ASSESSMENT OF THE COMMUNITY-DWELLING SENIOR FOR FALL PREVENTION: AN ANALYSIS OF AGREEMENT BETWEEN TWO COMMON PROTOCOLS OF THE TIMED UP AND GO TEST

by

Andrea Murphy

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
July 2014

Table of Contents

List of Tables

List of Figures

Abstract:

Purpose:

To determine if agreement exists between two common Timed Up and Go test (TUG) protocols (i.e., TUG_P and TUG_{SC}) in assessing community-dwelling seniors referred for home care physiotherapy services with balance and mobility impairments.

Rationale:

The TUG is recommended to assess balance and mobility as part of the assessment of falls prevention. Different testing protocols have emerged: i) for mobility: one practice trial plus one measured trial (TUG_P), ii) for falls risk: one practice trial, plus three measured trials, averaged together (TUG_{SC}). The TUG could be a valuable tool in fall prevention programs but variability in the protocols makes it challenging for researchers and clinicians to interpret and use literature findings. Knowledge of agreement between protocols can assist researchers and therapists in choosing, and interpreting, appropriate tests to include in falls prevention programs for seniors.

Materials and Methods:

Seniors, referred recruited for a Home Care Exercise Study [1] for balance problems, who completed baseline screening, including three experimental trials of the TUG.(N=19) Demographic variables were collected to describe the characteristics of the sample.

Analysis:

TUG_P and TUG_{SC} scores were calculated. Descriptive statistical analyses were performed on all variables. Four statistical methods were used to explore relationships between TUG_P and TUG_{SC}: Intra Class Correlation Coefficient($ICC_{(3,1)}$), Standard Error of Measurement (SEM), Spearman Rank Correlation (rho), and Limits of Agreement (LoA)[2].

Results:

Study participants, mean age 82.6 years, presented with a median of four co-morbidities, varied ability to use vestibular inputs, varied balance confidence, but no apparent cognitive impairment. Thirteen participants had a fall history. Half of the sample did not use a gait aid; the other half used either a walker or a cane. The range of TUG_p scores was 8.4-49.8 seconds (s), while TUG_{SC} ranged from 8.7-56.4 s. The $ICC_{(3,1)}$ equaled 0.99 ($p < 0.05$). The correlation analysis revealed rho=0.98 ($p < 0.05$), and a linear equation with a slope of 0.8, and intercept of 2.8. The LoA between the two protocols was 0.9 -1.2 s, and LoA plots revealed potential differences in TUG protocols for scores over 16 s. The SEM equaled 1.4 s and 1.7 s for TUG_P and TUG_{SC}, respectively, however, when outliers were removed, corresponding SEM values of 0.6 s and 0.5 s were observed.

Conclusions:

The ICC and Spearman Rank correlation indicated strong associations between the protocols. The line-of-best-fit analysis illustrated the two TUG protocols were interchangeable for those performing the test in less than 16 s. According to Bland-Altman analysis the scores obtained with TUG_P exceeded those of the TUG_{SC} by approximately one second, which was not a meaningful difference according to the minimal detectable difference. Agreement between protocols was apparent for those performing the TUG quickly, but did not hold true for the whole range of TUG scores.

List of Abbreviations Used

1XF - One-Time Fallers

ρ- rho Spearman's Correlation

ABC - Activities-Specific Balance Confidence Scale

AD - Alzheimer's Disease

ADLs - Activities of Daily Living

AL - Assisted Living

ANOVA - Analysis of variance

CD - Community-Dwelling

C.I. - Confidence Interval Cog.- Cognitive

CTSIB - Clinical Test of Sensory Interaction and Balance

DP - Day program

E.M.P. - Extra-Mural Program

F - Female

FF - Frequent Faller

HC - Healthy Comparison

HRF - High Risk Fall

HHIE-S - Hearing Handicap Inventory for the Elderly-Screening Version

IADLs - Instrumental Activities of Daily Living

ICC - Intraclass Correlation Coefficients

Indep. - Independent

LRF - Low Risk Fall

LTC - long-term care

M - Male

MDC - Minimal Detectable Change

MF - Multiple Fallers

MMSE - Mini-Mental State Exam

Mod. - Moderate

MVIC - Maximum Voluntary Isometric Contraction

NF - Non Faller

NR - Not Reported

P - Practice Trial

PD - Parkinson's Disease

Pro - Prospective

R= Retrospective

ROC - Receiver Operator Curves

ROP - Required Organizational Practices

r - Pearson Product Moment Correlation Coefficient

rho - Spearman Rank Correlation Coefficient

s - Seconds

SEM - Standard Error of Measurement

T - Timed Trial

TUG - Timed Up and Go

TUG_P - TUG protocol of one practice trial plus one measured trial

TUG_{SC} - TUG Protocol of one practice trial, plus three measured trials, averaged together

Yrs - Years

Glossary

Accreditation Canada is a not-for-profit, independent organization that provides health care organizations with an external peer review process to assess and improve the services they provide to their patients and clients based on standards of excellence.

Agreement is when the difference between measurements on one subject is small enough for the methods to be considered interchangeable.[3]

Analysis of Variance (ANOVA) is a statistical method to assess significant differences between two or more group means. [4]

Concurrent validity: denotes a type of criterion-related validity when the measurements are taken at the same time. This type of validity justifies a measures or tests validity in relation with a measure that has previously been validated indicating they are measuring the same or similar construct.[5]

95% Confidence Interval ($CI_{95\%}$) represents a 95 percent chance the individuals true score falls within ± 2 standard deviations of the mean score. [3]

Correlation is a statistical technique that can be used to represent measurement reliability of repeated measures and represents the relationship between two variables. [5]

Extra-Mural Program is a home care division of the Atlantic Health Science Corporation now known as Horizon Health located in Saint John, New Brunswick.

Intraclass Correlation Coefficient (ICC) is a statistical test often suggested to test agreement between or within raters, by comparing two or more measures at one time in order to produce an average correlation among the possible pairs of measurements. [5]It uses an ANOVA to calculate variance estimates and is thought to represent both agreement and correspondence among ratings.[3]

Minimal Detectable Change/Minimal Detectable Difference (MDC/MDD) indicates the smallest difference in performance of a task, over time, that represents a clinically significant difference in the test's performance. [6]

"Podsiadlo protocol" will be used to denote the TUG methodology with one practice and one timed trial (i.e., TUGP) as described by Podsiadlo and Richardson in their 1991 study.

Reliability: refers to the amount of consistency between successive measurements of the same variable, with the same subject under the same conditions.[5]

"Shumway-Cook protocol" will be used to describe the TUG methodology with one practice and three timed trials that are averaged (i.e., TUG_{SC}) used by Shumway-Cook et al. in their 2000 study.

Standard Error of Measurement (SEM) is a statistical tool that examines reliability through the concept of response stability.

Trimmed Sample is used to represent the population of 17 seniors in this study who performed the TUG under 30 seconds. (N=17)

Validity is defined as the degree to which an instrument measures what it is said to measure, the extent to which it fulfills its purpose. [5]

Whole Sample is used to represent the population of seniors in this study who had completed 3 experimental trials of the TUG. (N=19)

Acknowledgements:

This thesis is dedicated to my Mother and Father who, throughout my life told me I could be and do anything I dreamed.

I would like to thank Dr. Marie Earl who without her unfailing support, assistance and belief in me I may never have finished this journey.

Thank you to Dr Marilyn Mackay-Lyons, Dr. Elizabeth MacDonald, Dr. Cheryl Kozey for their suggestions, continuing with the project and encouraging me throughout this academic pursuit.

Thank you to Denise Hollway for being a great collaborator for this academic undertaking.

Thank you to my sister Jennifer, whose support and expert driving skills were much appreciated for my defense.

Lastly and most importantly to my husband Dale and my children, Ella and Audrey who loved, supported and encouraged me throughout the years so I could complete this personal goal.

Chapter 1-Introduction

BACKGROUND, RATIONALE AND OVERVIEW OF THE THESIS:

Background:

Falls are a common problem among seniors with serious consequences, not only in terms of personal health, but also in terms of health care resources. In Canada, the population is aging, with one in every three seniors living in the community experiencing at least one fall per year, with up to 50 percent of these individuals experiencing multiple falls. [7] Falls often have devastating consequences for seniors including: lengthy hospital admissions, fractures, fear of falling, disability, loss of independence, chronic pain, and even death. [8-11] Eighty-four percent of injury-related hospital admissions and 40 percent of admissions to nursing homes are due to falls. [7] The estimated cost of fall-related injuries among those aged 65 and older in Canada was over 2 billion dollars in 2004. [7, 8] The magnitude of these costs, on a personal and system level creates a need for effective strategies to prevent falls by seniors who live in the community, to protect their independence, and to reduce the load on health care resources. [12-14]

Likewise, falls prevention programs specifically for seniors who require home-care health services, due to the frequency, prevalence and serious consequences of seniors falling in their homes are needed. In a study of Canadian home care recipients, 27 percent of the sample reported falling one or more times in the previous 90 days. [15] The Canadian Community Health Survey data indicates that nearly 50 percent of falls by seniors in the home require hospitalization. [14] From 2008 to 2009, the average length of an acute hospital stay for a fall-related injury in seniors was 15 days, as compared to 9 days, for all other non-fall related hospitalizations. [8] Implementation of effective fall prevention

strategies through home-care health services has the potential to reduce this strain at a personal and system wide level.

Current evidence for preventing falls by seniors, dictates use of individualized interventions, based on comprehensive assessment of fall risk factors.[16, 17] Causes of falls in seniors are multi-factorial in nature, and include such risk factors as cognitive impairment, muscle weakness, visual impairment, foot problems and medication use, in addition to balance and gait abnormalities. [11, 18] As falls are multi-factorial, their prevention will be multi-factorial, with multidisciplinary interventions targeting multiple risk factors shown as effective in reducing the incidence and risk of falls in seniors. [16] Effective interventions known in reducing fall rates in seniors can include: muscle strengthening combined with balance retraining that is individually prescribed at home, home hazard assessment and modification, cardiac pacing for fallers with carotid sinus hypersensitivity and withdrawal of psychotropic medication.[13] Fall prevention interventions that are tailored to an individual's risk factors have been shown to be more effective than standard or group delivered programs.[16] Physiotherapists have a role in falls prevention programs through both assessment and treatment of the individual's falls risk factors such as muscle weakness, sensory integration, impaired balance and mobility. [11, 19-21]

A number of sources support the importance of appropriate assessments to determine individual risk factors in falls prevention programming.[17, 22-25] Accreditation Canada, a professional group that peer reviews healthcare organizations for standards of excellence, and the "Clinical Guideline for Prevention of Falls in Older Persons", through the collaborative effort of the American Geriatrics Society and the British Geriatrics

Society, place emphasis on the role of assessment to determine individual risk factors in falls prevention programs. [17, 23] Individually prescribed exercise programs with balance training that targets an individual's impairments, are more successful in reducing falls when based on assessment of individual risk factors rather than a generalized program. [16] Consequently, evidence-based falls prevention programs for seniors receiving home care physiotherapy that are conducted in the home need to include objective assessments of balance and gait, to inform and evaluate targeted interventions that will be effective in reducing the senior's risk of falling.

Rationale:

The Timed Up and Go test (TUG) could be a valuable tool in fall prevention programs administered in the home with seniors, but differences in testing protocols have challenged the interpretation of test results.[26-34] Many studies cite Podsaidlo and Richardson (1991) as the reference study for the TUG protocol and while the administration of the tool is standardized, there are a variety of numbers of practice and timed trials and ways to calculate the score without justification of the deviation from the reference study. [29, 31-41] Shumway-Cook et al. (2000) was the reference study using the TUG to identify fallers, rather than as a mobility tool, and used a modified number of trials with averaged scores to determine the outcome. [37] Though evidence for reliability and validity of the above two commonly quoted TUG protocols is apparent in the literature, little evidence exploring clinically meaningful difference is available and no evidence exploring relationships or agreement between the two protocols was found. Due to this lack of knowledge regarding the protocols of the TUG, it is difficult for clinicians and researchers to interpret and use literature with respect to the TUG and its use in the assessment of mobility in falls prevention programs.

Overview:

The purpose of this retrospective, methodological study was to determine if agreement exists between two common TUG protocols (i.e., TUG_P and $TUG_{SC)}$, for assessing community-dwelling, seniors referred for home care physiotherapy services.

Hypotheses:

Null Hypothesis: The results obtained using TUG_P agree with the results obtained using TUG_{SC}, when assessing community-dwelling seniors referred for home care physiotherapy services.

$$TUG_P = TUG_{SC}$$

Alternate Hypothesis: The score obtained using the TUG_P do not agree with the results obtained using the $TUG_{SC,}$ when assessing community-dwelling seniors referred for home care physiotherapy services.

$$TUG_P \neq TUG_S$$

Hypothesis testing, and associated descriptive analyses, were completed using the data set collected for our Home Care Exercise Study[1] in which I, as a co- investigator, explored the effect of progressive exercise on the balance and mobility of seniors who were referred to home care physiotherapy services due to balance impairment and increased risk for falls. Specifically, our study was a randomized control trial to evaluate the balance and mobility of seniors, with difficulties utilizing vestibular inputs, after an intervention program, targeting strength and balance. The Timed Up and Go test (TUG), was used to establish baseline functional mobility, and evaluate secondary outcomes of

the exercise program. Since the TUG protocol included one practice trial, followed by three recorded trials, it was possible to determine the TUG score according to the two common protocols: the Podsiadlo protocol, in which the test score was determined by the score on the first recorded trial (TUG_P), versus the Shumway-Cook protocol, in which the test score was calculated as the average of the three recorded trials (TUG_{SC}). Using data from the Home Care Exercise Study[1] was beneficial to research including secondary outcomes from the Home Care Exercise Study[1], for published literature and clinical decision making.

In this thesis, the relationships between the protocols were explored using four statistical methods, which are discussed in depth in Chapter 2 of this thesis, along with the analysis of the literature regarding the use of the TUG to evaluate dynamic balance versus fall risk of seniors who require home-care physiotherapy services. Chapter 3 contains the details of the methodology. Results follow in Chapter 4, with the Discussion and Conclusion in Chapter 5.

Chapter 2: Literature Review

Postural Control and Falls

Postural control is a complex motor skill derived from the interaction of multiple sensorimotor processes. [22] It involves controlling the body's position in space for the dual purposes of stability and orientation. [42] Postural control emerges from an interaction of the individual, the task with its inherent postural demands and the environment in which it occurs. [22, 42] All tasks, including gait, involve postural control. To maintain postural control complex interactions of both the musculoskeletal and neural systems are required. [22, 42]

Musculoskeletal components of postural control include joint range of motion (ROM), spinal flexibility, muscle properties, and the biomechanical relationships among linked body segments. [22, 42] Neural components of postural control include motor processes, sensory/perceptual processes, visual, vestibular and somatosensory systems, and higher-level processes that control anticipatory and reactive aspects of posture. [22, 42] External factors in the environment, can affect postural control, for example lighting, obstacles or ice. [22, 42] Balance impairments may arise in the context of a task, such as transfers and mobility, due to a problem in a person's internal system, for example, impaired vision, muscle weakness or dizziness, thus producing increased risk of falls. [22]

Deficits in static and dynamic postural control are major intrinsic risk factors for falls in seniors. [43] Physiological factors that show associations with falls risk include: a reduced ability to maintain stance, increased postural sway, reduced dynamic balance, reduced walking speed, decreased mobility, proprioception in the lower limbs, visual contrast sensitivity, reduced knee, hip, ankle strength and difficulty rising from a chair.

[10, 11, 44] Physiotherapists assess and address many of the noted physiological factors associated with falls in an individual's plan of treatment. An assessment to identify treatable and modifiable risk factors early is a main focus of fall prevention strategies. [45]

The Importance Of Assessment In Falls Prevention Programs For Community-Dwelling Seniors Receiving Home Care Services:

The importance of assessment in falls prevention is evident in the Required Organizational Practices (ROP), identified by Accreditation Canada for all health care teams, including those that provide home care. [23] One of the major tests of compliance, reviewed by accreditors of healthcare organizations within the ROP includes the implementation of a falls prevention strategy. This strategy must identify both the population at risk for falls and address the specific needs of that population at risk for falls. [23] Therefore, this link between assessment to identify issues and specific individualized interventions to modify the risk factors is inherent in the ROP.

The Guidelines for Prevention of Falls in Older Persons, published by The American Geriatrics Society in collaboration with the British Geriatrics Society, promote the use of a clinical algorithm, which links assessment processes to interventions for community dwelling older persons. These Best-Practice Guidelines identify assessment as an important element of falls prevention programs offered throughout primary health care services. [13, 17] According to these guidelines, seniors who present in a clinical setting are asked a short series of questions to establish their history of falls. These questions concentrate on falls that had occurred in the last 12 months, recurrent falls, potential mobility challenges or presentation to an emergency department due to a fall. Answering

"yes" to any of the above questions directs the practitioner to include an evaluation of gait and balance, using a tool such as the Timed Up and Go test (TUG), to help determine the factors that contributed to the individual's falls or risk of falls. [17] The results of the assessment are then used to design an intervention plan to address that individual's specific risk factors.[13, 16, 17]

The objective of using evidence based assessments to inform intervention in falls prevention, is to maintain a senior's quality of life and independence, and help reduce costs to the health care systems. Risk assessments, home safety assessments and targeted individualized interventions have shown reductions in fall rates, falls risks and resulted in cost savings. [13] Savings of 138 million dollars Canadian per year could be achieved by reducing falls and their outcomes among seniors by 20 percent. [7] For seniors who require home care services, application of these best practice falls prevention guidelines requires use of reliable, valid assessment strategies that can be completed in the home, as the basis for developing interventions to address individuals specific risk factors. The TUG is a potentially helpful tool in examining a seniors gait and mobility in the context of fall prevention.

THE TUG AS AN ASSESSMENT TOOL FOR FALLS PREVENTION PROGRAMS FOR COMMUNITY-DWELLING SENIORS RECEIVING HOME CARE SERVICES:

The TUG holds appeal as an assessment tool that can meet both accreditation standards and best practice guidelines for use in, in-home falls prevention programs, for community-dwelling seniors. The TUG is practical for use in the home, as it is scored by measuring the time taken to a complete a series of simple yet specific tasks: rising from a chair of standardized height, walking a specified distance, usually 3 metres, turning,

walking back to the chair and sitting down. As such, the TUG does not provide a comprehensive assessments of balance control, but does provide an objective measure of the performance of basic movements needed to function independently in the home including sit to stand transfers, mobilizing short distances, and negotiating turns. In addition, the equipment required is readily available in the home: a stopwatch, a chair of 46 centimetres seat height with arms and a three-metre area to walk. Clinicians easily use the TUG and the potential for fatiguing the client during assessment is reduced in comparison to some other tests, as it is completed in a relatively short period of time. [39]

The potential of the TUG as an assessment tool in fall prevention programs has prompted many studies to examine the reliability and validity of this tool. Interpretation of the results is challenging, as summarized in Appendix A, even within a designated population of seniors, due to differences in research methods and TUG testing protocols. Some researchers have used one practice trial, plus one recorded trial to score the test (TUG_P), as originally reported by Podsiadlo and Richardson (1991), to assess functional mobility of community-dwelling seniors. [26, 46, 47] Without rationale for the change in number of trials for scoring the test, Shumway-Cook and colleagues (2000) introduced the use of the TUG to screen for fall risk, but used one practice trial, plus the average of three recorded trials (TUG_{SC}), to score the test. Others have used the protocols interchangeably, [27, 28, 30, 48] and still others have varied the testing protocol in additional ways with little or no justification for varying the method. [29, 31, 33, 35, 36, 39, 40] As discussed in the following section of this literature review, these methodological differences make it difficult to interpret the reliability and validity of the tool, which can impact the use, and interpretation of the TUG in falls prevention programs. A detailed analysis of the

reliability and the various types of validity studied for the two commonly -cited TUG protocols are examined, followed by a summary of four statistical methods that may be of value to explore relationships and agreement between the two commonly cited TUG protocols.

RELIABILITY AND VALIDITY OF THE PODSIADLO PROTOCOL (TUG_P)

As illustrated in Appendix A, *Table A1*, the TUG_P protocol is reliable for a community dwelling senior population, supported by the Intraclass Correlation Coefficient (ICC) with excellent results. Intra-rater, inter-rater and test-retest reliability were found to be .99 with this protocol using an $ICC_{(2,1)}$. [47] Similar findings with seniors in residential facilities were found with an inter-rater reliability $ICC_{(1,1)}$ of .91, and an $ICC_{(3,1)}$ of .92 for intra-rater reliability. [49] Other studies involving institutionalized seniors have reported similar results when using Spearman rank correlations with an intra-rater reliability of .96 and inter-rater reliability of .91. [26] The TUG_P protocol therefore shows acceptable reliability in a similar population to home care clients as the studies contained mixed frail elderly populations.

When assessing validity of a tool different types are available for consideration, such as construct validity. The TUG_P protocol appears to have construct validity, given that TUG performance in community dwelling older people appears to be influenced by similar physiological factors associated with fall risk: such as sensorimotor function (r =.14 to .32), lower limb strength (r =-.43 and -.41), balance (r =.29 to .44), and reaction time (r =-.27 and -.37). [50] Studies using TUG_P have explored construct validity using Pearson correlations, revealing moderate associations with Activities of Daily Living

(ADLs)/Instrumental Activities of Daily Living (IADLs) scales (r =-.78) [47], BERG balance scale (r =-.81) [47] and gait speed (r=-0.61). [47]

Criterion-related validity with respect to mobility is demonstrated in the original Podsiadlo and Richardson (1991) study.[47] Significant differences were found in TUG scores between those who were freely independent, those who were independent in basic transfers and mobility (10 to 20 seconds), those in a "grey zone" who required further assessment to clarify their functional mobility (20-30 seconds) and those who were dependent for ADLs (30 seconds or more). [47] Bischoff and colleagues (2003), also explored the use of the TUG_P protocol to classify seniors by the location where they resided either: in the community or in a residential facility. [26] Appropriate statistical tests using Receiver Operating Curves(ROC) and positive and negative prediction values, indicated a person scoring 12 seconds or less, was a community-dwelling senior, with a 90 % confidence interval (C.I). [26]

The original validity studies by Podsiadlo and Richardson (1991) used a TUG_P protocol with the intent of assessing seniors to determine functional mobility norms. [47] Later studies have used this protocol to examine its validity for assessing seniors in relation to falls risk. [24, 27, 28, 30] Studies using the TUG_P protocol for falls risk examined criterion related validity and reported significant differences between groups of non-fallers and fallers. [24, 28] The studies examined different study populations: scores of 17.9 s for non-fallers and 25.1 s for fallers were reported for a falls clinic population, versus the findings from a post-hip-fracture-surgery population showing scores of 29.5 s and 42 s, for non-fallers, and fallers respectively. [24, 28] Despite the differences in findings, these studies demonstrate the potential ability of using the TUG_P protocol to

explore differences between fallers and non-fallers.

Studies examining the predictive validity of the TUG_P protocol for falls risk used ROC curves and reported criterion/cut off scores with sensitivity and specificity values. The participants in three of the four studies found had similarities to community dwelling seniors who use home-care services, while the fourth had a vastly different post surgical population. The reported criterion scores for the TUG_P protocol ranged from 10 to 24 s. Similar to the pattern observed for assessing mobility status with the TUG_P protocol[26], the sensitivities for assessing fall risk ranged from 73 to 96 percent and specificities from 32 to 39 percent. [24, 26-28, 30] The lack of agreement on a criterion score is potentially due to differences between the studies such as the population of interest [26-28, 30], the definition of a fall, and follow-up for tracking falls. [27, 28, 30] Therefore, the utility of a criterion related score to determine falls risk has not been demonstrated and an accepted cut off score has not been determined. [17]

RELIABILITY AND VALIDITY OF THE SHUMWAY-COOK PROTOCOL (TUG_{SC})

As illustrated in Table A2, acceptable reliability using the TUG_{SC} protocol was demonstrated by an $ICC_{(3, 3)}$ value of .98 for inter-rater reliability in community dwelling older adults with and without a history of falls. The population in the Shumway-Cook et al. (2000) study shares similarities to the Home Care Exercise Study [1], as they both contain community dwelling seniors over age 65, with or without a history of falls.[37]

Studies of criterion validity of the TUG that used the TUG_{SC} protocol with community-dwelling senior populations revealed significant differences in the scores of non-fallers versus fallers. [37, 51] Shumway-Cook et al. (2000) reported non-fallers scored 8.4 s

while fallers scored 22.2 s. Wrisely et al. (2010) found similar results with a score of 9.8 s for non-fallers, and 15.8 s for fallers. [37, 51] Efforts to identify criterion scores to distinguish fallers from non-fallers in community dwelling seniors with and without a falls history yield good sensitivity and specificity; a criterion of 12.3 s provided 96 percent specificity and 83 percent sensitivity, while a criterion of 13.5 s yielded 87 percent specificity, and 87 percent sensitivity. [37, 51]

CLINICALLY MEANINGFUL DIFFERENCES

An aspect of validity considered responsiveness, representing clinically meaningful differences referred to as minimal detectable difference (MDD), which is derived from the standard error of measurement (SEM). [6] The SEM represents the standard deviation of the measurement errors of the tool or test [5], while the MDD indicates the smallest difference in performance of a task, over time, that represents a clinically significant difference in the test's performance. [6] Some researchers have determined SEM and MDD values using a TUG_P protocol [30, 49, 52], while others have used variations of the protocol. [36, 38, 49, 53] No studies are found in the literature reporting SEM and MDD scores with the same TUG protocols or a study population of community-dwelling seniors. [36, 38, 49] Studies that are available use populations of persons living in long term care [49], with Parkinson's disease [38, 52, 54], Alzheimer's disease [36] and post stroke. [53]

Based on available evidence, collected with different TUG protocols, in populations of seniors living in the community with different chronic diseases, SEM and MDD's range between 1.1 s to 3.0 s, and 2.9 s to 4.8 s, respectively. [36, 52-54] MDD values that

varied up to 11 s were found in a population of seniors living in long term care with cognitive impairment who demonstrated considerable day-to-day variability in TUG. [49] While the majority of the available studies contain MDD's for specific populations of seniors with progressive chronic diseases, it is not known how these compare to the independently mobile community-dwelling elder, or those seniors who live in the community, with mobility problems. As the SEM and MDD rely on what appears to be population specific results there is a need to explore the SEM/MDD for other groups of seniors.

EVALUATING AGREEMENT BETWEEN TUG PROTOCOLS:

In community dwelling senior populations, despite differences in methods and populations, studies are available that demonstrate reliability and validity of both the TUG_P and TUG_{SC} protocols. [24, 28, 37, 47, 51] Studies are limited in recommending a criterion score to indicate falls risk for both common protocols as they contain different methods and study populations, and thereby produce different criterion scores. It appears that in the current available literature both the TUG_P and TUG_{SC} protocols may yield good sensitivity, but the TUG_P protocol potentially yields poor specificity in terms of falls risk as compared to the TUG_{SC} protocol. [24, 26-28, 37, 49, 51] To date there has been no agreement in a single criterion score with either protocol in determining fall risk of a community dwelling senior.

The research produced thus far demonstrates that the TUG is being used to examine mobility and falls risk with two related but distinct protocols, yet it has not been determined that the two common protocols can be used interchangeably. Since the TUG

has been recommended to assess balance, mobility and falls risk in seniors, [27, 28, 37, 39, 49], the agreement between the methods of TUG scoring requires consideration in light of the importance of fall prevention programs and their potential outcomes on the lives of seniors. Various statistical methods are available with some tools providing a measure of the association and correspondence between protocols, while others explore the extent of agreement between the protocols. The following is a discussion regarding advantages and disadvantages of four statistical methods that may be helpful in exploring the relationship of scores between the protocols.

