patient reported outcomes and expectations in the treatment of falls. The results of this investigation suggest that patients participating in falls rehabilitation present a number of limitations that far exceed the mobility domain. Participants' success criteria varied across domains, suggesting that, for this population, some domains are more important than others. In addition, participants had reasonable expectations, but considered change in the most affected domains most important.

Future work should be guided towards refining and validating the PPOQ. In addition, this instrument could be used to link clinical performance measures with patients' expectations and

success criteria. This way, the effectiveness of rehabilitation interventions can be determined using a mixed clinical/patient centered approach. To provide patients the treatment they deserve, clinicians need to understand the patient's expectations and goals. Patient satisfaction is an important goal for all services, including medical services. This investigation provides a first look at quantifying some of the areas needed to ensure patient's satisfaction with the treatment of falls.

Table 3-1. Demographics

		Number of participants
UF		12
VA		38
Total		50
Non-Compliant group		7
Compliant group		14
Age 74.1 (SD 10.9)		
Gender	Male	38
	Female	12

Table 3-2. Initial levels descriptive statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Mobility	50	0	95	47.64*	28.56
Self-care	50	0	95	23.68*	32.17
Interactions with people	50	0	100	20.80*	29.05
Community and social life	50	0	100	33.50	37.70
Energy and drive	50	0	100	53.60	29.67
Mental function	50	0	100	37.60	32.38
Emotional distress	50	0	100	40.14	34.41
Sensory	50	0	95	37.38	28.72
Pain	50	0	100	43.92	34.45

\* P= .006

Table 3-3. Success criteria (initial levels-success levels) descriptive statistics

	N	Minimum	Maximum	Mean	% Change	Std. Deviation
Mobility	50	0	70	24.87	51.86*	18.94
Self-care	50	0	90	14.66	35.52*	25.14
Interactions with people	50	0	75	9.15	22.27*	18.27
Community and social life	50	0	95	16.91	27.96*	27.05
Energy and drive	50	0	80	31.60	58.84*	23.82
Mental function	50	0	80	21.23	44.99	24.55
Emotional distress	50	0	85	25.13	48.24	26.11
Sensory	50	0	80	15.40	34.33	19.40
Pain	50	0	85	24.53	44.81	24.27

\* P= .006

Table 3-4. Treatment expectations criteria (initial levels-expectations levels)  
descriptive statistics

	N	Minimum	Maximum	Mean	% Change	Std. Deviation
Mobility	50	0	70	19.98	42.30*	18.34
Self-care	50	0	60	8.02	23.80*	13.35
Interactions with people	50	0	75	6.28	17.87*	14.00
Community and social life	50	0	95	15.43	26.78*	25.55
Energy and drive	50	0	70	23.38	42.49*	23.10
Mental function	50	0	65	14.51	34.44	17.90
Emotional distress	50	0	85	18.74	38.94	22.83
Sensory function	50	0	80	11.28	32.01	15.86
pain	50	0	80	13.98	27.92	19.19

\* P= .006

## CHAPTER 4 GENERAL SUMMARY AND CONCLUSIONS

Overall, the objectives of this dissertation were to investigate minimum detectable change in outcome measures used to assess individuals receiving rehabilitations services related to falls, and to explore treatment success criteria and expectations, in the same population. Recently, there is considerable debate in the medical and scientific community about what treatment outcomes constitute a meaningful change, and how to measure this change [125-127]. Clinical performance measures do not always capture all aspects of the patients' experience and often produce results that are difficult to interpret. In an attempt to simplify the interpretation of scores, single numbers are assigned to complex processes. In the name of objectivity, statistical analysis are performed on these scores, and the results are use to evaluate performance and assess change. Still, statistical significance does not necessarily equal meaningful change.

Patient reported outcomes (PRO) are receiving attention as legitimate outcome measures for clinical research. PRO instruments are used to measure treatment benefits by capturing concepts related to how a patient feels or functions with respect to his or her health or condition. This approach turns to the patient to investigate what is important to them. The ideas, activities, behaviors, or feelings measured by PRO instruments can be either verifiable in nature, such as walking, or can be non-observable, known only to the patient, such as pain, depression etc. Although these symptoms are highly dependent on the patients' perception, historically, these assessments were made by clinicians who observed and interacted with patients. Recently, these kinds of assessments are increasingly performed with PRO instruments. It seems intuitive to consider the patient's opinion when investigating what constitutes meaningful change.

