MEASURING PERCEIVED CHANGE IN MOBILITY AND BALANCE IN OLDER ADULTS: A MIXED-METHODS APPROACH

by

Elizabeth S. Hile

BS, Biochemistry, State University of New York at Geneseo, 1994

MPT, University of Pittsburgh, 1996

MS, Neuromuscular Physical Therapy, University of Pittsburgh, 1999

Submitted to the Graduate Faculty of

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This dissertation was presented

by

Elizabeth S. Hile

It was defended on

July 13, 2010

and approved by

Subashan Perera, PhD
Associate Professor of Medicine, Geriatrics
School of Medicine, University of Pittsburgh

Susan Zickmund, PhD
Assistant Professor of Medicine, General Internal Medicine
School of Medicine, University of Pittsburgh

Jennifer Brach, PhD, PT
Assistant Professor, Physical Therapy
School of Health and Rehabilitation Sciences, University of Pittsburgh

Committee co-Chair:
Stephanie Studenski, MD, MPH
Professor of Medicine, Geriatrics
School of Medicine, University of Pittsburgh

Committee co-Chair/ Dissertation Advisor:
Jessie VanSwearingen, PhD, PT
Associate Professor, Physical Therapy
School of Health and Rehabilitation Sciences, University of Pittsburgh

MEASURING PERCEIVED CHANGE IN MOBILITY AND BALANCE IN OLDER ADULTS: A MIXED-METHODS APPROACH

Elizabeth S. Hile, PhD

University of Pittsburgh, 2010

Background/ Purpose: A priority in healthcare for older adults is to detect declines in mobility and balance before falls occur with potential consequences of morbidity, disability, even mortality. Self-report and performance measures have varying degrees of respondent and administrative burden. We investigated the role of a single global self-rating in the detection of mobility decline. Change by repeated self-rated mobility state was compared to transition rating of self-perceived change over six months. Additionally, repeated state and transition reported change were compared with performance change. We anticipated discordance, and explored the support for potential theories to explain discordance.

Methods: Using a prospective, exploratory, observational cohort study with mixed-methods analysis, we focused on the natural history of age-related mobility change. Community-dwelling older adults provided state and transition global ratings of mobility and balance over six months, and completed questionnaires and performance tests of physical function. A subset of the older adults participated in semi-structured interviews to identify themes for domain definitions of mobility and balance, and the timeframe and frames of reference used for state ratings. Analyses included correlations for agreement between measures, and comparisons of means to investigate groups established based on discordance.

Results: Participants, n=104, had data at two consecutive time points, and 33 participated in interviews. Domain definitions and state timeframes varied. Two main frames of reference were identified. The

serial state and transition based rating of change were discordant, as were self-rated and performance change. Nearly 75 percent of those with gait speed decline of at least 0.10 m/s reported worsening by transition, and about 25 percent selected a lower state rating.

Conclusions: Transition ratings appear more sensitive than serial state ratings for detection of decline in gait speed, while decline by serial state rating may be more specific to larger performance changes. Self-ratings appear to communicate valuable information about mobility and balance not available from other measures, and we recommend an expanded use of open-ended questions in research and clinical practice.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

Older adults are at risk for declines in mobility and balance, and these declines can lead to morbidity, institutionalization, and even mortality. [1-7] Detecting subtle changes, particularly decline, in mobility function could be the key to preventing more serious consequences.[8, 9] Once detected, compensatory and preventative strategies can be implemented, whether the use of an assistive device, modifications to the home environment, or education in other reduction of fall risk.[10-12] Rehabilitative strategies aimed at reversing mobility deficits have also proven effective at reducing fall risk,[13, 14] and newer rehabilitative approaches are under investigation.[15, 16] The challenge lies in identifying the decline in its early stages, when it may not be obvious by watching an older adult walk into a healthcare clinic for a routine check-up.[17-19]

One method of detecting this change is by asking the individual directly. Evidence supports the validity of self-report for predicting subsequent changes in function, sometimes even before performance measures can detect the change. [20] However, just as respondent burden cannot be ignored in aging research,[21, 22] data suggests that respondent burden associated with self-report measures is greater in older adults.[23] Routine clinic use of comprehensive self-report questionnaires covering a range of activities and the reasons for any non-participation could be limited by both respondent and administrative burden,[24] and brief performance measures may be preferred.[8, 25, 26] A single question that would be useful as a preliminary screening tool to determine the need for administration of longer, more detailed self-report and performance-based measures would be invaluable in the clinical geriatric healthcare setting.[27]

Global questions (e.g. 'In general, how would you rate your *insert domain*?'), typically with a 5- or 7-item response scale (e.g. 'poor' to 'excellent') are used to obtain self-ratings for domains like health as part of larger, well-validated questionnaires including the Medical Outcomes Study 36-Item Short Form Survey Instrument (SF-36).[27-30] By themselves, these current state questions are lower burden than longer questionnaires and performance measures, and, if repeated at specific time intervals, they may have value for detecting change in a domain. Alternatively, transition-based global questions directly elicit ratings of perceived change from the individual using a Likert-style response scale, with various options of 'better' or 'worse' centered around an option of 'no change.'[31, 32] Such transition questions are used in anchor-based research methods for determining the minimal clinically important difference (MCID) of a measure,[33-36] and the most basic of questions (e.g. 'Are your symptoms better or worse since your last visit?') are used informally by clinicians to determine response to interventions.[32, 37] Transition questions require older adults to recall a previous time point for comparison, but memory is flawed in people of all ages,[38-40] and decline that occurs naturally may be too gradual to detect.[41, 42] As a result, some authors have recommended serial use of state measures over transition ratings to detect change, although the 'state' measures used are often longer questionnaires instead of a single global self-rating.[43-45] We hypothesize that serial use of a global state self-rating of mobility may be more effective for detecting decline in this domain than asking the older adult directly about self-perceived change.

We seek to use a mixed-methods approach to investigate self-ratings provided by community-dwelling older adults in response to both state and transition-based global questions for the domains of mobility and balance. In particular, we are interested in changes in state ratings over time, and the comparison of changes in successive or 'serial' state ratings to a more direct rating of self-perceived change measured using transition-based questions. Finally, we will examine the relationship between change in self-ratings obtained using each technique, and change in physical performance measures of mobility and balance over the same period. Ultimately, we hope to determine if either or both of these global self-ratings could

be used to detect change, especially decline, in a community-dwelling older adult cohort. We will focus on change occurring naturally, with the passage of time, rather than in response to an intervention.

1.2 AIMS AND HYPOTHESES

1.2.1. Specific Aim 1: We aim to determine and compare test-retest reliability of both state and transition-based global self-report questions for the domains of mobility and balance. Reliability will be determined in a community-dwelling older adult cohort using percent agreement and non-parametric correlations.

1.2.1.1 Hypothesis Aim 1: Both state and transition-based global ratings will demonstrate moderate test-retest reliability in a community-dwelling older adult cohort, with similar results for the domains of mobility and balance.

1.2.2. Specific Aims 2: We aim to describe the level of agreement or 'concordance' between transition- and state-based approaches to measuring change in self-ratings over a six-month period. This will be described for the domains of both mobility and balance.

1.2.2.1. Specific Aim 2A: We will use comparisons of means and distributions to determine whether concordance in self-rated change by transition and repeated state approaches can be explained by individual characteristics such as age, mood, cognitive test performance, views on aging, comorbid health conditions, and functional status.

1.2.3. Specific Aims 3: A mixed-methods approach with semi-quantitative interviews will be used to explore how community-dwelling older adults answer both current state and six-month transition-based global questions to self-rate mobility and balance.

1.2.3.1. Specific Aim 3A: We will use a qualitative approach to describe the meaning of the domains 'mobility' and 'balance' in this sample.

1.2.3.2. Specific Aim 3B: We will identify the comparison(s) drawn or 'frame(s) of reference' used most frequently by community-dwelling older adults when rating their own mobility and balance. Using means comparisons, we will determine whether the frames of reference used when self-rating global mobility and balance can be explained by individual characteristics including age, gender, comorbidity, views on aging, and functional status.

1.2.4. Specific Aims 4: We aim to quantify the relationship between change measured using self-ratings of global mobility and balance and measured change in performance-based tasks thought to represent the same domain.

1.2.4.1. Specific Aim 4A: We will use non-parametric correlations to determine whether the state- or transition-based approach to measuring change in global self-ratings relates more closely to measured change in performance-based tasks over the same six-month period. For the domain of mobility, we will compare self-rating change to change in performance using the measures of gait speed, Figure of 8 Walk,[46] and Timed 'Up and Go.'[47] For the domain of balance, we will compare self-rating change to performance change using the measures of gait speed, Figure of 8 Walk,[46] Timed 'Up and Go,'[47] and timed unilateral stance.

1.2.4.2. Specific Aim 4B: Contingency tables for change in self-ratings for the domain of mobility and change in mobility performance (measured using gait speed) will be examined to determine which

method of measuring self-rating change (state- or transition-based) more closely agrees with performance change over the same period. Groups will be established based on concordance of change in self-ratings and performance, and between-group comparisons will be used to determine whether individual characteristics (age, cognition, functional status) differ based on the method of assessing self-rating change that more closely agrees with change in performance.

1.3 BACKGROUND

Why global self-ratings?

Evidence suggests that self-report, even in the form of a single global self-rating, can be used by older adults to communicate valuable predictive information. In one study, 70-79 year old adults self-rated health at the present time in response to a single global question with a 5-point response scale ranging from 'poor' to 'excellent .'[48] Individuals who rated their health as "poor" or "bad" were 19.56 times more likely to die in three years than those who rated their health as "excellent." Among relatively healthier individuals (defined as having no more than one chronic disease), the adjusted odds ratio for mortality with a "poor" or "bad" response leaped to 93.51. The authors concluded that self-rated heath may be even more meaningful in healthy individuals than in their more frail counterparts, and that the time between successive medical check-ups should be shortened for patients who rate their health as poor.[48] Findings from a number of other studies support the predictive validity of self-rated health for the outcomes of mortality[49-51] and functional decline.[52]

The global health state-based question described above, and a global transition question about change in health over the past year, are two of the questions included in the Medical Outcomes Study SF-36 health survey.[28] While many have studied global self-ratings of health,[29, 30, 49, 50, 52-54] it remains unclear what the domain of health means to individual older adults. Global health has been found to reflect problems with

physical health more than mental health, [55-57] but qualitative exploration into global state self- ratings for the domain of health has revealed greater disparity in the frame of reference used by adults when selecting a response.[53] Some may think of specific health problems, while others consider more general physical functioning, or even health behaviors, and the frame of reference used appears to vary with age,[53] which could mean that the considerations of the older adult differ from those of the healthcare providers who review these ratings and use them to direct care As with health, conceptualization of domains like falls and successful aging have been shown to vary between older adults, and to disagree with the definitions held by healthcare providers.[58, 59] The use of self-ratings for the domains of mobility and balance will also require qualitative exploration into the meanings of the domains and the frames of reference used when selecting a response.

For many purposes, evidence supports the use of more specific self-report measures with a wider variety of items.[60-63] For example, when measuring the effects of interventions in those with a certain disease, disease-specific questionnaires are more responsive and valid.[64-66] While disease-specific questionnaires are preferred in most situations, change in a disease-specific questionnaire after respiratory rehabilitation was shown to correlate only weakly (r=0.19 to 0.28) with performance change on the 6-minute walk test.[64] Additionally, compared to one or two global questions, longer, more specific questionnaires are time-consuming and can be more challenging to administer to older adults who may have hearing or vision deficits that interfere with standard modes of administration.[67]

Rather than those with a specific condition, we are surveying the larger population of older adults who are at risk for age-related decline in mobility and balance.[68-73] We hypothesize that global ratings of mobility and balance could actually be more responsive to subtle change in abilities than longer questionnaires with more specific questions. Because no further specification is provided for the global terms, older adults may interpret the broader domains to encompass any areas in which they are having mobility-related difficulty, even tasks not typically explored by more specific questionnaires. Screening mobility

is important because mobility disability (sometimes defined as difficulty walking ½ mile and climbing stairs) or even difficulty has been shown to precede and predict disability with activities of daily living. [74-76] At a minimum, changes in a global self-rating of mobility or balance over time may be a useful screening tool to determine if further assessment, and ultimately referral to a mobility specialist, is warranted.

Detecting change using global self-ratings

Healthcare providers and researchers are often most interested in change, whether it be change over time (as with the natural history of aging),[77] or change in response to a rehabilitative intervention.[44, 78, 79] Standardized measures are available to detect change in mobility and balance performance,[25, 47, 80] but perceived change in the same domains may not correspond with performance change. Even when performance-based outcome measures demonstrate substantial improvement, self-ratings may decline from baseline. Such discordance was found with changes in ratings of balance confidence after exercise interventions aimed at improving physical performance and fall risk.[78, 81] We hypothesize that the rehabilitative process may make some individuals more aware of their residual deficits, shifting the focus from any improvement to the deficits that remain. Based on Bandura's self-efficacy theory[82, 83] and other work,[84] the perception of any improvement is more likely to impact changes in activity and social participation than performance-based improvement. Perera and colleagues have used the global mobility change rating as one of several clinical anchors to determine the magnitude of meaningful change in performance measures. [54] Global ratings of health are sometimes used to measure perceived change in mobility, as they have been shown to relate to mobility and physical function. [56, 57] A recent trial of interventions to improve balance in older women with osteoporosis used global self-ratings of change in the domain of health, comparing these ratings to outcomes on performance measures of balance, but participants were not specifically asked to rate change in balance or mobility.[85] The work of Fischer and

colleagues to determine outcomes after arthritis interventions using global ratings of change for the domain of physical limitation (disability) is closest to the domain of mobility. [44]

Change in global self-ratings can be measured using two different methods, one a comparison of **current state ratings** over two timepoints (sometimes called the 'serial' technique),[44] and the other a **transition** rating of the perceived change in a global domain relative to the earlier timepoint (sometimes called the 'global rating of change'). [31, 32, 86] The best way to assess change is unknown, particularly in older adults. This population brings unique considerations to self-ratings, including memory loss[87, 88] and a fear of losing independence[89] that may lead them to censor perceived decline from healthcare providers. Neither technique for measuring self-rating change is without flaw,[32, 44] and it is possible that the optimal method varies with certain characteristics of the individual, such as level of function or views on aging.