STATISTICAL METHODS TO EXPLORE AGREEMENT BETWEEN MEASUREMENTS

When considering agreement between protocols the measurements produced by a tool are rarely without error. Error can produce differences in scores and we need to consider that the magnitude of the error may contribute to differences in agreement. Observed scores of a tool are a function of a true score and an error component often classified into systematic errors and random errors. [5] Sources of possible measurement errors can include: the rater, the instrument and the variability of the characteristic being measured. [55] The ICC, SEM, Correlation and Bland-Altman methods were statistical tools used to examine the relationships between TUG_P and TUG_{SC} protocols. Review and discussion of each of these statistical methods will occur in the following section.

ICC

The ICC compares two or more measures at one time in order to produce an average correlation among the possible pairs of measurements. [5]The ICC is a preferred reliability coefficient over other tests, since it is thought to reflect both correspondence

and agreement among ratings through its use of the Analysis of Variance (ANOVA) [55] The ICC is calculated using variance estimates obtained through an ANOVA. [55] Advantages of using the ICC over other methods include the ability to assess reliability among two or more ratings without requiring the same number of raters per subject. [55]

ICC values closer to 1.00 indicate stronger reliability but the ICC value can range from 0.0 to 1.00. [55] Portney and Watkins (2009) suggest that values above .75 indicate good reliability, and those below .75 indicate poor to moderate reliability.[3] In the case of a clinical measure, such as the TUG an ICC result above 0.90 indicates acceptable reliability of the tool. [3]

Three ICC models are available and classified by two numbers in parentheses. [5] The first number indicates the model (1, 2, or 3) and the second number indicates the form (either 1 or k). For a single measurement (1) is used, with k equaling the number of scores used to obtain the mean, if the mean of several measurements is used. [55] ICC model 1 uses a standard one-way ANOVA with subjects as the independent variable. [55] In model 1 each subject is assessed by a different set of raters that are randomly chosen from a larger population of raters with a purpose of generalizing results. Model 2 uses a repeated measure ANOVA with rater being the independent variable. [55] In model 2 each subject is assessed by the same raters which are drawn from a larger population of similarly trained raters with the purpose of generalizing results to the larger population of trained raters. [55] The ICC model number 3 which also uses results obtained from a repeated measure ANOVA design, is indicated when the tested raters are the only raters of interest and when generalizing results are not indicated. [55] Models 2 and 3 are identical in their estimate of reliability, when systematic error is zero. [56] Refer to Table

1 below for a comparison of the various ICC models and forms.

USE OF THE ICC TO EXAMINE AGREEMENT BETWEEN TOOLS/PROTOCOLS

Typically, the ICC is used to examine differences in measurements made by different raters or within a rater. However, it is possible to consider both consistency and agreement between measurement tools using an ICC, by replacing the term 'rater' with "protocol" or "tool" to examine the agreement/consistency between the two methods. [57] Several factors determine the choice of ICC model: number of raters, number of measures, intent to generalize results, and purpose of its use agreement/consistency of measures. When there is only one rater and each protocol produces a single score, the $ICC_{(3,1)}$ is the appropriate choice: model 3 as there is one rater and form 1 because is each TUG protocol produces a single measure using a standardized protocol, which is considered the score for that protocol. With the $ICC_{(3,1)}$, test results are expected to reflect the consistency between the protocols, with little systematic error, using a two-way fixed effects ANOVA. Due to there being a single rater, the ICC model 3 is the only appropriate model for consideration for this study to examine the agreement between the two protocols of the TUG test.

Limitations of the ICC include being influenced by the range of measurements between the subjects and a reduced ability to detect agreement if the group becomes homogeneous, thus erroneously indicating poor reliability.[5] The limitation of the ICC representing agreement is that even though it uses an ANOVA to test for statistical differences between the means of the tools, it may still be composed of pairs of data that do not agree. [3] While the ICC can be sensitive to relationships, the index of the

reliability is not affected by a change or differences in the scale of measurement. When using the ICC to examine agreement the extent to which the two protocols may differ is unknown, the ICC is not reported in the unit of measurement of the tool used for data collection.

Table 1: Models and Forms of the ICC.

	ICC Model$_{(1,1)}$	ICC Model $_{(1,k)}$	ICC Model $_{(2,1)}$	ICC Model $_{(2,k)}$	ICC model $_{(3,1)}$	ICC model $_{(3,k)}$
Subject and rater condit-ion	Each subject is rated by multiple raters, raters randomly assigned, all subjects have the same number of raters [3, 5, 58]	Each subject is rated by multiple raters, raters randomly assigned, all subjects have same number of raters, reliability is for the mean of k ratings[58]	All subjects rated by the same raters , assumed to be a random subset of all possible raters[5, 58]	All subjects rated by the same raters assumed to be a random subset of all possible raters, reliability is for the mean of k ratings[58]	All subjects rated by the same raters assumed to be the entire population of raters [5, 58]	All subjects rated by the same raters assumed to be the entire population of raters, reliability is for the mean of k ratings. Assumes no subject by judge interaction[58]
ANOVA	one way random effects [3, 58]	one way random effects [3, 58]	Two way random effects Absolute agreement[3, 58] Recommended when you want to examine agreement of the judges are they interchangeable [58]	Two way random effects Absolute agreement [3, 58] Recommended when you want to examine agreement of the judges are they interchangeable [58]	Two way fixed effects Consistency [3, 58] Recommended when there is one rater and to examine consistency treating the judges as a fixed effect [58]	Two way fixed effects Consistency [3, 58] Recommended when there is one rater and to examine consistency treating the judges as a fixed effect [58]
Unit of analysis	single measure or rater [3, 58]	average of measure or raters[3, 58]	single measure or rater [58] Inter trial reliability (generalizability) for a single trial/rater/tool is considered 2,1 [56]	average of measure or raters [58] Inter trial reliability (generalizability) for an average of k trials is considered 2,k [56]	single measures or rater [58] Model 3 used when only one rater[58]	average of measure or raters [58]

CORRELATION

Correlation is a statistical technique that represents measurement reliability of repeated measures and reflects the relationship between two variables. [5] The variables are plotted on the X and Y-axes and are graphically depicted using a scatter plot. [4, 55] Correlation coefficients assess the degree to which variables are linearly related. [4, 59] The correlation reports a direction, positive or negative, and strength of the relationship. [4, 59] By convention in statistics, correlation coefficients range in value from -1.0 to +1.0, with values approaching 1.0 indicating a stronger association, whereas those approaching 0.0 indicate a weaker relationship or no systematic correlation. [55]

Correlations quantitatively describe the strength and direction of a relationship between two variables, but not necessarily agreement. [2, 5, 55] In order to determine agreement using correlation, the equation of the straight line must also be considered. [5] The equation of the straight line is expressed as: $y = a + bx$. For the measures to be identical, the slope (b) would equal 1, the intercept (a) would equal 0 and the r value would approach +1. [5]

Correlations are susceptible to testing effects, if the results contain a systematic error this will lead to a consistent change across measurements. [55] Using a correlation coefficient does not reflect these consistent changes across measurements, making this form of testing susceptible to bias. [55] The reliability coefficient remains the same if a bias occurs within the context of a correlation as it represents the measures proportionality to each other. [55] Often research of a clinical nature seeks confirmation that two measures are the same, not just proportional. [55] Using correlation allows only two variables at a

time to be examined (X vs. Y), which can be a disadvantage. [55] Correlation requires only pairs of data sets to be examined and this limits its use when multiple sets of data and data that may not be paired are being studied. [55] In the case of interchangeability of the TUG protocols, it would be unusual if the correlation did not show that the protocols are related, as they are both testing the same quality using the same measure, differing only by number of trials for scoring.

BLAND-ALTMAN METHOD

Using a Bland Altman approach to assess agreement between TUG protocols would provide an answer in the same unit of the measure, which may be more useful to answer the question of agreement. Bland and Altman (1986) argue that it is unlikely that different methods or tests will agree and give exactly the same results for all individuals, when examining agreement. [2] By using clinical interpretation and determining by how much one method is likely to differ from the other, the feasibility of replacing one method/tool with another or using them interchangeably can be examined. [2] The Bland Altman method involves first examining the correlation scatter plot, then plotting the differences between the methods against their mean.[2] By plotting the differences between methods against their mean allows examination of relationships between the measurement error and the true value, estimated to be the mean of the two measurements. [2] In the event that data is not normal in its distribution, or there is an apparent increase in variability of the differences as the measurement increases, the data should be log transformed. [60].If a consistently biased pattern is noted in throughout the plot of the limits of agreement (LoA) the tools cannot be considered interchangeable. [60]

In the literature no disadvantages on the use of the Bland Altman method to examine agreement are noted, only cautions to be consistent in following the steps as outlined by the Bland Altman method. [61, 62] Critics of the Bland Altman approach cite that the LoA are a reference interval and should not be used in isolation to conclude agreement between two methods or devices. [62] Bland and Altman recommended that clinical considerations, such as what would constitute an acceptable difference in the devices/tools, should be pre-contemplated and along with the LoA used to determine interchangeability. [2]

STANDARD ERROR OF MEASUREMENT (SEM)

The SEM examines the concept of response stability and may assist in conjunction with other statistical methods in exploring agreement through clinically meaningful differences, but used alone cannot determine agreement. [3, 5] The SEM calculation relies on a reliability score and the standard deviation of the sample mean. The SEM represents the standard deviation of the measurement errors of the tool or test.[5] If a measure is consistent in its measurement, the distribution becomes less variable, the errors are smaller, and the resultant SEM is smaller. [3] The smaller the SEM, the more reliable the measure is considered.[3] The SEM is used in the calculation of both the 95 percent confidence interval ($CI_{95\%}$), and Minimal Detectable Difference (MDD) score. The SEM is clinically valuable, as it can be used for examining performance over time, through calculation of the MDD score.[3]

An advantage of using the SEM is that the result is reported in the same unit of measurement of the tool used. Using SEM findings in conjunction with other statistical

methods to explore agreement helps with clinical considerations and judgments on differences that may be acceptable or clinically meaningful. A disadvantage of using the SEM is that its calculation is reliant on a reliability score and the standard deviation of the sample. If there is a lot of variability in subjects scores, this will produce a large standard deviation and resultant larger SEM, which may be biased due to the scores of a few in the sample, this becomes more apparent in studies with smaller sample sizes and subjects that have large variations in performance.

SUMMARY AND STATEMENT OF PURPOSE:

Evidence-based falls prevention programs for seniors receiving home care physiotherapy need to include screening tools and objective assessments of balance and gait, which can be completed in the client's home. The TUG can be a valuable tool to meet these needs, but differences in the testing protocols have challenged the use and purpose of the tool and interpretation of the results. The original TUG_P demonstrated reliability and the rationale for the protocol change for TUG_{SC} is not identified. Interchangeability of the protocols may be expected as the task itself is identical, but the possibility of repeated trials influencing scores if learning or fatigue change performance on subsequent trials is unknown. Knowledge of agreement between the protocols allows for comparisons between studies, and gives a broader evidence base to inform therapists in the appropriateness and application of the tool, in order to assess seniors in falls prevention programs. Therefore, the purpose of this study was to determine if agreement exists between two common TUG protocols (i.e., TUG_P and TUG_{SC}) in assessing community-dwelling seniors referred for home care physiotherapy services, considering four statistical analyses.

Chapter 3-Methods

STUDY DESIGN

RESEARCH DESIGN: This retrospective methodological research study was conducted, using quantitative methods, to assess agreement between scoring the TUG test performance based on the first recorded trial (TUG_P protocol), versus the average of the three recorded trials (TUG_{SC} protocol). Analysis was done using data collected for a Home Care Exercise Study [1].

BACKGROUND

The Extra-Mural Program Home Care Exercise Study [1] was a randomized control trial to evaluate the effect of exercise on balance and mobility of seniors referred to home-care services due to balance problems. The Extra-Mural Program is a home care division of the Atlantic Health Science Corporation, now known as Horizon Health, in Saint John, New Brunswick. Approval from both the Dalhousie University Ethics Board and the Atlantic Health Science Research Ethics Board was obtained including informing practice within the population studied. A grant from the Health Promotion Research Fund of the Atlantic Health Science Corporation provided funding for the paid research assistants. [1]

Recruitment of subjects for the Home Care Exercise Study [1] occurred in two ways: a waitlist review or a letter of information/invitation to physicians to refer to the study. EMP physiotherapy staff conducted a waitlist review of existing referrals for those over 65 years of age with balance and mobility issues and this practice continued throughout the study for ongoing referrals. Physicians

received a letter on two occasions outlining both the study and referral process. A copy of the letter to physicians is included in Appendix B.

After review of the referrals by EMP staff, appropriate referrals were forwarded to the research team for screening and, if appropriate, the potential subjects invited to participate. After the subject provided initial consent to participate, a member of the research team met with them and performed the initial screening. The research member who conducted the initial screening provided further information about participation in the study and obtained informed consent. The subjects' medical history encompassed areas such as age, height, co-morbidities, vision and hearing. Family physicians provided written medical clearance for a subject's participation in the study.

Trained research assistants collected baseline assessments of the Home Care Exercise Study [1] participants, which, included the following variables: Clinical Test of Sensory Interaction and Balance (CTSIB-to document sensory integration for standing balance), the TUG (to examine mobility), the Activities-Specific Balance Confidence Scale (ABC Scale-to examine balance confidence), Maximum Voluntary Isometric Contraction Strength Test (MVIC-to measure strength), Hearing Handicap Inventory for the Elderly-Screening Version (HHIE-S*- to reflect functional hearing impairment in older adults), the Mini-Mental State Exam (MMSE- to screen for cognitive impairment). Appendix B contains a flow chart of procedures illustrating participants and participation in the Home Care Exercise Study [1]. TUG scores were collected using one practice trial, followed by three recorded trials. A detailed description of each measure and procedure used for data collection is included in Appendix C, along with protocols completed during the baseline

assessment.

Random assignment of subjects to a strength training-only control group or a combined balance and strength-training group occurred after meeting the study criteria and obtaining medical clearance from the family physician. The paid research assistants completed re-assessments at 8 weeks and at 6 months post exercise training

PROCEDURE:
THE SAMPLE

Selection and Sample Size: Community-dwelling seniors referred for home care physiotherapy services for balance and mobility impairments were recruited for the Home Care Exercise Study [1]. The data used for this sub-study were drawn from those who completed the baseline screening, including three experimental trials of the TUG. Whether or not subjects participated in the exercise portion of the study was not relevant. (N=19) See Figure 1.

Inclusion criteria for the Home Care Exercise Study [1] included those participants who were over age 65, living at home, and had the ability to ambulate six metres and stand independently with or without an aid or assistance. This criterion was required in order to ensure that the senior population selected was able to perform the TUG successfully.

Subjects excluded from the Home Care Exercise Study [1] included unstable medical conditions, weight-bearing restrictions; pain with weight bearing; and acute osteoarthritis in their lower extremities. These restrictions were created for safety purposes. In addition excluding those with an MMSE less of than 23; those attending physiotherapy at time of

study; those who were legally blind, those with a diagnosis of progressive neurological condition those with an abnormal vestibular ocular reflex or evidence of nystagmus was done to reduce confounding effects. The Home Care Exercise Study [1] excluded seniors with a CTSIB test 5 score greater than 15 s as a score under 15 s was necessary to select seniors with difficulties using vestibular inputs. For this study, there was no exclusion of subjects based on ability to use vestibular inputs. Appendix **B** includes a description of the inclusion and exclusion criteria based for the Home Care Exercise Study [1].

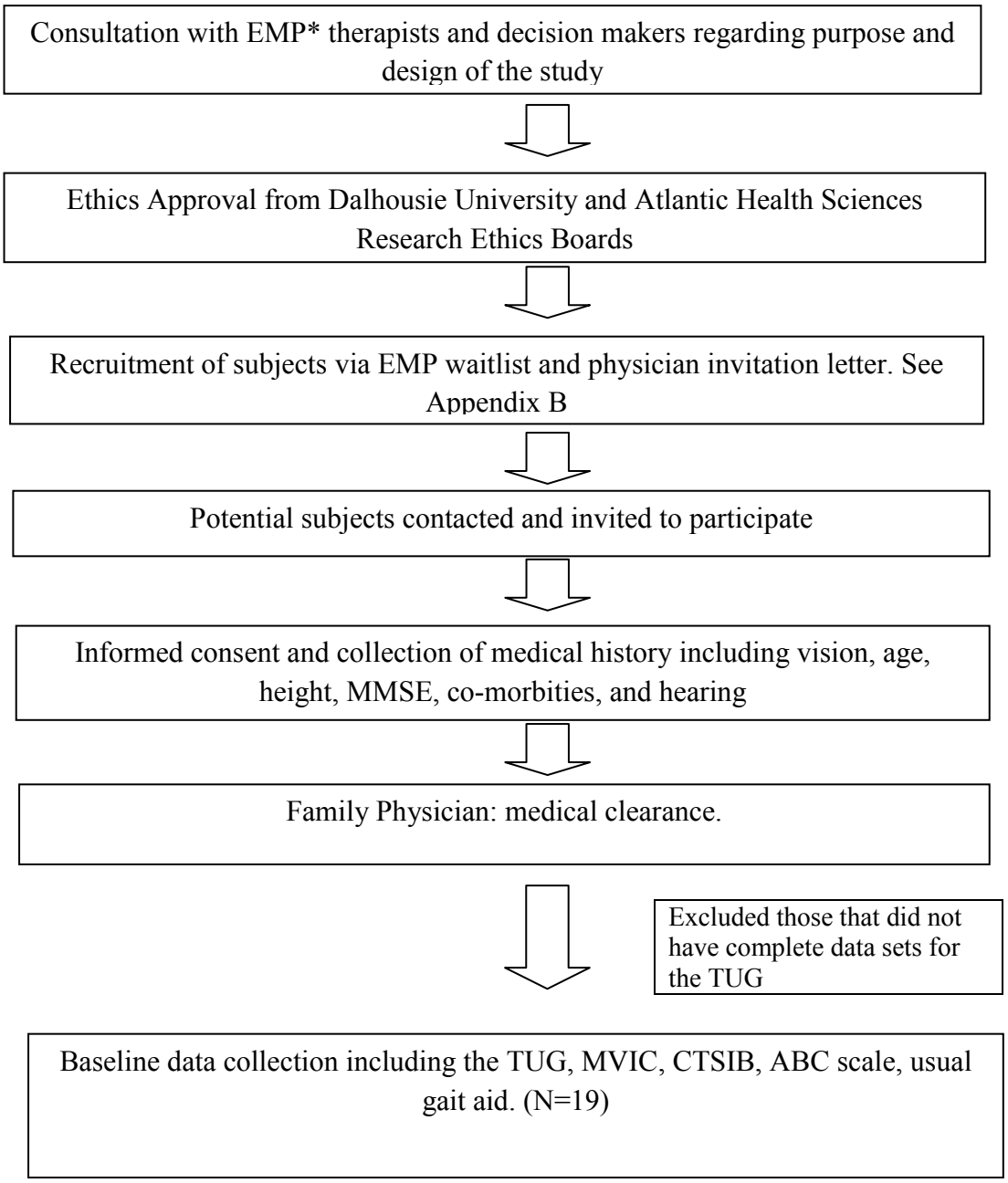


FIGURE 1: FLOWCHART OF RECRUITMENT AND BASELINE TESTING PROCEDURES.

MEASUREMENTS:

TIMED UP AND GO (TUG)

The TUG has been used as both a screening tool in assessing functional mobility and/or fall risk [47, 49], although it is not favored as a criterion based tool to screen fall risk across the continuum of care.[17] The time taken to complete the test is associated with the level of functional mobility. [47] Older adults (age 70 to 84 years), who complete the task in less than 20 s have been shown to be independent with transfers and Activities of Daily Living (ADL`s). [47] Those scoring under 20 s have also been shown to have high BERG scores and walk at gait speed that should be sufficient for community mobility.[47] Older adults with scores on the TUG of 30 s or longer tended to be more dependant in ADL`s, required assistive devices for mobility and score lower on the BERG. [47] TUG results between 20-30 s indicate a "grey zone," in which further evaluation is required in order to determine functional capabilities. [47] Shumway-Cook et al. (2000) indicated that the TUG is a valid screening test for falls risk in community dwelling older people. They determined that older adults who take longer than 13.5 s to complete the TUG were at a high risk for falls. [37]

The TUG has been used with several populations; including community dwelling seniors, long term care residents and those with and without cognitive impairments. [37, 49] Reported ICC's between .91-.99 lie within acceptable limits for intra and inter-rater reliability and test re-test conditions. Thus, it appears the TUG is a reliable measure as a functional mobility tool.[37, 47, 49]

The TUG holds appeal for use in home care, as it is quick and easy to administer and little equipment is needed (a chair, stopwatch, measuring tape and a 3 meter area to walk). Using the TUG as a screening tool for gait and balance issues may be an efficient use of resources to direct interventions.[24]

Refer to Appendix C for descriptions of the TUG and the standardized administration that was used.

THE ACTIVITIES-SPECIFIC BALANCE CONFIDENCE SCALE (ABC SCALE)

The ABC scale was developed to measure balance confidence /self-efficacy. [63] The ABC scale consists of a 16-item questionnaire that can be self administered or completed over the phone. Each item describes a specific activity that requires either a change in position or ambulating in progressively balance challenging situations.

The ABC scale appears to be a reliable scale that has internal consistency (Cronbach's alpha 0.96) and has shown test-retest reliability after a 2 week interval (r=0.92) in community dwelling elderly people. [63] The ABC scale demonstrates validity through Pearson correlations between the ABC scale and BERG of .72 and the ABC and TUG of .70. [64] Significant differences in ABC scores between fallers and non fallers have also been established with a suggested cut off score of 67 percent [63] on the ABC resulting in 84.4 percent sensitivity and 87.5 percent specificity in predicting future falls. [65]

The ABC scale is suited for use in home care as it appears to be a reliable and valid test, and it appears to correlate to performance on balance measures. No specialized equipment is required to complete the test.

Refer to Appendix C for descriptions of the ABC scale and the standardized

administration that was used.

CLINICAL TEST OF SENSORY INTERACTION AND BALANCE (CTSIB)

The CTSIB is a tool that assesses a person's ability to use sensory input for balance. [66] Test re-test reliability has been found to be .75 in older community dwelling adults, and test retest and inter-rater reliability being .99 in healthy young subjects.[67, 68] The CTSIB appears to be a valid method to examine older patients' ability to utilize somatosensory inputs and to remain upright. [69-71] Foam posturography (CTSIB) had a specificity of 90 percent and sensitivity of 95 percent for identifying those with vestibular dysfunction, as compared to platform posturography. [69] It appears that the ability to maintain balance for 30 s, during this test, and in all test conditions should be possible even in older adults. [71]

Refer to Appendix C for descriptions of the CTSIB and the standardized administration that was used.

MAXIMUM VOLUNTARY ISOMETRIC CONTRACTION STRENGTH TEST (MVIC)

Hand Held Dynamometry (HHD) is used to measure strength in community-dwelling seniors [72] It appears to be a reliable and valid tool with test re-test reliability in community dwelling elderly fallers, resulting in reported ICC $_{(2,1)}$ values ranging from .95 to .99 and ICC $_{(2,2)}$ values from .97 to 1.0. [73] In community dwelling seniors, results of HHD correlated well with those of the Biodex with an r-value of .91. [74]

Refer to Appendix C for descriptions of the MVIC and the standardized administration that was used.

HEARING HANDICAP INVENTORY FOR THE ELDERLY-SCREENING VERSION (HHIE-S)*

The HHIE-S is reported as a valid and reliable tool in detecting functional hearing impairment in older adults. [75-77] The HHIE-S consists of 10 questions which are scored for each question accordingly, "Yes"- equals 4 points; "Sometimes" equals 2 points, or "No" equals 0 points. [75, 77] The minimum score is zero, which indicates no handicap and up to 40 which indicates maximum handicap. [78] Scores from 0-8 indicate a 13 percent probability of hearing impairment, scores between 10-24 indicate a 50 percent probability of hearing impairment and score between 26 to 40 indicate an 84 percent probability of hearing impairment. [75, 76]

Refer to Appendix C for descriptions of the HHIE-S and the standardized administration that was used.

MINI-MENTAL STATE EXAMINATION (MMSE)

The MMSE is accepted as a reliable and valid tool in screening cognitive impairment in community-dwelling, hospitalized and institutionalized older adults. The MMSE is an 11-question tool that examines five areas of cognition: orientation, registration, attention and calculation, and recall and language. [79, 80] The MMSE is scored out of 30, with scores of 23 and below indicating the presence of cognitive impairment. [80]

Refer to Appendix C for descriptions of the MMSE and the standardized administration that was used.

DATA ANALYSIS

Data were analyzed using SPSS Version 22 (IBM Corporation, Armonk, New York).

Descriptive statistical analyses were performed on the following data, to summarize the characteristics of the participants: age (years), co-morbidities (number), sex (M/F), hearing handicap score (number reflecting score out of 40), MMSE score (number indicating score out of 30), ABC scores (percentage up to 100), CTSIB scores (s) and usual gait aid.

The above variables were plotted and examined for normal distribution and a Shapiro Wilk test of normality for continuous data were performed with an alpha level of 0.05. P-values were recorded to two decimal places, where applicable. The characteristics of the participants were summarized with minimum and maximum values, range, and the median for all variables; the mean and standard deviation were reported when a normal distribution was present.

TUG data were plotted by protocol and trial for each participant and examined for a normal distribution and outliers. When outliers were identified, analyses were carried out with (Whole Sample) and without outliers (Trimmed Sample) included. A Shapiro Wilk test of normality for continuous data was performed with an alpha level of 0.05. P-values were recorded to two decimal places, where applicable. Non-normally distributed data were log transformed for performance of statistical tests when a non parametric option was unavailable. The TUG scores were summarized with minimum and maximum values, range, and the median; the mean and standard deviation were reported when a normal distribution was present.

Four statistical methods were considered to explore relationships and agreement between the two TUG protocols, TUG_P and TUG_{SC}.

A) ICC

The TUG_P and TUG_{SC} protocols were examined for agreement using an $ICC_{(3,1)}$. Since the data for TUG_P and $TUGS_{SC}$ were not normally distributed, data were log transformed prior to the application of the statistical tool. In this study, there is one rater and each TUG protocol produces a single value, which is considered the score for that protocol. An $ICC_{(3,1)}$ was chosen as test results are expected to reflect the consistency between the protocols with little systematic error. It is calculated using results from a two-way, fixed effects ANOVA.

B) Correlation

TUG_P and TUG_{SC} data were plotted for normality. Non parametric statistical methods were used, as these data sets were not normally distributed and alpha levels were set at .05 with p values reported to two decimal places. A Spearman rank correlation (rho) was used to explore the relationship between TUG_P and TUG_{SC}, in conjunction with the equation of the straight line, represented by:

$$y = a + bx.$$

For the measures to be identical, the slope (b) would equal 1, the intercept (a) would equal 0 and the r value approaches +1. [5]

C) Limits of Agreement (LoA)

Bland-Altman (1986) methods were used for assessing agreement between TUG_P and

TUG_{SC} [2], for all participants, and with the outliers removed, [60] with the following steps:

i) Data were evaluated for normality of distribution using histograms and tests of normality as detailed previously. As the results for TUG_P and TUG_{SC} were not normal in their distribution, and demonstrate an increase in variability as the magnitude of the measurement increases, the data were log transformed (log_n) prior to analysis in order to examine the plot between the mean log_n difference score versus the mean log combined scores of the two protocols.[60] This was completed for all 19 participants (N=19) and for results with the 2 scores over 30 removed (N=17) as explained above.

ii) TUG_P was plotted against TUG_{SC}.

iii) Difference Scores were calculated with log transformed data,

Equation 1:

$$d_{meanlogn} = log_nTUG_P - log_nTUG_{SC}$$

iv) Mean Combined Score with log transformed data were calculated.