The two experiments described in this dissertation are aimed at investigating change in geriatric patients. Specifically, the group of patients investigated received rehabilitation

interventions related to mobility problems. The overall aim of this dissertation was to explore change in this group of patients employing two different approaches. First, minimum detectable change was investigated in clinical instruments used with this population. Second, a newly-developed PRO instrument was used to determine what constitutes successful treatment outcomes and expectations when participating in a rehabilitation program. A general conclusion for both studies in this dissertation follows.

### **Experiment I Summary**

The goal of this study was to investigate minimal detectable change (MDC) for two common instruments used to assess gait and balance in the elder population. The Berg Balance Scale [75] and the Dynamic Gait Index [82] were explored in this experiment. The procedure outlined by Stratford [86, 92] was used to calculate the MDC. Stratford proposed to use the standard error of measure (SEM) to calculate the amount of change in a given measure that must be obtained for a clinician to determine that true change has occurred. The MDC is expressed as a confidence interval around the SEM, indicating the values that are within the range of error attributable to the measuring instrument. The MDC is expressed in the same unit of the original instrument, providing clinicians useful and easy to understand criterion for change in patients' performance.

The results of this study indicated that for the Berg Balance Scale and the Dynamic Gait Index, 6.4 and 2.9 points respectively were required to be 95% confident that "genuine" change had occurred between 2 testing occasions. These results suggest that a significant amount of error is associated with these instruments. In addition, the results suggested that MDC values are not a constant feature of the instruments. MDC values for the high functional level group were 6.3 BBS points. In contrast, participants in the lower functional group presented MDC values of 7.3. That is, the values of MDC change based on the ability level of the persons assessed.

The above findings are relevant to present day clinical practice. Clinicians use the Berg Balance Scale and Dynamic Gait Index routinely to assess patients with mobility problems. An advantage of using the standard error of measure to determine minimal detectable change is that this method provides information about individual scores. Traditional methods of statistical significance rely on group differences to investigate the properties of the instruments. Group differences are relevant for researchers, but must be considered with caution when decisions must be made about an individual patient. The results of this investigation can be applied at the individual level. Knowing the amount of error associated with these instruments can help clinicians make decisions about individual's performance and monitor change overtime.

### **Experiment II Summary**

The primary aim of this study was to use a PRO questionnaire to investigate patient's success criteria and expectations when receiving rehabilitation services related to falls. More specifically, the patient's success criteria was assessed across several health domains including: mobility, self-care, interactions with people, community and social life, energy and drive, mental function, emotional distress, sensory function, and pain. In addition, a secondary aim was to investigate patient expectations for treatment across above mentioned domains.

In this study, participants with mobility problems leading to falls demonstrated significant levels of interference across several of the health domains measured. Participants reported considerable initial levels of impairment in domains such as energy and drive (53/100), and pain (44/100). The mobility domain also received high scores (indicating impairment) (47/100). Lower scores were seen in domains such as interaction with people (21/100) and community and social life (33/100). These findings suggest that, in this population, domains with a strong social component were not as affected as domains with a strong physical component.

Participants in this study required significant improvement in health domains to consider their treatment successful. Domains such as mobility; and energy and drive, required significantly larger reductions than the community and social life; and interactions with people domains. This provides information about what is important to patients receiving this intervention. Furthermore, these findings could lead to developing rehabilitation strategies guided towards areas of greatest concern to the patient.

This study also explored participants' expectations with treatment outcomes. Participants expected mobility to change the most as a result of this intervention (52%). However, similar finding was reported in the domain of energy and drive. An interesting finding is that across domains, participant's expectation was that the treatment would *not* meet their success criteria. This indicates that, for this population of elder individuals with mobility problems, residual levels of impairment in the measured domains are expected after treatment. Compliance was also investigated in this group. No differences were found between compliant and non-compliant groups based on treatment expectations.

The results of this study point out that a number of health domains are significantly affected in this population. Patients receiving rehabilitation services related to falls have treatment expectations that far exceeds the mobility problems for which they are been treated. In exploring meaningful change in patients receiving rehabilitation interventions, the patient's expectations and success criteria must be considered. By linking existing clinical instruments with patient reported outcomes researchers and clinicians can be sure that therapies used achieve a meaningful change. Physical rehabilitation strategies must take into consideration the complex and multidimensional nature of falls and provide interventions that target the different domains that affect this population.

## **General Conclusions and Future Directions**

The question of what constitutes a meaningful change in the population investigated in this dissertation remains unanswered. However, several conclusions can be drawn from the work presented. First, clinical instruments provide valuable information about patients' performance and can help clinicians and researchers evaluate patient's ability and monitor improvement. However, when these instruments are used at the individual level, clinicians must be aware that all change is not genuine change. That is, instrument error must be considered when reporting patient change at the individual level. Second, elder patients receiving mobility-related rehabilitative services expect their treatment to produce changes in a number of health domains that extend beyond mobility improvement.