Because they require patients to recall a prior state and then compare it to their current opinion of their status in that domain, transition questions are plagued by memory phenomena including 'transience' and 'recall bias' in people of all ages, [38, 39] but these may be compounded by memory loss in the older adult population.[87, 88] Even when memories are not distorted or forgotten, there is evidence that patients are heavily influenced by their current health state when they answer transition questions,[31, 32] and one author concluded that transition ratings may actually reflect current state rather than change.[90] These pheonomena may be problematic, particularly if geriatric practitioners wish to use self-rated global change to detect decline between office visits that are spaced out by six months or even one year.

Alternatively, use of global state ratings to measure change in a healthcare setting requires at least one prior contact with the individual, and a contact during which the same question was asked and the answer is now available for review. This is not the case during most initial evaluations, when a physician or therapist meets a patient for the first time and would like to establish whether the patient's function has recently declined and may be amenable to intervention.[17, 80, 91] Even when remote answers are available for comparison, 'response shift' literature suggests that changes in internal standards, values, or

conceptualization over time could confound the comparison with a more recent response, suggesting change in ability when only internal perceptions have changed.[92, 93] In the rehabilitation setting, this could explain some of the observations described previously, when individuals whose performance has clearly improved over the course of an intervention report no improvement, or even decline, in their status using self-ratings.[78, 81] By making them more aware of the extent of their initial deficits, and of any deficits that remain, participation in clinical or research rehabilitation protocols could 'shift' responses by changing internal standards.

We seek to determine which method of assessing self-rating change over time best corresponds with the natural history of mobility change in older adults. Published findings that focus on the communication of change in response to an intervention or after a discrete event[44, 94, 95] may not apply to more gradual, age-related change, and we are not aware of any published work that has answered this question using global self-ratings of mobility and balance.

Relationship between Self-Report and Performance

The relative contribution of self-report measures and performance-based tests to the assessment of health, mobility, and function in older adults has long been debated. [25, 96-101] In one study, both self-report (e.g. perceived ability to walk ¼ mile or to get up from a chair without use of the arms) and performance measures (e.g. 400-meter walk of endurance, narrow walk) designed to discriminate higher-functioning older adults were independent predictors of walking endurance.[101] However, correlations between the two types of measures were only modest (r=0.13-0.35), suggesting that each reflects different aspects of physical function.[101] Studenski and colleagues[25] studied almost 500 males age 65 and older in a primary care clinic environment, comparing gait speed and lower extremity performance on a battery of tests to self-rated health on a 5-point Likert scale and self-rated functional status using two different

questionnaires. They found the combination of self-report and performance to better predict decline in health over twelve months than either measure alone, although the physical report measures in this study independently predicted decline in function better than self-report measures.[25]

There is strong evidence for the value of self-report measures in the assessment of function and disability. For older women with gait speeds above 0.6 m/s, self-report of mobility provides additional predictive information about incident difficulty and disability not provided by walking speed alone, and it is suggested that the two be used in combination for greatest specificity.[20] These same authors identified a physiologic basis for self-reported function by demonstrating that physical performance, measured by walking speed, balance, strength and exercise tolerance, decreased in association with decreasing self-reported mobility function for the tasks of walking ½ mile and climbing ten steps.[20] After comparing results of self-report of physical function to performance-based measures of walking, stance, and chair rise in 221 community-dwelling older adults ranging in age from 60-102 years, Alexander and colleagues conclude that compared to other self-report items, "self-reported walking ability may be the best indicator of ADL and mobility performance in community-dwelling older adults, many of whom have ADL and mobility difficulty," and they suggest that it replace walking speed as an indicator of mobility function in large-scale public health assessments.[102] A study of mobility in mid-life to older adults with knee osteoarthritis resulted in only moderate (r=0.44) correlation between self-report and performance-based measures, and the authors concluded that performance results alone do not adequately reflect functional mobility. [103]

While some data may support a preference for one mode of testing over another,[8, 104] most authors advocate use of both self-report and performance-based measures clinically, depending on the goals of the assessment .[99, 102, 105, 106] Self-report measures in the arthritis population were found to relate most strongly to reported pain, while performance measures related to self-efficacy for mobility tasks.[107] Suggestions have been made not only to include both self-report and performance measures, but also to

include both disease-specific and more general health-related quality of life self-reports.[107, 108] Reuben and colleagues "urge caution in relying solely on one measure of physical function" to screen, diagnose, or monitor individual patients after finding only weak to moderate associations between multiple self-report measures and a single performance-based measure of function that they designed to measure the same construct of physical function. [100]

Both self-report and performance-based measures have faults. As Jette notes in a report validating his own self-report measure of functional disability, performance measures limit the scope of a functional assessment because observation is usually restricted to simpler functional tasks in order to minimize respondent burden and enhance feasibility for the clinician.[109] Even then, physical performance measures are not always acceptable to patients and research participants. Out of 182 people able and willing to fill out a 50-item ADL questionnaire, only 54% agreed to attempt at least one of 14 performance based tasks carried out in their own homes.[99] Although self-reported measures of function obtained via interview or self-administered questionnaires are easier, faster, and less expensive, Jette cautions that they must be highly structured in order to yield valid and reliable information.[109] Other authors have cited self-report measures for limited validity and reliability, including discrepancy between patients' perceptions and their actual performance abilities,[96] lack of responsiveness to change,[96] and limited ability to characterize higher levels of function.[105]

While standardization of administration is important for both self-report and performance-based measures, training multiple testers to administer a performance test is sometimes the bigger challenge, and healthcare professionals who use the measures to make clinical diagnoses and establish plans of care may vary widely in level of training. Even when properly administered, results of performance measures may lack generalizability to the participants' daily function.[97, 99] Most of these measures time, or require an observer to rate, performance on a limited set of tasks over a short period of time, and they are usually administered in an artificial environment not reflecting the environment in which the individual typically

performs functional tasks.[109] Additionally, performance on the day of testing may not reflect the individual's average daily performance on those same tasks, but the average performance may be reflected in self-report. This could help explain why self-report of abilities often correlates only weakly with performance on clinical measures,[96, 99-101] and why self-report has been shown to provide additional information that is independent of performance, as referenced above.[20, 25, 103]

Citing the propensity of published comparisons between self-report scales of basic and instrumental activities of daily living (BADL/IADL) and performance-based measures that differ in content, rating format, and scoring procedures, Myers et al. carried out one of the only studies to compare self-reported difficulty with IADL's to performance of those same tasks scored by a rater blinded to self-assessment results.[99] They found categorical observer ratings of task performance and self-ratings of difficulty more reliable than timed ratings over two weeks. Overall, they did not find performance-based measures to be more acceptable to patients, more reproducible, or more sensitive to change than a 50-item self-report IADL measure, and they, like many other authors,[102, 105, 106] conclude by arguing to combine self-report and performance tests 'rather than using one method to the exclusion of the other.' [99]

While the credibility of self-report measures in older adults could be questioned based on concerns with cognition, particularly memory, [87, 88] there is sufficient evidence to support use of such measures in this population. Dorevitch and colleagues[110] used the Barthel index to examine self-ratings of ability to perform basic activities of daily living in 150 elderly patients with mean age of 81 years (upper limit of 96 years) who attended a geriatric day hospital. When compared to therapists' direct observations of their performance, self-ratings were found to be more accurate than informant-ratings obtained by report from another care provider, who consistently underestimated capacity for BADL's, and self-ratings of independence predicted task performance most consistently.[110] Other studies have concluded that older adults either under-report [111] or over-report[112] their functional abilities, but our work aims to measure change in self-ratings over time, so the actual rating is not the primary variable of interest.

Although much work has been published comparing self-report of status to results of performance-based measures thought to reflect the same domain, much of this work has been cross-sectional rather than prospective.[96, 99, 100, 103] Ultimately, the debate about relative contributions of the two methods for measuring functional ability rages on, and the role that each should play in drawing conclusions in both clinical and research settings is unclear. Additionally, no performance measures can be identified from the literature as 'best' in reflecting the domains of balance and mobility as defined by community-dwelling older adults.

Discordance between change in self-report and performance-based change

As described previously, agreement expressed as the degree of correlation between self-report and performance on measures thought to represent the same domain is often modest at best.[96, 99-101] In some cases, discordance or lack of agreement is clear, as when self-report directly opposes performance, demonstrated when 13.6% of people who could not complete an 8 foot walk reported that they could walk ½ mile without help.[105]

When global self-report of both current state and any change over the past six months (transition) are obtained along with performance at those two successive points in time, and all three are used to assess change over the time period of interest, discordance can occur between any two of these three representations of change. Discordance between the two methods for assessing change in self-ratings can occur if clients report change in one direction (e.g. 'worse') by transition, but communicate no change (i.e. select the same state), or even change in the opposite direction (e.g. improve from 'fair' to 'good'), over that same time period. An example of discordance between self-ratings and performance change would be measured improvement in gait speed over the same period for which an older adult reports that her mobility has worsened by either transition or serial state ratings.

Fischer and colleagues[44] explored the relationship between successive state and transition assessments of change in health status, specifically the domains of pain and disability. Individuals with arthritis were surveyed before and after one of three interventions, ranging from education to surgery. Pain was assessed using a visual-analog pain scale, and disability using the Health Assessment Questionnaire[113]. The difference in responses for each measure was compared to Likert-scale global ratings of change in pain and physical limitation obtained after intervention. Based on these comparisons, sensitivity to change in each domain was determined. Agreement between state and transition methods for determining change was reported as poor, averaging 29% agreement (range 0% to 50%), with correlations between the two methods of measuring change as low as 0.13 for disability ratings in the group with smallest change (education group). Consistent with the findings of previous groups who studied outcomes after interventions that included surgery,[94, 95, 114, 115] transition ratings were found to be more sensitive than change in serial measures. These retrospective transition ratings also correlated more strongly with patient satisfaction.[44] For the education group, transition ratings were found to correlate only weakly ($r<0.20$) with age, education level, disease duration, and expectations for improvement. One-week test-retest correlations were lower for retrospective ratings (0.58 for pain; 0.81 for disability) than for cross-sectional pain and disability scores (0.85 and 0.88 respectively). The authors conclude that 'retrospective measurements are not capricious or random but rather are detecting a particular perception of outcome.' They suggest that serial assessments are limited in relevance by 'temporal precision,' whereas retrospective transition ratings reflect 'a more composite appraisal over time,' but they ultimately argue to use retrospective ratings to supplement rather than replace serial measures in clinical practice.[44]

Any discordance in these two methods of measuring change in self-ratings, or between self-rating and performance-based change must be further investigated, as the discrepancy could threaten the value of global self-ratings in both research and clinical settings. While some may use discordance as ammunition to diminish the value of self-ratings in healthcare settings, it is more likely that inconsistencies suggest that we, as geriatric researchers and healthcare providers, do not understand how our older adult clients

perceive our questions, and ultimately how to interpret these data. By comparing two different methods of measuring change in self-ratings with change in performance measures for the domains of mobility and balance, we hope to shed light on situations or individual characteristics for which one method of determining change in self-rating may be preferred over the other. Unlike the work of Fischer and others,[44, 94, 95, 114, 115] we are investigating the natural history of change, which may be more difficult to detect than change in response to an intervention, so may bear greatest similarity to his education group.

Even if agreement with performance measures representing the same domains is only modest, it is self-ratings that are more likely to impact activity and social participation.[116, 117] Self-ratings reflect confidence and self-perceived ability in the domains of interest, and Bandura's theories suggest that people's performance is determined more by their beliefs about what they can do than by their actual capacity.[82, 83] Therefore, a decline in perceived mobility or balance may have significant implications for future performance decline if it leads people to decrease their level of activity and social participation.[118, 119]

Responses to global self-rating of change questions are often used as the 'gold standard' in anchor-based research analyses to establish the minimal clinically important difference (MCID) for other self-report or performance-based outcome measures,[33-36] but how meaningful are global self-ratings when used to detect mobility and balance change in the older adult population? While significant improvement after intervention or significant decline after an injury may be recognizable, can older adults recognize and report more gradual, somewhat subtle change in ability that occurs without an identifiable precipitating factor as they age?

Theories to explain discordance between the two methods of determining change in self-ratings, or between change in self-ratings and performance

A number of theories may explain discordance between the two methods of measuring change in self-ratings, or either method for determining self-rating change when compared to change in performance. These theories include:

1) Intentional Censoring

While older adults in rehabilitation settings have been found to under-report their functional abilities,[111] the community-dwelling population in which we are interested may be more likely to over-report their mobility and functional independence. This could reflect fear that reporting their deficits may lead a healthcare provider to push them to use an assistive device for ambulation, receive in-home services, or even make a recommendation for a move out of their own home.[89] We have termed this 'Intentional Censoring,' and it could lead to two different types of discordance. Between methods for determining change in self-ratings, fear of reporting decline may lead to transition reports of no change in mobility while participants may be unaware that a change in their state ratings over the same period reflects decline. It is more likely that discordance would occur between self-ratings and performance, with reports of no change or even improvement that are inconsistent with a decline in performance.

2) Response Shift

Response shift is a recognized phenomenon that describes the recalibration of internal standards of measurement.[93, 120] This could occur not only in response to interventions or participation in a research protocol, but also with the passage of time. Self-perceived personality changes in values have been described as women transition from mid-life to older adult, and in other personality domains (dominance, social recognition, and play) as men make the same transition.[121] We hypothesize that mobility-related definitions and frames of reference also change during this period, based on the evolution of physical activity, social participation, and functional ability for each individual over the lifespan. Younger adults

may view mobility as the activities of running and jumping, never considering transfers from bed or a chair, yet these may be the first activities to come to mind when an older adult selects a self-rating. Comparisons to others and participation in research and rehabilitation protocols may shift responses over a period of minutes. An older adult may select a mobility rating of 'fair' because he no longer plays tennis or jogs two miles a day, but if asked to consider only his ability to walk and stand up from a chair, his response may immediately shift to 'very good.' Alternatively, a wife may rate her balance as 'poor' because her husband can stand on one leg for 30 seconds in exercise class while she can hold the position for only five, but after visiting a friend in a nursing home and watching the residents struggle to walk down the hall without falling, her rating may shift to 'excellent.' Response shift would result in discordance due to a change in self-ratings, particularly serial state ratings, with no corresponding change in performance.