Equation 2:

$$\text{Mean log combined score} = \frac{(\text{lognTUG}_P + \text{lognTUGsc})}{2}$$

v) Mean log Difference Score was plotted versus Mean log Combined Score

vi) The limits of agreement (LoA) were calculated using

Equation 3:

$$LoA = d_{meanlogn} \pm 2(sd_{meanlogn})$$

Where:

$d_{meanlogn}$ =mean of log_n calculated Difference Scores

$sd_{meanlogn}$= standard deviation of mean log_n calculated difference scores

$$\text{Lower LoA} = d_{meanlogn} - 2(sd_{meanlogn})$$

$$\text{Upper LoA} = d_{meanlogn} - 2(sd_{meanlogn})$$

vii) Data were then reverse log transformed, and the limits of agreement calculated in the correct unit of measurement (seconds).

Equation 5:

$$LoA = d_{\text{mean reverse log}} \pm 2(sd_{\text{mean reverse log}})$$

Where:

$d_{\text{mean reverse log}}$ = mean reverse log of the calculated Difference Scores

$sd_{\text{mean reverse log}}$= standard deviation of mean reverse log calculated difference scores:

$$\text{Lower LoA} = d_{\text{mean reverse log}} - 2(sd_{\text{mean reverse log}})$$

$$\text{Upper LoA} = d_{\text{mean reverse log}} - 2(sd_{\text{mean reverse log}})$$

D) SEM, CI 95 % AND MDD

Used in conjunction with other statistical methods to examine clinically relevant differences of agreement. The following formulas were used in the calculations:

Equation 5: Standard Error of Measurement (SEM):

$$SEM = S_x\sqrt{1-r_{xx}}$$

Where:

S_x is the standard deviation of the observed test scores for each of the two protocols,

and

A reliability score is needed for the SEM to be calculated. The test -retest reliability results available in the literature and the $ICC_{(3,1)}$ and Spearman Rho results from this study were examined to determine the reliability score to be used. With test - retest reliability results of the TUG in literature ranging from .97 to .99 [34, 36, 37, 47] a value of .98 was chosen as the reliability coefficient (r_{xx}) value for the SEM calculation.

r_{xx} = .98 was used as the reliability coefficient for the measurement

Equation 6: $CI_{95\%}$ was calculated using the following formula:

$$CI_{95\%} = [\text{mean} \pm 1.96 * \text{SEM}]$$

Minimal Detectable Difference (MDD) is defined as the amount of change required in a variable that must be achieved to reflect a true difference.[3]

Equation 7: MDD:

$$\text{MDD} = 1.96 * \sqrt{2} * SEM$$

Chapter 4-Results

SAMPLE CHARACTERISTICS (N=19):

The characteristics of the sample are summarized in Table 2 and Table 3. Age followed a normal distribution (W=.98, p= .92), as did the number of co-morbidities (W=.94, p=.24) and the ABC scores (W=.94, p=.27), so these variables were summarized using the mean and standard deviation. All other variables are reported using the median, minimum, maximum values and range, since the data did not demonstrate normal distribution ($p < 0.05$). There was incomplete data regarding fall history with 2 participants having an unknown fall history (Participant Identification Codes -EM 03, EM 22) and 4 non fallers (Participant Identification Codes - EM 02, EM 06, EM 34 and EM 37) while the rest of the participants were fallers. As noted in Table 3 below MMSE scores ranged from 21-30. The only MMSE score below the exclusion cut off was one MMSE score of 21. The score of 21 obtained by participant EM 03 was questioned for accuracy as the participant had a questioned level of literacy and the MMSE was not administered in his native language and therefore in terms of with accepted practice of scoring the MMSE and medical clearance it was determined he was able to provide informed consent and was appropriate for inclusion in the study.[81] Refer to Appendix D for the complete results of the normality tests, including histogram plots of the data; the complete data sets of participant characteristics and TUG scores by trial and protocol for each participant are also included.

TABLE 2: SUMMARY OF STUDY PARTICIPANT CHARACTERISTICS: SEX, FALLS HISTORY, VESTIBULAR ISSUE AND GAIT AID

Characteristic	**N=19**			**N=17**	
Sex	**Female**	**14**	**(73.7%)**	**Female**	**12(70.6%)**
	Male	**5**	**(26.3%)**	**Male**	**5 (29.4%)**
Falls history (Prior)	Non faller	4	(21.0%)	Non faller	4 (23.5%)
	Faller	13	(68.4%)	Faller	11(64.7%)
	Unknown	2	(10.5%)	Unknown	2 (11.8%)
Vestibular issue	Yes	10	(52.6 %)	Yes	8 (47.0 %)
	No	9	(47.4%)	No	9 (53.0%)
Usual Gait aid	None	10	(52.6%)	None	10(58.8%)
	Cane	3	(15.8%)	Cane	3 (17.6%)
	2-wheeled walker	4	(21.1%)	2-wheeled walker	3 (17.6%)
	4-wheeled walker	2	(10.5%)	4-wheeled walker	1 (5.9%)

TABLE 3: SUMMARY OF STUDY PARTICIPANT CHARACTERISTICS: AGE, ABC SCORE, CTSIB, HHIE-S, MMSE AND CO-MOBIDITIES

Characteristic	**Mean (sd) N=19**	**Median (min-max) Range**	**Mean (sd) N=17**	**Median (min-max) Range**
Age (years)	**82.4 (6.6)**	**82.0 (71-97) 26**	**81.9 (5.6)**	**81.0 (71-97) 26**
ABC score (max=100)	**63.2 (28.1)**	**71.5 (0-99.4) 99.4**	**68.8 (5.7)**	**74.7 (27.5-99.4) 71.9**
CTSIB (s)		**24.6 (0-30.0) 30**		**24.6 (0-30.0) 30**
Hearing Handicap		**6 (0-26) 26**		**6 (0-26) 26**
MMSE		**29 (21-30) 9**		**29 (21-30) 9**
Number of co-morbidities	**3.5 (1.5)**	**4.0 (1-7)**	**3.5 (0.4)**	**4.0 (1-7)**

TUG DATA BY PROTOCOL

TUG data were plotted by trial, and by protocol, TUG_P and $TUG_{SC,}$ for each participant, as illustrated in Figures 1 and 2. The raw data used for these plots is included in Appendix D, Table D-1. Two of the 19 participants had scores above 30 s (participant identification codes EM 19 and EM 33) for both the TUG_P and TUG_{SC} protocol as illustrated in Figures1 and 2,. By inspection, for both of these participants, the TUG_P (trial 1) score was their fastest score, with subsequent trials showing an increase in time taken to complete the trial. TUG_{SC} scores were slower than TUG_P for these two participants scoring over 30 s. For those scoring under 30 s, on both the TUG_P and $TUG_{SC,}$ the slowest score usually occurred on the first trial (TUG_P), or the trials were so close that they were indistinguishable. The two participants with scores exceeding 30 s (EM 19, EM 33) were considered outliers. The data were examined with the Whole Sample (N=19) and with the data from those scoring over 30 s using the TUG_P or TUG_{SC} protocol removed, the Trimmed Sample (N=17). The summary statistics are presented in Table 4.

SEM, MDD and $CI_{95\%}$ are used in conjunction with other tests as a benchmark to explore agreement. Results of the SEM, MDD and $CI_{95\%}$ are summarized in Table 4, for the whole sample and the trimmed sample. Appendix D contains the calculations of the SEM, MDD and $CI_{95\%}$.

TABLE 4: MEAN, SD, MEDIAN, MINIMUM, MAXIMUM, CI $_{95\%}$, SEM, MDD AND NORMALITY OF DISTRIBUTION BY PROTOCOL AND POPULATION.

	Mean (s)	Sd (s)	Median (s)	Min (s)	Max (s)	**CI $_{95\%}$ (s)**	**SEM (s)**	**MDD (s)**	Shapiro -Wilk (W)
TUG_P (N=19)	na	na	14.6	8.4	49.8	15.1 – 20.7	1.4	4.0	0.70 p=.00
TUG_{SC} (N=19)	na	na	13.7	8.7	56.4	14.56 – 21.2	1.7	4.7	0.63 p=.00
TUG_P (N=17)	na	na	14.1	8.4	26.1	13.70 – 15.9	0.6	1.6	0.88 p=.02
TUG_{SC} (N=17)	**14.1**	**3.7**	13.4	8.7	23.7	13.14 – 15.1	0.5	1.4	0.90 P=.06

na=Not applicable

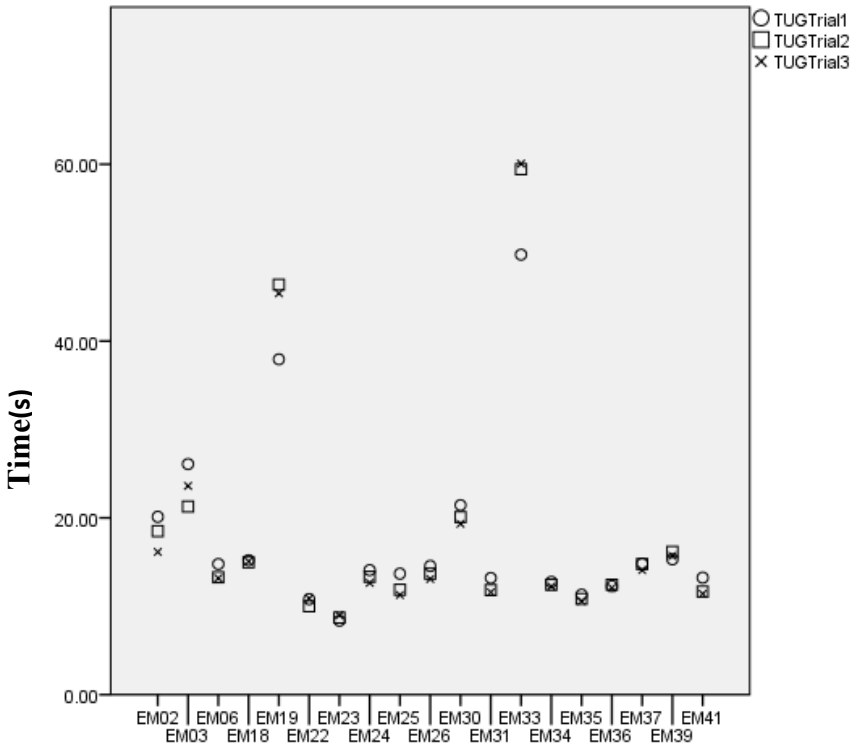


FIGURE 2: TUG PERFORMANCE BY PARTICIPANT AND TUG TRIAL.

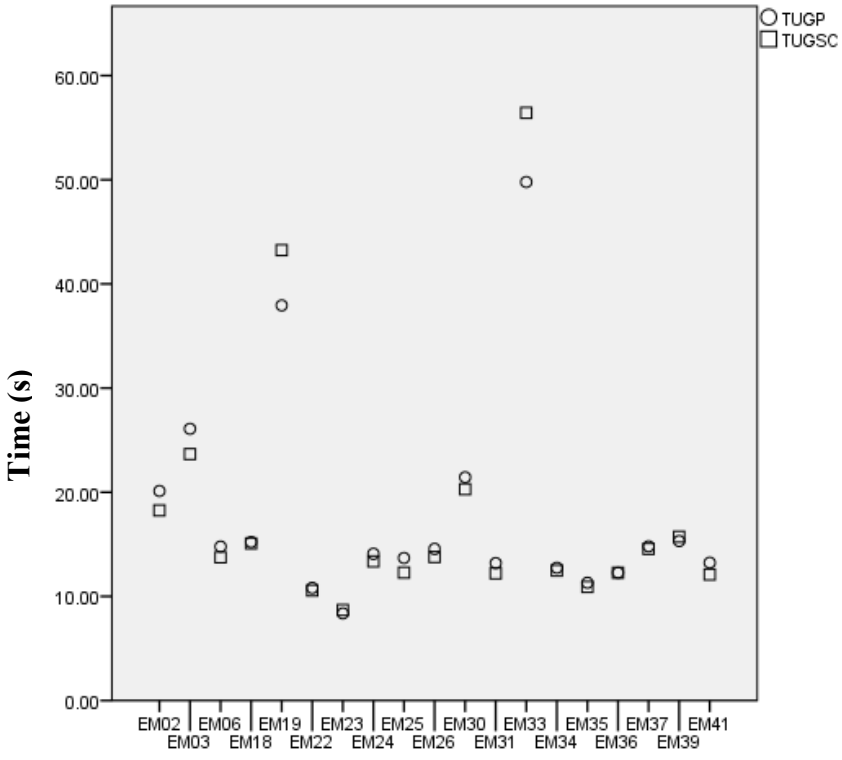


FIGURE 3: TUG_P AND TUG_{SC} SCORES, BY PARTICIPANT.

EVALUATING AGREEMENT, METHOD A): ICC MODEL (3,1)

Results of the agreement/consistency analysis using $ICC_{(3,1)}$ are summarized in Table 5, for the Whole Sample and the Trimmed Sample. Appendix F contains the SPSS output tables with the ANOVA and ICC results.

TABLE 5: $ICC_{(3,1)}$ RESULTS DATA WERE LOG TRANSFORMED.

Participants	$ICC_{(3,1)}$	CI 95%	$ICC_{(3,k)}$	CI 95%
(N=19)	.99	.97 - 1.00	.99	.98 - 1.00
(N=17)	.98	.96 - 1.00	.99	.98 - 1.00

EVALUATING AGREEMENT, METHOD B) CORRELATION:

Correlation coefficients were computed for TUG_P and TUG_{SC} using Spearman rank correlation techniques, since the data were not normally distributed and therefore non-parametric techniques were warranted. For the WS, the Spearman rho was found to be 0.98 (p=.00). For the TS, the Spearman rho was found to be 0.97 (p=.00). See Figures 4 and 5, for scatter plots of the Spearman rank correlation (ρ), with the corresponding equations. SPSS output tables are included in Appendix G.

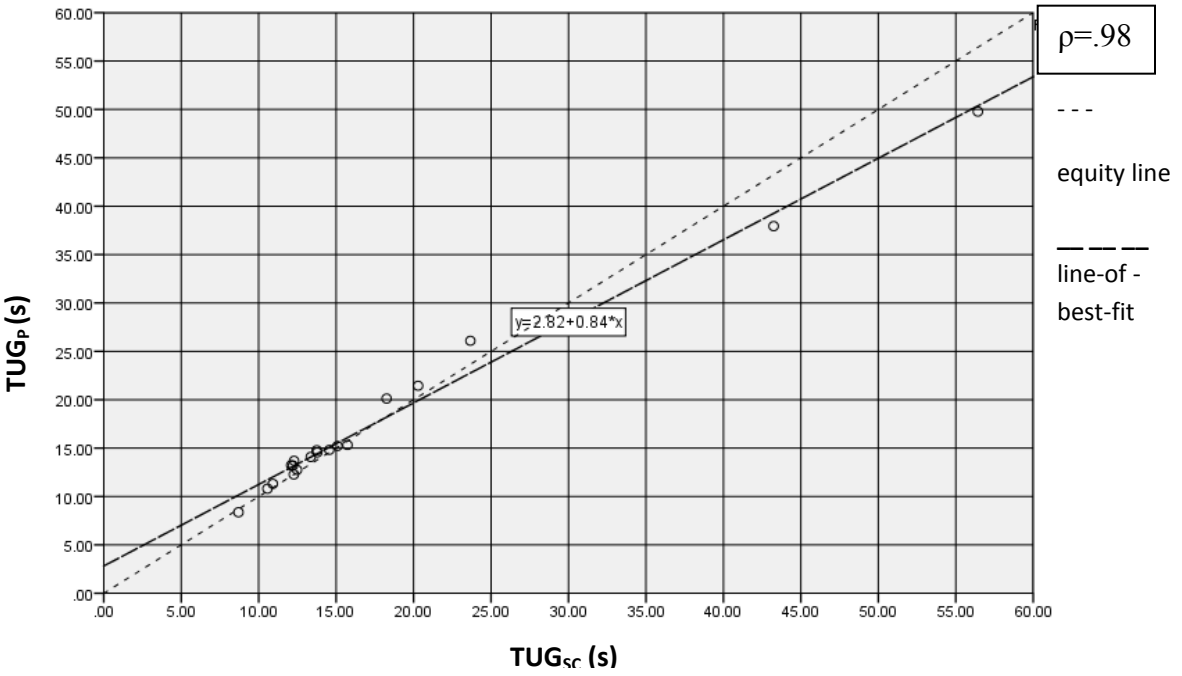


FIGURE 4: SCATTER PLOT OF TUG_P VERSUS TUG_{SC} (N=19).

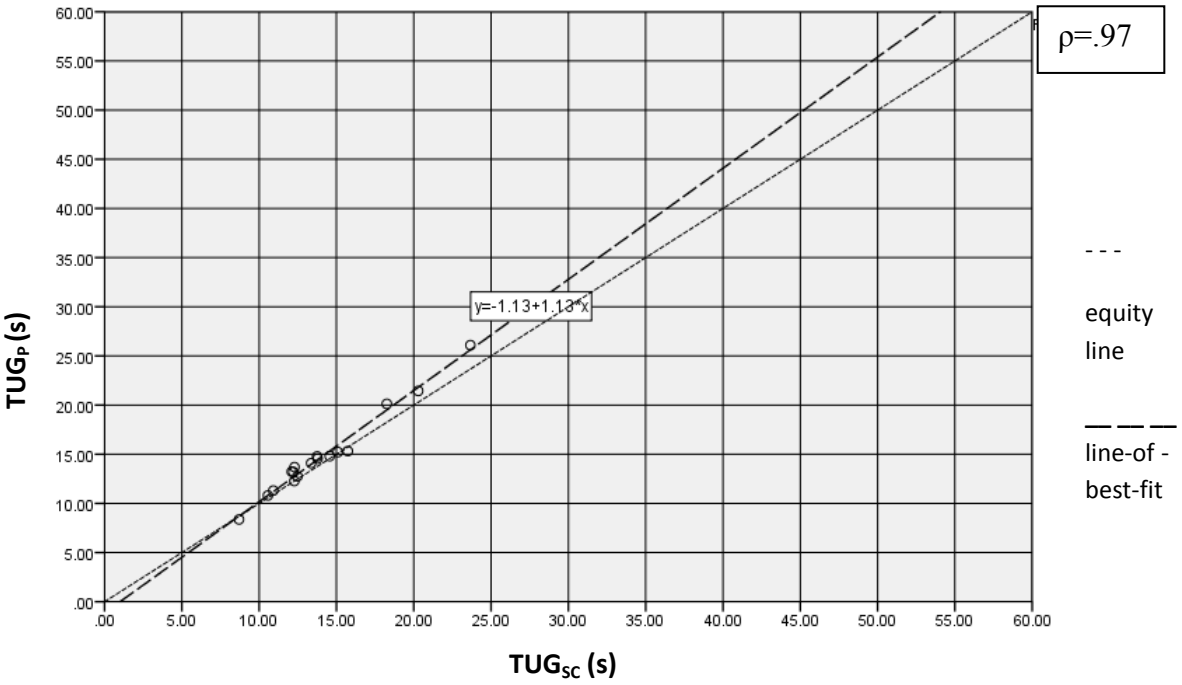


FIGURE 5: SCATTER PLOT WITH TUG_P VERSUS TUG_{SC} WITH SCORES OVER 30 S REMOVED AND EQUATION OF THE STRAIGHT LINE (N=17).

EVALUATING AGREEMENT, METHOD C) BLAND ALTMAN:

With methods recommended by Bland and Altman (1999), the LoA were calculated for TUG_P and TUG_{SC}. [60] Since the data sets for TUG_P and TUG_{SC} were not normally distributed and showed increasing variability as TUG scores increased the data were log transformed. The LoA were then calculated and plotted for the Whole Sample and for the Trimmed Sample. See Table 6 for the results of the LoA. See Figures 6 through 9 for the Bland Altman plots of log transformed and log reversed data. Appendix H contains the plots of raw data and SPSS output tables.

TABLE 6: RESULTS OF LIMITS OF AGREEMENT WITH AND WITHOUT LOG TRANSFORMATION, OUTLIERS WITH AND WITHOUT REVERSE LOG TRANSFORMATION.

Study participants	LoA Log transformed upper	LoA Log transformed lower	LoA Log reversed upper (s)	LoA Log reversed lower (s)
(N=19)	0.2	-0.1	1.2	0.9
(N=17)	0.1	-0.0	1.1	0.9

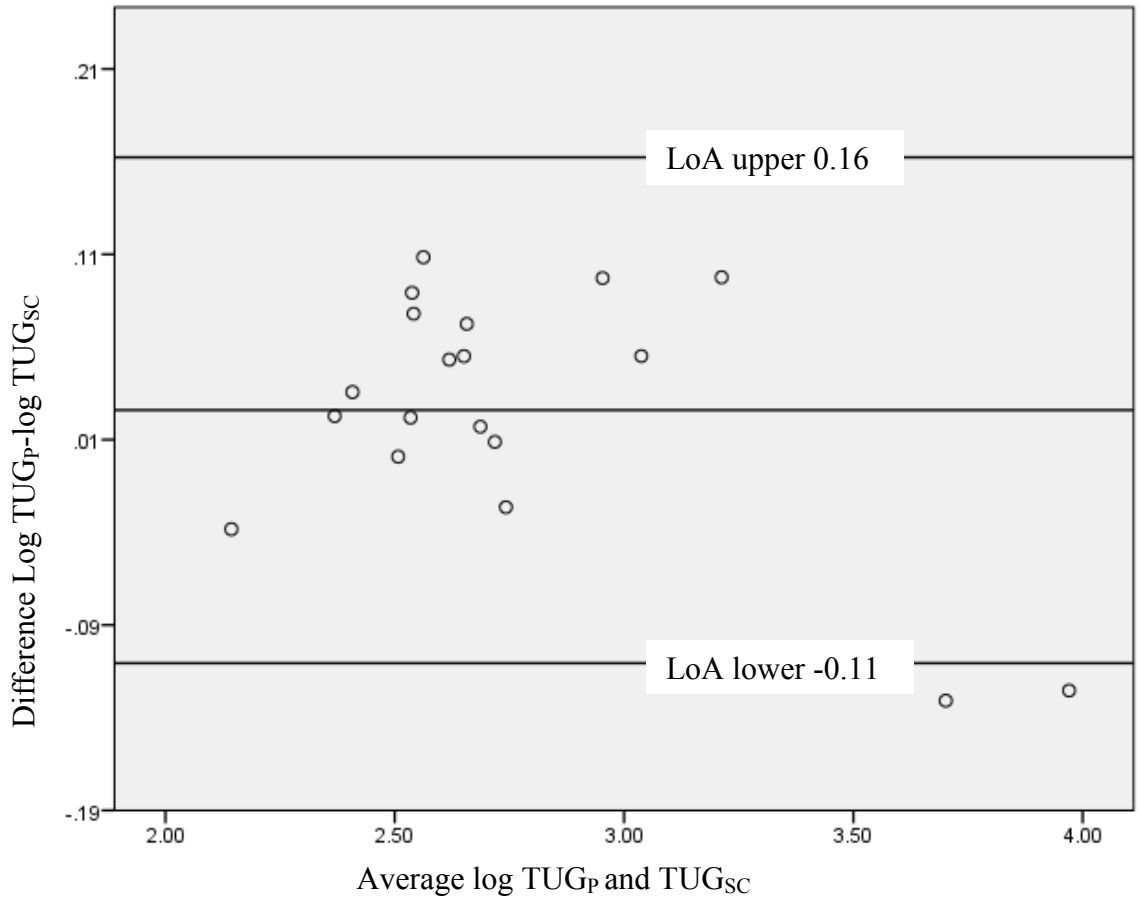


FIGURE 6: SCATTER PLOT SHOWING LIMITS OF AGREEMENT WITH DATA LOG TRANSFORMED (N=19).

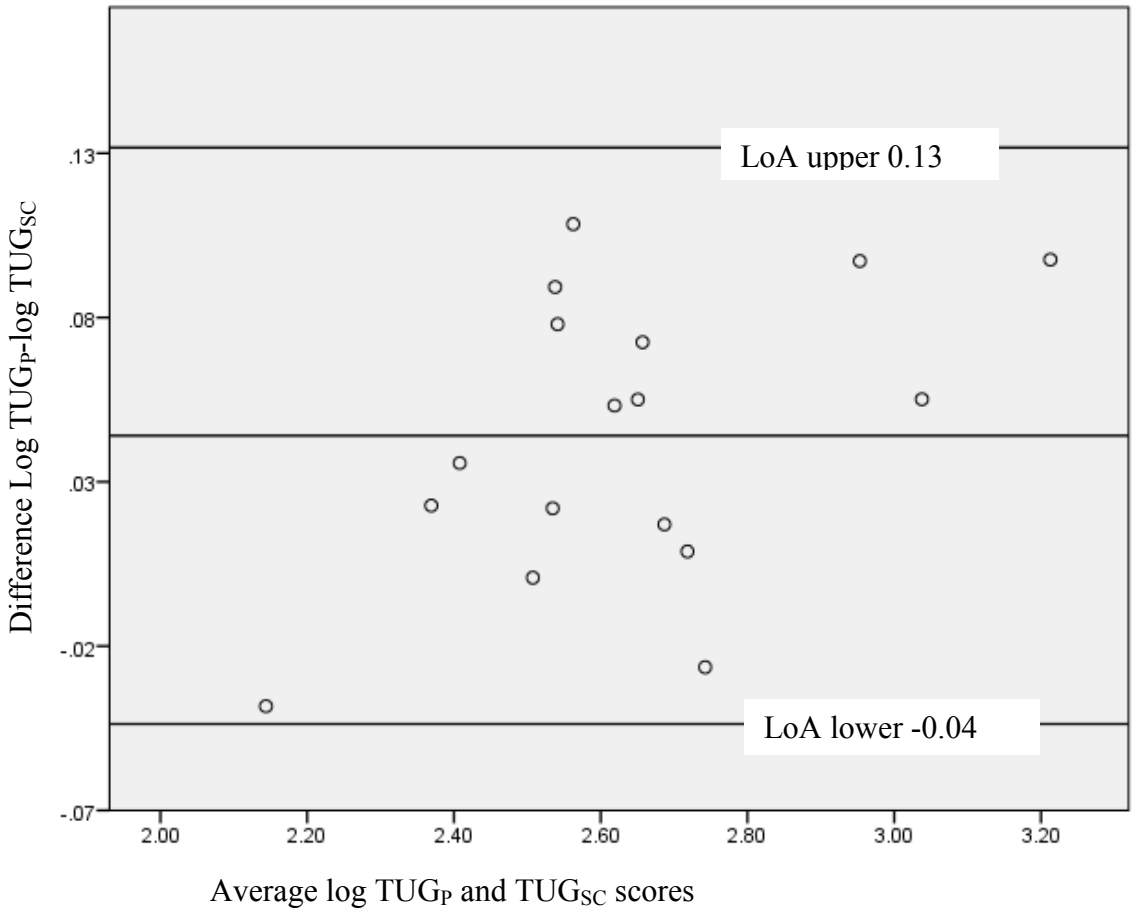


FIGURE 7: SCATTER PLOT SHOWING LIMITS OF AGREEMENT WITH DATA LOG TRANSFORMED (N=17).