Patient reported outcomes could serve as a bridge to link clinical practice with meaningful patient-centered treatment results. There is a growing movement for patients to take an active role in their medical care and be involved in making decision about treatment options [128]. Patients demand services that meet their needs and expect treatments to address nonclinical aspects that affect their day to day life. This view is receiving attention from the research community and regulatory agencies. A number of randomized clinical trials are now including patient reported outcomes as important endpoints in addition to traditional clinical measures [129-130]. Regulatory agencies are also recognizing this need and establishing criteria for the use of patient reported outcomes [126]. Incorporating the patient's view in the rehabilitation process ensures that interventions meet the patient's needs and therefore play a role at empowering the patient and making him or her responsible for actively participating in the rehabilitative process.

A number of interesting questions have arisen as a result of this dissertation. Further work is warranted at investigating minimal detectable change levels associated with different

functional levels of the patients investigated. This is particularly important for high risk patients for which small change in performance, indicating improvement or deterioration, can have serious consequences. In this functional group, knowing the exact functional level of the patient can help clinicians take appropriate measures to avoid possible injuries, for example prescribe an assistive or protective device.

Future work should also be conducted to investigate issues of compliance in this population. In this dissertation, possibly due to small sample size, no significant differences between compliant and non-compliant groups were found. Still, low compliance levels were seen in the investigated group. The traditional model of care, where clinicians tell patients what to do and try to motivate them to change, may not be the most effective method of intervention. Empowering the patient by considering their specific needs and addressing their expectations might result in better compliance with treatment. PRO instruments can be used to incorporate the patient's perspective in the rehabilitative process. These instruments can serve as an anchor to compare clinical instruments against meaningful patient-centered outcomes.

Collectively, this series of studies promotes our understanding of significant change in patients receiving rehabilitation services related to falls. The results obtained indicate that current rehabilitation programs must consider the limitations of available instruments and take into consideration the needs and expectations of patients. Ultimately, this research aims to influence treatment by providing information to help clinicians select the best tools available for the rehabilitation of falls, and suggests the inclusion of the patient's perspective as one of the outcome measures in their treatment plans.

APPENDIX  
PATIENT'S PERSPECTIVE OUTCOME QUESTIONNAIRE (PPOQ)

We have identified some common problems people can experience as they get older. Specifically, older people who have balance problems and may have suffered a fall often face problems of **mobility, self-care, interactions with people, community and social life, energy and drive, mental function, emotional distress, sensory function, and pain** (see the last page for an extended explanation of these problem-areas). We would like to ask you a few questions to see how important these problems are to you.

FIRST, WE WOULD LIKE TO KNOW YOUR USUAL LEVELS OF MOBILITY, SELF-CARE, INTERACTION WITH PEOPLE, COMMUNITY AND SOCIAL LIFE, ENERGY AND DRIVE, MENTAL FUNCTION, EMOTIONAL DISTRESS, SENSORY FUNCTION, AND PAIN.

On a scale of 0 (none/not affected) to 100 (worst imaginable/most affected), please indicate your usual level (during the past week) of

- mobility \_\_\_\_\_
- self-care \_\_\_\_\_
- interactions with people \_\_\_\_\_
- community and social life \_\_\_\_\_
- energy and drive \_\_\_\_\_
- mental function \_\_\_\_\_
- emotional distress \_\_\_\_\_
- sensory function \_\_\_\_\_
- pain \_\_\_\_\_

PATIENTS UNDERSTANDABLY WANT THEIR TREATMENT TO RESULT IN DESIRED OR IDEAL OUTCOMES. UNFORTUNATELY, AVAILABLE TREATMENTS DO NOT ALWAYS PRODUCE DESIRED OUTCOMES. THEREFORE, IT IS IMPORTANT FOR US TO UNDERSTAND WHAT TREATMENT OUTCOMES YOU WOULD CONSIDER SUCCESSFUL.

On a scale of 0 (none/not affected) to 100 (worst imaginable/most affected), please indicate the level each of these areas would have to be at to consider treatment successful.

- mobility \_\_\_\_\_
- self-care \_\_\_\_\_
- interactions with people \_\_\_\_\_
- community and social life \_\_\_\_\_
- energy and drive \_\_\_\_\_
- mental function \_\_\_\_\_
- emotional distress \_\_\_\_\_
- sensory function \_\_\_\_\_
- pain \_\_\_\_\_

NOW, WE WOULD LIKE TO KNOW WHAT YOU EXPECT YOUR TREATMENT TO DO FOR YOU.