We are also using the term 'response shift' to encompass a phenomenon we are coining '**timeframe incongruency**.' This occurs when the older adult rates change by transition while unintentionally referring to a time period other than the one intended by the interviewer. In some cases, this may be due to an intervening event that bears greater significance to the individual than the previous timepoint of interest to the interviewer. For example, an older adult may be asked to rate change over the past six months, and the healthcare provider may observe the patient's mobility to have declined in that period. However, having undergone surgery three months ago, the individual may instead reference her status immediately after surgery, and subsequently indicate improvement by transition rating, even though she has not yet returned to her original level of function six months earlier. This would create discordance between the two methods of determining change in self-rating, and between self-rating and performance change. This phenomenon would be a form of response shift that is also related to recall bias as described below.

3) Recall Bias

This is a documented phenomenon by which people of all ages unintentionally distort the past,[38-40] in this case their abilities at some point in the past, while trying to draw a comparison with their current status in order to answer a transition question. Even without absolute loss of the memory, evidence suggests that adults give more consideration to the present than to the past when answering transition questions.[31, 32] Recall bias could create discordance between change in state- and transition-based self-ratings, and also between transition ratings and change in performance. In cases of recall bias, serial use of state ratings may be the preferred method for determining change in self-perceived ability.

4) Consistency Bias (and Central Tendency)

The bias toward avoid the end options of a Likert response scale and instead choosing the central option, often the option of 'neutral' or 'no change,' is well-established,[122] and we hypothesize that the potential for this bias exists for both our state and transition-based global questions. 'Good' is the central option of our 5-point state response scale, and 'about the same' is the center of our 7-level Likert response scale. The tendency toward the central response may be heightened by our use of 'about the same' rather than 'the same,' however, this was intended to make the selection of 'a little' better or worse even more meaningful in this study of change. Selection of the central transition response option, or avoidance of the end options, when performance change has occurred will create discordance between self-rating and performance change. This same discordance would result from consistency bias, a tendency for people to re-create the past for greater consistency with the present,[38] thus communicating no change by transition and repeated state ratings, even when change has occurred.

5) Gradual Change Theory

Aging is inevitable, and often gradual across the individual's lifetime.[42] In the absence of a specific event such as a fall or hospitalization due to illness, decline in mobility and balance may be slow and

insidious, therefore difficult to perceive. Consistency bias would result in the selection of the same state rating over successive timepoints, and a transition rating of the 'about the same,' even though performance shows subtle decline.

6) Low Expectations of Aging

This theory could be considered the opposite of the gradual change theory. With low expectations of aging,[123] an individual may determine that decline is an inevitable consequence of aging, and may report being 'worse' at almost every timepoint, even if no change in performance has occurred. Likewise, they may consistently choose the lowest rating provided on the state scale, thereby creating a floor effect that would lead to discordance between the two methods of assessing change in self-ratings, and between self-rating and performance-based change.

7) Pre-clinical Disability

There is evidence that older adults can recognize change, particularly decline, before it manifests as a change in their performance measures. Fried and colleagues argue for a "pre-clinical, intermediate phase of disablement which might develop in parallel with progression of underlying disease and precede and predict disability."[124] Difficulty with a task predicts eventual dependence,[125] however, it may be possible to identify older individuals at risk even for the onset of difficulty before it actually occurs. Self-report of the need to modify the method or change the frequency of task performance as a result of underlying health problems without actual difficulty in the task (defined as preclinical disability) has previously been associated with a level of functional performance that falls between that of individuals with actual difficulty and those reporting neither difficulty nor modification.[17] Fried and colleagues[124] validated this concept of 'preclinical disability' by demonstrating that community-dwelling women ages 70-80 years with preclinical disability were three to four times more likely to progress to disability than those in the high functioning group. The ability to detect preclinical mobility disability through a single global self-

rating would be invaluable as a predictor of future mobility-related disability, allowing individuals to be targeted for intervention in the clinic setting and for sample selection in research. In our work, preclinical disability could be reflected by a transition rating of 'a little worse' with no corresponding change in report of BADL function or performance change, and possibly the selection of the same state rating as in previous visits.

8) **Disparity in Domain Definitions and Frames of Reference**:

Evidence supports distinctions between patient and provider definitions for a number of domains other than mobility and balance.[53, 58, 59] With global questions even more than specific questionnaires, it is possible that older adults are not rating their balance or mobility using the same definitions that predominate among healthcare providers, or that coincide with the performance measures clinicians or researchers are using to represent the same domain. For example, a decline in mobility state rating and a transition report of worse mobility may reflect difficulty arising from a chair due to knee pain that is not present when walking. As a result, no corresponding decline in gait speed is found, resulting in discordance between change measured using self-ratings and performance.

In order to select a global self-rating option from a scale ranging from 'poor' to 'excellent,' we hypothesize that older adults draw a comparison or use a 'frame of reference.' This is based on anecdotal experience administering global questions of mobility and balance to older adults in both clinic and research settings. It is not uncommon for older adults to ask whether they should rate their status relative to their age, or on a spectrum of people of any age group, in which case they may indicate that the rating will be lower. Additionally, some comment that they are rating themselves in comparison to their own abilities at some point in the past, although generally the more remote past than the prior timepoint to which they are asked to compare their current status when responding to transition-based questions.

Through semi-quantitative exploration of the definitions and frames of reference used by older adults when rating their abilities for the global domains of mobility and balance, we may be able to discover approaches that will provide us with more meaningful data. For example, different stems may be provided as part of a global question in order to better standardize responses. These stems could assist older adults in defining the domain on which they are rating themselves, or the frame of reference by which they should draw a comparison. Alternatively, performance measures that are more representative of older adults' perceptions of these domains may be developed or identified.

Influence of cognition and depression in self-report:

The influence of cognition and mood, specifically depression, will be considered as covariates in our work. Cress and colleagues[96] found a modest association ($p < 0.01$) between gait speed and depression in both nursing home and community-dwelling participants, but found depression to affect a self-report measure (the physical subscore of the Sickness Impact Profile) even more than gait speed. Ultimately, depression was concluded to be an independent predictor of self-perceived function among nursing home residents.[96]

While discrepancies between self-report and performance-based measures have been attributed to cognition,[102] other authors have found the role of cognitive function to be only minor. Kempen and colleagues included subjects with MMSE scores as low as 17 out of a possible 30 in their studies of frail elderly,[98] even though scores less than 24 are generally accepted to indicate cognitive impairment[126] In their work, depression was associated more strongly than cognition with self-reported activities of daily living. Individuals with lower perceived physical competence and higher levels of depression reported lower ADL function than was suggested by their performance.[98]

2 AGREEMENT BETWEEN TWO METHODS OF DETERMINING CHANGE IN SELF-RATINGS OF MOBILITY AND BALANCE

2.1 INTRODUCTION

A common goal in rehabilitation of older adults is to improve balance and mobility, thereby decreasing the risk for falls. While we strive to quantify improvement using standardized measures of physical performance, perceived benefit from the perspective of the older adult may be equally important. In fact, self-perceived improvement may have greater impact on physical activity and community participation.[82-84] Even if it surpasses the threshold for clinical significance, performance-based improvement may bolster confidence and impact behavior only when perceived by the older adult. Alternatively, in cases where performance improvement is not detected by our tools, an increase in confidence could still overcome self-induced restriction for mobility-based activities.

While sometimes faulted for limitations in reliability, correlation with performance, and responsiveness to change,[96, 105] self-report continues to be valued in clinical and research settings.[20, 25, 99, 101-103, 106] Global self-ratings are used to determine outcomes of interventions,[31, 32, 44, 95, 127] and they serve as the gold standard in some statistical techniques for determining meaningful change in other measures.[33-36] Certainly, many self-report questions have high face validity, and it is perceived improvement that determines satisfaction with healthcare,[44] but there is also evidence to support the importance of self-ratings in capturing intangibles that we as healthcare providers have yet to quantify.[20, 48] An interesting study demonstrated the ability of older adults to judge their own health status. Individuals who rated their own health as 'poor' or 'bad' on a 5-point Likert scale were 19.56 times more likely to die in the next 3 years than those who rated themselves as 'excellent,' and the adjusted odds ratio jumped to 93.51 for those with no more than one chronic condition.[48] It appears that these individuals may recognize health deficits that we cannot quantify using measures such as comorbidity scales.

In addition to self-rating of current status, we are interested in how older adults perceive their own change in response to intervention, or even just the passage of time. The best way to assess perceived change is unknown. Repeated self-report of current status (**state rating**) can be compared at two time points (sometimes referred to as the '**serial technique**'),[44] or people can indicate how much they have changed relative to some point in the past using a change scale (**transition rating** or **retrospective technique**). [31, 32, 86]

Transition ratings require older adults to recall their status at a previous time point, which could be six months or even one year ago, and then compare to current status. Unfortunately, memories are transient, and often distorted in a well-documented phenomenon of recall bias, and these memory flaws[38, 39] may help to explain why transition ratings are heavily influenced by current state.[31, 32] However, serial assessments that use repeated ratings of current state have their own limitations. First, they require a prior response, which does not inform healthcare providers who may wish to measure self-perceived change (particularly decline) during the first encounter with a new patient. Additionally, when repeated, state ratings are affected by the phenomenon termed 'response shift' by which individuals experience a change in their internal standards.[93]

We are interested in measuring perceived change in mobility and balance in response to the passage of time. In this chapter, we report discordance between transition and state approaches to measuring change in balance and mobility over time in an older adult population, and discuss some theories about why the two methods may not agree. Any discordance identified between the two methods of determining change in self-ratings is concerning because some may use it to discredit the value of self-report in older adults. However, we believe that the discordance can be explained and informative in many cases, and that it actually highlights our need to better understand how to obtain, interpret and use these data.

2.2 METHODS

2.2.1 Participants: One hundred nineteen community-dwelling adults age 65 and older participated in an observational study to refine protocols and develop measures for the assessment of balance and mobility in older adults. Participants completed a battery of self-report and performance-based measures upon enrollment, and then approximately six and twelve months later. In addition, 46 participants returned about one week after baseline for a test-retest reliability session. The cohort represented a range of functional abilities, but all were independently ambulatory for at least household distances using an assistive device if needed, and could rise from a chair without assistance.

Participants were recruited from the University of Pittsburgh Claude D. Pepper Older American's Independence Center's (OAIC) Research Registry. The Registry is comprised of community-dwelling adults who agree to be contacted for participation in studies of balance and mobility. Effort was made to sample a wide range of functional mobility and balance by pre-screening over the phone for self-reported mobility and perceived balance. Individuals were excluded for unstable medical conditions (cancer with current treatment, angina), or for progressive or persistent neuromuscular conditions (stroke, Parkinson's) or pain that restrict movement.

2.2.2 Measures:

<u>Primary Measures</u>:

Self-Reported Mobility and Balance:

Global state and transition self-ratings of balance and mobility were obtained every 6 months for one year, along with a four hour battery of other self-report and performance-based measures felt to represent

primarily the domains of balance and mobility. Participants were not reminded of the self-report ratings they had given six months earlier, and global self-ratings were obtained before any performance-based tests of balance or mobility were administered in that session. The latter was done out of concern that first attempting high-level balance tests such as unsupported tandem and unilateral stance could influence balance and mobility self-ratings to reflect test performance rather than real-world functional ability. State questions were modeled after the Medical Outcomes Study SF-36 question of current health status[28] and simply asked participants to rate their 'level of *mobility* (alternatively *balance*) in general' with no further description of each domain or the specific period of time. Response options were provided on a 5-point scale of 1) Excellent, 2) Very Good, 3) Good, 4) Fair, and 5) Poor (Figure 2.1). Participants were then asked the transition question, 'In the past six months, has there been any change in your overall mobility (alternatively 'balance')? Response options were provided on a 7-point Likert-based scale of 3) Much Better, 2) Somewhat Better, 1) A Little Better, 0) About the Same, -1) A Little Worse, -2) Somewhat Worse, and -3) Much Worse.

Self-Report Measures Used To Examine Differences Between Groups:

15-item Geriatric Depression Scale (GDS):[128] The 15-item version of this basic screening measure for depression in older adults was used. For each item, participants are assigned one point if the 'Yes/No' response option selected corresponds to lower mood, for total scores ranging from 0 to 15. When used clinically, scores higher than 5 suggest depression and signal the need for further evaluation.

Digit Symbol Substitution Test:[129] This paper and pencil test from the Wechsler Adult Intelligence Scale measures perceptual processing speed. Participants are given a coding key with nine boxes. The upper half of each box contains a number from 1-9, and the bottom contains the symbol that corresponds to that number. The rest of the page is filled with rows of split boxes. The top of each box contains randomly ordered numbers (ranging from 1-9), and the coding key is used to fill in the bottom half of each box with the corresponding symbol. Participants fill in as many of the boxes as possible in 90 sec.

Performance declines with age.[130] When combined with a gait speed of less than 1.0 m/s, DSST scores of less than 27 have been associated with increased mortality and incident disability in an older adult cohort of mean age 70 years.[131]

Expectations Regarding Aging (ERA-12)[60]- A twelve-item self-report tool designed to measure expectations regarding aging in the domains of physical health, mental health, and cognitive function. Responses have demonstrated acceptable levels of reliability and construct validity.[60]

Activities-Specific Balance Confidence Scale (ABC):[132] Participants were asked to rate their confidence that they will not fall or lose their balance during the performance of each of 16 mobility-related activities. A response of 0 indicates no confidence and 100 is complete confidence. The total score is calculated as the average confidence reported for all items and is expressed as a percentage out of a possible 100. Scores of greater than 80% are common in high functioning, physically active older adults.[132, 133]

Survey of Basic and Instrumental Activities of Daily Living (BADL/ IADL Questionnaire): A self-report measure of basic and instrumental activities of daily living (BADL/ IADL) taken from the National Health Interview Survey (NHIS) was used to assess function for daily activities.[134] The BADL/ IADL scale measured self-reported ability to perform 16 tasks, including getting up from bed or chairs, walking, stair climbing, getting outside the home, and shopping. The total reported here indicates the number of tasks out of a possible 16 identified by the participant as either difficult or not performed for reasons related to health and physical function. Interrater and test-retest reliability ICC's above 0.9 have been reported in community-dwelling older adults.[25]

Survey of Activities and Fear of Falling in the Elderly (SAFFE):[135] This self-report measure assesses physical activity and fear of falling relative to eleven specific mobility-related activities, ranging from bending over and reaching overhead to walking for exercise and going out in slippery environments.