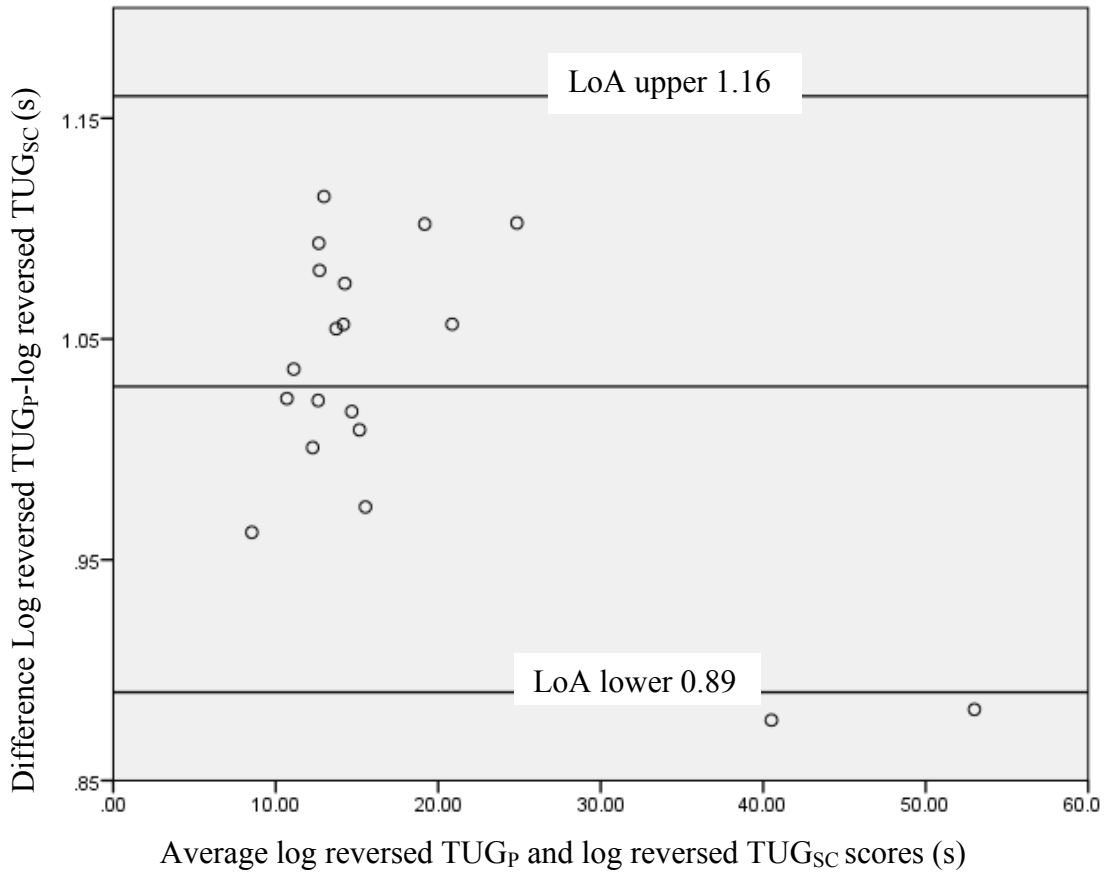


FIGURE 8: SCATTER PLOT SHOWING LIMITS OF AGREEMENT WITH DATA REVERSE LOG TRANSFORMED (N=19).

* This may be interpreted as a biased pattern, as the magnitude of the score increases over 30 s the TUG_{SC} measures the participant higher than the $TUG_{P.}$ In the 16 s to 30 s range, TUG_P measures participants higher, while scores under 16 s appear unbiased.

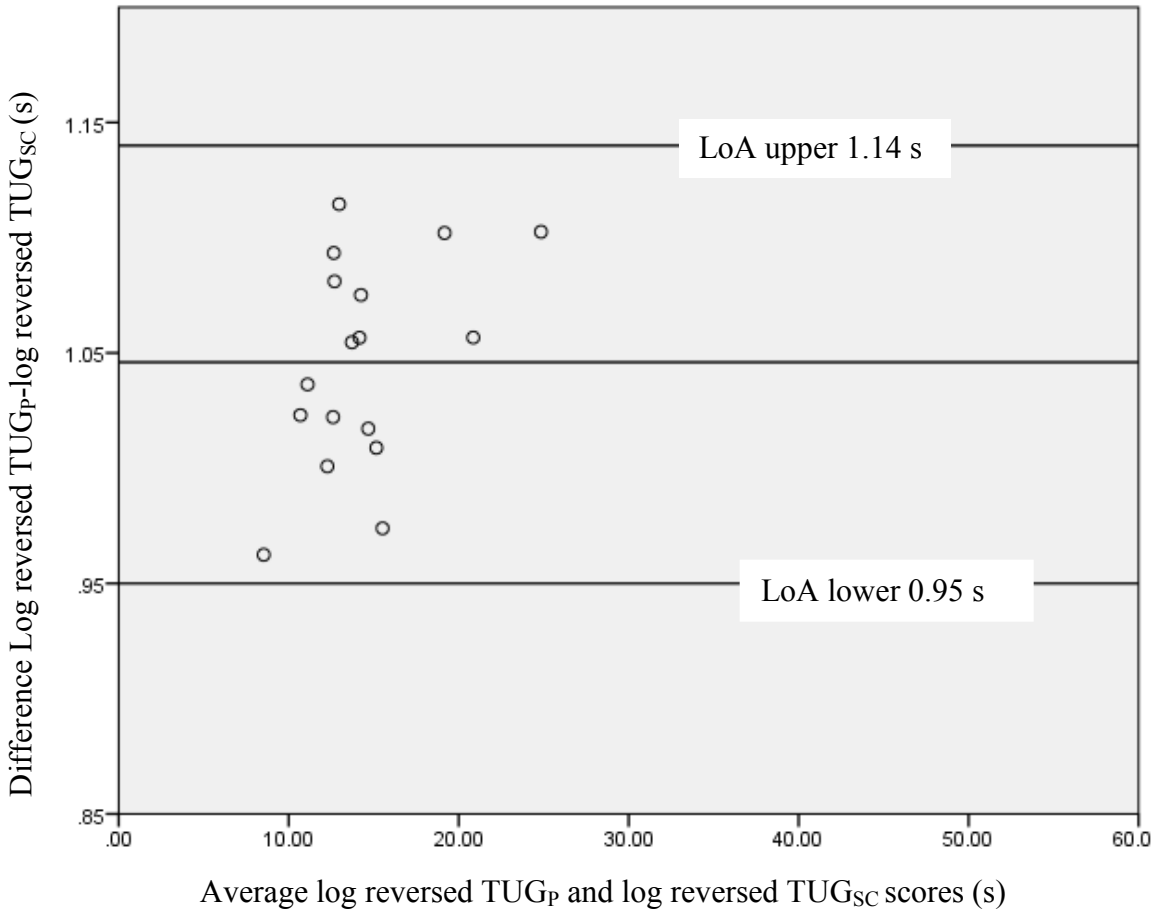


FIGURE 9: SCATTER PLOT SHOWING LIMITS OF AGREEMENT WITH DATA REVERSE LOG TRANSFORMED (N=17).

Chapter 5-Discussion:

This retrospective methodological research study was done to examine the agreement between two commonly cited protocols of the Timed Up and Go (TUG) test in seniors who were referred to home care physiotherapy services for balance and mobility issues. The opportunity to evaluate the agreement between two common TUG protocols was possible through the participation and baseline screening of a select group of seniors, who were recruited as part of the Home Care Exercise Study [1]. As detailed in the following sections, the results of this analysis illustrate the merits of different statistical methods that are used to examine agreement between different measurement protocols, and inform the methods used to evaluate the mobility of the participants in the Home Care Exercise Study.[1] As well, the results have implications for the use and interpretation of the TUG in research and clinical practice related to falls prevention programs for the frail, community-dwelling senior.

CHARACTERISTICS OF THE PARTICIPANTS

Although the study population did not demonstrate cognitive impairment, according to Mini Mental State Exam scores [80], the characteristics of the participants resembled those of seniors who receive home-care physiotherapy services, in several ways. [82-84] Participants' ages ranged from 71 to 97 years, and 73 % were female. [82, 84] Co-morbidities ranged from one to seven with a median of four [82, 83]. Of the 19 subjects, 13 had a history of falls in the previous year. Balance confidence varied across the sample, from virtually no confidence, to a high degree of confidence, as evidenced by the

scores on the Activities Specific Confidence Scale (ABC). [63, 65] The sample included those with and without difficulties using vestibular inputs according to the scores on test 5 of the CTSIB. [71] Half of the sample did not use a gait aid, while the other half used either a two or four wheeled walker or a cane, which is similar to reported mobility in senior populations receiving home care in Canada.[84].

Although few studies have reported TUG scores specifically of seniors who require home-care, the TUG scores of the participants, as measured by either protocol, fall within the range of scores reported for other groups of community-dwelling seniors. [26, 27, 47, 85] Some were able to complete the test as quickly as non fallers and freely mobile seniors cited in literature [24, 26, 27, 37, 47] ,while the scores of several others were more consistent with the scores of those who would be expected to have limited mobility and dependence for ADL's[47]. Two participants within the sample had TUG scores that exceeded 30 s, which gave rise to the possibility that they should be considered as outliers, in comparison to the rest of the sample.

The distribution, range, Standard Error of Measurement (SEM), and Minimal Detectable Difference of the TUG scores were greatly influenced by two participants who had scores over 30 s. The TUG score data of the Whole Sample (N=19) were not normally distributed, being skewed to the slower TUG scores. For the Trimmed Sample (N=17), the TUG_{SC} data was normally distributed, and the distribution of the TUG_P data improved with disappearance of the skew, but data did not normalize to a significant p value. With inclusion of the two participants who had scores over 30 s, the range of TUG scores was more than double that observed with the scores over 30 s removed from the sample. This pattern was also apparent in the calculations of the SEM and MDD, with a substantial

reduction in these estimates of meaningful differences in the TUG scores, when those with scores over 30 s were excluded.

The personal characteristics of the participants, whose TUG scores exceeded 30s, appeared to differ from others in the sample, lending support to the notion that they should be considered outliers relative to the other participants in the study. Both subjects who scored over 30 s had a history of falls, difficulties utilizing vestibular inputs, used a walker as a gait aid and had very low ABC scores of zero percent and 32 percent, indicating no to low balance confidence.[63, 65] They also demonstrated large differences between the score obtained on the first and second trials, indicating the possibility that fatigue influenced their performance as successive trials were completed; this pattern was not observed in the other participants.

Due to the characteristics of these two participants, and the effects of their scores on TUG score ranges, SEM and MDD without appreciable effects on the TUG median scores, the two participants who scored over 30 s were considered to be outliers in comparison to the remainder of the Sample. Therefore, the statistical analyses, including the SEM and MDD calculations, were completed on both the WS and the TS to allow for exploration of the influence of outliers on agreement between the two protocols.

SEM AND MDD AS INDICATIONS OF CLINICALLY MEANINGFUL DIFFERENCES IN TUG SCORES

No studies reporting SEM and MDD values for using the TUG with a home care population were found in the literature. The SEM and MDD from this study are the first reported scores for the home care population. The SEM and MDD of the WS are within ranges found in the literature for other TUG protocols and were obtained from

seniors who were frail, cognitively impaired or living with progressive neurological conditions. [36, 52-54] SEM scores in the literature for the TUG, range from 1.1 to 3.0 s [36, 52-54], while the findings of this study for the WS were 1.4 s and 1.7 s, for TUG_P and TUG_{SC} respectively. The MDD in the literature range from 2.9 s to 4.8 s [36, 52-54], while the MDD values of the WS were 4.0 s and 4.7 s, for TUG_P and TUG_{SC} respectively.

The SEM and MDD values, calculated excluding outliers (TS), are below the range expected from the literature. The differences are likely due to more variable TUG performances with resultant larger standard deviations found in the literature as compared to those found in the TS from this study. [36, 52-54] The TS contained a majority of participants scoring under 20 seconds and included no scores over 30 s with resultant SEM values of 0.6 s and 0.5 s for TUG_P and TUG_{SC} respectively. The MDD values observed for the TS were 1.6 s and 1.4 s, for TUG_P and TUG_{SC} respectively. When performance of the TUG was faster and less variable, the smaller differences were meaningful.

As expected, the diverse characteristics of the sample influenced the SEM and MDD since the calculation of these variables is based on the standard deviation of the sample score. With the inclusion of all participants, the study sample combined those who completed the test quickly with little variability and those whose performance was slower and influenced by the use of multiple trials. The SEM and MDD values of the WS were large, indicating a difference of approximately 4 s as being required to signal meaningful change over time, and disagreement between protocols. By comparison, the SEM and the MDD of the TS, the group of participants who were able to complete the test relatively

quickly and with less variability between trials, indicate that approximately 1 s signals a meaningful difference between scores for seniors in this subset of the population.

Statistical Method and Agreement Between TUG Protocols

Each statistical method contributed to the exploration of agreement between the TUG protocols in different ways. Both the $ICC_{(3,1)}$ and Spearman rho correlation coefficient, captured correspondence between the two protocols. The SEM and MDD provided a clinical benchmark in exploring the extent of agreement. The statistical methods that provided clinical meaningful ways to explore the extent of agreement in this study included the line-of-best-fit, the plots of the difference for the LoA and the LoA. Below is a discussion, by statistical method, of the results regarding agreement between the two protocols, including the influence of sample heterogeneity on the results.

Evaluating Agreement with ICC

The ICC does not represent the reliability for the individual protocols as it is based on variance across both protocols. [3] The ICC compares two or more measures at one time, in order to produce an average correlation among the possible pairs of measurements to test reliability between the pairs of scores. [5] The ICC is often a preferred reliability coefficient over other tests, as it may be considered to reflect both the correspondence and agreement among ratings through its use of the ANOVA. [55] While the use of the ANOVA within the ICC may indicate the means of the two protocols are not significantly different, the problem with determining agreement with this method is the distribution

may still be composed of pairs of data with no agreement. [3] In order to determine if there is an interaction between a subject and the protocol such as non-agreement of a subpopulation, the inspection of scatterplots is needed. The ICC is a reliability coefficient represented by a single number, without units, between 0-1.0; by convention, the coefficient is judged to represent a strong association between measures, when the magnitude of the ICC is 0.90 or greater. [3]

The $ICC_{(3,1)}$ results in this study indicate a strong association between the two commonly cited protocols of the TUG [3], for both the Whole Sample (N-19), and the Trimmed Sample (N-17). A strong association, above the .90 expected for clinical studies was observed between the two protocols within both the WS and TS.[3] The correlation coefficient and 95 percent Confidence Intervals ($CI_{95\%}$) differed only slightly when the outliers were excluded from the analysis. Although the $ICC_{(3,1)}$ analysis does not readily indicate the extent of the agreement in clinically meaningful units, it does convey the strong association between the two protocols.

EVALUATING AGREEMENT WITH CORRELATION ANALYSIS

To examine the extent of agreement using correlation, both the correlation coefficient and the equation of the line-of-best-fit are considered together to avoid errors in assumption of agreement between the protocols based on an association alone. [5] Correlation is a statistical technique used to represent measurement reliability of repeated measures and represents the relationship between two variables. [5] While the correlation coefficient indicates the strength of the relationship between the TUG protocols, the use of the scatter plot, and the line-of-best-fit in relation to the line of equity, provide information on the extent of the relationship between the protocols. The equation of the line-of-best-fit is represented by: $y = a + bx$. The two TUG protocols would be identical when the slope (b) equals 1, the intercept (a) equals 0 and the correlation coefficient (r) approaches +1. [5]. While correlation coefficients are unit-less and do not reflect a meaningful result in relation to the measurement unit of the tool, the line-of-best-fit does provide a meaningful way to examine the extent of agreement when used in conjunction with the line of equity, and a clinical benchmark such as the SEM.

The values of the Spearman rho, the correlation coefficient used in this study, calculated for the Whole Sample, indicate a strong association between the two protocols. The rho of the Trimmed Sample, while not identical to that of the Whole Sample, also indicates a strong association between the protocols. As is the case when using an ICC to explore agreement, the clinical relevance of the strong association found is limited, as its relevance in terms of unit of measurement of the tool is unknown.

However, consideration of the line-of-best-fit, in conjunction with the SEM findings of

this study, revealed that the pattern of the relationship between the scores of the two protocols seemed to be influenced by the inclusion of the outliers in the data set. As anticipated, the line-of-best-fit for the TUG data in this study revealed that the scores of one protocol are not in absolute agreement with the scores of the other [5]. When the SEM of the WS was considered, the line-of-best-fit appeared to deviate from the equality line, for scores greater than 27 s; in particular, the TUG_{SC} exceeded the TUG_P score beyond expected error of the measurement for scores greater than 27.5 s. When the outliers were removed from the data set, and the SEM was considered, the line-of-best-fit appeared to deviate from the line of equality for TUG scores greater than 16s, but in this case, the TUG_P scores exceeded those of the TUG_{SC} protocol.

In summary, this study illustrates the necessity of considering both the correlation coefficient and the line-of-best-fit equation, to examine questions of agreement to avoid errors in assumptions by using the correlation coefficient alone. The combination of the correlation coefficient and the line-of-best fit, indicate the scores of the two TUG protocols were strongly related. However, these preliminary findings using the SEM as the clinical benchmark for interpreting the line-of-best-fit relative to the equity-line, revealed differences in protocol scores for those scoring over 16 s, or over 30s, depending on the exclusion or inclusion of the outliers, respectively.

EVALUATING AGREEMENT WITH THE BLAND-ALTMAN METHOD

Bland and Altman (1986) recommend using clinical interpretation and calculated Limits of Agreement (LoA) to determine whether there is sufficient agreement between two tools to use them interchangeably, as it is unlikely that different methods or tools will

give exactly the same results for all individuals.[2] To examine agreement with the Bland-Altman method, plots of the differences between the methods against their mean are used, and the LoA are calculated in the same unit of measurement of the tool. [2] Examination of the plots of the mean differences is required in conjunction with the LoA to ensure no bias is present. If a strongly biased pattern exists in the plot of the LoA, the tools are not considered interchangeable. [60]

A limitation of using the LoA is the necessity of a criterion, or benchmark, to examine the appropriateness of the clinical interpretation as to whether the findings of the LoA are clinically relevant. The MDD's from this study were used as the clinical criterion to examine LoA results and examine extent of agreement between the TUG protocols. The MDD was chosen since both the LoA and MDD represent a measure spanning two standard deviations from the mean.

The calculated LoA appeared robust with a heterogeneous sample of community-dwelling seniors referred for home care physiotherapy services. The LoA of 0.9 to 1.2 s observed for the WS, is not appreciably different from the LoA of 0.9 to 1.1 s observed for the TS, given that TUG is measured with a manual stopwatch, and voluntary reaction time is expected to be in the range of .200 to .350 s. [86-88] Minimal differences with outlier inclusion are in keeping with Bland and Altman (1999), who do not recommend excluding outliers, when using the LoA method to examine agreement. [60] While the calculated LoA between the two protocols indicates that scores from the two protocols are expected to differ by approximately 1 s, the LoA are small in comparison to the observed MDD indicating possible interchangeability of the protocols. However, the inspection of the LoA plots for bias is also a necessary part of the determination of agreement.

Inspection of the LoA scatter plots examining the difference scores between the protocols reveals equal distribution of difference scores of each protocol above and below the mean difference for those who score less than 16 s. The equal distribution of the difference scores of the protocols above and below the mean indicates random error of the protocols with no apparent biased pattern; this provides evidence that the two protocols agree within approximately 1 s, and could be used interchangeably. There are only a few participants whose scores were over 16 s, but it is of interest to note that from 16 s through to 30 s on the x-axis, 3 participants' difference scores fall within the LOA, but above the mean difference between the two protocols. On the contrary, the difference scores for the two participants with average TUG scores greater than 30 s were below the mean difference between the two protocols, and outside the limits of agreement. These patterns indicate a potential bias within this data set: toward TUG_P values for those scoring from 16 s to 30 s, and toward TUG_{SC} over 30 s. However, a greater sample size is needed to see if the bias is present in the whole population, or a chance finding in the sample recruited for this study.

SUMMARY EVALUATING AGREEMENT BETWEEN TUG_P AND TUG_{SC}

The results from the different statistical methods demonstrate different aspects of agreement. Two methods, the $ICC_{(3,1)}$ and the Spearman Rank Correlation, show strong associations between the TUG protocols, which remained strong when the two outliers were included. Two statistical methods were used to examine the extent of agreement between the two common TUG protocols: the line-of-best-fit and the LoA. Both revealed similar conclusions indicating the likelihood of differences in scores that is protocol-based is due to the speed of TUG performance. For this sample of seniors referred for

home care physiotherapy services due to mobility and balance impairments, it would appear, for those scoring under 16 s, either protocol could be used to score the TUG. Further studies are needed to determine if TUG performance over 16 s represents a true difference in score based on protocol, or is a result of the sample population in this study.

TUG Protocol Influence on Home Care Exercise Study [1] Participants

When considering the results and secondary outcomes of the Home Care Exercise Study [1] 4 of the 7 participants who completed the exercise portion of the study, scored over 16 s on the first experimental trial for both pre- and post-exercise TUG tests. As the Home Care Exercise Study [1] contained participants with scores over 16 s three experimental trials should be analyzed to examine the differences due to protocol use and to examine for performance changes from trial to trial of the TUG post exercise and balance training. Until further studies are completed with greater sample sizes and seniors with variability in TUG scores to confirm and clarify protocol differences in those scoring over 16 s, the collection and analysis of the three experimental trials is warranted for comparison of results with those of Shumway-Cook regarding fall risk. For interpretation of the results with respect to mobility concerns, the Podsiadlo protocol is appropriate.

Recommendations for Future Research

Further studies that determine normative data and potential presence of subpopulations in a community-dwelling senior population referred for home care physiotherapy services are needed. The senior home care population contains individuals with a range of home care needs, capabilities and mobility levels making it a diverse population of health care

users.[82] With the diverse nature of senior home care users, many capabilities may occur within the same sample, as noted in typical home care populations and in the sample for this study. [82, 84]This study contained only two seniors that scored over 30 s, who appeared to be outliers in the sample used for this study. As the home care population is diverse, examining agreement between the two TUG protocols, using a larger sample of seniors with longer and more variable scores is needed to confirm if the speed of the TUG score is affected by protocol use as noted in this study.

Further studies with a larger, more diverse sample of seniors referred to home-care physiotherapy for balance and mobility concerns, are also needed to clarify expected normative and meaningful differences in TUG scores for this population. There was little research in the literature reporting SEM, and MDD values for community-dwelling seniors, with no studies found involving those receiving home care physiotherapy services. The SEM and MDD values obtained in this study for community dwelling seniors with mobility and balance concerns referred for home care physiotherapy services were lower than those available in the current literature. [36, 49, 52-54] Comparisons of findings in the literature to the findings in this study support the theory that sub-populations of seniors differ with respect to variability in performance of the TUG.[30] As this is the first time results for a home care population have been reported, further studies with greater sample size including those with variability in TUG scores should be conducted to establish appropriate values of SEM and MDD of the TUG, to inform clinical, and research decisions regarding meaningful change in TUG score performance.

This retrospective study provided preliminary evidence regarding the agreement between two different TUG protocols, as tested by a single rater. Further prospective studies with

larger, more diverse sample of participants, and using multiple raters will also be important to obtain results that can be generalized more broadly for improvement in the use of the TUG in clinical practice and research. Furthermore, due to the number of TUG protocols in use in studies, more research regarding the agreement of the numerous TUG protocols is needed. Research into the agreement of the many variations of the TUG and numbers of trials used for its scoring in the literature is not available. The influence of the difference in other protocols beyond the two common protocols used in this study affecting the TUG score has not been noted in the literature and therefore requires further investigation.

CLINICAL IMPLICATIONS

Consideration of multiple statistical methods allows the clinician to gain a greater understanding of the complexities of relationship and extent of agreement between TUG protocols that may not have been apparent with consideration of one method alone. The ICC and correlation coefficient contribute to the determination of association and correspondence between the protocol scores with the downfall that the correlation co-efficient is not quantified in the unit of measurement, limiting ability to comment on clinically meaningful differences. The line-of-best fit, and the LoA with difference plots of the Bland Altman method, coupled with SEM and MDD scores allows the clinician to evaluate the extent of agreement between scores of the protocols in the relevant unit of measurement and allows for better clinical interpretation with respect to the performance of the individual by protocol. By examining multiple statistical analyses, clinical decisions around how and when to use a certain TUG protocol can be made with greater confidence.

This study provides preliminary evidence that for those with fast TUG scores, corresponding to good mobility skills, these two TUG protocols appear to be interchangeable. The results also raise doubt on the interchangeability of the protocols for those with slow TUG scores and/or large trial-trial variability in performance of the TUG. Agreement between the two protocols appears influenced by the speed at which the TUG is performed and as TUG times increase, the variability of the score increases, which is noted both the literature and in this study. [26, 49] Despite the observed protocol differences based on speed of performance, when using the TUG to screen balance and mobility as part of a falls prevention program, there is little evidence to suggest benefit of completing the three trials to determine mobility. Based on the findings of this study, considered in conjunction with characteristics of the sample, use of the Podsiadlo protocol with the mobility classifications of scores under 10 s, 10 to 20 s, while 20-30 s and over 30 s for, appears sufficient for screening mobility in falls prevention programs for seniors referred for home care physiotherapy services

Consideration of the differences in the two protocols in terms of the effect of repeated trials on the test performance, could give the clinician some insight into the effects of either fatigue or adaption, particularly when assessing the balance or mobility of seniors who take longer than 20 s to complete the task. In this study, participants who completed the test between 16 s to 30 s were often slowest on the first experimental trial and subsequent trials were quicker, suggesting the influence of learning or warm-up effects. For those needing more than 30 s to perform the TUG, the first experimental trial produced the fastest TUG score, with slower subsequent trials, which is an indication that factors such as fatigue influenced the TUG performance. Therefore, collecting three

experimental trials could provide useful insight into underlying impairments that limit more sustained mobility tasks, while performance on the first trial may signal issues related to initial physiological adaptation, and execution of basic mobility tasks. The finding of difficulties sustaining mobility with repeated trials may prompt the clinician to perform appropriate tests to explore endurance, for example, the 6-minute walk test. [89]

The results of this study also have implications for clinical interpretation of the literature regarding the use of the TUG to assess seniors who are referred for balance and mobility problems. When appraising studies for community dwelling seniors who are independent ambulators scoring under 16 s on the TUG studies with either TUG_P or TUG_{SC}, it may be appropriate to consider the protocols to be equivalent, and to consider studies collectively. For studies containing different TUG protocols, different populations, and large ranges of scores or variable performance of the TUG from trial to trial, caution should be used in grouping TUG outcomes together for comparison and interpretation.

LIMITATIONS

Although the demographics of home care populations demonstrate variability in mobility and daily functional activities, this study only included a small sample of seniors receiving home care physiotherapy, and there were only 5 participants who scored over 16 s on the TUG. With the inclusion of more home care senior participants who score over 16 seconds on the TUG, examination of protocol differences due to the magnitude of the TUG performance can be more completely explored.