On a scale of 0 (none/not affected) to 100 (worst imaginable/most affected), please indicate the levels you expect following treatment.

- mobility \_\_\_\_\_
- self-care \_\_\_\_\_
- interactions with people \_\_\_\_\_
- community and social life \_\_\_\_\_
- energy and drive \_\_\_\_\_
- mental function \_\_\_\_\_
- emotional distress \_\_\_\_\_
- sensory function \_\_\_\_\_
- pain \_\_\_\_\_

FINALLY, WE WOULD LIKE TO UNDERSTAND HOW IMPORTANT IT IS FOR YOU TO SEE IMPROVEMENT IN YOUR MOBILITY, SELF-CARE, INTERACTIONS WITH PEOPLE, COMMUNITY AND SOCIAL LIFE, ENERGY AND DRIVE, MENTAL FUNCTION, EMOTIONAL DISTRESS, SENSORY FUNCTION, AND PAIN FOLLOWING TREATMENT.

On a scale of 0 (not at all important) to 100 (most important), please indicate how important it is for you to see improvement in your...

- mobility \_\_\_\_\_
- self-care \_\_\_\_\_
- interactions with people \_\_\_\_\_
- community and social life \_\_\_\_\_
- energy and drive \_\_\_\_\_
- mental function \_\_\_\_\_
- emotional distress \_\_\_\_\_
- sensory function \_\_\_\_\_
- pain \_\_\_\_\_

### **Problem areas explained:**

**Mobility:** This term refers to the ability to change location or transfer from one place to another. It also includes actions such as carrying, moving or manipulating objects and capacity to walk, run or climb. Lastly, mobility also refers to the ability to use various forms of transportation.

**Self-care:** This problem-area is about caring for oneself, washing and drying oneself, caring for one's body and body parts, dressing, eating and drinking, and looking after one's health.

**Interactions with people:** This area is about the ability to socially interact with strangers, friends, relatives, family members and significant others.

**Community and social life:** This area is about the actions and tasks required to engage in organized social life outside the family, in community, social and civic areas of life. Examples include: participating in religious or spiritual activities and participation in leisure or recreational activities.

**Energy and drive:** This problem area refers to feelings of fatigue, motivation and energy level.

**Mental function:** This problem area includes issues related to memory, attention, concentration, and decision making.

**Emotional distress:** This problem area includes feelings of depression, anger, anxiety, and frustration.

**Sensory function:** This area includes problems of vision and hearing.

**Pain:** This area includes all types of pain, including chronic and acute.

## LIST OF REFERENCES

1. Miaskowski C. The impact of age on a patient's perception of pain and ways it can be managed. *Pain Manag Nurs* 2000; 1:2-7.
2. Centers for Disease Control and Prevention, Department of Health and Human Services. Healthy Aging for Older Adults [online]. (2007) [cited 2008 Apr 14]. Available from URL: <http://www.cdc.gov/aging/>
3. Seematter-Bagnoud L, Wietlisbach V, Yersin B, Büla CJ. Healthcare Utilization of Elderly Persons Hospitalized After a Noninjurious Fall in a Swiss Academic Medical Center. *Journal of the American Geriatrics Society*, 2006; 54(6),891–897.
4. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS) [online]. (2006) [cited 2008 Apr 14]. Available from URL: [www.cdc.gov/ncipc/wisqars](http://www.cdc.gov/ncipc/wisqars)
5. Hausdorff JM, Rios DA, Edelber HK. Gait variability and fall risk in community-living older adults: a 1-year prospective study. *Archives of Physical Medicine and Rehabilitation* 2001; 82(8):1050–6.
6. Englander F, Hodson TJ, Terregrossa RA. Economic dimensions of slip and fall injuries. *Journal of Forensic Science* 1996; 41(5):733–46.
7. World Health Organization. International Classification of Functioning (ICF), Disability and Health. Geneva: WHO; 2001.
8. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977; 196(4286):129-36.
9. Kirshner B, Guyatt GH. A methodological framework for assessing health indices. *J Chronic Dis* 1985; 38:27–36.
10. Guyatt GH, Jaeschke R, Feeny DH, Patrick DL. Measurements in clinical trials: choosing the right approach. In: B. Spilker, Editor, *Quality of life and pharmacoeconomics in clinical trials*, Lippincott-Raven Publishers, Philadelphia;1996; 41–49.
11. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*1993; 118:622–629.
12. Wyrwich KW, Metz S, Babu AN. The reliability of retrospective change assessments. *Qual Life Res* 2002; 11:636.
13. Sim J, Arnell P. Measurement validity in Physical Therapy research. *Phys Ther* 1993; 73: 102-15.

14. Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992; 47:76–83.
15. Portney LG, Watkins MP. Reliability. In: Portney LG, Watkins MP. *Foundations of Clinical Research: applications to practice*. 2nd ed. New Jersey: Prentice Hall Health; 2000: 53-68.
16. Walmsley RP, Amell TK. The application and interpretation of intraclass correlation in the assessment of reliability in Isokinetic dynamometry. *Isokinet Exerc Sci* 1996; 6:117-24.
17. Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992; 47:76–83.
18. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985; 38: 27-36.
19. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance *J Chronic Dis*. 1986; 39:897–906.
20. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disabil Rehabil* 2005; 27:315–340.
21. Jaeschke R, Singer J, Guyatt GH. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10:407–415.
22. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics* 1995; 15:141–155.
23. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; 16:139–144.
24. Norman GR, Sridhar FG, Guyatt GH, Walter SD. The Relation of Distribution- and Anchor-Based Approaches in Interpretation of Changes in Health Related Quality of Life. *Medical Care* 2001; 39:1039–1047.
25. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, and the Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings* 2002; 77:371–383.
26. Deyo RA, Inui TS, Leininger J, Overman S. Physical and psychosocial function in rheumatoid arthritis: clinical use of a self-administered health status instrument. *Arch Intern Med* 1982; 142:879.

27. Johnson PA, Goldman L, Orav EJ, Garcia T, Steven D, Pearson, et al. Comparison of the medical outcomes study short-form 36-item health survey in black patients and white patients with acute chest pain. *Med Care* 1995; 33:145–160.
28. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *New Engl J Med* 1996; 28:835–840.
29. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; 59:12–19.
30. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics* 1995; 15:141–155.
31. Revicki DA, Allen H, Bungay K, Williams GH, Weinstein MC. Responsiveness and calibration of the general well-being adjustment scale in patients with hypertension. *J Clin Epidemiol* 1994; 47:1333–1342.
32. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10:407-415.
33. Idler EL, Angel RJ. Self-rated health and mortality in the NHANES-I epidemiology follow-up study. *Am J Public Health* 1990; 80:446–452.
34. Farar JT. What is clinically meaningful: outcome measures in pain clinical trials. *Clin J Pain* 200; 16(2 Suppl):S106-12.
35. Lydick F, Yawn BP. Clinical interpretation of health-related quality of life data. In: M.J. Staquet, R.D. Hays and P.M. Fayers, Editors, *Quality of life assessment in clinical trials: methods and practice*. Oxford University Press, Oxford 1998:299–314.
36. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000; 53:459–468.
37. G.R. Norman, P. Stratford and G. Regehr. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997; 50:869–879.
38. Liang NH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985; 28:542–547.
39. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Control Clin Trial* 1991; 12:142S–158S.
40. Byrk AS, Raudenbush SW. *Hierarchical linear models: applications and data analysis methods*. Newbury Park (CA), Sage; 1992.

41. Speer C, Greenbaum PD. Five methods for computing significant individual client change and improvement rates: support for an individual growth curve approach. *J Consult Clin Psychol* 1995, 63:1044–1048.
42. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 7; 2002:147–177.
43. Cohen, J. *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates 1988:34.
44. Kazis LE, Anderson JJ, Meenan RS. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27: Suppl 3:S178–S189.
45. Fayers PM, Machin D. *Quality of life: assessment, analysis and interpretation*. John Wiley&Sons, Chichester; 2000.
46. Clancy CM, Eisenberg JM. Outcomes research: measuring the end results of health care. *Science* 1998; 282(5387):245-6.
47. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; 52:861–873.
48. McHorney CA, Tarlov A. Individual-patient monitoring in clinical practice: are available health status surveys adequate. *Qual Life Res* 1995; 4:293–307.
49. Anastasi A, Urbina S. *Psychological Testing*. 7th ed. Upper Saddle River, NJ: Prentice-Hall; 1997.
50. Hageman WJ, Arrindell WA. Establishing clinically significant change: increment of precision and distinction between individual and group level of analysis. *Behav Res Ther* 1999; 37:1169–1193.
51. Roebroek ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. *Phys Ther* 1993; 73 (6): 386-95.
52. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977; 196(4286):129-36.
53. Engel GL. The clinical application of the biopsychosocial model. *Am J Psychiatry* 1980; 137: 535-544.
54. U.S. Food and Drug Administration. *Guidance for Industry. Patient-Reported Outcome measures : use in medical product development to support labeling claims. Draft guidance*. [cited 2008 June 20]. Available from: <http://www.fda.gov/cder/guidance/5460dft.pdf>