SAFFE activity is scored as the number of activities performed out of 11, and SAFFE restriction is calculated as the number of activities reported as performed less frequently now than five years ago. Both activity and restriction scores range from 0 to 11, with higher scores indicating greater activity or greater restriction, respectively. To determine SAFFE fear score, participants are asked how worried they are about falling with each activity performed. Total fear is calculated as the mean response across only the activities that are currently performed, ranging from 0 (not at all worried) to 3 (very worried).

Performance Measures Used To Examine Differences Between Groups:

Short Physical Performance Battery (SPPB):[105] The SPPB quantifies functional mobility using three categories: standing balance, walking, and repeated chair stands. Performance in each category is timed, and the time is converted to an ordinal score ranging from 0 (unable to perform) to 4 (best performance), used to calculate a summary score (maximum 12). The full protocol and training instructions can be downloaded from the NIH website. Test-retest and inter-rater reliability of each of the three categories range from 0.73 to 0.97,[104, 136, 137] and the reliability of the summary scale was established using internal consistency (Chronbach's alpha = 0.76). [105] Originally validated in over 5,000 adults age 71 and older, summary scores correlated strongly with self-reported ADL disability for tasks like walking across a room and transferring from bed to chair; and with self-reported ability to walk up and down steps or walk ½ mile without help.[75] Summary scores were also found to strongly predict mortality and nursing home admission.[105]

Gait Speed: Gait speed was measured in meters per second (m/s) using an instrumented walkway (GaitMat II) of approximately 6 meters in length. To avoid including acceleration and deceleration, recordings were taken only from the central 4 meters of the walkway. Up to four 'passes' were performed according to participant tolerance, and gait speed was averaged over all passes. Participants

used their own assistive device if needed and were closely spotted by research staff, but were not allowed physical assistance from another person. If GaitMat gait speed could not be determined (due to equipment malfunction or administration of study measures in the home), the mean time to complete the SPPB timed walk was converted to velocity and substituted for the GaitMat data. Because the SPPB 4-meter timed walk is measured from a standing start, a conversion factor was developed by our statistician using both the GaitMat and SPPB timed walk data available on all participants at baseline. Gait speed in older adults is highly predictive of adverse outcomes, including institutionalization and mortality.[138-141]

Timed Up and Go:[47] The time in seconds required to stand from a chair with armrests, walk 3 meters using any assistive device, turn, walk back, and sit down was measured and averaged over two trials. Normal, healthy elderly usually complete the task in under 10 seconds, while completion times greater than 30 seconds are consistent with functional dependence and use of an ambulatory assistive device.[47]

2.2.3 Statistical Analysis

PASW® Statistics version 18.0 (IBM SPSS, Inc., Chicago, Illinois) was used for all statistical analyses. Appropriate descriptive statistics (means, standard deviations, percentages) were used to summarize participant characteristics.

Change in State Rating from 6 to 12 months was expressed as the rating (From Excellent [1] to Poor [5]) reported at the 6 month time point minus the rating selected at 12 months. As a result, state change could range from -4 to +4, with negative values indicating selection of a 'worse' state rating at the latter timepoint, while positive values indicated that a more favorable rating was selected. **Transition Rating at 12 months** was expressed as the raw value (-3 to +3) selected by the participant at the 12 month timepoint. Negative values therefore indicate a self-reported worsening of status.

Each method of determining change in self-report (either using state or transition responses) was re-coded to create 3-level categorical measures reflecting whether the participant felt worse, the same, or better compared to six months earlier. Thus, change values were collapsed such that change of one or more levels of state ratings in the corresponding direction was considered better (e.g. 'poor' at 6 months to 'fair' at 12 months) or worse (e.g. 'excellent' at 6 months to 'very good' at 12 months). If the same state rating was selected at both timepoints (change in state rating = 0), the participant was coded as the same for that domain. For transition ratings, any response other than 'about the same' in the corresponding direction (i.e. +1 to +3 or -1 to -3) was considered better or worse, respectively. Transition ratings of 'about the same' or '0' were categorized as same. We constructed 3×3 contingency tables to summarize the results for change based on transition ratings and repeated state ratings. Individuals whose change fell into the same category (Better, Same, Worse) when state and transition-based methods were compared were categorized as **concordant** for that domain. **Discordance** was identified when state and transition-based change ratings fell into two different categories, such as an improvement in state rating from 6 to 12 months, but decline or no change by transition report at 12 months.

The discordant group was further divided according to the nature of the discordance. Some individuals reported being worse than six months ago by transition, but selected the same, or an even better, state rating than the state selected six months earlier. This was termed '**Discordant: Transition More Negative**.' Also categorized into this group were individuals who selected a better state rating than six months earlier, but reported being 'about the same' by transition. Therefore, members of this group gave transition ratings at the later timepoint that were either negative (indicating a worsening) or neutral (0, indicating status 'about the same'). Another dicordance group was identified and titled 'Transition More Positive.' This group reported being better than six months ago by transition, but selected the same, or an even worse, state rating than the rating they had selected six months earlier. Others in this group reported being 'about the same' by transition, but selected a better state rating. Members of the '**Discordant:**

Transition More Positive' group had therefore given transition ratings that were positive or neutral at the later timepoint.

Here we focus on data from 101 subjects at the 12-month period because of greater incidence of reported change in the past six months at this timepoint compared to reported change from baseline at the 6-month study visit. Those volunteering to enroll in a research study tend not to be in a state of flux with respect to balance and mobility limitations, and we felt that the 6 to 12- month period was more representative of a 6-month change in a community dwelling cohort.

For each time period of potential change, we used Kendall's tau and percent agreement to quantify the level of agreement between the difference in state ratings selected at successive study visits and the transition ratings given at the later visit. Kendall's tau and percent agreement were also used to compare 1-week repeated assessments for reliability. One-way ANOVA comparisons were used to examine differences in demographic characteristics, comorbidity, depression, cognition, aging expectations and mobility-related function between the three groups established based on the concordance or discordance between transition ratings and repeated state ratings, with discordance distinguished based on which method of determining change in self-report yielded a more negative result. Fisher's Least Significant Difference (LSD) post-hoc testing was used to examine pairwise comparisons for significance at the .05 level. Analyses were repeated using the baseline to 6-month period instead of the 6 to 12 month period to assess the robustness of the results.

2.3 RESULTS

Of the original 119 participants, 104 had data for at least two of the three timepoints of interest (baseline, 6-months and 12- months). The most common reason for missing data at six months was travel out of the area. By 12 months, home visits minimized missing data to only cases of serious illness or death. Baseline characteristics for these 104 participants are summarized in Table 2.1.

Test-Retest Reliability: For the purposes of 1-week test-retest reliability correlations, transition ratings were collapsed into three levels by direction (better, same, worse). The clinical meaning of the degrees of better and worse ('a little,' 'somewhat,' and 'much') are unknown, and we report here on discordance by direction without attention to magnitude, so reliability was determined relative to this approach to discordance . Kendall's tau indicated greater 1-week agreement for state ratings (.604 and .662 for mobility and balance, respectively) than transition ratings (.443 and .428). For a more complete picture of the distribution of the repeated measures, 1-week test-retest results are also presented in Table 2.2 as percent agreement by individual response level for the 5-level state ratings and the 7-level transition scale. Based on these results, 61% to 74% of older adults selected the same state or transition response option at both baseline visits, and 89% to 100% selected a response within one level of the response chosen one week earlier.

Relationship between Mobility and Balance: Agreement in baseline ratings between the domains of mobility and balance are shown in Table 2.3. Self-report ratings for the two domains at baseline were correlated at 0.576 (Kendall's tau) for state ratings, and .560 for transition when collapsed to three levels (better, same, worse), however percent agreement provides more detail for all levels of both ratings scales. About 50% of state ratings were in exact agreement for the two domains, and 90% fell within one level of agreement. Greater exact agreement was found for self-perceived change, as 74% selected the same transition rating for both mobility and balance at baseline, and 96% of ratings agreed within one level.

Change in Ratings with Time: Tables 2.4 (State) and 2.5 (Transition) give the distribution of self-ratings at all timepoints (baseline, 6 months, and 12 months) for the domains of mobility and balance. While some regression to the mean may have occurred in state ratings for the domain of balance between baseline and 6 months, the distribution for mobility appears stable across all timepoints. Still, change was seen in individual ratings, with about 33% of participants selecting a different state rating for mobility

over each of the two time periods. Selection of floor and ceiling state response options was infrequent, but did seem to vary by domain. About 10% of participants rated their mobility as 'excellent,' but the specific participants selecting this rating varied by timepoint. 'Excellent' was used less often to self-rate balance, selected by 2 to 7% of the cohort. Conversely, state ratings of 'poor' were slightly more frequent for balance (4 to 8%) than for mobility (1 to 3%).

Distribution of transition ratings shifts slightly for both domains at 12 months, with a greater percentage of participants reporting a worsening of status from 6 to 12 months than in the 0 to 6 month time period. This trend is also evident in Tables 2.6-2.9, contingency tables for comparison of the change in state rating over two successive study visits with the transition rating of perceived six-month change provided at the later study visit. Tables are provided for each domain (mobility and balance), and at each of the two time periods of potential change. Based on transition ratings, greater perceived change in mobility and balance occurred between 6 and 12 months than between 0 and 6 months, and worsening was reported more often than improvement. At 12 months, 41.6 and 33.7% of participants perceived worsening of their mobility and balance, respectively, compared to 29.4 and 24.5% at 6 months. Compared to transition reports of worsening mobility at both timepoints, only about half as many participants selected a lower mobility state rating at the later timepoint. However, participants were as or more likely to report being better by state rating than transition, and this was even more apparent for the domain of balance, where it was four to five times more common to select a better state rating than to report being better by transition.

This differential reporting of change in status between study visits when determined using two different approaches (state and transition-based), translated into discordance in about 50% (46% to 53%) of the cohort, and this finding held true for both timepoints and both domains, although the same individuals were not consistently discordant by domain and timepoint. For example, of the 48 who were concordant for change in mobility self-report from 0 to 6 months, 21 (44%) were **discordant** in mobility ratings at 12

months. Additionally, 38% of the mobility concordant group at 6 months were discordant in balance at that same time period. 'Better/ worse discordance,' defined as reporting change in one direction by transition, but then selecting a state rating indicative of change in the opposite direction, was rare. Thus, the majority of discordance fit these two scenarios: 1) Selection of a better or worse state rating than selected six months ago, but then reporting being 'about the same' as six months ago by transition, or 2) Reporting better or worse status by transition, but then selecting the same state rating as six months earlier. Correlations between the transition rating and the change in state rating from the previous study visit (six months earlier) ranged from .196 to .338. All correlations were significant at the .05 level. For each of the Tables 2.6 through 2.9, the three cells comprising the diagonal (shaded in grey) represent concordance, and the six surrounding cells represent discordance. Discordance was further broken down into two types as described in the methods section, based on whether the transition rating was more positive or more negative than any corresponding change in state rating. Participants represented by the yellow cells located to the upper right of the diagonal populate the 'Discordant: Transition More Positive' group, and those to the lower left (blue cells) are the 'Discordant: Transition More Negative' group. Refer to Table 2.10 for the discordance frequencies by type for both domains and time periods.

Results of between-group comparisons for all three discordance groups for the 6 to 12 month time period are presented in Table 2.11 for the domains of both mobility and balance. For the balance domain, only GDS scores were significantly different between groups. The mean number of comorbid conditions of the discordant: transition more negative group (3.5) was higher than the concordant and discordant: transition more positive groups by one comorbid condition. Unlike balance, a number of comparisons were significant for the domain of mobility. The discordant group with transition report more negative than any change in state appears to restrict two more activities than the concordant group and report difficulty with two additional daily activities. Their self-selected walking speed is 0.15 m/s slower (0.10 m/s is reported to be moderate meaningful change),[35, 142] TUG time is about 3 seconds longer, and their SPPB scores are lower by more than 1 point, the threshold for moderate meaningful change.[35, 142] These

mobility findings may explain the 16% lower balance confidence and greater fear of falling in this group. For most measures, performance of the 'discordant transition more positive' group fell between those of the concordant group and the 'discordant transition more negative' group. Results for the baseline to 6-month period were qualitatively similar for most measures between the concordant and discordant: transition more negative groups (data not shown), but the comparisons did not reach statistical significance for either domain.

Tables 2.12 (mobility) and 2.13 (balance) detail the distribution of state and transition ratings by 12-month discordance group. The distributions of state ratings appear qualitatively similar across discordance groups, but there may be a trend toward lower ratings in both discordant groups for the domain of mobility. In the 12-month transition more negative group, 27 of 32 (84%) reported worse mobility by transition, but the majority of these same participants (24 of 27) selected the same state ratings as six months earlier. Only 16% of the 'transition more negative' group reported being 'about the same' by transition but selecting a higher state rating. At the 6-month timepoint, the 'transition more negative' discordance group was more evenly split (60%/40%) between participants who reported worse mobility by transition and those who report mobility 'about the same.' (data not shown)

2.4 DISCUSSION

When comparing two different techniques for measuring change in self-rated balance and mobility over a six-month period, we found about 50% of our community-dwelling older adult sample to be discordant. While this finding was robust across domains and at both time periods of potential change, the same individuals were not discordant across domains and timepoints, indicating that discordance is not a phenomenon restricted to a select sub-group of our cohort, but may have more to do with perceived change over time and differences in the state and transition ratings scales.

First, and most importantly, we recognize that we are comparing responses to two very different questions. Interpolating perceived change from ratings of current state given at two different time points is not the same as asking an individual the degree to which they have changed over the same period. Since our study visits were spaced six months apart, participants are not expected to remember ratings given in the previous session.[38] Thus, the rating given for current state would reflect perception of state at that time and is not likely to reflect any attempt to communicate change relative to the state rating given six months earlier. As a result, we have found discordance when comparing **self-rated change by transition** to a **change in self-rating of current state**, and this is really the basis for all other theories we present below to explain the discordance. Disclosing the rating from the previous time point before obtaining the current rating of state would likely have generated greater concordance between methods, although it would not have been our primary aim. We wanted to determine whether people who report perceived change in the previous six months would also communicate change in the same direction indirectly by providing a different rating of current state than they had six months earlier. Eventually, we would like to understand more about what our response options for global state questions mean to older adults and how they choose one option instead of another.