The small sample size in this study limits the ability to generalize results to the larger home care population. Generalizability of the results is also limited as an ICC model (3,1)

was used. Larger sample sizes can assist with generalizability in addition to exploration of SEM, MDD and LOA for this population of senior home care recipients.

The retrospective nature of this study resulted in the inclusion and exclusion criteria for the Home Care Exercise Study [1]being applied to this study. The exclusion criteria were necessary for exercising participants safely and to reduce confounding factors, which were not a necessity to examine agreement between TUG protocols

CONCLUSION

According to best practice guidelines, falls prevention programs for seniors need to include assessments of gait and mobility. [17] The TUG is a tool to assess gait and mobility that is commonly used, but with many different protocols. The purpose of this study was to examine the agreement between two common TUG protocols (i.e., TUG_P and TUG_{SC}) used to test a small sample of community-dwelling seniors in the context of falls prevention who were referred to home care physiotherapy services for mobility and balance concerns.

This study adds new evidence regarding the use of the TUG to assess seniors with mobility and balance impairments, who are referred for home-care physiotherapy services. The values of the SEM and MDD for the study population of seniors with balance and mobility impairments referred for home care physiotherapy services have not been previously reported in literature. Since differences were noted between the SEM and MDD in this study population versus values currently available in the literature, further research is warranted to identify appropriate SEM and MDD values for different sub-groups of seniors.

Further research, involving greater numbers of seniors with diverse mobility skills, and multiple raters, is required to verify the generalizability of these results, and to guide best practices regarding the use of the TUG test for the assessment of balance and mobility in fall prevention programs for seniors who are referred for home care physiotherapy services. Within this small sample, a strong association between the scores of the two protocols was observed. Based on the line-of-best-fit analysis, evidence that the two TUG protocols were interchangeable was strongest for those performing the test in less than 16 s. According to the Bland-Altman analysis, in general, the scores obtained with the Podsiadlo protocol exceeded those of the Shumway-Cook protocol by approximately one second, which is not a meaningful difference in relation to the minimal detectable difference for the TUG. Agreement between the protocols was apparent for those performing the TUG quickly, but this did not hold true for the whole range of TUG scores of all participants.

Appendix A: Summary of Papers Regarding the TUG

Table A1: Summary of Studies Utilizing the "Podsiadlo Protocol" to Administer the TUG

Table A2: Summary of Studies Utilizing the "Shumway- Cook Protocol" to Administer the TUG

Table A3: Summary of Studies Utilizing the "Unique" Protocols to Administer the TUG

Table A4: Summary of Studies Using Normative Data to Describe TUG Results

TABLE A1: SUMMARY OF STUDIES UTILIZING THE "PODSIADLO PROTOCOL" TO ADMINISTER THE TIMED UP AND GO.

See definitions for abbreviations, at the end of the table.

Author	Podsiadlo and Richardson, 1991 **[47]**		Bischoff 2003 **[26]**		Kristensen 2009 **[46]**	Arnold 2007 **[27]**	Large 2006 **[48]**	Kristensen 2007 **[28]**	Nordin 2008 **[30]**	Nordin 2006 **[49]**	Kwan et al 2001 **[50]**	Whitney 2005 **[24]**
Purpose	Classify mobility, Reliability Validity		Classify Mobility		Classify Mobility	Classify FR	Predict FR[18]	Predict FR	Predict FR[18]	Reliability MDC[11]	Validity	Screening tool
Popula-tion	CD Frail day hospital	HC	Mixed : CD	LTC	Inpatient Patients post hip fracture surgery	CD Hip OA	Acute care in-patients	Inpatient Post hip fracture with surgery	RF frail in LTC	RF Cognitive impairment	CD MMSE above 19	CD Fall clinic
Sample	F= 37 M=23 N=60	F=4 M=6 N=10	n=413	n=78	F=90, M=36 N=126	F=77, M=29 N=106	F=1476 M =912 N= 2388	F=45, M=14 N=59	F= 134 M=49 N=183	F=49 M=29 N=78	F=120 M =160 N=280	F=83 M=27 N=110
Age (yrs) Mean Std Dev range	79.5 nr 60-90	75.0 nr 70-84	73.0 ±3.2 65-85	79.7 ±3.7 65-85	74.8 ±12.7 nr	74.4 ± 6.2 65-88	82.1 ± 7.7 nr	81.0 nr 42-97	84.3 ±6.6 66-98	84.8 ±5.7 66-97	74.9 ± 6.4 65-91	79.3 ±7.2 63-95
Design						R	Pro	Pro	Pro			
TUG Protocol	1 P 1 T		1 P 1 T		1 P 1T 2 types gait aids verbal cues	1 P 1T	1 P 1 T	1 P 1 T	1 P 1 T20	1 P 1 T verbal cues	1 P 1 T	1 P 1 T

Author	Podsiadlo and Richardson, 1991 **[47]**	Bischoff 2003 **[26]**	Kristensen 2009 **[46]**	Arnold 2007 **[27]**	Large 2006 **[48]**	Kristensen 2007 **[28]**	Nordin 2008 **[30]**	Nordin 2006 **[49]**	Kwan et al 2001 **[50]**	Whitney 2005 **[24]**
Reliabil-ity	Intra-rater ICC = .99 n=22 Inter-rater ICC = 99	Intra-rater ρ = .96 n=23 Inter-rater ρ =.91				Intra-rater ρ = .91 n=nr Inter-rater ρ =.91		Intra-rater $ICC_{(3,1)}$= .92 n=78 Inter-rater $ICC_{(1,1)}$= .91		
Validity	BERG r =.81 Barthels r =.78 gait speed r =.61							MDC [95%CI] 10s [7- 13] 40s [26- 61]		
TUG score (s)	HC:8.5 range 7-10	CD: 8.3± 1.9 LTC: 28.2 ± 23.0	Group one Walker: 48.5±20.9 Rollator: 38±16.3 Group two Crutches: 24.3±13.4 Rollator: 20.8±9.2	All 12.8 ±5.3	All Median = 28 range 19-41	NF Median =29.5 Fallers median= 42	All Median = 25.5 Range= 17.6-35.9	All Day 1-30±17.4 Day 2 29.9±17.7 Day 3 28.0±15.1		All Median =19.6 H FR Median= 25.1 LFR Median= 17.9

Author	Podsiadlo and Richardson, 1991 [47]	Bischoff 2003 [26]	Kristensen 2009 [46]	Arnold 2007 [27]	Large 2006 [48]	Kristensen 2007 [28]	Nordin 2008 [30]	Nordin 2006 [49]	Kwan et al 2001 [50]	Whitney 2005 [24]
Criterion Score (s)	FM: 10 IM: <20 D: > 30	12		10 F	10N F	24	15			High Risk 15
Sensi-tivity (%)				73	81	95	96			81
Speci-ficity (%)				35	36	35	32			39
Falls definition				Uninten-tional fall to ground or lower level			Uninten-tional coming to rest on floor or ground	Unexpected falls from standing position to the ground or floor		

F=Female; M=Male; CD= community-dwelling; OA- osteoarthritis; NF=non faller; F = Faller; FR=falls risk, HRF=High Risk Fall, LRF=Low Risk Fall
Indep. = independent; LTC =long-term care; RF residential facility/care; HC healthy comparisons; cog. =cognitive;
FM = Freely mobile; IM = Independent. Mobility; D = Dependent
nr = not reported; MDC= Minimal Detectable Change; C.I. = 95 % confidence interval; r= Pearson correlation; ρ = Spearman correlations m=metres
SP= Study participants; P=practice Trial; T=Timed Trial
Pro =Prospective; R=Retrospective

TABLE A2: SUMMARY OF STUDIES UTILIZING THE "SHUMWAY- COOK PROTOCOL" TO ADMINISTER THE TIMED UP AND GO.

See definitions of abbreviations at end of table.

Author	Shumway-Cook 2000 **[37]**	Wrisley, D.M. and N.A. Kumar 2010 **[51]**
Purpose	Classify Falls risk Validity	Predict Falls risk
Population	CD with and without Hx of 2 or more falls	CD with no Hx of falls
Sample	F=18 M=12 N= 30 No falls 2+ falls n=15 n=15	F=18 M=17 N=35
Age(years) **range**	78±6 86.2±6 65-85 76-95	 60-90
Design	Retrospective	Prospective
TUG Protocol	1 P 3 T average of 3 T 3 conditions : TUG ,TUG manual, TUG cognitive	1 P 3 T, average of 3 T
Reliability	Inter-rater ICC $_{(3,3)}$ =.98	
TUG score (s)	NF: 8.4±1.7 Fallers: 22.2± 9.3	NF: 9.8±1.6 Fallers: 15.8±8.2
Criterion Score (s)	13.5	11 Optimum12.3
Sensitivity(%)	87	83 83
Specificity(%)	87	86 96
Falls definition	Unintentional fall to ground or lower level	Unintentional fall to lower level below person height. Explained or unexplained

F=Female; M=Male; CD= community dwelling; Hx=history; NF=non faller; P= practice trial; T= timed trial

TABLE A3: SUMMARY OF STUDIES UTILIZING THE "UNIQUE PROTOCOLS" TO ADMINISTER THE TIMED UP AND GO.

See definitions of abbreviations at end of table.

Author	Gunter 2000 [31]	Dite 2002 [33]			Lin 2004 [39]	Ries 2009 [36]	Rockwood 2000 [32]	Steffen 2008 [38]	Boulga-rides 2003 [35]	Morris 2007 [29]	Vicarro 2011[40]
Purpose	Classify FR	Classify FR			Classify FR Reliability, validity	Reliability, Validity and MDC AD	Reliability, construct validity	Reliability/ Validity; MDC; PD	Predict FR	Predict FR	Predict FR
Population	CD; Indep. ;AL	CD; MF; 1XF; HC			CD	AD;LTC; AL;DP	CD;LTC cog. impaired	CD;PD	CD; Indep	CD; F with vertebral fracture	CD Presented to care clinics
Sample	F=131, M=26 n=157	F=42 M=39 N=81 MF n=27	1XF n=27	HC n=27	F= nr M= nr N=1200	F=34, M=17 N=51	F=1431, M=874 N= 2305	F=11, M=26 N=37	F=60, M=39 N=99	F=86 N= 86	F=199 M=258 N=457
Age (years)											
Mean	77.4	74	74	74.1	73.4 nr	80.71	78.1 nr	71	74	78	74
SD	± 5.4	±5.7	±6.0	±6.0	nr	± 8.77	69- 104	±12	± 5.64	±7	± nr
Range	nr		nr			nr		nr	65-90	63-91	nr
Design	R	R			Pro				Pro	Pro	Pro

Author	Gunter 2000 [31]	Dite 2002 [33]	Lin 2004 [39]	Ries 2009 [36]	Rockwood 2000 [32]	Steffen 2008 [38]	Boulga-rides 2003 [35]	Morris 2007 [29]	Vicarro 2011[40]
TUG Protocol	1 P 2 T, fastest of 2 T	0 P Trial repeated if unsuccessful	?1 T	2 T mean of 2 T Cuing allowed	1 P 2T best of 2T Armless chair	1 P 2 T mean of 2 T	1 P 2 T mean of 2 T	1 P 1 T 5 m TUG	2 T, mean of 2 T
Reliability	nr ? for all gait tests .68 to .93 n=26		ICC Intra-rater and inter-rater = .93-.99 n=60	ICC $_{(2,2)}$ test - retest= .985-.988 n=51	ICC Test - retest =.56 n=1115	ICC $_{(3,2)}$ Test- retest =.85 n=36			
Validity		step test ρ=.79 FSST ρ= .88 FR ρ=.47	Tinneti r= .53 Gait Speed r= .66 ADL r= .45	MDC =4.1 s		MDC=11 s			
TUG score (s)	NF : 7.54±1.2 1XF:8.91±1.34 FF 9.21±1.3	NF 12 MF 16.68 HC 10.0	13.3	Mild to mod. 19.9±9.8 Mod. To severe 28.0±17.5	all 14 cog. intact 12 cog. Impaired 15	All 15 ±10	NF:8-10 MF:9-13	Median : 18.7 range 13.7-29.8	12.3 ± 5.5
Criterion Score (s)		MF[6] 13	NR[9]					30	

Author	Gunter 2000 [31]	Dite 2002 [33]	Lin 2004 [39]	Ries 2009 [36]	Rockwood 2000 [32]	Steffen 2008 [38]	Boulga-rides 2003 [35]	Morris 2007 [29]	Vicarro 2011[40]
Specificity (%)		89							
Sensitivity (%)		93						13	
Falls definition	Unintentional fall to ground or lower level	Unintentional fall to ground or lower level					hip, knee, hand coming to rest on other surface.	Unexpec-ted event	

F=Female; M=Male ;CD= community dwelling: Indep.= independent : NF=non faller ; MF=multiple fallers; IXF=one time fallers; HC healthy comparison ; nr not reported ;AL= Assisted Living ;FF=frequent faller ;AD=Alzheimer's Disease; MDC= Minimal Detectable Change; Mod. Moderate; LTC =long term care; cog.=cognitive; SEM=standard error of measurement; PD= Parkinson's Disease; C.I. = 95 % confidence interval ; yrs=years ;FR =Fall risk ; m=metres; P=practice trial; t = Timed trial ; R= Retrospective ;Pro =Prospective ; DP=Day program

TABLE A4: SUMMARY OF STUDIES REPORTING NORMATIVE DATA FOR TUG.

See definitions of abbreviations at end of table.

Author	**Bohannon 2006 [90]**	Steffen 2002 **[34]**	Pondal 2008 **[91]**
Purpose	Normative data for TUG by age, functional mobility	Normative data for common clinical tests	Normative data for TUG by age functional mobility
Population		CD	CD
Sample	Meta-analysis of 21 studies consisting of normal TUG values	F=59, M=37 n=96	F= 137, M= 171 n=308
Age range(years)	NR	61-89	71-99
TUG Protocol	Variable according to study	1 P 2T , mean of 2T	1 P 1 T Armless chair
Reliability		test retest $ICC_{(2.1)}$= .97 n=96	
TUG score (s)	Overall mean 9.4 [8.9-9.9] 60- 69 yrs: 8.1 [7.1-9.0] 70 to 79 yrs: 9.2 [8.2-10.2] 80 to 99 yrs: 11.3 [10.0-12.7]	60-69 yrs M=8±2 C.I· 7-8 (n=37) F=8±2 C.I. 7-9 70-79 yrs M=9±3, C.I.· 7-11 (n=36) F=9±2, C.I. 8-10 80-89 yrs M=10±1, C.I. 9-11 (n=23) F=11±3 C.I. 9-12	Grand mean: 10.2 ±3 (5 - 25) 71-75 yrs[1]: 9.5 ±2.5 76-80 yrs[1]: 9.9 ±3 81-85 yrs: 11.2±3.6 86-99 yrs: 12 ± 3.8
Criterion Score (s)	9.4		10.2

yrs=years; F=Female; M=Male; CD= community dwelling; NR not reported; C.I. = 95 % confidence interval; P=practice trial; T= timed trial

Appendix B: Home Care Exercise Study [1] Documents

Figure B-1: Flowchart of Recruitment and Participation in the Home Care Exercise Study.

Table B1: Inclusion and Exclusion Criteria for the Home Care Exercise Study [1]

Letter of Explanation to Physicians for Patient Participation in Home Care Exercise Study [1]

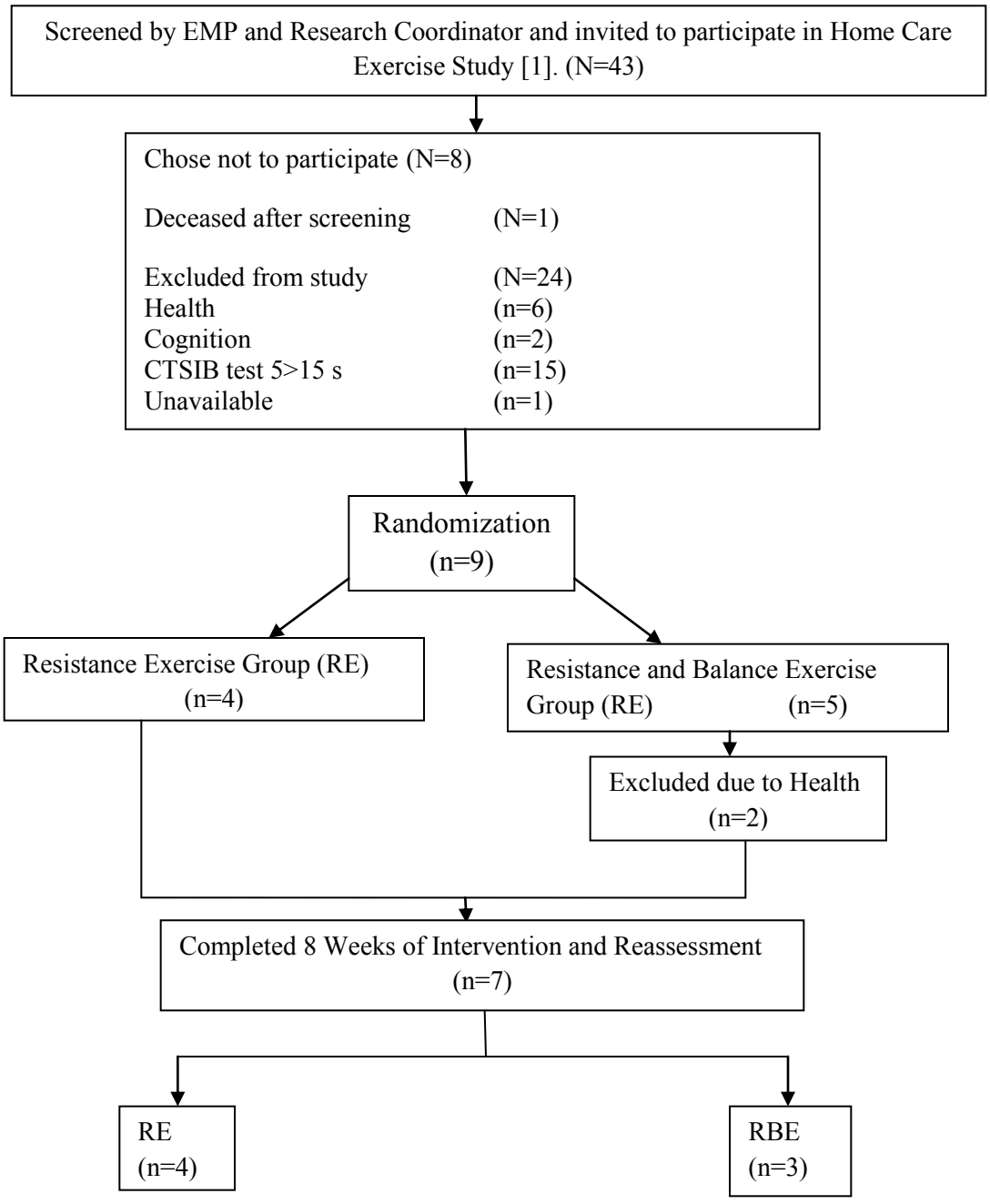


FIGURE B-1: FLOWCHART OF RECRUITMENT AND PARTICIPATION IN THE HOME CARE EXERCISE STUDY. [92]

TABLE B1: INCLUSION AND EXCLUSION CRITERIA FOR THE HOME CARE EXERCISE STUDY
[1, 93]

Inclusion Criteria
65 years or older
Ambulating at least in own home with/without aid for 6 m; able to stand independently without aid
Living in community or independent retirement home
Able to provide informed consent
Exclusion Criteria
Unstable medical conditions including poor control of chronic medical conditions
MMSE < 23; moderate to severe dementia
Weight bearing restrictions
Pain on weight bearing
Receiving physiotherapy at time of study
Acute osteoarthritis in lower extremities that limits maximal muscle contractions
Legal blindness
Diagnosis of progressive neurological condition that would influence balance or muscle strength such as Parkinson's, ALS, MS or stroke within past year
Abnormal VOR or evidence of nystagmus
CTSIB test 5 > 15s

LETTER OF EXPLANATION TO PHYSICIANS FOR PATIENT PARTICIPATION IN HOME CARE EXERCISE STUDY [1]:

Summary of the Research Project for Extramural Physiotherapists and Physicians
An intervention study to determine if balance impaired older adults can improve ability to utilize sensory inputs with progressive sensory integration training. A secondary purpose of the intervention is to investigate the contribution of sensory inputs and lower extremity (LE) strength to mobility in this population.

Older adults living in the community are at increased risk for falls. There is evidence that older adults who fall have difficult utilizing sensory inputs and have decreased strength and mobility. Older adults with reduced ability to utilize sensory function have not been well studied. Older adults in research settings have improved balance measures after receiving balance training that manipulates sensory inputs. It is not known if providing balance training that manipulates sensory inputs in a home setting will improve balance. There is evidence that high intensity strength training in a gym setting can improve balance control in balance impaired older adults living in the community. Again it is not known if this type of strength training in a home setting will provide similar results. The ability to utilize sensory inputs, and decreased muscle strength, can be modifiable risk factors for falls in balance impaired older community dwelling adults. The **primary purpose** of this study is to examine the influence of an individualized progressive exercise program, which manipulates sensory inputs, on the ability of this population to use vestibular inputs to maintain balance. The secondary objective of this study is to assess the relative contribution of ability to use vestibular inputs, and LE muscle strength on the mobility of this population.

The **sample** population for this study is older adults, 65 year and over, who have been referred to Extra-Mural Program (EMP) physiotherapy in Health Region 2, New Brunswick for increased fall risk (i.e. decreased balance and or LE strength). Patients who have been referred will be contacted by EMP physiotherapy to ask permission for the research coordinator to contact them regarding the study. If the patient agrees, the research coordinator will meet with the patient, fully explain the study and give the patient an opportunity to ask questions. Individuals who volunteer for the study, and who meet the inclusion and exclusion criteria below, will review the informed consent forms with the research coordinator and sign informed consent forms. After the consent form is signed, an MMSE will be done by the research physiotherapist as part of the initial screening process.

The family physician will be asked to provide medical clearance for the patients to participate in the study. The Clinical Test of Sensory Interaction and Balance (Foam and Dome Test) will be used to assess the ability of the participants to utilize sensory inputs for balance control, lower extremity muscle strength will be assessed with portable dynamometry and mobility will be assessed with the Timed Up and Go Test. Participants will be randomly assigned to a Control Group (progressive resistive exercise for lower limb muscle strength, delivered with 1:1 supervision by a research physiotherapist from the Extra-Mural Program) or to a Combined Exercise Group with progressive exercise for both balance and lower limb strength, delivered with 1:1 supervision by the research coordinator.

Assessment will be conducted in the participant's home by a member of the research team. Results of the assessments will be made available to EMP physiotherapists and

physicians upon request of the participant.
Thank you for your time and assistance with this research project.
Sincerely,
Denise Hollway, Research Coordinator
Physiotherapist
Student, MSc. Rehabilitation Research- Physiotherapy
School of Physiotherapy, Dalhousie University
Phone 506-849-0245
Email dhollway@nb.sympatio.ca

Inclusion Criteria

65 years or older Able to provide informed consent
Ambulating at least in own home with/with out aide for 6 m; able to stand independently without aide
Living in community or independent retirement home

Exclusion Criteria

Unstable medical conditions including poor control of chronic medical conditions
Diagnosis of progressive neurological condition that would influence balance or muscle strength such as Parkinson's, ALS, MS
MMSE score of < 23 (Moderate to advanced dementia)
Acute osteoarthritis in lower extremities that limits maximal muscle contractions
Pain on weight bearing Weight bearing restrictions
Receiving physiotherapy at time of study Abnormal VOR or evidence of nystagmus
Legal blindness Stroke within past year
MMSE = Mini mental State Examination; VOR=vestibular ocular reflex; ALS=amyotrophic lateral sclerosis,

Appendix C: Standard Tool Descriptions and Procedures:

Timed up and Go (TUG)

The Activities-Specific Balance Confidence Scale (ABC Scale)

Clinical Test of Sensory Interaction and Balance (CTSIB)

Maximum Voluntary Isometric Contraction (MVIC) Strength Test

Hearing Handicap Inventory for the Elderly-Screening Version (HHIE-S)*

Mini-Mental State Examination (MMSE)

TIMED UP AND GO (TUG):

The TUG has been used as a screening tool to assess risk of falls and/or as a measure of functional mobility. [47, 49] The time taken to complete the test can be correlated to the level of functional mobility. [47] Older adults (age 70 to 84 years) who are able to complete the task in 20 s have been shown to be independent in transfer tasks involved in Activities of daily living (ADL), have high Berg scores and walk at gait speed that should be sufficient for community mobility.[47] Whereas those with scores of 30 seconds or longer tend to be more dependant in ADLS, require assistive devices for ambulation and score lower on the Berg Balance scale. [47]

Shumway-Cook et al. (2000) indicated that the TUG is a valid screening test for both the level of functional mobility and risk of falls in community dwelling elderly people. They found that older adults who take longer than 14 seconds to complete the TUG have a high risk for falls. [37] The TUG has clinical utility as a falls risk screening tool it cannot provide detailed information regarding impairments that contribute to falls risk and therefore provides limited information on how to target intervention strategies.[24]

The TUG has been used on several populations including community dwelling seniors and nursing home populations and those with and without cognitive impairment. [37, 49] The inter and intra-rater reliability appear to fall within acceptable limits (3 studies reported ICC's between .91-.99 for intra and inter-rater reliability and test retest conditions) thus it appears TUG is a reliable measure as a risk assessment tool.[37, 47, 49] It appears that more work needs to be done around the TUG as an outcome measure and as a predictive tool. [24, 37]

The TUG has appeal for use in home care as it is quick to administer and little equipment is needed – a chair, stopwatch, measuring tape and a 3 meter area to walk. Using the TUG as a screening tool may be an efficient use of resources to prompt further assessments to direct interventions.[24]

TIMED UP AND GO TEST PROTOCOL USED IN THE STUDY SCREENING PROCESS:

[93]

Instructions to the participant:

"The purpose of this test is to time the length of time that it takes for you to get up from this chair, walk to the mark on the floor, turn around, come back to the chair and then sit down. For this test you may wear your regular foot wear and use your regular walking aide. You will not be given any physical assistance with this test however, someone will be near you to prevent you from falling. For added safety, you are required to wear a safety belt for this test. You will be given a practice run with this test that is not timed to familiarize you with this test. You will be allowed to rest for 1 minute between the practice test and the timed test"

Instructions to the tester:

- Allow the participant to practice the test and rest for at least 1 minute. When the participant is ready, complete one timed trial of the test.
- Start timing when the participant initiates sit to stand movement and stop timing when the participant's back come to rest against the back rest of the chair.
- Record the time taken to complete the TUG test to the nearest 0.01 second
- Repeat the last two steps for a total of 3 timed trials
- One practice trial was performed and three timed trials were recorded.

THE ACTIVITIES-SPECIFIC BALANCE CONFIDENCE SCALE (ABC SCALE):

The ABC scale was developed to measure balance confidence /self-efficacy. [63] It is a 16 item questionnaire that can be self-administered or over the phone. Each item describes a specific activity that requires a position change or walking in more difficult situations.