55. Hippocrates. *The Book of Prognostics, Part 2* (approx. 400 B.C.). As translated by Francis Adams (Great Books Index, 1977-99). [cited February 2008]. Available from: <http://classics.mit.edu/Hippocrates/prognost.mb.txt>
56. FDA Consumer Magazine. *The Importance of Patient-Reported Outcomes ... It's All About the Patients 2006*; Vol 40: 6. Available from: [http://www.fda.gov/fdac/606\\_toc.html](http://www.fda.gov/fdac/606_toc.html).
57. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, Rothman M. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual. Life Res.* 2000; 9:887–900.
58. Patrick DL, Erickson P. *Health status and health policy: quality of life in health care evaluation and resource allocation.* Oxford University Press, New York; 1993.
59. Carmines E G, Zeller RA. *Reliability and validity assessment.* Newbury Park: Sage Publications; 1991.
60. Ostir G, Granger C, Black T, Roberts P, Burgos L, Martinkewiz P, et al. Preliminary Results for the PAR-PRO: A Measure of Home and Community Participation. *Arch Phy Med and Rehab* 2006; 87(8):1043-51.
61. Willkea RJ, Burkeb LB, Ericksonc P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Controlled Clinical Trials* 2004; 25(6):535-52.
62. Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, Rothman M. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res.* 2000; 9(8):887-900.
63. Zerhouni E. NIH Roadmap. *Science* 2003; 302(5642):63-72.
64. Stucki G. International classification of functioning, disability and health (ICF): A promising framework and classification for rehabilitation medicine. *Am J Phys Med Rehabil* 2005; 84:733–740.
65. World Health Organization. *International Classification of Diseases, Disorders and Injuries, 10th revision.* Geneva: World Health Organization, 1994.
66. Degner L, Sloan J. Symptom distress in newly diagnosed ambulatory cancer patients and as a predictor of survival in lung cancer. *J Pain Symptom Manage* 1995; 10:1–8.
67. Sloan JA, Loprinzi CL, Kuross SA, Miser AW, O'Fallon JR, Mahoney MR, et al. Randomized comparison of four tools measuring overall quality of life in patient with advanced cancer. *J Clin Oncol* 1998; 16:3662–3673.
68. Frost MH, Huschka M. Quality of Life from a Patient's Perspective: Can We Believe the Patient? *Current Problems in Cancer.* 2005; 29(6):326-331.

69. Hodgkins M, Albert D, Daltroy L. Comparing patients' and their physicians' assessment of pains. *Pain* 1985; 23: 273–277.
70. Patel KK, Veenstra DL, Patrick DL. A Review of Selected Patient-Generated Outcome Measures and Their Application in Clinical Trials. *Value in Health* 2003; 6(5):595-603(9).
71. Davidhizar R, Giger JN. A review of the literature on care of clients in pain who are culturally diverse. *International Nursing Review* 2004; 51(1):47-55.
72. Atkinson MJ, Lennox RD. Extending basic principles of measurement models to the design and validation of Patient Reported Outcomes. *Health Qual Life Outcomes* 2006; 4: 65.
73. Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002; 11:193–205.
74. Medley A. Predicting the probability of falls in community dwelling persons with brain injury: a pilot study. *Brain Inj.* 2006; 20(13-14):1403-8.
75. Berg K, Maki B, Williams JI, Holliday PJ, Wood-Dauphinee SL. Clinical and laboratory measures of postural balance in an elderly population. *Archives Of Physical Medicine and Rehabilitation* 1992a; 73:1073-1080.
76. Berg KO, Wood-Dauphinée SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. *Can J Public Health.*1992b; 83(suppl 2):S7–S11.
77. Wolf SL, Catlin PA, Gage K, Gurucharri K, Robertson R, Stephen K. Establishing the reliability and validity of measurements of walking time using the Emory Functional Ambulation Profile. *Phys Ther.*1999; 79:1122–1133.
78. Liston RA, Brouwer BJ. Reliability and validity of measures obtained from stroke patients using the Balance Master. *Arch Phys Med Rehabil.*1996; 77:425–430.
79. Berg K, Wood-Dauphinee S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. *Physiother Can.* 1989c; 41:304–311.
80. Shumway-Cook A, Woolhicoii MH. *Motor Control: Theory and Practical Applications*, Philadelphia, Pa: Lippincott Williams & Wilkins; 2001:401, 405-406,
81. Whitney S, Wrisley D, Furman J. Concurrent validity of the Berg Balance Scale and the Dynamic Gait Index in people with vestibular dysfunction. *Physiother Res Int* 2003; 8(4):178-86.
82. Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. *Physical Therapy* 1997; 77(8):812–9.