In addition to the comparison of direct with indirect ratings to measure change, we also used two very different response scales to obtain the self-ratings. It is possible that change of one full level on the 5-point scale for current state (for example, from 'fair' to 'good') requires an individual to perceive a greater degree of change than is reflected by a rating of 'a little better' or even 'somewhat better' on the 7-point transition scale. This could account for the discordance seen for 31% of the sample for the domain of mobility and 20% for balance.

Two major types of discordance were identified, based on the different direction of change indicated by the transition rating and consecutive state ratings: 1) Discordance with transition rating more negative than the change in consecutive state ratings, or 2) Discordance with transition rating more positive.

Between-group comparisons of means on measures thought to represent or impact mobility and balance, or self-ratings in general, revealed significant differences between the concordant group and 'discordant: transition more negative group' for the domain of mobility only. At 12 months (the time of the transition rating and second state rating), the 'transition more negative' group appears to have lower mobility, less balance confidence, greater difficulty with daily activities, and greater restriction of activity in the past five years, but these between-group differences were less obvious for mobility discordance between 0 and 6 months. One explanation is the difference in the composition of the 'transition more negative groups' at each of these timepoints. At 12 months, most (84%) felt their mobility had worsened in the past 6 months, while just over half of the 6-month 'transition more negative' group reported worsening. This means that the 'transition more negative' group at 6 months really reflects a fairly even split between those who felt their mobility is worse and those who felt they have remained the same but then report a higher mobility state. This distinction from a group in which nearly all (84%) perceive that they have worsened could explain the differential ANOVA findings for mobility across the two time periods of change. Note that although substantially different than the concordant group, the mean mobility performance for the 'transition more negative' group at 12 months is still relatively high (0.92 m/s gait speed and SPPB score almost 9), which may explain why ~65% of this group still rate themselves as 'Good' or better at 12 mo.

Cognition and depression are known to impact self-report,[96, 98, 102] but were not significantly different between any of the three groups in the discordance comparisons, with one exception. The 'discordant: transition more positive' group had higher GDS scores than the concordant group for the domain of mobility only, but the mean of 2.3 for the discordant group is still well below the score of 6 that signals the need for referral for possible depression,[128] and the individual with the highest GDS score (8 out of 15) fell into the concordant group for balance, and reported better mobility by transition. Scores on the ERA-12 scale also do not appear to explain discordance. We had theorized that some older adults may perceive that their status has worsened with the passage of time simply because they have grown older.

These same people may not select the lower state rating, particularly if they are already on the lower end of the scale, because they may still believe that they have further room to decline. Although age expectation does not appear to explain discordance in our sample, it could be a reason why about 20% of the sample reported worsening mobility by transition at both 6 and 12 months, and 14% for balance at both timepoints, and it may explain why improvement was indicated less frequently by transition for both domains. Without some form of intervention, older adults may not anticipate improvement in domains that tend to decline gradually with age.[18, 68, 69, 72] There is evidence that older adults anticipate functional decline and associated loss of independence,[89] but they may not anticipate such change in a period as short as six months.

Unlike mobility, the between-group comparisons based on 12 month discordance were not significant for the balance domain. It is possible that the measures selected for comparison between groups are not representative of the activities that impact their own balance perceptions, and our results do suggest distinctions in perceptions of the domains of balance and mobility. Moderate correlation was found in self-perception between the domains of balance and mobility. Exact agreement was better for transition than state, possibly because of the popularity of the 'about the same' response option. For both of the six month time periods, more change was reported by transition for the domain of mobility than for balance, and this may be partially explained if balance is considered one factor that contributes to the greater concept of mobility. This would allow changes in pain, strength, and endurance to impact mobility, resulting in a change even when balance remains stable. Additionally, although the extremes of 'excellent' and 'poor' were used infrequently (11% average over all timepoints and both domains) to describe current status, 'poor' ratings were seen slightly more often for balance, and 'excellent' ratings were used more often for mobility.

Consistency bias is a documented phenomenon by which individuals tend to report being the same, often re-constructing the past to achieve similarity with the present.[38] The central option of our transition

question is '**about** the same,' making our results even more prone to consistency bias and central tendency bias.[122] However, we prefer this scale because responses of 'a little better' or 'a little worse' may be even more likely to represent meaningful and persistent change than if the central option were simply 'the same.' If our cohort is prone to consistency or central tendency bias, then the more indirect serial method may be better at detecting change in self-perceived status. This may explain discordance for the 11 (25% of total discordance at 12 months) participants who demonstrated serial change for mobility while reporting being 'about the same' by transition, and 25 (50% of total discordance) for balance. Although the central option of each scale ('good' for state and 'about the same' for transition) was the one selected most often across domains and timepoints, and the floor and ceiling options were least commonly selected, there was still an acceptable distribution across response options to suggest that older adults consider the entire scale rather than just the central option.

Use of the transition response scale was not limited to the central option, and concordance was not limited to those participants who report being 'about the same' over a given time period. While 66% of concordant selected the central transition option, 34% reported being either better or worse compared to six months ago. Alternatively, older adults who report being 'about the same' were often discordant because they selected a better or worse state option at the later timepoint.

One limitation of our repeated measures design is the potential for response shift.[93, 120] At baseline, participants self-rated mobility and balance before participation in a number of timed performance measures, including high-level balance (standing on one leg) and mobility activities (walking while saying every other letter of the alphabet) that they may not have attempted in years, if ever. At 6 and 12 months, ratings were again obtained before performance measures, but knowledge of performance at previous visits could still have impacted the selection of responses. In fact, a few participants commented immediately after the performance tests at the baseline visit that they would like to change their answers to the global questions, and some reported 'practicing' the more difficult tasks in preparation for the visits

6 and 12 months later. Response shift would be seen between primarily between baseline and 6 months, and distribution of mobility responses (Table 2.4) is almost identical, while balance responses appear to regress toward the mean. It is possible that our protocol had greater impact on self-perceived balance than mobility, which may depend more on daily activities than special tests. Depending on performance relative to initial self-perceived ability, the response could shift in either direction. For balance, people with initial 'excellent' or 'very good' ratings may have decreased if they did not 'ace' every test. Indeed, 4 of 6 who rated 'excellent' at baseline and returned at six months decreased their rating to 'very good,' and 9 of 16 (56%) who rated 'very good' decreased to 'good' or 'fair.' Alternatively, those who rated 'poor' or 'fair' may have gained new perspective if able to complete some of the tasks. Four of 7 who rated balance as 'poor' at baseline and returned for six months increased their rating to 'fair,' and 10 of 31 (32%) 'fair' ratings improved to 'good' while only one worsened to 'poor.'

Response shift may also have impacted 1-week test-retest reliability ratings, although state ratings were somewhat more reliable than transition for both domains. Both types of self-report questions were found to be moderately reliable, and the somewhat lower reliability for transition could be due to the response scale for this question. The transition response scale included seven items, compared to only five for the state scale. There is therefore more opportunity to select a different option one week later, and the distinction between 'about the same' and 'a little' better or worse may be less obvious than the distinction between a single level on the state scale (e.g. 'good' to 'very good' or 'fair') Additionally, global transition questions may have somewhat lower reliability as a result of the recall bias associated with these ratings.[38-40] For both state and transition, it is also possible that differences in ratings over the one-week period represent actual perceived fluctuations in these domains. From the ranges described in Table 2.1, it is evident that a few lower functioning participants were enrolled. In fact, two lower functioning individuals passed away before the 6-month visit, so it is possible that some of the older adults in this cohort did perceive fluctuations in their mobility and balance over one week. Stability of state ratings may depend on the timeframe used when selecting a response, as 'current status' was

assumed, but no timeframe was specified. A few of our participants with chronic conditions have reported that their balance and mobility can change from day to day with fluctuations in blood pressure, timing of medications, inner ear disorders, pain or viral illness, and mobility disability has been shown to fluctuate in older adults.[71] Thus, older adults who fluctuate and consider only their status on the day of the rating could legitimately rate themselves a little lower or higher one week later.

The tendency for recall bias in response to transition questions could support a serial approach for assessing change in self-report. Based on anecdotal reports from our data collectors, we have recognized a form of recall bias that we refer to as 'timeframe incongruency.' While responding to our global questions, some individuals who had experienced a significant event such as a serious illness over the timeframe of interest indicated that they were now rating themselves relative to that period, even after being redirected to compare themselves to their previous study visit, six months ago. When asked to rate change retrospectively, they indicated that they were drawing a comparison to the lowest point they had reached after the intervening event. They subsequently reported improvement, even though they appeared to still be far below their status six months earlier. This phenomenon would be restricted to a small and specific group of individuals, and may explain self-report discordance reported for those who select one of the 'better' transition response options, less than 10% of our sample. In these cases, the current state rating may be less biased as it does not require the individual to compare to a previous time point; however, it is also possible that the intervening event will cause a response shift in which the individual re-calibrates the response categories provided in the state question. Timeframe incongruency may actually pose a greater threat to concordance between self-reported change and change in performance over a given period, and this will be investigated in an upcoming chapter.

Floor and ceiling effects of the state response options could explain the discordance that occurs when change is reported by transition with no corresponding change in state rating, however, infrequent selection of the 'excellent' and 'poor' response options makes this an unlikely explanation for this cohort.

In fact, across both time periods, only one instance of discordance can be attributed to a ceiling effect (for the domain of balance from 6 to 12 mo), and all three discordant episodes that can be explained by floor effects were also for the balance domain (two from 0 to 6 mo; one from 6 to 12 mo. – A single participant was responsible for two of the three occurrences.).

Our results reveal considerable self-reported change over a six month period for the domains of both mobility and balance in community-dwelling older adults without intervention. This was true when change was assessed using both state (32.7% mobility; 44.6% balance) and transition ratings (52.5% mobility; 39.6% transition). Of those who do report change by transition, people are about four times more likely to report worsening than improvement. We did not anticipate a relatively high-functioning, community-dwelling older adult cohort to so readily report worsening of mobility and balance over a six month period. The fact that the central option of our transition scale was 'about the same' rather than 'the same' makes this observation even more significant. It provides evidence against our initial theory that community-dwelling older adults may be reluctant to report a decline in the domains of balance and mobility, out of fear of losing their independence. We hypothesized that they may indicate the decline only indirectly, through serial ratings, but the greatest worsening was communicated by transition for the domain of mobility in our cohort. It is possible that this fear would not be a factor in a study of this nature, but would instead manifest only in the presence of family or healthcare professionals known to make recommendations for the individual's safety.

The generalizability of our findings may be limited to community-dwelling older adults with moderate to high functional mobility. However, while mean values are consistent with a moderate to high functioning older adult cohort, there is a range of mobility ability evident from the minimum values of 'usual pace' gait speeds, SPPB and TUG scores. The gait speed of 0.45 m/s is far below the desired value of 1.2 m/sec for community ambulation,[143, 144] and TUG scores greater than 30 seconds are consistent with functional dependence.[47] The 0-12 ordinal SPPB scale was established based on quartiles of performance in the

original older adult validation cohort,[105] and scores in successively lower quartiles have been shown to be predictive of greater risk for incident mobility disability, nursing home admission, and death.[75, 105, 140] Our mean score of 9.5 is not in the highest quartile of performance, and our lowest score of 3 (scored by two participants) indicates very low mobility function.[75, 105, 140] On average, the cohort had difficulty with only 2.5 basic or instrumental activities of daily living and performed almost 9 of 11 SAFFE activities, however, most participants restricted at least a few activities, and a few in the cohort approached the ceiling for difficulty or limitation on both questionnaires.

It must also be noted as a limitation that, as a result of our 3-level categorical recoding of any change in self-report, the determination of concordance or discordance was based on agreement in direction but not necessarily magnitude of any change. For example, a transition rating of 'Much Worse' (-3) is considered concordant with 1-level decline in state rating (e.g. from 'Excellent' to 'Very Good') over the same period. Furthermore, the transition more negative discordant group did not necessarily indicate a worsening of status, but rather the rating method yielding results described as 'more negative' could have actually yielded a 'no change' transition status in comparison to an improvement detected by repeated state ratings. Combining these two different types of discordance may have confounded the results of the between-groups comparisons, however, the 'transition more positive' group is already limited in power by its small group size.

The majority (27 out of 32) of individuals in the 'transition more negative' group perceived worsening mobility, but their state ratings remained stable (88%) or more rarely improved, even with no floor effects. While it appears that their mobility was worse than that of the concordant group (and the other discordant group) at 12 months, it is not clear if performance declined, remained stable, or even improved over that same period. Regardless of their performance change, one explanation for the self-report discordance is that the perceived decline could be reported as 'a little worse' by transition, but was not enough to lower their state rating one full level, and no intermediate levels were available. Although

there was slightly more discordance in mobility self-report during the 0 to 6 month time period, there was also less reported change by transition, and it is possible that some of the shifts in state rating during the 0 to 6-month period reflect response shift induced by study participation.

Our findings for change in the domains of mobility and balance with the passage of time parallel prior findings for change in pain and physical limitation in response to rehabilitative interventions.[44] Fischer and colleagues[44] employed these two different methods of self-reporting change in a group of individuals after intervention for arthritis. They monitored change in the domains of pain and physical limitation, and found considerable discordance between the two methods for both domains. Agreement between state and transition methods was only 29%, with correlations between the two methods as low as 0.13. In all cases, retrospective measures demonstrated a greater percentage change than serial comparisons of state ratings, and therefore were concluded to be more sensitive. They also correlated more strongly with patient satisfaction. Correlations between the transition rating and the change in successive state ratings in our study ranged from 0.196 to 0.338.[44] Rather than using these discordance findings to discredit the ability of older adults to perceive and report change, we urge fellow clinicians and researchers to consider that we are not yet asking the right questions and do not understand how to best interpret the responses. This discordance must be further investigated if we are to know which of the two methods should be used to determine response to intervention or changes occurring as part of the natural history of aging.

We found greater percent agreement between the transition and serial techniques for assessing self-rated change (50-55% agreement across domains) than reported by Fischer's group (29%).[44] In addition to sampling a different population (individuals with arthritis) and focusing on different domains (pain and physical performance), their study also measures change rated after an intervention. It is possible that regular participation in an intervention with rehabilitative goals could create a greater response shift than participation in a three-visit, measurement-only research protocol. If response shift affects state more than transition ratings, a change in state rating that reflects a shift in internal standards rather than true

perceived change in ability would create a scenario of discordance for both self-report and the comparison of self-report to performance.