The ABC scale appears to be a reliable scale that has internal consistency (Cronbach's alpha 0.96) and has shown test retest reliability after a 2 week interval (r=0.92) in the Community dwelling elderly. [63] The ABC scale demonstrates validity through the relationship with the performance on balance measures and Pearson correlations between the ABC scale and Berg of .72 and the ABC and TUG of .698. [64] Significant differences in ABC scores between fallers and non fallers have also been found with a suggested cut off score of 67% [63] on the ABC resulted in 84.4 % sensitivity and 87.5% specificity in predicting future falls. [65]

The ABC scale is suited for use in home care as it appears to be a reliable and valid test and appears to correlate to performance on balance measures. No specialized equipment is required.

ABC PROTOCOL USED IN THE STUDY SCREENING PROCESS
[93]

Instructions to the tester:

Participants should be queried concerning their understanding of instructions, and probed regarding difficulty answering specific items.

Instructions to participants prior to testing:

"For each of the following, please indicate your level of confidence in doing the activity without losing your balance or becoming unsteady from choosing one of the percentage points on the scale form 0% to 100%. If you do not currently do the activity in question, try and imagine how confident you would be if you had to do the activity. If you normally use a walking aid to do the activity or hold onto someone, rate your confidence as it you were using these supports. If you have any questions about answering any of these items, please ask "

Instructions for Scoring:

The ABC is an 11-point scale and ratings should consist of whole numbers (0-100) for each item. **Total the ratings (possible range = 0 – 1600) and divide by 16 to get each subject's ABC score.** If a subject qualifies his/her response to items #2, #9, #11, #14 or #15 (different ratings for "up" vs. "down" or "onto" vs. "off"), solicit separate ratings and use the lowest confidence of the two (as this will limit the entire activity, for instance the likelihood of using the stairs.)

THE ACTIVITIES-SPECIFIC BALANCE CONFIDENCE (ABC) SCALE [94]

For each of the following activities, please indicate your level of self-confidence by choosing a corresponding number from the following rating scale:
0% 10 20 30 40 50 60 70 80 90 100%

not confidence completely confident

"How confident are you that you will not lose your balance or become unsteady when you…

…walk around the house? ____%

…walk up or down stairs? ____%

...bend over and pick up a slipper from the front of a closet floor ____%

…reach for a small can off a shelf at eye level? ____%

…stand on your tiptoes and reach for something above your head? ____%

…stand on a chair and reach for something? ____%

…sweep the floor? ____%

…walk outside the house to a car parked in the driveway? ____%

…get into or out of a car? ____%

…walk across a parking lot to the mall? ____%

…walk up or down a ramp? ____%

…walk in a crowded mall where people rapidly walk past you? ____%

…are bumped into by people as you walk through the mall?____%

… step onto or off an escalator while you are holding onto a railing? ____%

… step onto or off an escalator while holding onto parcels such that you cannot hold onto the railing? ____%

…walk outside on icy sidewalks? ____%

CLINICAL TEST OF SENSORY INTERACTION AND BALANCE (CTSIB)

The CTSIB is a tool that assesses a person ability to use sensory input for balance. [66] Test retest reliability has been found to be .75 in older community dwelling adults with test- retest and inter-rater reliability being .99 in healthy young subjects.[67, 68] The CTSIB appears to be a valid way to examine older patients' ability to utilize somatosensory inputs and remain upright. [69-71] It has been found that foam posturography (CTSIB) had a specificity of 90 % and sensitivity of 95% as compared to the gold standard-platform posturography for identifying those with vestibular dysfunction. [69] It appears that an ability to maintain balance during this test for 30 seconds in all test conditions should be possible even in older adults. [71]

CTSIB PROTOCOL USED IN THE STUDY SCREENING PROCESS

[93]

Instructions to the participant prior to testing:[93]

"The purpose of this test is to test your ability to stand during 6 different balance tests. For the first test, you will be standing on the floor with your feet shoulder width apart with your eyes open. For the second test, your eyes will be closed and for the third test, you will wear a dome that blocks your vision. You can be in your stocking feet or be in your bare feet for all of the trials. For the next fourth test, you will stand on a large piece of foam with your feet shoulder width apart with your eyes open. For the fifth test, you will stand on a piece of foam with your eyes closed and for the sixth test you will wear a dome over your head block your vision while standing on the piece of foam. You will be asked to complete each test three times. Each trial will last up to 30 seconds. If you can hold your balance for 30 seconds then you will be asked to repeat that trial with your feet

together. You will be allowed to rest between trials for at least 30 seconds. To ensure that you are safe, there will be someone standing next to you as a spotter in case you lose your balance. We will require you to wear a safety belt for the whole test. Please tell us if you would like to stop for any reason."

Instructions during testing:

"Stand with your arms folded across your waist with your hands above your elbows, and look straight ahead. Hold this position until I tell you to stop"

Instructions to the tester:

1. Repeat the above instructions, having the participants close their eyes for Tests 2 and 4, wearing the visual conflict dome for Tests 3 and 6, and standing in the center of the foam for Tests 4 through 6.
2. Rotate the foam platform 90^0 and flip it over between trials
3. Record the time, to the nearest 0.01seconds, that the participant is able to maintain their balance during each trial. Stop timing if the participant's hands move, knees bend, heels or toes lift off the floor or if the participant takes a step to correct their balance.
4. If the participant is able to maintain the position for 30 seconds, record their score as 30 seconds for any subsequent trials of that test.
5. If the participant is able to complete the first trial for 30 seconds with their feet apart, have them repeat the trial with their feet together.
6. The final score will be the average of the three scores for each test.

MAXIMUM VOLUNTARY ISOMETRIC CONTRACTION STRENGTH TEST (MVIC)

Overall Hand Held Dynamometry (HHD) does appear to be a reliable and valid tool to measure strength in community dwelling seniors. [72] Test retest reliability in community dwelling elderly fallers using ICC $_{(2,1)}$ ranged from and .95 to .99 and ICC $_{(2,2)}$ 97 to 1.0. [73] In community dwelling seniors results of HHD correlated well with those of the Biodex (the gold standard) with an r of .91. [74]

MVIC PROTOCOL USED IN THE STUDY SCREENING PROCESS

[93]

Instructions to participant:[93]

"The purpose of this test is to record the force that you can produce with the muscles in your legs. I will be taking measures of the muscles that move your hip, your knee and your ankle. I will go through a practice run with you for each movement so you can warm up the muscles I am testing. I will go through three practice runs with you for each muscle being tested. I will then test each muscle three times. For each test, I would like you to push as hard as you can against the instrument for 6 seconds. I will tell you when to start and when to stop pushing by saying 'go' and 'stop'. It is important that you breathe properly during the muscle testing by taking a deep breath in before you push and breathing out as you push. I don't want you to hold your breath as you are pushing. You can rest for as long as you wish between tests. Please let me know if you have any pain or discomfort during the test, because the tests should not be painful"

Instructions to the tester:

- Record the moment arm length (to the nearest 0.1 cm), from the joint axis to the centre of the paddle on the dynamometer
- The total hold time for each contraction is 6 seconds including the ramp time. Encourage the participant to build up to their maximal contraction and hold for 6 seconds. Record the peak force, to the nearest 0.1 Newton that is produced.
- Provide three practice sets of sub maximal trials for each muscle group being tested.
- Perform three test trials with at least 2 minutes between each trial. Rest time can be longer depending on the needs of the participant.
- During the trial encourage the participant to breath properly
- Encourage proper technique with a ramp build up and monitor accessory movement of the body.
- Monitor dynamometer and strap placement during the trials to ensure that moment arm length is the same for all three trials.
- Encourage the participant to report and pain or fatigue between trials.

HEARING HANDICAP INVENTORY FOR THE ELDERLY-SCREENING VERSION (HHIE-S)*

The HHIE-S has been reported to be a valid and reliable tool to detect functional hearing impairment in the older adult. [75-77] The HHIE-S consists of 10 questions and are scored for each question accordingly , yes-4 points; sometimes-2 points; or no-0 points.[75, 77] The minimal score is 0 (no handicap) up to 40 (maximum handicap). (Yueh, Shapiro et al. 2003) Scores from 0-8 indicate 13 % probability of hearing impairment, 10-24 a 50 % probability and 26-40% an 84 % probability of hearing loss. [75, 76]

HEARING HANDICAP INVENTORY FOR THE ELDERLY-SCREENING VERSION (HHIE-S)* [76]

1. Does a hearing problem cause you to feel embarrassed when meeting new people?
2. Does a hearing problem cause you to feel frustrated when talking to members of your family?
3. Do you have difficulty hearing when someone speaks in a whisper?
4. Do you feel handicapped by a hearing problem?
5. Does a hearing problem cause you difficulty when visiting friends, relatives, or neighbors?
6. Does a hearing problem cause you to attend religious services less often than you would like?
7. Does a hearing problem cause you to have arguments with family members?
8. Does a hearing problem cause you difficulty when listening to TV or radio?
9. Do you feel that any difficulty with your hearing limits or hampers your personal or social life?

10. Does a hearing problem cause you difficulty when in a restaurant with relatives or friends?

MINI-MENTAL STATE EXAMINATION (MMSE)

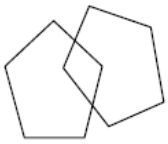
The MMSE is accepted as a reliable and valid screening tool for cognitive impairment with community dwelling, hospitalized and institutionalized older adults. It is an 11 question screening tool that examines five areas of cognition: orientation, registration, attention and calculation, recall and language. [79, 80] It is scored out of 30 with scores of 23 and lower indicating cognitive impairment. [80]

The scoring key, as well as, the exam are provided below. [80]

MMSE[80]

Mini-Mental State Examination (MMSE)

Maximum Score	Score	
		ORIENTATION
5	( )	What is the (year) (season) (date) (day) (month)?
5	( )	Where are we: (state) (county) (town or city) (hospital) (floor).
		REGISTRATION
3	( )	Name 3 Common objects (eg, "apple," "table," "penny"). Take 1 second to say each. Then ask the patient to repeat all 3 after you have said them. Give 1 point for each correct answer. Then repeat them until he/she learns all 3. Count trials and record. Trials:
		ATTENTION AND CALCULATION
5	( )	Spell "world" backwards. The score is the number of letters in correct order (D__L__R__O__W__).
		RECALL
3	( )	Ask for the 3 objects repeated above. Give 1 point for each correct answer. [Note: recall cannot be tested if all 3 objects were not remembered during registration]
		LANGUAGE
2	( )	Name a "pencil," and "watch." (2 points)
1	( )	Repeat the following, "no ifs, ands, or buts." (1 point)
3	( )	Follow a 3-stage command: "Take a paper in you right hand, fold it in half, and put it on the floor." (3 points)
		Read and obey the following:
1	( )	Close your eyes. (1 point)
1	( )	Write a sentence. (1 point)
1	( )	Copy the following design. (1 point)



No construction problem.

Total score _____

Appendix D: Histograms/Distribution, and Summary Tables by Particpant

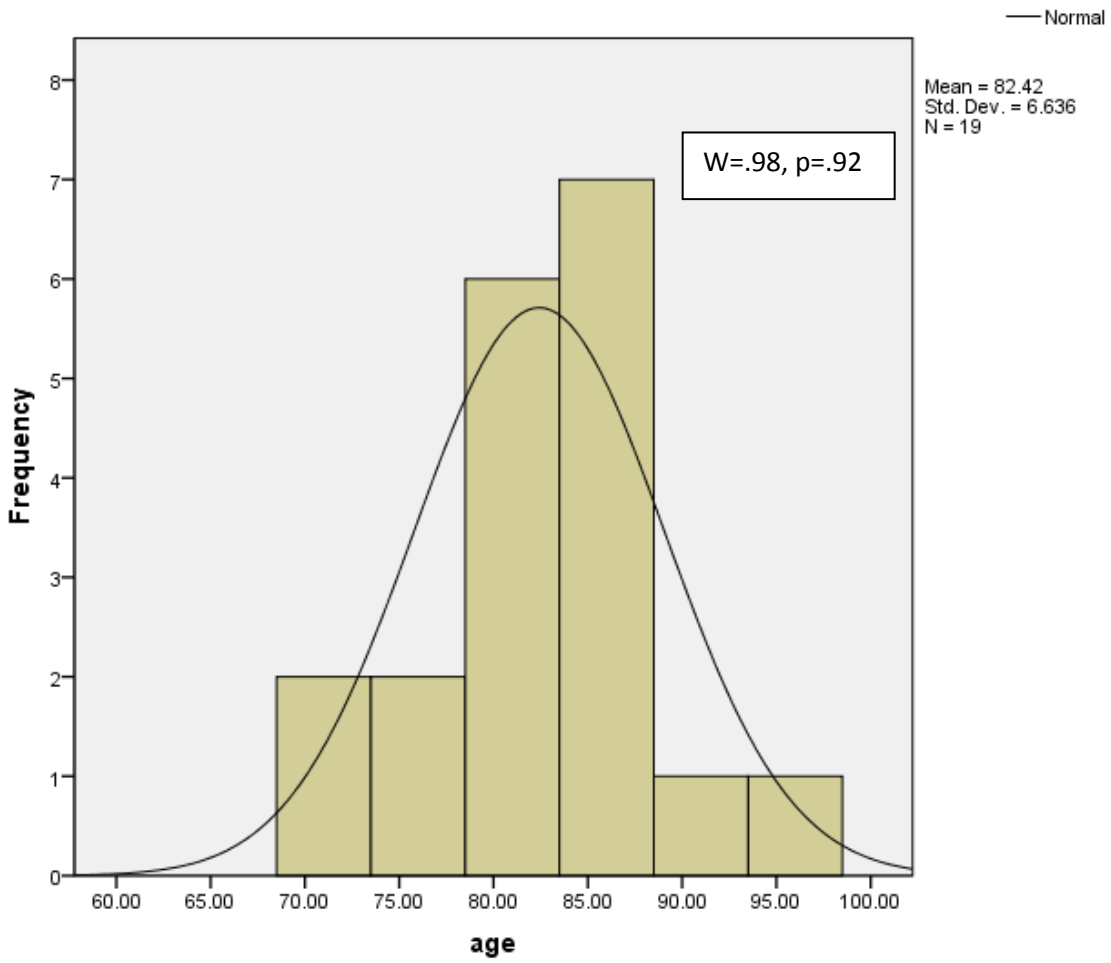


FIGURE D-1: HISTOGRAM WITH DISTRIBUTION CURVE OF AGE.

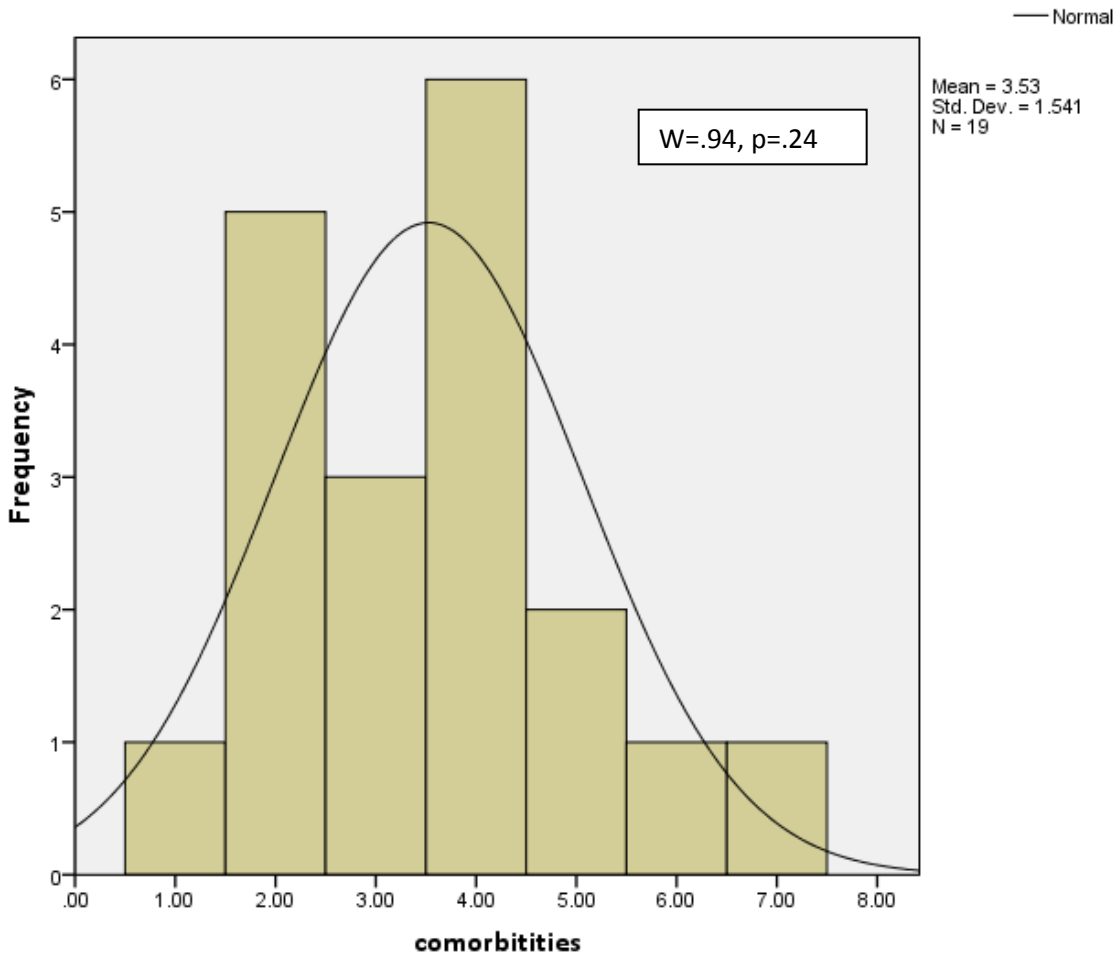


FIGURE D-2: HISTOGRAM WITH DISTRIBUTION CURVE OF CO-MORBIDITIES.

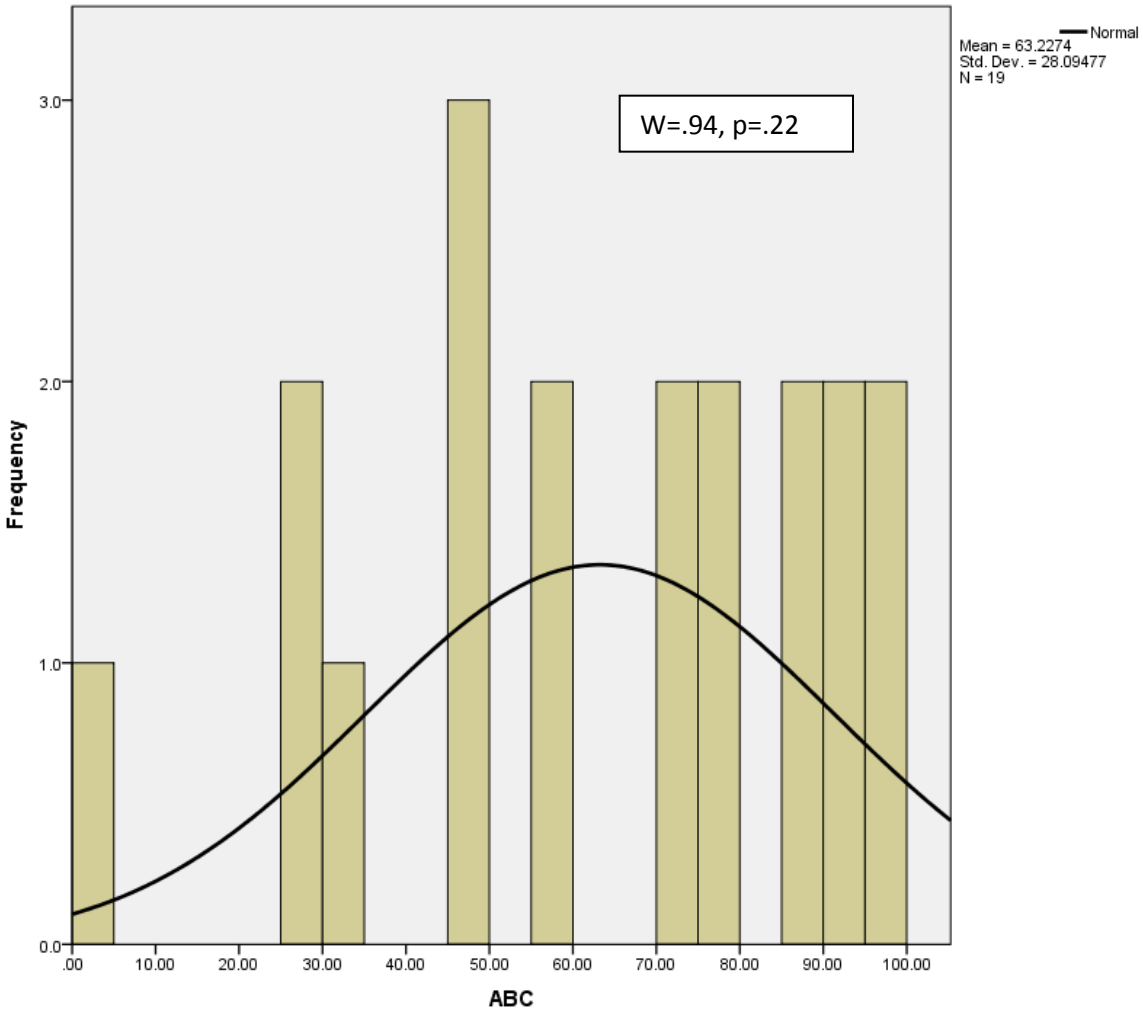


FIGURE D-3: HISTOGRAM WITH DISTRIBUTION CURVE OF ABC SCORES.

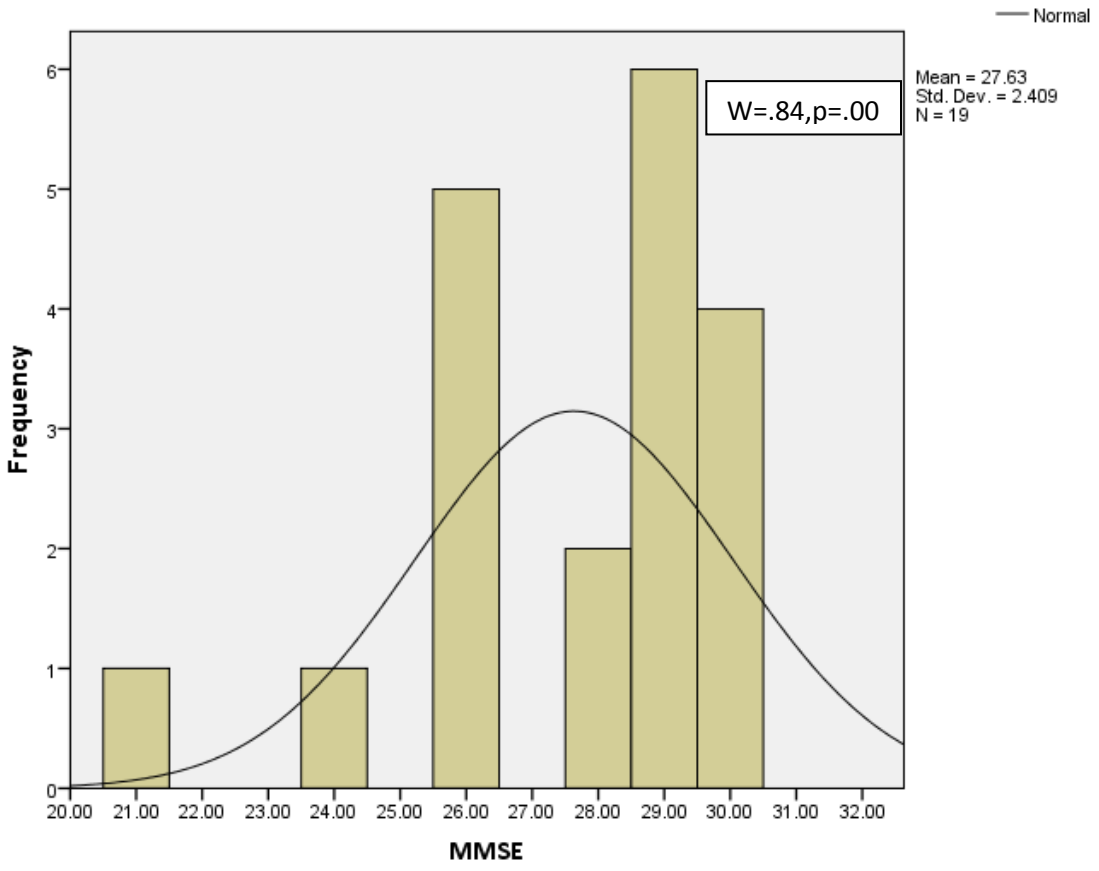


FIGURE D-4: HISTOGRAM WITH DISTRIBUTION CURVE OF MMSE SCORES.

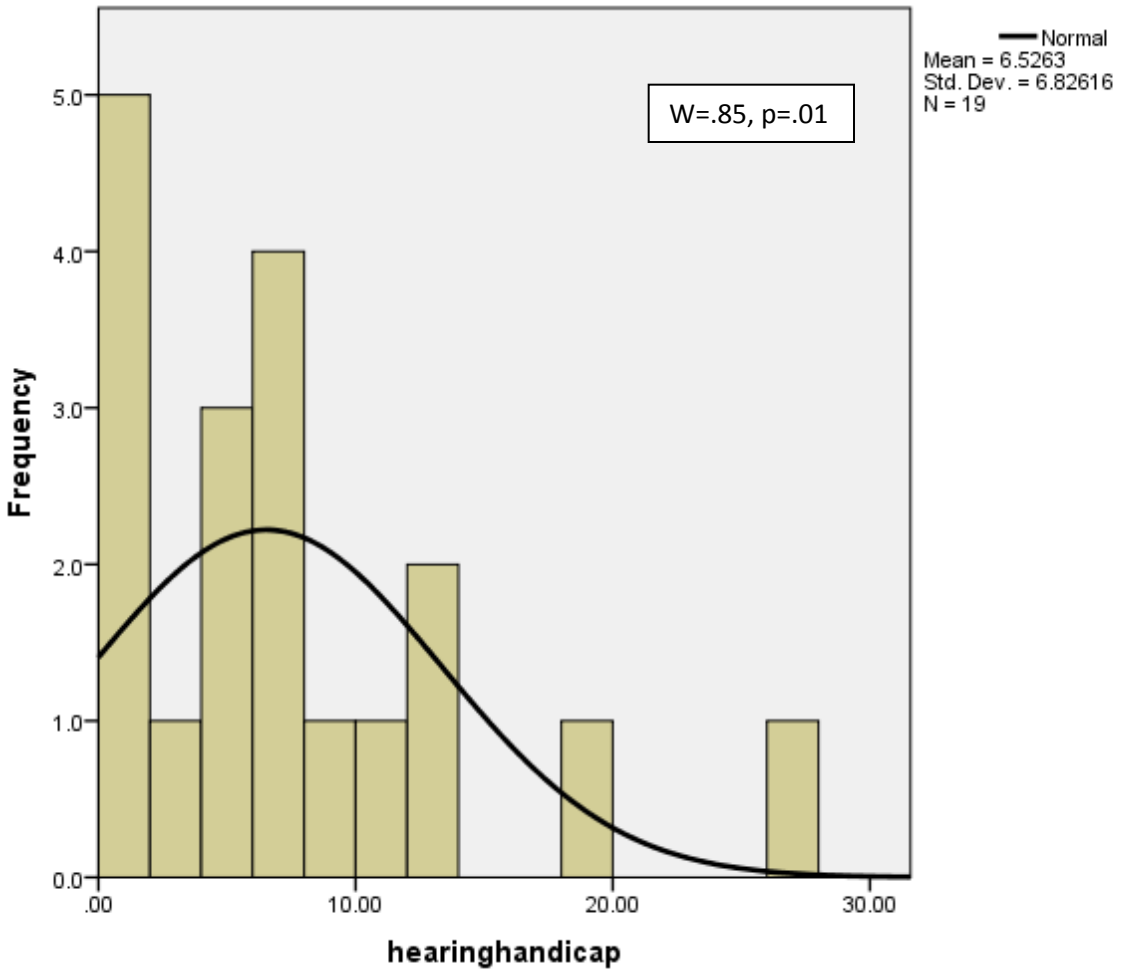


FIGURE D-5: HISTOGRAM WITH DISTRIBUTION CURVE FOR HEARING HANDICAP INVENTORY SCORES.