83. Bogle Thorbahn LD, Newton RA. Use of the Berg Balance Test to predict falls in elderly persons. *Phys Ther* 1996 Jun;76(6):576-83; discussion 584-5.
84. Whitney SL, Hudak MT, Marchetti GF. The dynamic gait index relates to self-reported fall history in individuals with vestibular dysfunction. *J Vestib Res*. 2000; 10:99-105.
85. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol*. 1992; 45:1341–1345.
86. Stratford PW. Reliability: consistency or differentiating among subjects (editorial). *Phys Ther* 1999; 69: 299–300.
87. Shikiar R, Harding G, Leahy M, Lennox RD. Minimal important difference (MID) of the Dermatology Life Quality Index (DLQI): results from patients with chronic idiopathic urticaria. *Health Qual Life Outcomes*. 2005; 20;3:36.
88. Birmingham TB, Kramer JF, Speechley M, Chesworth BM, MacDermid J. Measurement variability and sincerity of effort: clinical utility of isokinetic strength coefficient of variation scores. *Ergonomics* 1998; 41: 853–863.
89. Bland JM, Altman DG. Measurement error. *BMJ* 1996; 313:744.
90. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. *Phys Ther* 1993; 73(6):386-95.
91. Stevenson TJ. Detecting change in patients with stroke using the Berg Balance Scale. *Australian Journal of Physiotherapy* 2001; 47:29-38.
92. Stratford P, Finch W, Solomon E, Binkley P, Gill J, Moreland C. Using the Roland-Morris questionnaire to make decisions about individual patients. *Physiotherapy Can* 1996, 48:107–110.
93. Berg K, Wood-Dauphinee S, Williams JI. The Balance Scale: reliability assessment with elderly residents and patients with an acute stroke. *Scandinavian Journal of Rehabilitation Medicine* 1995; 27:27-36.
94. Tyson SF, DeSouza LH. Reliability and validity of functional balance tests post stroke. *Clin Rehabil* 2004; 18: 916–923.
95. Mao HF, Hsueh IP, Tang PF, Sheu CF, Hsieh CL. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke* 2002; 33: 1022–1027.
96. Ottonello M, Ferriero G, Benevolo E, Sessarego P, Dughi D. Psychometric evaluation of the Italian version of the Berg balance scale in rehabilitation inpatients. *Eur Med Phys* 2003; 39:181–189.

97. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Psychiat. Res* 1975; 12(3):189-198.
98. Guyatt G, V Montori PJ, Devereaux H, Schünemann H, Bhandari M. Patients at the centre: In our practice and in our use of language. *ACP Journal Club* 2004; 140: A-11.
99. Chang JT, Morton SC, Rubenstein LZ, Mojica W, Maglione M, Suttorp M, et al. Interventions for the prevention of falls in older adults: Systematic review and meta-analysis of randomized controlled trials. *British Medical Journal* 2004; 328:680-683.
100. Gillespie LD, Gillespie W J, Robertson MC, Lamb SE, Cumming RG, Rowe BH. Interventions for preventing falls in elderly people (Cochrane Review). London: Wiley 2001.
101. Skelton D, Todd C. What are the main risk factors for falls amongst older people and what are the most effective interventions to prevent these falls? How should interventions to prevent falls be implemented? Health Evidence Network Synthesis. Copenhagen, Denmark: World Health Organization 2004.
102. Robertson M C, Devlin N, Gardner MM, Campbell A J. Effectiveness and economic evaluation of a nurse delivered home exercise program to prevent falls: Randomized controlled trial. *British Medical Journal* 2001; 322:697-701.
103. Stevens M, Holman CD, Bennett N, de Klerk N. Preventing falls in older people: Outcome evaluation of a randomized controlled trial. *Journal of the American Geriatrics Society* 2001;49:1448-1455.
104. Robinson M, Brown J, George S, Edwards P, Atchison J, Hirsh A, Waxenberg L, Wittmer V, Fillingim R. Multidimensional success criteria and expectations for treatment of chronic pain: The patient perspective. *Pain Medicine* 2005; 6(5):336-45.
105. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS). [cited 2008 Jan 15]. Available from URL: [www.cdc.gov/ncipc/wisqars](http://www.cdc.gov/ncipc/wisqars).
106. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003; 56:395–407.
107. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002; 14:109–14.
108. De Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006; 4:54–59.