The best method for measuring self-report of change in balance and mobility in older adults remains unknown. Future studies should include exploration of how older adults define mobility and balance, how they determine the direction and magnitude of change reported over time, and how these reports relate to performance change. While our future plans include a comparison of change determined by self-rating to change in performance-based measures thought to reflect the same domains, this will not necessarily answer the question of which self-rating method is 'best.' In accordance with Fried's concept of 'pre-clinical disability,' it is possible that our older adults can perceive change before we can measure it clinically.[20, 145] Additionally, self-ratings may be a more integrated measure of function over time, so someone may report worsening of mobility as an integrated phenomenon that is not demonstrated on the day of our performance testing, or on the isolated tasks they are asked to perform in our artificial environment.[109]

Even with further exploration, we anticipate that neither approach will be preferred in all situations. When responding to a transition question, people are aware that they are rating change. This would then be the option with greater face validity for the concept of self-perceived change, and the one that would correlate most with satisfaction with care. However, because of documented issues like recall bias and consistency bias, a less direct approach may warrant consideration for use in specific circumstances. For example, a serial comparison of state responses may be better at uncovering change that is harder for the individual to recognize as occurring over a specific time period, such as a slow decline in balance or mobility.

STATE QUESTION:

Would you say your level of mobility in general is:

1. Excellent
2. Very Good
3. Good
4. Fair
5. Poor

TRANSITION QUESTION:

In the last 6 months, has there been any change in your overall mobility?

3 Much Better

2 Somewhat Better

1 A Little Better

0 About the Same

-1 A Little Worse

-2 Somewhat Worse

-3 Much Worse

FIGURE 2.1: Global Questions

TABLE 2.1: Baseline Characteristics of the Sample (n=103)

CHARACTERISTIC	Mean (SD)	Range
Age (yrs)	77.4 (6.0)	64 - 92
Gender (% Female)	74%	—
Race (% Black)	12.5%	—
Education (% Who Attended College)	70%	—
Comorbidities (0-17)	2.9 (1.4)	0 - 6
15-Item GDS	1.48 (1.7)	0 - 8
DSST (# in 90 sec)	48.1	16-70
ERA (0-100)	53.1	11.1-97.2
Gait Speed (m/sec)	0.97 (0.24)	0.45 – 1.52
TUG (sec)	10.5 (4.1)	6.0-26.0
SPPB Score (0 -12)	9.5 (2.3)	3 - 12
BADL/IADL difficulty (0 - 16)	2.5 (2.9)	0 - 14
SAFFE Activity (0 – 11)	8.7 (1.5)	3 - 11
SAFFE Fear (0-3)	0.13 (0.19)	0-0.78
SAFFE Restriction (0 -11)	2.9 (2.8)	0 - 9

TABLE 2.2: Test-Retest Reliability of Global Items (Percent Agreement; n=46)

Agreement	Global Items	MOBILITY		BALANCE	
		State	Transition	State	Transition
Exact		34 (73.9%)	28 (60.9%)	28 (60.9%)	32 (69.6%)
Within 1 Level		42 (91.3 %)	42 (91.3%)	46 (100%)	41 (89.1%)
Within 2 Levels		46 (100%)	44 (95.7%)		46 (100%)
Kendall's Tau		0.604	0.443 (3 level)	0.662	0.428 (3 level)

TABLE 2.3: Percent Agreement for Mobility and Balance Self-Ratings at Baseline (n=119)

Agreement	Global Items	COMPARISON OF MOBILITY and BALANCE RATINGS	
		State	Transition
Exact		57 (47.9%)	88 (73.9%)
Within 1 Level		107 (89.9 %)	114 (95.8%)
Within 2 Levels		114 (98.3%)	118 (99.2%)
Kendall's Tau		0.576	0.560 (3 level)

TABLE 2.4: Comparison of State Ratings for the Domains of Mobility and Balance Across All Timepoints

RATING	DOMAIN AND TIMEPOINT					
	MOBILITY			BALANCE		
	BL1 N = 119	6 mo N = 103	12 mo N = 112	BL1 N = 119	6 mo N = 103	12 mo N = 112
Excellent	13 (10.9%)	11 (10.7%)	11 (9.8%)	8 (6.7%)	2 (1.9%)	5 (4.5%)
Very Good	38 (31.9%)	30 (29.1%)	30 (26.8%)	19 (16.0%)	23 (22.3%)	15 (13.4%)
Good	42 (35.3%)	38 (36.9%)	41 (36.6%)	48 (40.3%)	38 (36.9%)	46 (41.4%)
Fair	25 (21.0%)	23 (22.3%)	27 (24.1%)	34 (28.6%)	36 (35.0%)	41 (36.6%)
Poor	1 (0.8%)	1 (1.0%)	3 (2.7%)	10 (8.4%)	4 (3.9%)	5 (4.5%)

TABLE 2.5: Transition Ratings Across All Timepoints

RATING	DOMAIN AND TIMEPOINT					
	MOBILITY			BALANCE		
	BL1 N = 119	6 mo N = 103	12 mo N = 112	BL1 N = 119	6 mo N = 103	12 mo N = 112
Much Better	1 (0.8%)	1 (1.0%)	2 (1.8%)	0 (0%)	1 (1.0%)	1 (0.9%)
Somewhat Better	4 (3.4%)	3 (2.9%)	2 (1.8%)	2 (1.7%)	0 (0%)	2 (1.8%)
A Little Better	10 (8.4%)	3 (2.9%)	7 (6.3%)	4 (3.4%)	5 (4.9%)	3 (2.7%)
About the Same	74 (62.2%)	65 (63.1%)	56 (50.0%)	89 (74.8%)	71 (68.9%)	70 (62.5)
A Little Worse	29 (24.4%)	24 (23.3%)	35 (31.3%)	19 (16.0%)	20 (19.4%)	28 (25.0%)
Somewhat Worse	1 (0.8%)	5 (4.9%)	9 (8.0%)	5 (4.2%)	5 (4.9%)	7 (6.3%)
Much Worse	0 (0%)	2 (1.9%)	1 (0.9%)	0 (0%)	1 (1.0%)	1 (0.9%)

TABLE 2.6: Agreement for Change in Self-Rated **Mobility** Between **6 and 12 Months; Tau .327****

		TRANSITION			
		Worse (-1, -2, -3)	**Same (0)**	**Better (+1, +2, +3)**	**Total**
STATE	**Worse (-1, -2)**	15 (14.9%)	**6 (5.9%)**	**0 (0%)**	21 (20.8%)
	Same (0)	**24 (23.8%)**	37 (36.6%)	**7 (6.9%)**	68 (67.3%)
	Better (+1)	**3 (3.0%)**	**5 (5.0%)**	4 (4.0%)	12 (11.9%)
	Total	42 (41.6%)	48 (47.5%)	11 (10.9%)	101 (100%)

KEY for Tables 6-9: Bold text indicates disagreement in category (worse/ same/ better) or 'Discordance' using the two approaches. Values in red text indicate disagreement in opposite directions or 'Better/Worse Discordance.' Cells shaded in gray are 'Concordant in Self-Report.' Yellow = 'Discordant: Transition More Positive'; Blue = 'Discordant: Transition More Negative'

** $p \leq 0.05$

TABLE 2.7: Agreement for Change in Self-Rated **Mobility** Between **0 and 6 Months; Tau=.196***

		TRANSITION			
		Worse	**Same**	**Better**	**Total**
STATE	**Worse**	6 (5.9%)	**11 (10.8%)**	**0 (0%)**	17 (16.7%)
	Same	**23 (22.5%)**	40 (39.2%)	**5 (4.9%)**	68 (66.7%)
	Better	**1 (1.0%)**	**14 (13.7%)**	2 (2.0%)	17 (16.7%)
	Total	30 (29.4%)	65 (63.7%)	7 (6.9%)	102 (100%)

*$p \leq 0.05$

TABLE 2.8: Agreement for Change in Self-Rated **Balance** Between **6 and 12 Months; Tau = .217***

		TRANSITION			
		Worse	**Same**	**Better**	**Total**
STATE	**Worse**	12 (11.9%)	**11 (10.9%)**	**2 (2.0%)**	25 (24.8%)
	Same	**19 (18.8%)**	36 (35.6%)	**1 (1.0%)**	56 (55.4%)
	Better	**3 (3.0%)**	**14 (13.9%)**	3 (3.0%)	20 (19.8%)
	Total	34 (33.7%)	61 (60.4%)	6 (5.9%)	101 (100%)

*p≤0.05

TABLE 2.9: Agreement for Change in Self-Rated **Balance** Between **0 and 6 Months; Tau=.338****

		TRANSITION			
		Worse	**Same**	**Better**	**Total**
STATE	**Worse**	11 (10.8%)	**12 (11.8%)**	**1 (1.0%)**	24 (23.5%)
	Same	**13 (12.7%)**	40 (39.2%)	**1 (1.0%)**	54 (52.9%)
	Better	**1 (1.0%)**	**19 (18.6%)**	4 (3.9%)	24 (23.5%)
	Total	25 (24.5%)	71 (69.6%)	6 (5.9%)	102 (100%)

*p≤0.05

TABLE 2.10: Summary of Discordance for Domains and Time Periods

Discordance		MOBILITY		BALANCE	
		0-6 Months n = 102	6-12 Months n = 101	0-6 Months n = 102	6-12 Months n = 101
Total Discordance (n) **(% of sample)**		**54** **(52.9 %)**	**45** **(44.6%)**	**47** **(46.1%)**	**50** **(49.5%)**
Transition More Negative (n) (% of total discordance)		38 (70.3%)	32 (71.1%)	33 (70.2%)	36 (72%)
Transition More Positive (n) (% of total discordance)		16 (29.6%)	13 (28.9%)	14 (29.8%)	14 (28%)
Better/ Worse Discordance (n) (% of total discordance)		1 (1.9%)	3 (6.7%)	2 (4.3%)	5 (10.0%)

TABLE 2.11: Comparison of 12 Month Concordance Groups by Individual Characteristics at 12 Months

Characteristic	MOBILITY (Group Mean)			BALANCE (Group Mean)		
	Concordant n=56	**Discordant Transition More Neg** n=32	**Discordant Transition More Pos** n=13	**Concordant** n=51	**Discordant Transition More Neg** n=36	**Discordant Transition More Pos** n=14
Age (yrs)	78.9	78.6	75.2	77.7	79.3	77.9
Comorbidities (0-17)	2.7	3.1	3.2	**2.5**	**3.5****	**2.6**
GDS (0-15)	1.1*	1.8	2.3*	1.4	1.8	1.2
DSST (# in 90 sec)	47.0	49.4	49.0	47.7	49.2	46.2
ERA (0-100)	54.6	50.9	55.3	54.0	50.5	60.3
Gait Speed (m/s)	**1.07***	**0.92***	**0.98**	1.03	0.99	1.03
TUG (sec)	**9.53***	**12.51***	10.55	10.9	10.4	10.0
SPPB Score (0-12)	**10.1***	**8.8***	9.9	9.6	9.7	9.9
BADL/IADL (0-16)	**1.8***	**4.0***	**3.5**	2.7	2.8	2.6
ABC (0-100%)	**81.1****	**65.1**	**61.7**	75.4	69.2	76.7
SAFFE Act (0-11)	**9.0**	**7.9****	**8.1**	8.6	8.4	8.4
SAFFE Fear (0-3)	**0.07****	**0.24**	**0.30**	0.16	0.14	0.20
SAFFE Restriction (0-11)	**2.6**	**4.6****	**2.8**	**2.9**	**3.8**	**2.9**

*Significant difference ($p \leq .05$) between the concordant group and **only one** discordant group, also denoted by*

** Significant difference ($p \leq .05$) between this group and **both** other groups.

Bold text indicates mean differences thought to be clinically meaningful although not statistically significant.

TABLE 2.12: Distribution of MOBILITY State Ratings at 6 and 12 Months By 12 Month Discordance Group

	Concordant N=56	**Discordant: Transition More Negative** N=32	**Discordant: Transition More Positive** N=13
MOBILITY State Rating at 12 Months			
Excellent	9 (16%)	1 (3%)	0 (0%)
Very Good	18 (32%)	6 (19%)	2 (15%)
Good	17 (30%)	14 (44%)	5 (38.5%)
Fair	10 (18%)	11 (34%)	5 (38.5%)
Poor	2 (4%)	0 (0%)	1 (8%)
MOBILITY State Rating at 6 Months			
Excellent	10 (18%)	0 (0%)	0 (0%)
Very Good	22 (39%)	3 (9%)	4 (31%)
Good	17 (30%)	15 (47%)	6 (46%)
Fair	6 (11%)	14 (44%)	3 (23%)
Poor	1 (2%)	0 (0%)	0 (0%)
MOBILITY Transition Rating at 12 Months			
Better	4 (7%)	0 (0%)	7 (54%)
About the Same	37 (66%)	5 (16%)	6 (46%)
Worse	15 (27%)	27 (84%)	0 (0%)

TABLE 2.13: Distribution of BALANCE State Ratings at 6 and 12 Months By 12 Month Discordance Group

	Concordant N=51	**Discordant: Transition More Negative** N=36	**Discordant: Transition More Positive** N=14
BALANCE State Rating at 12 Months			
Excellent	1 (2%)	3 (8%)	**1 (7%) Ceiling**
Very Good	9 (18%)	4 (11%)	0 (0%)
Good	22 (43%)	13 (36%)	5 (36%)
Fair	17 (33%)	15 (42%)	6 (43%)
Poor	2 (4%)	**1 (3%) Floor**	2 (14%)
BALANCE State Rating at 6 Months			
Excellent	1 (2%)	0 (0%)	1 (7%)
Very Good	13 (25.5%)	3 (8%)	6 (43%)
Good	25 (49%)	8 (22%)	5 (36%)
Fair	10 (20%)	23 (64%)	2 (14%)
Poor	2 (4%)	2 (6%)	0 (0%)
BALANCE Transition Rating at 12 Months			
Better	3 (6%)	0 (0%)	3 (21%)
About the Same	36 (71%)	14 (39%)	11 (79%)
Worse	12 (23%)	22 (61%)	0 (0%)

3. INSIGHTS GAINED FROM SEMI-QUANTITATIVE INTERVIEWS OF OLDER ADULTS

3.1 INTRODUCTION

Global self- ratings of specific domains such as pain and function are commonly used to determine response to an intervention,[44, 94, 95] but another goal in healthcare is tracking change with the passage of time.[68, 72, 77] This is of particular interest in the field of geriatrics, where balance and mobility declines can be insidious in onset and have catastrophic results.[1, 2, 54, 68, 146] Global self- or observer-current state ratings of balance and mobility using terms like 'fair' or 'very good' can be used to characterize older adults on a continuum of function, and are sometimes the only information available to healthcare providers when determining the need for more formal evaluation and rehabilitation. Self-ratings can also be used by researchers to determine inclusion in studies of balance and mobility. In all settings, self-ratings may provide a picture disparate with classifications imposed by observers, or with performance on measures thought to represent the same domain.[85] For example, one older adult who requires a walker to ambulate and is limited to household distances may rate her mobility as 'very good,' while another who needs no device and walks a mile a day may rate himself as only 'fair.' One explanation for this discordance is a disparity in perspectives among older adults,[53] and the healthcare providers who may also be rating them.[58, 59]

Certainly global ratings of current status made using an ordinal scale ranging from poor to excellent are 'relative' and based on some sort of comparison, but relative to what? We suspect that whether an individual is rating herself or a healthcare provider is rating a patient, selection of a response is made after consultation with some internal frame of reference. Seasoned geriatric healthcare providers can reference their past experience with hundreds of older adults who range from immobile to having no observable

mobility deficit. Performance at the 'floor' and 'ceiling' of standardized measures could command ratings of 'poor' and 'excellent' respectively, but what if an older adult's self-rating is based only on a comparison with adults of much lower mobility encountered while volunteering at a nursing home, while another of the same ability considers only the high functioning teammates on a senior softball league? Maybe some older adults do not draw a comparison to others at all, but hold an internal standard of perfect balance and mobility that they believe they have never met.