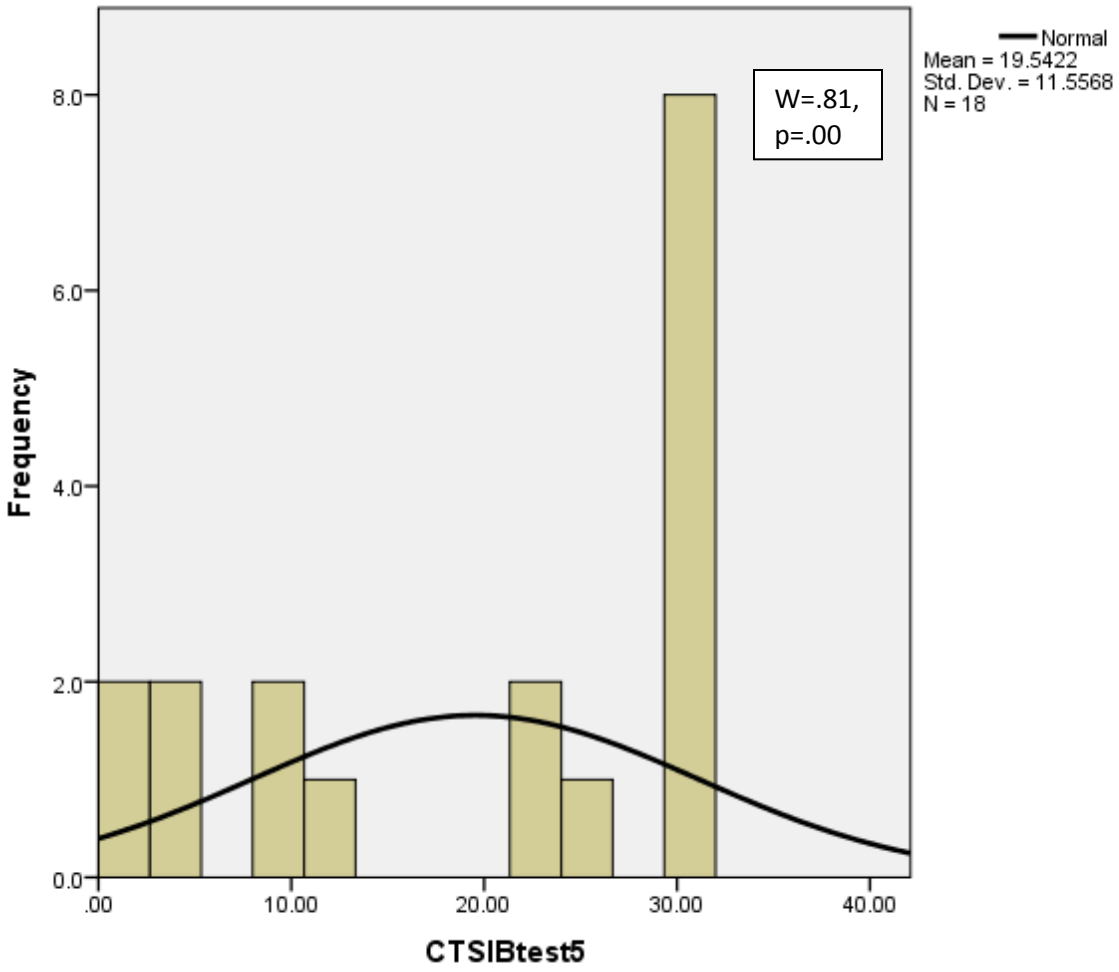


FIGURE D-6: HISTOGRAM WITH DISTRIBUTION CURVE FOR CTSIB TEST 5 SCORES.

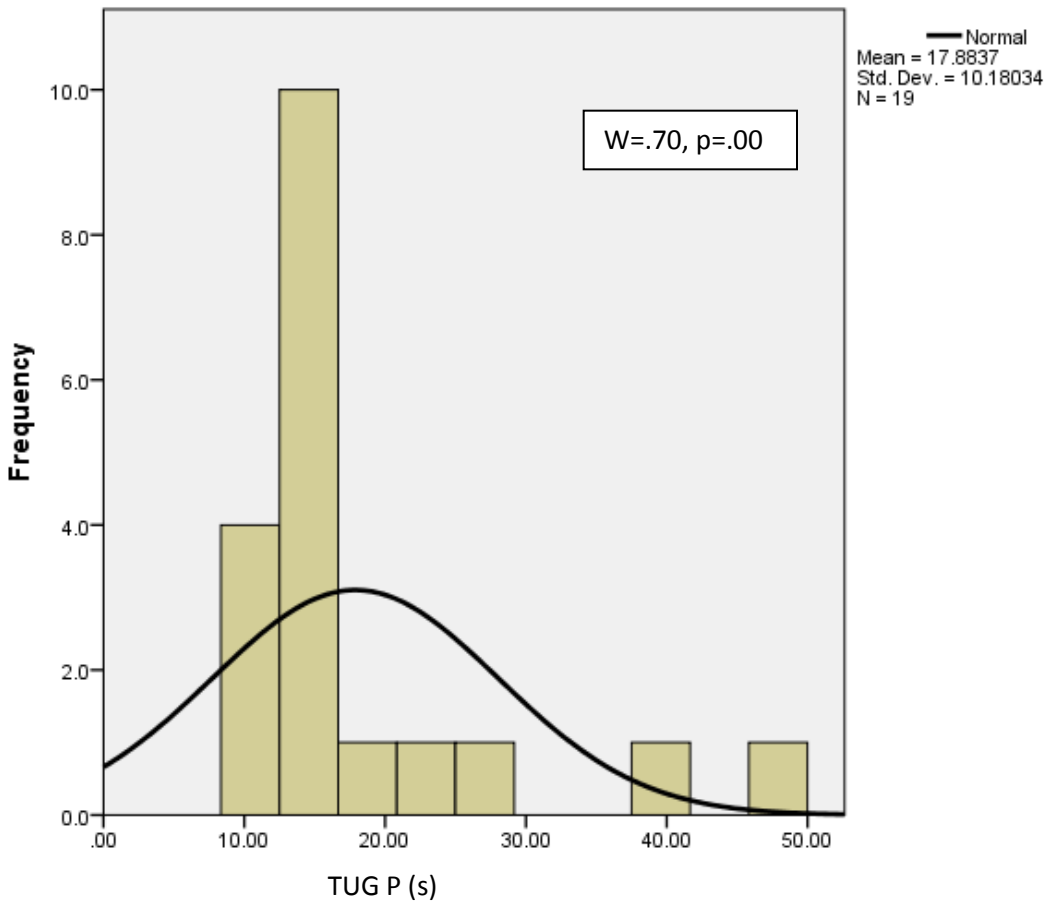


FIGURE D-7: HISTOGRAM WITH DISTRIBUTION CURVE OF TUG_P SCORES (N=19).

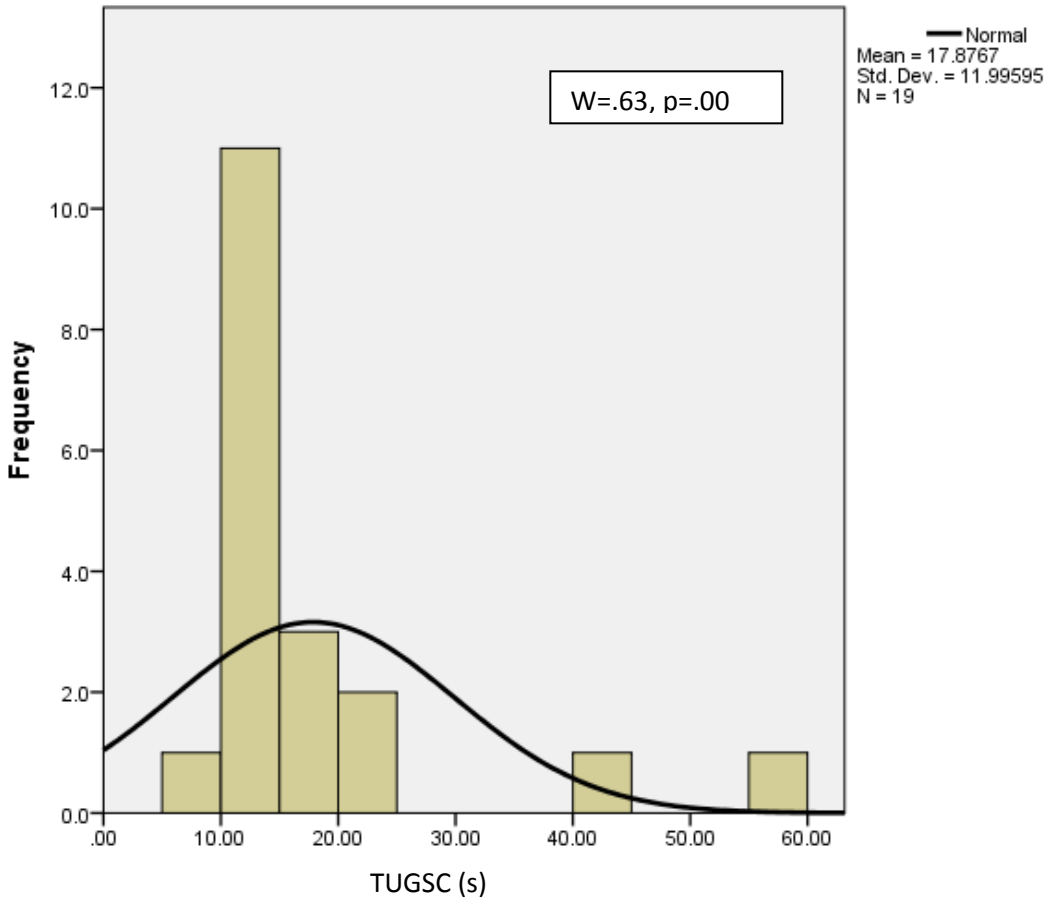


FIGURE D- 8: HISTOGRAM WITH DISTRIBUTION CURVE OF TUG_{SC} SCORES (N=19).

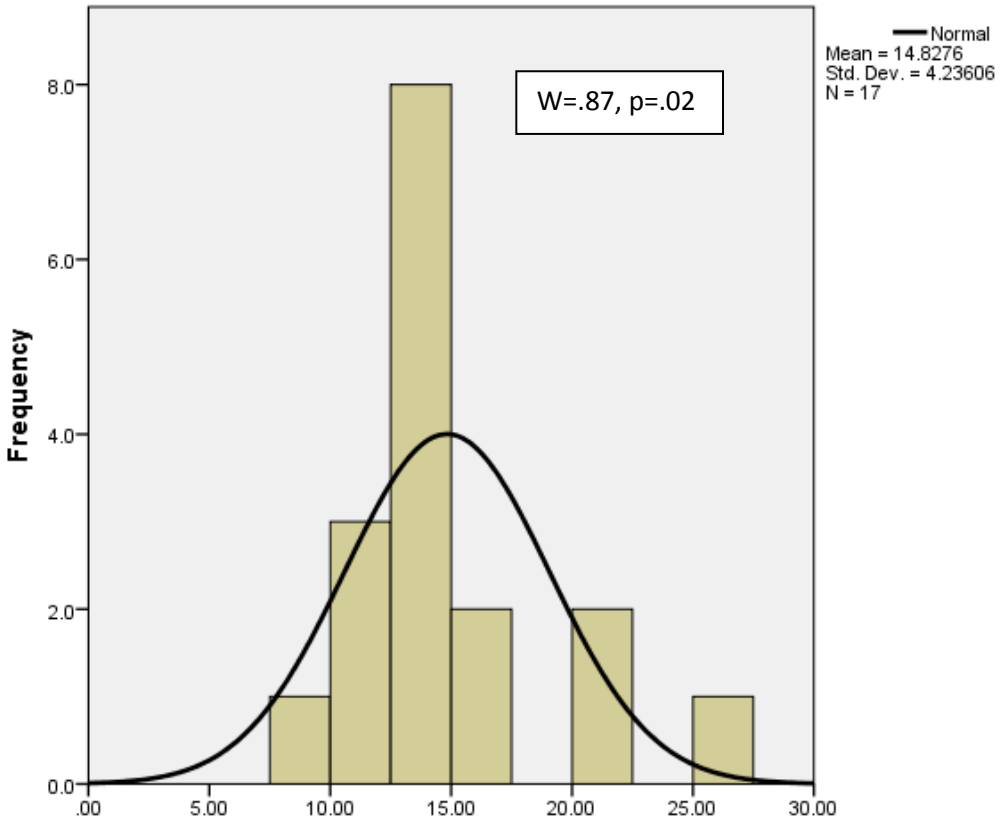


FIGURE D-9: HISTOGRAM WITH DISTRIBUTION CURVE OF TUG_P SCORES OF THOSE BELOW 30S (N=17).

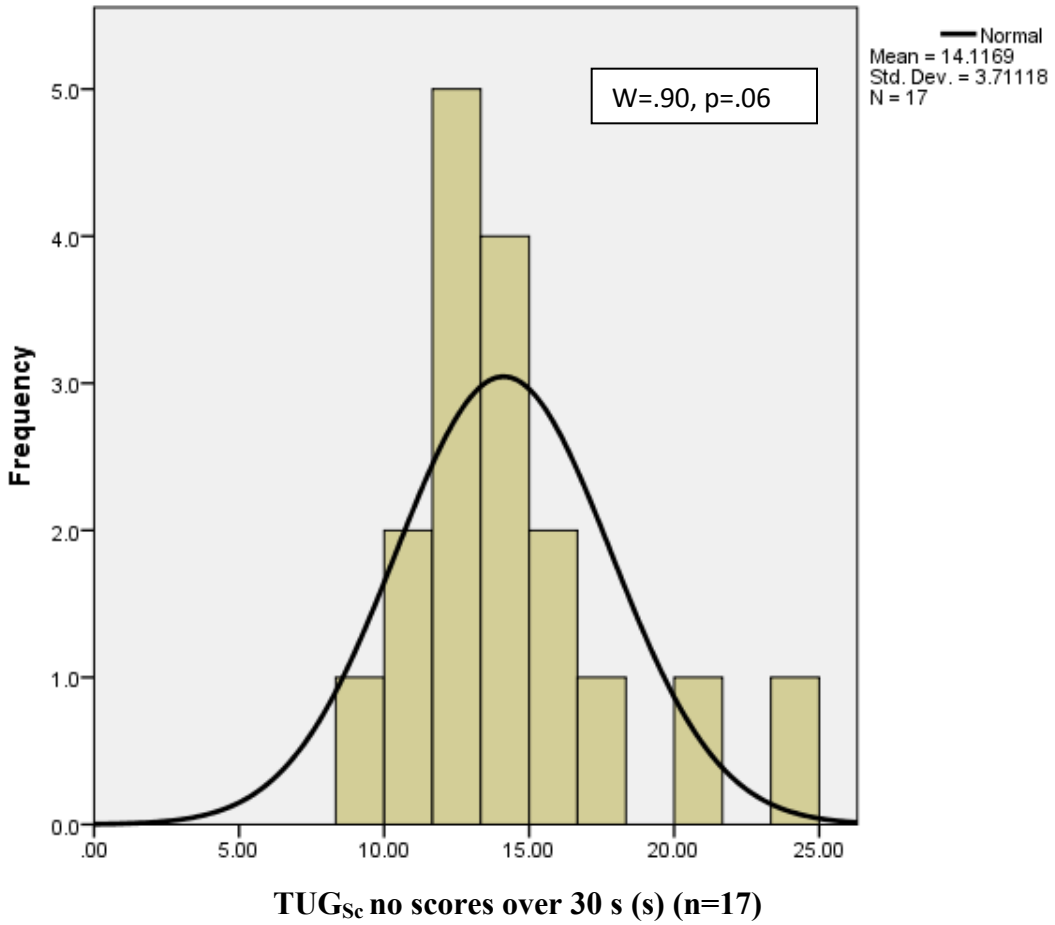


FIGURE D-10: HISTOGRAM WITH DISTRIBUTION CURVE OF TUG_{SC} SCORES OF THOSE BELOW 30S (N=17).

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TUGP	.336	19	.000	.699	19	.000
TUGSC	.308	19	.000	.627	19	.000

a. Lilliefors Significance Correction

FIGURE D-11: SPSS OUTPUT OF THE TEST OF NORMALITY RESULTS FOR THE TUG_P AND TUG_{SC} PROTOCOLS.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
age	.095	19	.200*	.978	19	.923
MMSE	.241	19	.005	.839	19	.004
comorbitities	.169	19	.159	.938	19	.240
ABC	.142	19	.200*	.941	19	.273
hearinghandicap	.215	19	.021	.846	19	.006

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CTSIBtest5	.250	18	.004	.810	18	.002

a. Lilliefors Significance Correction

FIGURE D-12: SPSS OUTPUTS OF THE TEST OF NORMALITY RESULTS FOR THE SAMPLE CHARACTERISTICS.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
falls prior to screen	.492	16	.000	.484	16	.000

a. Lilliefors Significance Correction

FIGURED-13: SPSS OUTPUT OF THE TEST OF NORMALITY RESULTS FOR THE SAMPLE CHARACTERISTICS.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TUGPnoout	.277	17	.001	.872	17	.024
TUGSCnoout	.183	17	.132	.896	17	.058

a. Lilliefors Significance Correction

FIGURE D-14: SPSS OUTPUT OF THE TEST OF NORMALITY RESULTS WITH NO OUTLIERS (N=17) FOR THE TUG_P AND TUG_{SC} PROTOCOLS.

Descriptive Statistics

	N	Range	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
TUGP	19	41.41	17.8837	2.33553	10.18034	103.639
TUGSC	19	47.73	17.8767	2.75206	11.99595	143.903
TUGPnoout	17	17.72	14.8276	1.02740	4.23606	17.944
TUGSCnoout	17	14.97	14.1169	.90009	3.71118	13.773
Valid N (listwise)	17					

FIGURE D-15: SPSS OUTPUT OF THE MEAN AND STANDARD DEVIATION RESULTS WITH (N=19) AND WITHOUT OUTLIERS (N=17) FOR THE TUG_P AND TUG_{SC} PROTOCOLS.

TABLE D-1: SUMMARY OF TUG RESULTS BY PROTOCOL AND TRIAL

Participant ID	TUG_P/ Trial 1 (s)	TUG trial 2 (s)	TUG trial 3 (s)	TUG_{SC} (s)
EM06	14.78	13.28	13.18	13.75
EM18	15.22	14.97	15.07	15.09
EM19	37.94	46.38	45.41	43.24
EM22	10.81	10.00	10.89	10.57
EM23	8.37	8.72	9.00	8.70
EM25	13.69	11.88	11.28	12.28
EM31	13.20	11.84	11.59	12.21
EM35	11.31	10.84	10.59	10.91
EM36	12.28	12.44	12.09	12.27
EM37	14.81	14.78	14.09	14.56
EM39	15.32	16.17	15.70	15.73
EM41	13.23	11.66	11.41	12.10
EM02	20.12	18.50	16.15	18.26
EM03	26.09	21.28	23.62	23.66
EM24	14.09	13.34	12.65	13.36
EM26	14.56	13.69	13.09	13.78
EM30	21.44	20.09	19.34	20.29
EM33	49.78	59.44	60.06	56.43
EM34	12.75	12.43	12.24	12.47

TABLE D-2: SUMMARY OF CHARACTERISTICS BY PARTICIPANT

Participant ID	Age	Sex	MMSE	ABC	Co-morbities	Gait aid	Fall history	Vestibular issue	CTSIB test 5	Participant in the exercise study
EM06	81.00	F	30.00	27.50	5.00	4 wheeled walker	no	no	30.00	no
EM18	75.00	F	29.00	78.00	6.00	2 wheeled walker	yes	no	24.63	no
EM19	92.00	F	26.00	32.00	2.00	4 wheeled walker	yes	yes	Unable to complete	no
EM22	79.00	F	29.00	94.40	7.00	none	yes	no	29.49	no
EM23	84.00	M	29.00	87.50	3.00	none	unknown	no	30.00	no
EM25	71.00	F	26.00	98.00	4.00	none	yes	no	30.00	no
EM31	80.00	F	26.00	71.50	2.00	cane	yes	no	30.00	no
EM35	87.00	F	30.00	74.70	4.00	none	yes	no	30.00	no
EM36	80.00	M	29.00	99.40	4.00	none	yes	no	30.00	no
EM37	72.00	F	30.00	48.70	4.00	2 wheeled walker	no	no	30.00	no
EM39	88.00	F	24.00	77.00	2.00	2 wheeled walker	yes	yes	21.61	no
EM41	97.00	F	29.00	85.60	1.00	none	yes	yes	22.17	no
EM02	75.00	F	26.00	55.60	3.00	cane	no	yes	10.49	yes
EM03	80.00	M	21.00	49.00	2.00	none	unknown	yes	.00	yes
EM24	86.00	F	26.00	59.00	4.00	none	yes	yes	10.37	yes
EM26	88.00	M	28.00	90.00	2.00	none	yes	yes	4.81	yes
EM30	84.00	F	30.00	27.80	4.00	cane	yes	yes	1.83	yes
EM33	82.00	F	28.00	.00	5.00	2 wheeled walker	yes	yes	4.18	yes
EM34	85.00	M	29.00	45.62	3.00	none	no	yes	12.18	yes

Appendix E: SEM, 95 % CI and MDD

Calculations of SEM, CI $_{95\%,}$ MDD

FigureE-1: SPSS Output for Mean and Standard Deviation for (N=19).

Figure E-2: SPSS Output for Mean and Standard Deviation for (N=17).

Figure E-3: SPSS Output for Mean and Standard Deviation for (N=19) and (N=17)

Average Between Protocol TUG_P and TUG_{SC} Scores

CALCULATIONS OF SEM, CI $_{95\%,}$ MDD

Using ICC model 3 results:

a) SEM=$S_x\sqrt{1-r_{xx}}$

Where SEM=$S_x\sqrt{1-r_{xx}}$

b) Calculation : CI $_{95\ \%}$

95 % CI = observed score± 1.96 (SEM)

c) MDD = $1.96 * \sqrt{2} * SEM$

TUG_P (N=19) r_x=.98

SEM= 10.18034 (.14)

a) SEM =1.44

b) 95% CI

=17.88±2.82

=15.06 to 20.70 seconds

c) MDD $= 1.96 * \sqrt{2} * SEM$

=1.96*1.4*1.44

=3.99s

TUG_{SC} (N=19)

SEM= 12.00(.1414)

a) SEM= 1.70

95 % Confidence Interval

= 17.88±3.32

b) =14.56 to 21.20

c) MDD= 1.96*1.4*1.70

=4.72s

For sample with scores above 30 removed TUG_P (n=17)

SEM=4.23(.134)

a) SEM=.57

95 % Confidence interval

14.82±1.11

b) =13.70 to 15.94 seconds

c) MDD= 1.96*1.4*.57

=1.58s

For sample with scores above 30 removed TUGSC (n=17)

SEM= 3.71(0.134)

A) SEM= .50

b) 95% Confidence interval

=14.12± .98

=13.14 to 15.10 seconds

c) MDD= 1.96*1.4*.50

=1.38 s

Descriptive Statistics

	Mean	Std. Deviation	N
TUG trial 1	17.8837	10.18034	19
TUGSC	17.8767	11.99595	19

FIGURE E-1: SPSS OUTPUT FOR MEAN AND STANDARD DEVIATION FOR (N=19).

Descriptive Statistics

	Mean	Std. Deviation	N
TUG1noout	14.8276	4.23606	17
mean3trialnoout	14.1169	3.71118	17

FIGURE E-2: SPSS OUTPUT FOR MEAN AND STANDARD DEVIATION FOR (N=17).

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
TUGPTUGSCave	19	44.57	8.53	53.10	17.8802	11.06887
nooutTUGPTUGSCave	17	16.34	8.53	24.88	14.4723	3.96421
Valid N (listwise)	17					

FIGURE E-3: SPSS OUTPUT FOR MEAN AND STANDARD DEVIATION FOR (N=19) AND (N=17) AVERAGE BETWEEN PROTOCOL TUG_P AND TUG_{SC} SCORES.

Appendix F: ICC Model (3,1)

Figure F-1: SPSS Output for ICC Model 3 ANOVA Table (N=19).

Figure F-2: SPSS Output for ICC Model 3 ICC Table (N=19).

Figure F-3: SPSS Output for ICC Model 3 ANOVA Table (N=17).

Figure F-4: SPSS Output for ICC Model 3 ICC Table (N=17).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig
Between People		7.254	18	.403		
	Between Items	.006	1	.006	2.742	.115
Within People	Residual	.042	18	.002		
	Total	.048	19	.003		
Total		7.302	37	.197		

Grand Mean = 2.7660

FIGURE F-1: SPSS OUTPUT FOR ICC MODEL 3 ANOVA TABLE (N=19).

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.989[a]	.970	.996	173.128	18	18	.000
Average Measures	.994[c]	.985	.998	173.128	18	18	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

FIGURE F-2: SPSS OUTPUT FOR ICC MODEL 3 ICC TABLE (N=19).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig
Between People		2.064	16	.129		
	Between Items	.016	1	.016	17.147	.001
Within People	Residual	.015	16	.001		
	Total	.032	17	.002		
Total		2.096	33	.064		

Grand Mean = 2.6401

FIGURE F-3: SPSS OUTPUT FOR ICC MODEL 3 ANOVA TABLE (N=17).

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.985[a]	.960	.995	134.202	16	16	.000
Average Measures	.993[c]	.979	.997	134.202	16	16	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

FIGURE F-4: SPSS OUTPUT FOR ICC MODEL 3 ICC TABLE (N=17).

Appendix G Correlation:

Figure G-1: SPSS Output of the Means and Standard Deviation for TUG_P and TUG_{SC} (N=19).

Figure G- 2: SPSS Output of the Spearman Rank Correlation for TUG_P and TUG_{SC} (N=19).

Figure G-3: SPSS Output of the Means and Standard Deviation for TUG_P and TUG_{SC} (n=17).

Figure G-4: SPSS Output of the Spearman Rank Correlation for TUG_P and TUG_{SC} (n=17).

Descriptive Statistics

	Mean	Std. Deviation	N
TUG trial 1	17.8837	10.18034	19
TUGSC	17.8767	11.99595	19

FIGURE G- 1: SPSS OUTPUT OF THE MEANS AND STANDARD DEVIATION FOR TUG_P AND TUG_{SC} (N=19).

Correlations

			TUG trial 1	TUGSC
Spearman's rho	TUG trial 1	Correlation Coefficient	1.000	.977**
		Sig. (2-tailed)	.	.000
		N	19	19
	TUGSC	Correlation Coefficient	.977**	1.000
		Sig. (2-tailed)	.000	.
		N	19	19

**. Correlation is significant at the 0.01 level (2-tailed).

FIGURE G-2: SPSS OUTPUT OF THE SPEARMAN RANK CORRELATION FOR TUG_P AND TUG_{SC} (N=19).