109. Schünemann HJ, Guyatt GH. Goodbye MCID! Hello MID, Where Do You Come From? *Health Serv Res* 2005; 40(2):593–597.
110. Ward SE, Gordon DB. Patient satisfaction and pain severity as outcomes in pain management: a longitudinal view of one setting's experience. *J Pain Symptom Manage*. 1996; 11:242-251.
111. Pellino TA, Ward SE. Perceived control mediates the relationship between pain severity and patient satisfaction. *J Pain Symptom Manage* 1998; 15:110-116.
112. Calkins E, Boulton C, Wagner E, Pacala JT. New ways to care for older people. Building systems based on evidence. New York: Springer; 1999.
113. Bergland A, Wyller TB. Risk factors for serious fall-related injury in elderly women living at home. *Injury Prevention* 2004; 10:308-313.
114. Sattin RW. Falls among older persons: A public health perspective. *Annual Review of Public Health* 1992; 13:489-508.
115. Tinetti ME, Doucette J, Claus E, Marottoli R. Risk factors for serious injury during falls by older persons in the community. *Journal of the American Geriatrics Society* 1995; 43:1214-1221.
116. International Classification of Functioning (ICF), Disability and Health. World Health Organization. Vol. 2002, 2001.
117. Gillespie LD, Gillespie WJ, Cumming R, Lamb SE, Rowe BH. Interventions to reduce the incidence of falling in the elderly. *Cochrane Library* 1997; 4:1-29.
118. Lichstein KL, Means MK, Noe SL, Aguillard RN. Fatigue and sleep disorders. *Behaviour Research and Therapy* 1997; 35:733–740.
119. Karlsen K, Larsen JP, Tandberg E, Jorgensen K. Fatigue in patients with Parkinson's disease. *Movement Disorders*, 1999; 14:237–241.
120. Nail LM. Fatigue in patients with cancer. *Oncol Nurs Forum* 2002; 29:537–546.
121. Swain MG. Fatigue in chronic disease. *Clin Sci* 2000; 99:1–8.
122. Peat TG, Harris LH, Wilkie R, Croft PR. The prevalence of pain and pain interference in a general population of older adults: cross-sectional findings from the North Staffordshire Osteoarthritis Project (NorSTOP). *Pain* 2004; 110:361–368.
123. Blyth FM, Cumming R, Mitchell P, Wang JJ. Pain and falls in older people. *European Journal of Pain* 2007; 11(5):564-571.

124. Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the International Classification of Functioning; Disability and Health. *J Rehabil Med* 2002; 34: 205-10.
125. Middel B, van Sonderen E. Statistically significant change versus relevant or important change in (quasi) experimental design. *Int J Integr Care* 2002; 2: 1–22.
126. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, O'Neill R, Kennedy DL. Patient-Reported Outcomes to Support Medical Product Labeling Claims: FDA Perspective. *Value in Health* 2007; 10(s2):S125–S137.
127. Rothman ML, Beltran P, Pharm D, Cappelleri JC, Lipscomb J, Teschendorf B. The Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group Value in Health 2007; 10(s2):S66–S75.
128. Wagner EH, Austin BT, Von Koroff M. Improving outcomes in chronic illness. *Managed Care Quarterly* 1996; 4(2):12-25.
129. Revicki. FDA draft guidance and health-outcomes research. *The Lancet* 2007; 369(9561):540-542 D.
130. Fairclough DL. Patient reported outcomes as endpoints in medical research. *Stat Methods Med Res* 2004; 13:115-38.

## BIOGRAPHICAL SKETCH

Sergio Romero was born and raised in the South of Spain. He moved to the USA in 1991 to pursue his studies in the area of exercise and sport science. In 1996, he received a bachelor's degree in exercise and sport science from the University of Florida. He continued his studies at University of Florida and received a master's degree in exercise and sport science in 1998. The next few years, he worked in a variety of departments at this university and developed an interest in geriatrics, and more specifically in falls prevention and falls rehabilitation. After taking a few courses in geriatrics during 2002 and 2003, he was officially accepted in the rehabilitation science doctoral program in the college of Public Health and Health Professions in spring 2004. He specialized in movement disorders and continued working in the area of falls in the geriatric population. He was involved in a variety of research projects. In 2007, he received a pre-doctoral national fellowship from the Veterans Administration (VA) to work on his dissertation project. The financial support he received from the VA allowed him to dedicate a full year to the completion of this dissertation project. He received a Doctor of Philosophy degree from the University of Florida in August 2008.