Variations in personal definitions of balance and mobility as global terms could also lead to a discrepancy between self-ratings, provider ratings, and performance.[53] Does mobility include getting out of bed, or is the focus on higher-level tasks like walking in gravel parking lots and running? Some may rate only their ability to perform activities that they do regularly, leaving more difficult activities that they do not attempt or cannot perform out of the picture. Alternatively, an individual may base a 'poor' rating on one or two high-level tasks they can no longer perform, without considering that they perform many more tasks with no difficulty.

In our experience administering self-report measures to older adults in both clinic and research settings, some have difficulty rating their current global status on a scale from poor to excellent, because their answer varies with the frame of reference used. Participants will sometimes ask whether they should rate their status relative to their age, or on a spectrum of people of any age group, in which case they often indicate that the rating will be lower. Additionally, some will comment that they are rating themselves in comparison to their own abilities at some point in the past, and this is essentially what they are asked to do when responding to transition-based questions comparing their ability over a finite period of time using terms like 'better,' 'the same,' or 'worse.'

Purpose: The purpose of this study is to explore using a mixed-methods approach how a group of community-dwelling older adults answer both current state and transition-based global questions to rate their own balance and mobility. We will describe the meaning of the global terms, the timeframes

considered, and the most common 'frames of reference' used by older adults when rating their own balance and mobility. Finally, we will examine whether the frame of reference used can be explained by individual characteristics that may influence self-ratings of balance and mobility.

3.2 METHODS

3.2.1 Participants

This was a cross-sectional sub-study of larger longitudinal study. The primary aim of the larger study was to develop and refine a battery of self-report and performance-based measures of balance, mobility, and function in a sample of community dwelling older adults representing a range of mobility. Participants were recruited from the University of Pittsburgh Claude D Pepper Older American's Independence Center (OAIC) Pepper Registry, a group of community-dwelling older adults who have consented to be contacted for participation in studies of mobility and aging at the Pepper Center. Adults age 65 and older were included if they could walk independently for household distances, with or without an assistive device. Primary exclusions were signs of medical instability that may contraindicate participation in performance-based tests, or neuromuscular conditions (Parkinson's disease, residual stroke) that significantly impair movement. Study visits occurred every six months for one year, with a smaller sub-group invited to participate in an abbreviated home visit at 18 months. The semi-structured interviews described here were conducted only at the 12 or 18-month timepoint. Participants were offered the voluntary interviews based on the timing of their follow-up visits.

3.2.2 Measures

Interviews and Self-Ratings:

During all study visits, participants completed a three to four hour battery of both self-report and performance-based measures of balance, mobility and function. The semi-quantitative interviews were conducted first in the sequence of measures, prior to any performance-based tests. Interviews lasted about 15 minutes. Participants first self-rated global mobility and balance on a 5-point scale (1 Poor, 2 Fair, 3 Good, 4 Very Good, 5 Excellent) in response to the current status prompt, '***Would you say your level of mobility (alt: balance) in general is. . .***? No descriptors were provided to clarify the terms balance and mobility, and no timeframe was given, although participants were then asked the transition question, '***Has there been any change in your mobility (alt: balance) over the past six months***?' No frame of reference was provided to participants while rating current status. Participants were simply told to select the response from the 5-point scale provided that best applied to them. For the transition-based question, a seven-item Likert scale was provided ranging from 'Much Worse' (-3) to 'Much Better' (+3). Adjectives of 'Somewhat' and 'A little' were used for the intermediate options, and the central option was 'About the same.'

The current state and transition global ratings were followed by scripted, semi-quantitative interviews designed to clarify the process used by older adults when providing global ratings. The interviews were organized as follows: 1) a discussion of the definitions of mobility and balance, 2) clarification of the timeframe used to rate current status and the process used to rate any change over the past six months, and 3) an inquiry into any frames of reference used when rating current state using the 5-point ordinal scale provided. At the end of the interview, participants were invited to provide any additional narrative information they felt would clarify the process of rating one's own balance and mobility using either current status or transition-based questions. Following the 'think-out-loud' qualitative method,[147] we

recorded and analyzed any comments viewed as relevant to the previously named topics, including those made while rating current state and six-month change. To minimize biasing responses toward our own theories, open-ended responses were obtained before closed-ended questions with specific response options were introduced. Most closed-ended questions allowed for multiple answers, or the participant could refuse all provided options, substituting their own response if desired. Refer to **Appendix** for the specific script used to obtain both open-ended and closed-ended responses for each of the topics above.

In addition to the interviews, self-report and performance-based measures used as explanatory variables in these analyses include:

Explanatory Self-Report Measures:

15-item Geriatric Depression Scale (GDS):[128] The 15-item version of this basic screening measure for depression in older adults was used. For each item, participants are assigned one point if the 'Yes/No' response option selected corresponds to lower mood, for total scores ranging from 0 to 15. When used clinically, scores higher than 5 suggest depression and signal the need for further evaluation.

Digit Symbol Substitution Test:[129] This paper and pencil test from the Wechsler Adult Intelligence Scale measures perceptual processing speed. Participants are given a coding key with nine boxes. The upper half of each box contains a number from 1-9, and the bottom contains the symbol that corresponds to that number. The rest of the page is filled with rows of split boxes. The top of each box contains randomly ordered numbers (ranging from 1-9), and the coding key is used to fill in the bottom half of each box with the corresponding symbol. Participants fill in as many of the boxes as possible in 90 sec. Performance declines with age.[130] When combined with a gait speed of less than 1.0 m/s, DSST scores of less than 27 have been associated with increased mortality and incident disability in an older adult cohort of mean age 70 years.[131]

Expectations Regarding Aging (ERA-12)[60]- A twelve-item self-report tool designed to measure expectations regarding aging in the domains of physical health, mental health, and cognitive function. Responses have demonstrated acceptable levels of reliability and construct validity.[60]

Activities-Specific Balance Confidence Scale (ABC):[132] Participants were asked to rate their confidence that they will not fall or lose their balance during the performance of each of 16 mobility-related activities. A response of 0 indicates no confidence and 100 is complete confidence. The total score is calculated as the average confidence reported for all items and is expressed as a percentage out of a possible 100. Scores of greater than 80% are common in high functioning, physically active older adults.[132, 133]

Survey of Basic and Instrumental Activities of Daily Living (BADL/ IADL Questionnaire): A self-report measure of basic and instrumental activities of daily living (BADL/ IADL) taken from the National Health Interview Survey (NHIS) was used to assess function for daily activities.[134] The BADL/ IADL scale measured self-reported ability to perform 16 tasks, including getting up from bed or chairs, walking, stair climbing, getting outside the home, and shopping. Because we are interested in the domains of balance and mobility, we excluded from analysis the two items related to bowel and bladder function, but retained the toilet transfer item. Therefore, the total indicates the number of tasks out of a possible 14 reported as either difficult or not performed for reasons related to health and physical function. Interrater and test-retest reliability ICC's above 0.9 have been reported in community-dwelling older adults.[25]

Survey of Activities and Fear of Falling in the Elderly (SAFFE):[135] This self-report measure assesses physical activity and fear of falling relative to eleven specific mobility-related activities, ranging from bending over and reaching overhead to walking for exercise and going out in slippery environments. SAFFE activity is scored as the number of activities performed out of 11, and SAFFE restriction is calculated as the number of activities reported as performed less frequently now than five years ago. Both activity and restriction scores range from 0 to 11, with higher scores indicating greater activity or greater

restriction, respectively. To determine SAFFE fear score, participants are asked how worried they are about falling with each activity performed. Total fear is calculated as the mean response across only the activities that are currently performed, ranging from 0 (not at all worried) to 3 (very worried).

Explanatory Performance Measures:

Short Physical Performance Battery (SPPB):[105] The SPPB quantifies functional mobility using three categories: standing balance, walking, and repeated chair stands. Performance in each category is timed, and the time is converted to an ordinal score ranging from 0 (unable to perform) to 4 (best performance), used to calculate a summary score (maximum 12). The full protocol and training instructions can be downloaded from the NIH website. Test-retest and inter-rater reliability of each of the three categories range from 0.73 to 0.97,[104, 136, 137] and the reliability of the summary scale was established using internal consistency (Chronbach's alpha = 0.76). [105] Originally validated in over 5,000 adults age 71 and older, summary scores correlated strongly with self-reported ADL disability for tasks like walking across a room and transferring from bed to chair; and with self-reported ability to walk up and down steps or walk ½ mile without help.[75] Summary scores were also found to strongly predict mortality and nursing home admission.[105]

Gait Speed: Gait speed was measured in meters per second (m/s) using an instrumented walkway (GaitMat II) of approximately 6 meters in length. To avoid including acceleration and deceleration, recordings were taken only from the central 4 meters of the walkway. Up to four 'passes' were performed according to participant tolerance, and gait speed was averaged over all passes. Participants used their own assistive device if needed and were closely spotted by research staff, but were not allowed physical assistance from another person. If GaitMat gait speed could not be determined (due to equipment malfunction or administration of study measures in the home), the mean time to complete the

SPPB timed walk was converted to velocity and substituted for the GaitMat data. Because the SPPB 4-meter timed walk is measured from a standing start, a conversion factor was developed by our statistician using both the GaitMat and SPPB timed walk data available on all participants at baseline. Gait speed in older adults is highly predictive of adverse outcomes, including institutionalization and mortality.[138-141]

3.2.3 Statistical Analysis

A mixed-methods approach was used. Qualitative data were analyzed using a quasi-statistical method as developed by Crabtree and Miller.[148] Interviews were transcribed verbatim by the primary author, and open-ended responses were coded using ATLAS.ti version 6.0 (Scientific Software Development, Berlin, Germany). Themes were developed for each of the following areas: 1) Mobility definition, 2) Balance definition, 3) Timeframe, and 4) Frame of Reference. Codes for each domain are expressed as frequencies and percentages of the total. After reviewing the code book developed by the primary author, one of the co-authors independently coded all interviews, and inter-coder reliability was calculated using the Kappa statistic.

Responses to closed-ended questions in each of the four areas above were tabulated using descriptive statistics. A response option was considered to be 'chosen' only if directly selected by the participant from the available list, regardless of responses to open-ended questions or 'think out loud' comments made during the interview. For the frame of reference domain, after tabulating responses to the three closed-ended options provided (Past-Self, Age-Expectation, Other Comparison), a dichotomous summary code was assigned for each of the three themes based on a combination of the participant's closed-ended selections and the open-ended codes that emerged from the interview. For each of the frames of 'Past-Self' and 'Other Comparison,' the resulting two groups (present/absent) were compared on individual characteristics such as age and gait speed. Because 'Past-Self' was identified as a frame of reference for the majority of participants, and some insisted that it was their only frame of reference, another set of

comparisons was run between two groups established based on whether 'Past-Self' was the **only** frame of reference that emerged during the interview. For example, a participant may have selected only 'Past-Self' from the three-item list originally provided, but if the same participant reported comparing himself to a friend while rating his own mobility, that participant was categorized in the 'Other Comparison' group for the first analysis and the 'Past-Self Plus' group for the second analysis. Comparisons were made using t-tests for continuous variables, or ordinal variables that can be treated as continuous (such as SPPB scores).[105] Chi-square comparisons of observed and expected values were used for categorical or ordinal variables with only two levels, and the Mann-Whitney U Test compared 5-level SPPB balance scores and 3 to 5-level self-ratings between groups.

Finally, any change in global self-ratings between the time of the interview and the visit six-months prior were examined for discordance, and compared based on the frames of reference identified during the interviews. To do this, transition ratings provided at the time of the interview were collapsed into three categories (better, same, worse) for each of the domains of mobility and balance. State ratings for each domain were compared to the state ratings provided at the previous study visit, and the difference between the two ratings was collapsed to a three-level state-based change rating (better, same, worse). Transition ratings at the time of interview were then compared to the change in state rating over the two study visits, and participants were defined as either concordant or discordant in self-rating for each domain based on whether the two methods of measuring change in self-rating agreed in direction. For example, a participant who rated her mobility as 'worse' or 'about the same' by transition, but gave a current state rating one or more levels higher than six months ago (e.g. from 'fair' to 'good') was categorized as discordant for the domain of mobility. Concordant and discordant groups were compared by frame of reference using the same three dichotomous groupings described above ('Other Comparison,' 'Past-Self,' and 'Past-Self **only**.') For all comparisons, a type-I error rate of 0.05 was used.