Descriptive Statistics

	Mean	Std. Deviation	N
TUG1noout	14.8276	4.23606	17
mean3trialnoout	14.1169	3.71118	17

FIGURE G-3: SPSS OUTPUT OF THE MEANS AND STANDARD DEVIATION FOR TUG_P AND TUG_{SC} (N=17).

Correlations

			TUG1noout	mean3trialnoout
Spearman's rho	TUG1noout	Correlation Coefficient	1.000	.968**
		Sig. (2-tailed)	.	.000
		N	17	17
	mean3trialnoout	Correlation Coefficient	.968**	1.000
		Sig. (2-tailed)	.000	.
		N	17	17

**. Correlation is significant at the 0.01 level (2-tailed).

FIGURE G- 4: SPSS OUTPUT OF THE SPEARMAN RANK CORRELATION FOR TUG_P AND TUG_{SC} (N=17).

Appendix H: Bland Altman:

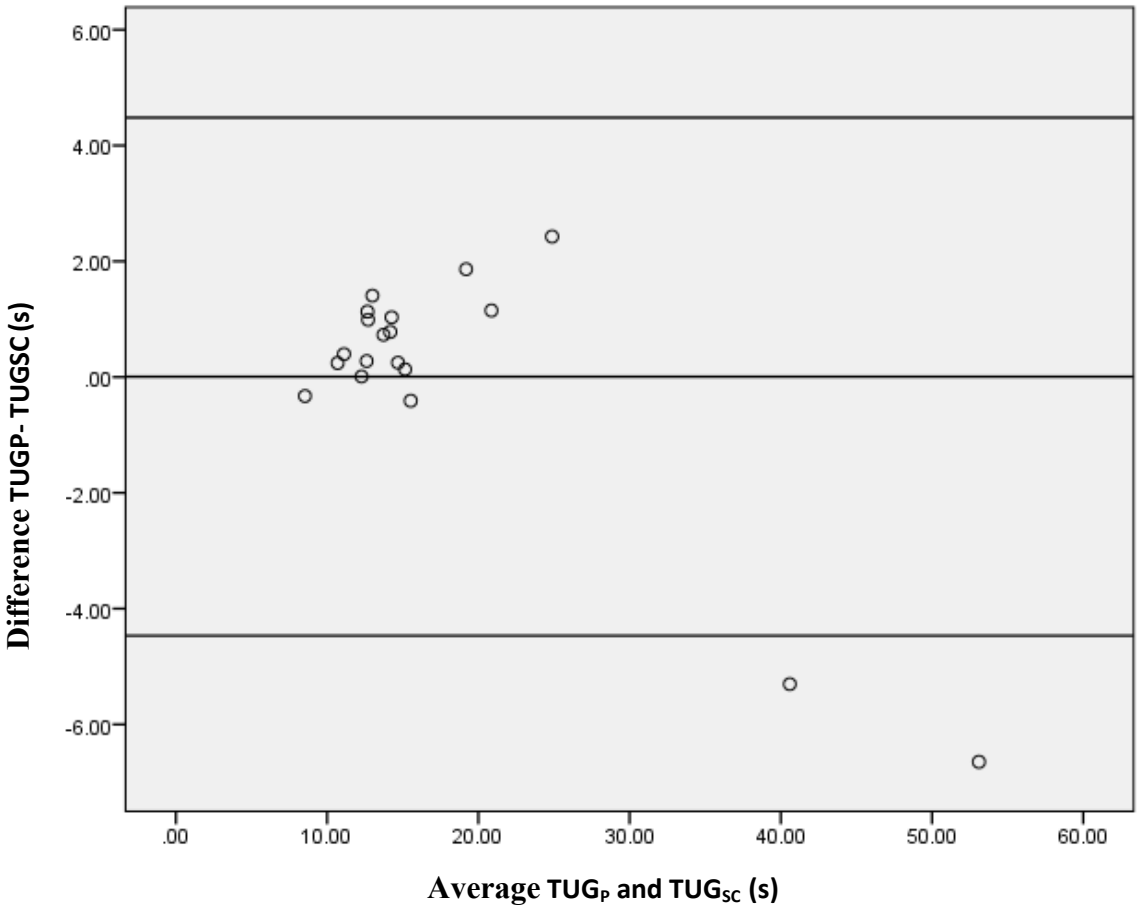


FIGURE H-1: SCATTER PLOT WITH LIMITS OF AGREEMENT SHOWN FOR ALL DATA -RAW, WITHOUT LOG TRANSFORMATION (N=19)

Limits of agreement:

Upper = 4.48

Lower = -4.47

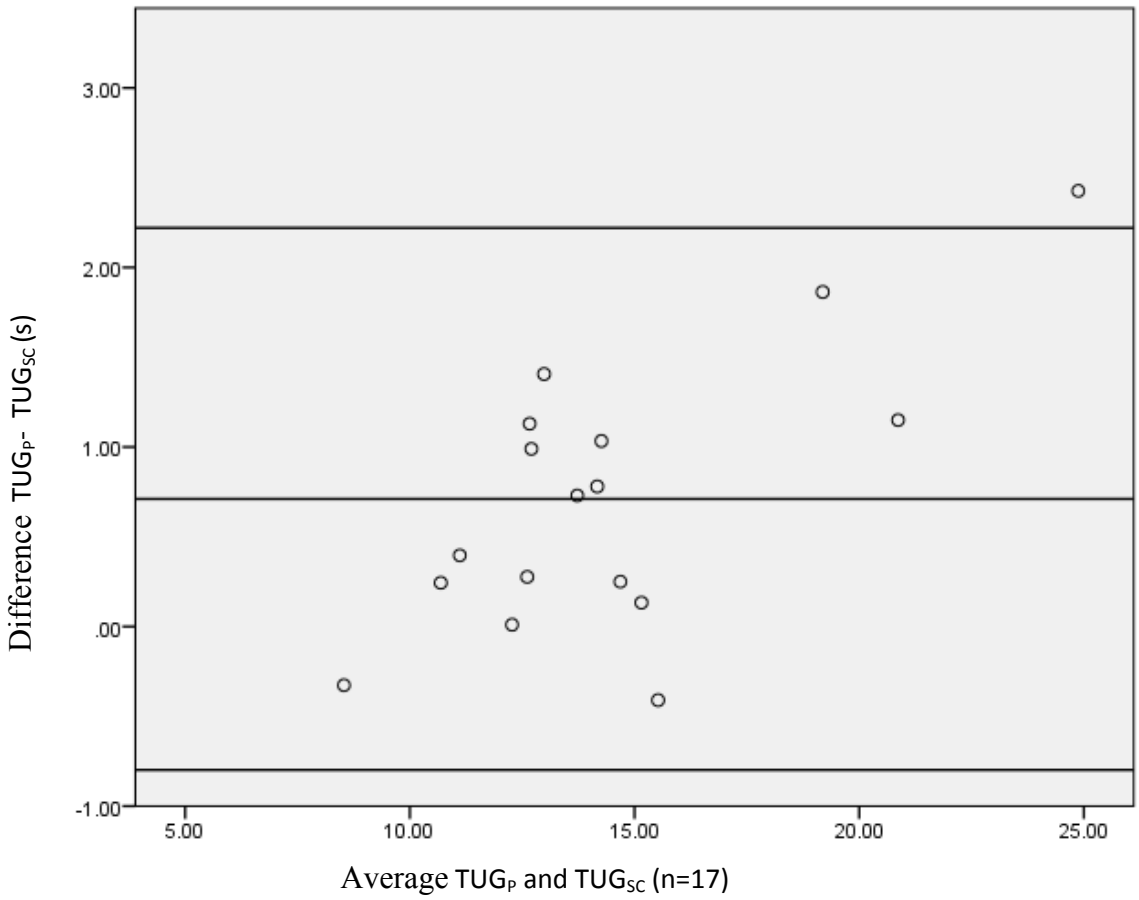


FIGURE H-2: SCATTER PLOT OF RAW DATA LIMITS OF AGREEMENT WITHOUT OUTLIERS (N=17).

Limits of agreement =.7108±1.51556

Upper= 2.23

Lower= -0.80

Statistics

		TUGPTUGSCdiff	nooutTUGPTUGSCdiff
N	Valid	19	17
	Missing	53	55
Mean		.0070	.7108
Std. Deviation		2.23705	.75778

FIGURE H-3: SPSS TABLE DEPICTING MEAN AND STANDARD SEVIATION FOR DIFFERENCE DCORES BETWEEN TUG_P AND TUG_{SC} WITH AND WITHOUT OUTLIERS NO TRANSFORMATION (N=19).

Limits of agreement = 0.00±4.4741

Upper 4.4811

Lower -4.4671

(n=17) no transformation

Limits of agreement = .7108±1.1556

Upper =2.22636

Lower= -0.80476

Statistics

		lognnouttugptugscdiff	LOgTUGPDiffTUGSCall
N	Valid	17	19
	Missing	55	53
Mean		.0440	.0259
Std. Deviation		.04385	.06823

FIGURE H-4: SPSS TABLE DEPICTING LOGN MEAN AND STANDARD DEVIATION

Limits of agreement with data logn transformed (N=19)

Limits of agreement= 0.0259± 0.16236

Upper =0.16236

Lower= -0.11056

Limits of agreement with data logn transformed (n=17)

Limits of agreement=0.0440± 0 .0877

Upper =0.1317

Lower= -0.0437

Statistics

		reverselogdiifnoout	reverselogdiifall
N	Valid	17	19
	Missing	55	53
Mean		1.0460	1.0285
Std. Deviation		.04561	.06779

FIGURE H-5: SPSS OUTPUT DEPICTING LOG REVERSED MEAN AND STANDARD DEVIATION OF MEAN DIFFERENCE DATA

All participants (N=19)

Limits of agreement 1.0285 ±0.13558

Lower=0.89292

Upper= 1.16408

No outliers (n=17)

Limits of agreement1.0460±.09122

Lower=.95478

Upper = 1.13722

References

1. Hollway, D., E.M. Earl, A. Murphy, M, Mackay-Lyons, E. MAcDonald, *Progressive Exercise to Address Impaired Balance and Mobility of Older Adults Referred for Home Care Physiotherapy Services: Is it Beneficial to Target Vestibular Control and Lower Limb Muscle Strength?* 2008.
2. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **1**(8476): p. 307-10.
3. Portney, L.G. and M.P. Watkins, *Foundations of clinical research : applications to practice*. 3rd ed. 2009, Upper Saddle River, N.J.: Pearson/Prentice Hall. xix, 892 p.
4. Diekhoff, G., *Statistics for the social and behavioral sciences : univariate, bivariate, multivariate*. 1992, Dubuque, IA: Wm. C. Brown Publishers. xviii, 418 p.
5. Norkin, C.C. and D.J. White, *Measurement of joint motion : a guide to goniometry*. 3rd ed. 2003, Philadelphia: F.A. Davis. xvi, 404 p.
6. Rydwik, E., et al., *Psychometric Properties of Timed Up and Go in Elderly People: A Systematic Review.* Physical & Occupational Therapy in Geriatrics, 2011. **29**(2): p. 102-125.
7. Canada, C.H., *Healthy aging : prevention of unintentional injuries among seniors. 2002, [Ottawa]: Health Canada. 18, 20 p.*
8. Scott, V., et al., *A public health approach to fall prevention among older persons in Canada.* Clin Geriatr Med, 2010. **26**(4): p. 705-18.
9. Fortinsky, R.H., et al., *Extent of implementation of evidence-based fall prevention practices for older patients in home health care.* J Am Geriatr Soc, 2008. **56**(4): p. 737-43.

10. Tinetti, M.E., et al., *Risk factors for serious injury during falls by older persons in the community.* J Am Geriatr Soc, 1995. **43**(11): p. 1214-21.

11. King, M.B. and M.E. Tinetti, *Falls in community-dwelling older persons.* J Am Geriatr Soc, 1995. **43**(10): p. 1146-54.

12. Canada. Health Canada., *Healthy aging : prevention of unintentional injuries among seniors*. 2002, [Ottawa]: Health Canada. 18, 20 p.

13. Gillespie, L.D., et al., *Interventions for preventing falls in older people living in the community.* Cochrane Database Syst Rev, 2012. **9**: p. CD007146.

14. Public Health Agency of Canada., *Report on seniors' falls in Canada*. 2005, Public Health Agency of Canada: Ottawa, Ont. p. 1 electronic text (64 p.

15. Fletcher, P.C. and J.P. Hirdes, *Risk factors for falling among community-based seniors using home care services.* J Gerontol A Biol Sci Med Sci, 2002. **57**(8): p. M504-10.

16. Gillespie, L.D., et al., *Interventions for preventing falls in elderly people.* Cochrane Database Syst Rev, 2003(4): p. CD000340.

17. Kenny, R.M., et al., *Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons.* J Am Geriatr Soc, 2011. **59**(1): p. 148-57.

18. Huang, M., et al., *Performance of balance impaired elders on three balance tests under two visual conditions.* J Geriatr Phys Ther, 2006. **29**(1): p. 5-9.

19. Tinetti, M.E., M. Speechley, and S.F. Ginter, *Risk factors for falls among elderly persons living in the community.* N Engl J Med, 1988. **319**(26): p. 1701-7.

20. Rose, D.J. and S. Clark, *Can the control of bodily orientation be significantly improved in a group of older adults with a history of falls?* J Am Geriatr Soc, 2000. **48**(3): p. 275-82.

21. Hess, J.A. and M. Woollacott, *Effect of high-intensity strength-training on functional measures of balance ability in balance-impaired older adults.* J Manipulative Physiol Ther, 2005. **28**(8): p. 582-90.

22. Horak, F.B., *Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls?* Age Ageing, 2006. **35 Suppl 2**: p. ii7-ii11.

23. AccreditationCanada. *http://www.accreditation.ca/accreditation-programs/qmentum/required-organizational-practices/*. 2012,2013 [cited 2013 January 20, 2012 , June 16, 2013, September 16, 2013].

24. Whitney, J.C., S.R. Lord, and J.C. Close, *Streamlining assessment and intervention in a falls clinic using the Timed Up and Go Test and Physiological Profile Assessments.* Age Ageing, 2005. **34**(6): p. 567-71.

25. Gates, S., et al., *Multifactorial assessment and targeted intervention for preventing falls and injuries among older people in community and emergency care settings: systematic review and meta-analysis.* Bmj, 2008. **336**(7636): p. 130-3.

26. Bischoff, H.A., et al., *Identifying a cut-off point for normal mobility: a comparison of the timed 'up and go' test in community-dwelling and institutionalised elderly women.* Age Ageing, 2003. **32**(3): p. 315-20.

27. Arnold, C.M. and R.A. Faulkner, *The history of falls and the association of the timed up and go test to falls and near-falls in older adults with hip osteoarthritis.* BMC Geriatr, 2007. **7**: p. 17.

28. Kristensen, M.T., N.B. Foss, and H. Kehlet, *Timed "up & go" test as a predictor of falls within 6 months after hip fracture surgery.* Phys Ther, 2007. **87**(1): p. 24-30.

29. Morris, R., et al., *A comparison of different balance tests in the prediction of falls in older women with vertebral fractures: a cohort study.* Age Ageing, 2007. **36**(1): p. 78-83.

30. Nordin, E., et al., *Prognostic validity of the Timed Up-and-Go test, a modified Get-Up-and-Go test, staff's global judgement and fall history in evaluating fall risk in residential care facilities.* Age Ageing, 2008. **37**(4): p. 442-8.

31. Gunter, K.B., et al., *Functional mobility discriminates nonfallers from one-time and frequent fallers.* J Gerontol A Biol Sci Med Sci, 2000. **55**(11): p. M672-6.

32. Rockwood, K., et al., *Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging.* J Gerontol A Biol Sci Med Sci, 2000. **55**(2): p. M70-3.

33. Dite, W. and V.A. Temple, *A clinical test of stepping and change of direction to identify multiple falling older adults.* Arch Phys Med Rehabil, 2002. **83**(11): p. 1566-71.

34. Steffen, T.M., T.A. Hacker, and L. Mollinger, *Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds.* Phys Ther, 2002. **82**(2): p. 128-37.

35. Boulgarides, L.K., et al., *Use of clinical and impairment-based tests to predict falls by community-dwelling older adults.* Phys Ther, 2003. **83**(4): p. 328-39.

36. Ries, J.D., et al., *Test-retest reliability and minimal detectable change scores for the timed "up & go" test, the six-minute walk test, and gait speed in people with Alzheimer disease.* Phys Ther, 2009. **89**(6): p. 569-79.

37. Shumway-Cook, A., S. Brauer, and M. Woollacott, *Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test.* Phys Ther, 2000. **80**(9): p. 896-903.

38. Steffen, T. and M. Seney, *Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health survey, and the unified*

Parkinson disease rating scale in people with parkinsonism. Phys Ther, 2008. **88**(6): p. 733-46.

39. Lin, M.R., et al., *Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people.* J Am Geriatr Soc, 2004. **52**(8): p. 1343-8.

40. Viccaro, L.J., S. Perera, and S.A. Studenski, *Is timed up and go better than gait speed in predicting health, function, and falls in older adults?* J Am Geriatr Soc, 2011. **59**(5): p. 887-92.

41. Okumiya, K., et al., *The timed "Up & Go" test and manual button score are useful predictors of functional decline in basic and instrumental ADL in community-dwelling older people.* J Am Geriatr Soc, 1999. **47**(4): p. 497-8.

42. Shumway-Cook, A. and M.H. Woollacott, *Motor control : translating research into clinical practice*. 3rd ed. 2007, Philadelphia, PA: Lippincott Williams & Wilkins. x, 612 p.

43. Granacher, U., et al., *Age-related effects on postural control under multi-task conditions.* Gerontology, 2011. **57**(3): p. 247-55.

44. Rogers, M.E., et al., *Methods to assess and improve the physical parameters associated with fall risk in older adults.* Prev Med, 2003. **36**(3): p. 255-64.

45. Axer, H., et al., *Falls and gait disorders in geriatric neurology.* Clin Neurol Neurosurg, 2010. **112**(4): p. 265-74.

46. Kristensen, M.T., et al., *Timed up & go test score in patients with hip fracture is related to the type of walking aid.* Arch Phys Med Rehabil, 2009. **90**(10): p. 1760-5.

47. Podsiadlo, D. and S. Richardson, *The timed "Up & Go": a test of basic functional mobility for frail elderly persons.* J Am Geriatr Soc, 1991. **39**(2): p. 142-8.

48. Large, J., et al., *Using the timed up and go test to stratify elderly inpatients at risk of falls.* Clin Rehabil, 2006. **20**(5): p. 421-8.

49. Nordin, E., E. Rosendahl, and L. Lundin-Olsson, *Timed "Up & Go" test: reliability in older people dependent in activities of daily living--focus on cognitive state.* Phys Ther, 2006. **86**(5): p. 646-55.

50. Kwan, M.M., et al., *Sensorimotor function, balance abilities and pain influence Timed Up and Go performance in older community-living people.* Aging Clin Exp Res, 2011. **23**(3): p. 196-201.

51. Wrisley, D.M. and N.A. Kumar, *Functional gait assessment: concurrent, discriminative, and predictive validity in community-dwelling older adults.* Phys Ther, 2010. **90**(5): p. 761-73.

52. Huang, S.L., et al., *Minimal detectable change of the timed "up & go" test and the dynamic gait index in people with Parkinson disease.* Phys Ther, 2011. **91**(1): p. 114-21.

53. Flansbjer, U.B., et al., *Reliability of gait performance tests in men and women with hemiparesis after stroke.* J Rehabil Med, 2005. **37**(2): p. 75-82.

54. Dal Bello-Haas, V., et al., *Psychometric Properties of Activity, Self-Efficacy, and Quality-of-Life Measures in Individuals with Parkinson Disease.* Physiother Can, 2011. **63**(1): p. 47-57.

55. Portney, L.G. and M.P. Watkins, *Foundations of clinical research : applications to practice*. 2nd ed. 2000, Upper Saddle River, NJ: Prentice Hall. xiv, 768 p.

56. Stratford, P.W., *On "test-retest reliability and minimal detectable change on balance..." Steffen T, Seney M. Phys Ther. 2008;88:733-746.* Phys Ther, 2008. **88**(7): p. 888-90; author reply 891.

57. Eliasziw, M., et al., *Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example.* Phys Ther, 1994. **74**(8): p. 777-88.

58. Shrout, P.E. and J.L. Fleiss, *Intraclass correlations: uses in assessing rater reliability.* Psychol Bull, 1979. **86**(2): p. 420-8.

59. Green, S.B. and N.J. Salkind, *Using SPSS for Windows and Macintosh : analyzing and understanding data*. 6th ed. 2011, Boston: Prentice Hall. xv, 459 p.

60. Bland, J.M. and D.G. Altman, *Measuring agreement in method comparison studies.* Stat Methods Med Res, 1999. **8**(2): p. 135-60.

61. Myles, P.S. and J. Cui, *Using the Bland-Altman method to measure agreement with repeated measures.* Br J Anaesth, 2007. **99**(3): p. 309-11.

62. Hamilton, C. and J. Stamey, *Using Bland-Altman to assess agreement between two medical devices--don't forget the confidence intervals!* J Clin Monit Comput, 2007. **21**(6): p. 331-3.

63. Powell, L.E. and A.M. Myers, *The Activities-specific Balance Confidence (ABC) Scale.* J Gerontol A Biol Sci Med Sci, 1995. **50a**(1): p. M28-34.

64. Hatch, J., K.M. Gill-Body, and L.G. Portney, *Determinants of balance confidence in community-dwelling elderly people.* Phys Ther, 2003. **83**(12): p. 1072-9.

65. Lajoie, Y., A. Girard, and M. Guay, *Comparison of the reaction time, the Berg Scale and the ABC in non-fallers and fallers.* Arch Gerontol Geriatr, 2002. **35**(3): p. 215-25.

66. Shumway-Cook, A. and F.B. Horak, *Assessing the influence of sensory interaction of balance. Suggestion from the field.* Phys Ther, 1986. **66**(10): p. 1548-50.

67. Wrisley, D.M. and S.L. Whitney, *The effect of foot position on the modified clinical test of sensory interaction and balance.* Arch Phys Med Rehabil, 2004. **85**(2): p. 335-8.

68. Whitney, S.L. and D.M. Wrisley, *The influence of footwear on timed balance scores of the modified clinical test of sensory interaction and balance.* Arch Phys Med Rehabil, 2004. **85**(3): p. 439-43.

69. Weber, P.C. and S.P. Cass, *Clinical assessment of postural stability.* Am J Otol, 1993. **14**(6): p. 566-9.

70. Cohen, H., C.A. Blatchly, and L.L. Gombash, *A study of the clinical test of sensory interaction and balance.* Phys Ther, 1993. **73**(6): p. 346-51; discussion 351-4.

71. Murray, K.J., et al., *A pilot study of falls risk and vestibular dysfunction in older fallers presenting to hospital emergency departments.* Disabil Rehabil, 2005. **27**(9): p. 499-506.

72. Stark, T., et al., *Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review.* Pm r, 2011. **3**(5): p. 472-9.

73. Wang, C.Y., S.L. Olson, and E.J. Protas, *Test-retest strength reliability: hand-held dynamometry in community-dwelling elderly fallers.* Arch Phys Med Rehabil, 2002. **83**(6): p. 811-5.

74. Martin, H.J., et al., *Is hand-held dynamometry useful for the measurement of quadriceps strength in older people? A comparison with the gold standard Bodex dynamometry.* Gerontology, 2006. **52**(3): p. 154-9.

75. Yueh, B., et al., *Screening and Management of Adult Hearing Loss in Primary Care: Scientific Review.* JAMA, 2003. **289**(15): p. 1976-1985.

76. Lichtenstein, M.J., F.H. Bess, and S.A. Logan, *Validation of screening tools for identifying hearing-impaired elderly in primary care.* JAMA, 1988. **259**(19): p. 2875-8.

77. Ventry, I.M. and B.E. Weinstein, *Identification of elderly people with hearing problems.* ASHA, 1983. **25**(7): p. 37-42.

78. Yueh, B., et al., *Screening and management of adult hearing loss in primary care: scientific review.* Jama, 2003. **289**(15): p. 1976-85.

79. Folstein, M.F., S.E. Folstein, and P.R. McHugh, *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.* J Psychiatr Res, 1975. **12**(3): p. 189-98.

80. Kurlowicz, L. and M. Wallace, *The Mini-Mental State Examination (MMSE).* J Gerontol Nurs, 1999. **25**(5): p. 8-9.

81. Andrea Vertesi, J.A.L., D. William Molloy, MB, Brett Sanderson, Irene Tuttle, Laura Pokoradi, Elaine Principi, *Standardized Mini-Mental State Examination Use and interpretation.* Canadian Family Physician October 2001 vol. 47 no. 10 2018-2023 2002.

82. Association, C.H., *Canadian Healthcare Associ. Home Care in Canada: From the Margins to the Mainstream. Ottawa: Author.* 2009, Canadian Healthcare Association.

83. Canadian Institute for Health Information. and Canadian Electronic Library (Firm), *Health care in Canada, 2011 a focus on seniors and aging*. 2011, Canadian Institute for Health Information,: Ottawa, Ont. p. 1 electronic text (xi, 152 p.).

84. Rotermann, M.H.a.M., *Seniors' use of and unmet needs for home care, 2009*. 2009, Statitics Canada.

85. Hunter, K.F., et al., *Lower urinary tract symptoms and falls risk among older women receiving home support: a prospective cohort study.* BMC Geriatr, 2013. **13**: p. 46.

86. Fujiyama, H., et al., *Age-related differences in corticomotor excitability and inhibitory processes during a visuomotor RT task.* J Cogn Neurosci, 2012. **24**(5): p. 1253-63.

87. Kandel, E.R., J.H. Schwartz, and T.M. Jessell, *Principles of neural science*. 4th ed. 2000, New York ; Toronto: McGraw-Hill, Health Professions Division. xli, 1414 p.

88. Marsden, C.D., et al., *Feedback control of voluntary movements in man.* Electroencephalogr Clin Neurophysiol Suppl, 1978(34): p. 507-10.

89. Guyatt, G.H., et al., *The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure.* Can Med Assoc J, 1985. **132**(8): p. 919-23.

90. Bohannon, R.W., *Reference values for the timed up and go test: a descriptive meta-analysis.* J Geriatr Phys Ther, 2006. **29**(2): p. 64-8.

91. Pondal, M. and T. del Ser, *Normative data and determinants for the timed "up and go" test in a population-based sample of elderly individuals without gait disturbances.* J Geriatr Phys Ther, 2008. **31**(2): p. 57-63.

92. Hollway, D., Earl, E.M., Murphy, A., MacKay-Lyons, M., and MacDonald, E., *"Progressive Exercise for Balance Impaired Older Adults Referred for Home Care Physiotherapy: Is It Beneficial to Target Vestibular Control and Lower Limb Muscle Strength?"* Physiotherapy Canada, 2010. **Physiotherapy Canada 2010, 62(S1):28**.

93. Hollway, D., E.M. Earl, A. Murphy, M, Mackay-Lyons, E. MAcDonald, *Reference# 2008-1704:Progressive Exercise to Address Impaired Balance and Mobility of Older Adults Referred for Home Care Physiotherapy Services*. 2008, Dalhousie University Ethics Board.

94. http://web.missouri.edu/~proste/tool/Activities-specific-Balance-Confidence-Scale.rtf. October 9, 2007].