A formal power analysis is not appropriate for an exploratory study of this nature, however, grounded theory qualitative research method requires as few as 30 participants to develop relevant concepts,[149] and thematic saturation when using a homogenous patient population, as included in this analysis, can be achieved between 15-20 interviews.[150]

3.3 RESULTS

Thirty-three community-dwelling older adults (mean age 77.2 $\pm$ 6.3 years, range 65-90 years; 73% female, 12% black) completed semi-structured audiotaped interviews. Nine of these were completed as part of an 18-month home visit.

Domain definitions for mobility and balance

Open-ended definitions of both mobility and balance most often involved mention of specific activities, however, the themes that emerged differed slightly for the two domains. Six activity themes were identified for each domain, but differences were noted. One theme ('Balance tests') was exclusive to the balance domain, while the theme 'IADL' emerged only for the mobility domain. IADLs identified included shopping, cleaning, and driving or travel out of town, while only BADLs (showering, standing on one leg to don pants) were identified for balance. Even within the same theme, the specific activities referenced by older adults varied by domain. This was true for 'Basic,' 'Advanced,' 'Recreation,' 'Transfers,' and 'BADL.' For example, recreational activities for mobility included swimming and seated exercise, while those mentioned for balance included roller skating and bowling.

The full list of activity themes for each domain is presented in Figures 3.1 (mobility) and 3.2 (balance). Both figures show the number of respondents for whom the specified theme appeared at least once during the interview. Inter-coder reliability ranged from 0.77 to 1.0 (Kappa), and thematic saturation was reached for all topics, indicating the identification of new themes through completion of additional

interviews is unlikely . Regardless of the number of specific activities or quotations categorized under the same theme, each theme or code is reflected no more than once per participant in the frequency results, but specific activities represented by each theme are listed in the keys to Figures 3.1 and 3.2.

In addition to the activity themes, non-activity themes also emerged from the open-ended interviews. For both domains, these included the themes of '*outside*' as an environment, '*assistance*' of either an assistive device or another person to remain upright or complete a task, and '*falling*.' Additionally, the theme of '*balance*' emerged among the definitions of mobility, and '*fear*' for the domain of balance. Frequencies for each of these themes are presented for both domains in Figure 3.3.

Figure 3.1 and Table 3.1 report closed-ended response selections for each domain. As illustrated in Figure 3.1, two activity themes (Advanced Mobility and IADL) and one other theme (Balance) were not anticipated by the authors, so were not available for selection as closed-ended responses for the domain of mobility. The most popular closed-ended response option used to define mobility was 'walking in all environments that I encounter daily,' but 10 (30.3%) participants indicated that all responses apply, including the closed-ended responses corresponding to themes like 'BADL' that emerged infrequently (twice) in open-ended responses. 'Being steady on my feet' (n=29) and 'Confident that I will not fall' (n=30) were closed-ended responses each chosen by nearly all of the older adults to describe balance, and the corresponding open-ended themes of 'Basic' balance activities and 'Falling' were identified in 22 and 29, respectively . Balance (n = 24; 73%) was recognized by more of the cohort than mobility (n=19; 58%) as involving only activities where one is up on his feet, with quotes including, 'Anything that entails 'being up' to me identifies or pertains to mobility. If I'm in bed, I'm not mobile. I'm just there.' However, other participants referenced activities like swimming as mobility activities during which they are not on their feet. One participant explained, 'I would think like sitting and reading a paper. . . or eating, I would think those are mobility because you're moving around at the table and doing different things.' According to another, 'I could lay on the floor and have mobility.'

Timeframe Domain

Participants reported timeframes ranging from today (n=3) to the past 15 years (n=1) when providing the state-based self-rating in response to the prompt, 'Would you say your mobility (alt: balance) is. . .' The most frequent response (n = 9) was 'one year,' although seven participants were unable to specify a timeframe at all, most making broader comments that they were rating their status 'in general.' Refer to Figure 3.4 for frequencies of all responses to the timeframe query. While responses were skewed toward longer timeframes, two participants (#108 and 114) specified that their ratings of mobility and balance today differ from those they would have given a day or two earlier. One of those two was experiencing a recurrence of her episodic vertigo and was unable to specify a timeframe when asked. The other, who had fallen in the past year and tied for the lowest SPPB score (7/10) of all 33 participants, specified a timeframe of the past month for her ratings.

When providing a global transition rating of any change in their balance or mobility over the past six months, 26 (79% of the sample) reported thinking more of how they are doing now than of how they were doing six months ago, and 5 more could not specify a preference. Still, only 4 (15%) of those who emphasize the present reported having difficulty remembering how they were doing six months ago. We are considering this effect of memory transience to fall under the discordance theory of recall bias.[38] Interestingly, one of these 4 participants with recall bias was among seven (21% of the total sample) who made specific reference at some point during the interview to the experience of their prior study visits, often to their performance on specific measures. Four participants reported thinking of both the present and past timepoints equally while rating transition, and only two felt that the past more heavily influenced their rating.

Although '**recognized** recall bias' was rare, the theme of 'timeframe incongruency' or 'TFI' emerged. This code was assigned when a participant indicated later in the interview that they had based the transition rating on some timeperiod other than the six months specified in the transition question script.

Seven participants were coded as having a TFI, two of whom also had recognized recall bias. None of the seven appeared to recognize the incongruency statement they made, and it was not brought to their attention by the interviewer. Although it was not specifically asked of all participants, six volunteered that they had experienced an 'Intervening Event' since the last study visit, like a significant illness or the diagnosis of a new health condition such as polymyalgia rheumatica. Only one of these six demonstrated a timeframe incongruency.

Frame of Reference Domain

Perspectives shared by participants confirmed our suspicions about the importance of the frame of reference used when providing self-ratings. When asked how she selects one of five current state response options, one participant responded 'That's a really good question because I don't know. What am I comparing it with? With what I used to be, or with what other people my age are, or other young people? That's why I'm skeptical about these forms.' (108)

A breakdown of the selection of closed-ended frame of reference response options is shown in Table 3.2. A comparison to one's self in the past ('Past-Self') was selected most frequently, followed by an age-based expectation ('Age Expectation') and a comparison to others ('Other Comparison'). As with all other domains, open-ended responses about frame of reference were obtained before the closed-ended options were revealed, and the open-ended responses supported these same three themes, with no other themes identified. However, the distribution of the themes upon analysis of open-ended responses was very different. (Table 3.2) Closed-ended response option 3, 'a specific person I know,' was selected only by 4 participants (12%), however, a total of 16 (48%) mentioned at some point during the interview that they compare themselves to others when rating their own status. In a few cases, participants stated they did not pick this 'Other Comparison' closed-ended option because the comparison they made is not to a single person. For others, the comparison was made to people they see regularly (for example, at the gym), but do not know well on a personal level. Because the comparison to another person who is

perceived by the participant as having either better or worse mobility or balance is of greater interest than the compared individual's true abilities or relationship to the participant, primary focus will be placed on the results of the open-ended responses for this theme. Table 3.2 includes relevant excerpts from the interviews representing each of the themes. Essentially, 'Past-Self' is a comparison to one's own mobility or balance performance at any point in the past. 'Age Expectation' was identified as a theme if self-ratings were described as being at a certain level based on the participant's age, or compared to some theoretical standard held for people of the same or another age group. Finally, 'Other Comparison' was identified when references were made to another person or group of people with whom the participant has any sort of relationship, ranging from occasional visual contact to a longstanding personal association with direct communication. Thus, a general or theoretical comparison to 'people in their 20's' would be assigned the code 'Age Expectation,' while a more specific comparison to a group of younger people who attend their exercise class would be coded as 'Other Comparison.'

As mentioned previously, these three frames of reference were not mutually exclusive. While 11 (41%) of those identified with 'Past-Self' and 4 (27%) of those with 'Age Expectation' mentioned only that single frame of reference when both their open-ended and closed-ended responses were reviewed, only one of 16 coded as 'Other Comparison' were exclusive to this theme. Even within a single statement, more than one frame of reference could emerge. For example, the statement 'I didn't choose excellent because I used to be able to keep up with the younger members of the family' was identified as both 'Past-Self' and 'Other Comparison.'

Dichotomous groups were established as described under methods, using responses to both closed- and open-ended questions, and the results of the between-group comparisons for the 'Past-Self' frame of reference are presented in Table 3.3. Comparisons between those who do and do not use a 'Past-Self' frame of reference at all are limited by cell size, with only six people identified as not considering themselves in the past. Still, significance ($p = .004$) at the 0.05 level was found when comparing the

number of tasks reported as difficult on the NHIS scale (1.6 for the 'Past-Self' group vs. 0.33 for those with 'no Past-Self' theme), and the number of activities performed on the SAFFE (8.9 and 10.2, respectively, p =.014). Mean age approached significance (p = .069), higher for the 'Past-Self' group (78.19 compared to 73.0 years).

When comparing those who consider 'Past-Self **only**' (n=11) to those who use the 'Past-Self **plus**' at least one other frame of reference (n =16), no statistically significant differences were found, although the proportion who attended college approached significance at p = .054. Everyone in the 'Past-Self only' group attended college, compared to 62.5% of those in the 'Past-Self plus' group. Additionally, the distribution of SPPB balance sub-scale scores differs between the two groups, with a greater frequency of lower scores in the 'Past-Self only' group. SPPB balance scores range from 0 (Unable to stand unsupported with feet side-by-side) to 4 (able to hold a tandem stance for the upper limit of 10 seconds), and the distribution of scores for each of the two groups is provided in Table 4. About 75% of the participants who use other frames of reference in addition to 'Past-Self' achieved the highest score on this test of balance, while 3 (27%) of the 'Past-Self only' group scored 2 or less, indicating that they cannot hold a tandem stance position for at least 3 seconds without support, and one cannot hold a semi-tandem position for 10 seconds.

The exploration into differences between those who compare themselves to others and those who do not yielded only one result that even approached statistical significance (data not shown). Mean SAFFE activity score was 9.6 for those who compare to others, nearly one point higher than the mean of 8.8 for those who draw no such comparison (p=.04)

When questioned about how they selected a response category for the frame of reference domain, a number of participants shared their own internal definitions for some or all of the five response options, ranging from 'poor' to 'excellent.' Twelve (36%) indicated that the ceiling response option of 'excellent' would imply perfect mobility or balance, and some appeared to view this to be impossible to achieve.

Selected quotations revealing personal definitions of the self-rating response options are presented in Table 3.5, along with the actual current state self-ratings provided by these same participants at the time of the interview. The selection of quotes is taken from eight participants who represent a wide variety of current state self-ratings for the domains of mobility and balance. Quotations are organized from highest to lowest state. In addition to illustrating the perception held by some that excellent equates to 'perfect' status, it is also clear that conflicting definitions of a single rating exist. Statements highlighted in bold in Table 3.5 illustrate the disparity among definitions of the response option 'Good.' While some participants viewed a rating of 'good' as unfavorable, explaining that 'good is just getting by,' and 'good starts to imply a little bit of negativity, specific problems,' for another participant, this same rating seemed desirable: 'Good means that I'm able to do what I want to do when I want to do it.'

Results for the distribution of self-report ratings across dichotomous frame of reference groups are presented in Tables 3.6, 3.7 and 3.8. Analysis of transition ratings provided at the time of interview reveal a trend toward more reports of worsening mobility for those who use a past-self frame of reference compared to those who use no such frame. With power limited by a cell size of only six for the 'No Past-Self' group, the distributions were not significantly different. Similar trends toward decline by self-rating were seen when change in state ratings was examined. Compared to the self-reported state rating provided for mobility or balance at the previous visit (approximately six months earlier), only participants who use a past-self frame of reference selected a rating at least one level lower during the interview.

No such trends were seen when those who use **only** a past-self frame of reference were compared to those who use past-self plus another frame of reference, or when groups were compared based on the use of an 'Other Comparison' frame of reference.

Results for comparison of concordant and discordant self-report groups by frame of reference failed to reveal meaningful trends. While those in the 'Past-Self Only' group appear more concordant (80%) than discordant (20%) when reported change by transition is compared to change in successive state ratings

over the same period, this observation did not persist for the balance domain. Another observation, that all five participants who did not use a past-self frame of reference were concordant, also held true only for the domain of mobility.

3.4 DISCUSSION

Through semi-structured interviews and mixed-methods analyses, insight was gained into how community-dwelling older adults self-rate global balance and mobility, and change in both domains over six months. Four primary findings emerged: 1) Definitions of balance and mobility focus on general activity themes that overlap between participants, but specific activities vary in degree of difficulty and appear to reflect relative ability and preferences. 2) Interpretations of current state response options vary, such that one person's definition of 'excellent' may be another's definition of 'good.' 3) A number of timeframes are considered when responding to a 'current status' question, and these tend to be longer than anticipated. 4) Three frames of reference appear to be used by older adults when gauging their balance and mobility status, and most people use more than one of these frames. To our knowledge, this pilot study is the first to specifically focus on global self-ratings in the older adult population for the specific domains of balance and mobility. These results could provide insight into the discordance often found between self-report and performance,[100, 103] and between self- and observer-ratings,[99, 112] as well as informing structural changes to the questions if standardized responses are desired.

Mobility and Balance Domain Definitions

Both open and closed-ended responses for definitions of mobility indicate that older adults may consider only certain activities while rating their own status, and in some cases, these activities reflect their own level of social participation and ability. While not prompted by the script, fifteen participants (45%)

volunteered that they selected closed-ended mobility definitions based on their own participation and abilities. For example, five participants who did not select the recreation response option noted that they do not perform any such activities, while six others who selected the option mentioned the specific recreation they perform. Six participants said that they were not selecting the transfer, BADL or assistance options because they have no difficulty or need for assistance in these areas. According to one woman who rated mobility as 'excellent 'and balance as 'very good,' participates in aerobics and pilates, and walks for exercise daily, the transfer option 'doesn't even need to be here.' Another participant who golfs, attends a group exercise program for seniors and works part-time noted, 'Transfers, moving in and out of bed – I don't have trouble with any of that stuff. Getting dressed, bathing – now that's ridiculous as far as I'm concerned.' This supports our initial reflection that older adults of similar ability may choose very different rating options, perhaps because they are focusing on different activities, or they have different internal definitions of the response options. Both individuals quoted previously have no difficulty with basic mobility tasks and detailed different recreational activities, but their mobility ratings range from 'excellent' to 'good.' Alternatively, those who clearly differ in mobility or balance may choose the same rating option because they are considering activities at opposite ends of the difficulty spectrum. For example, a participant selected the same 'good' mobility rating as the golfer above, but stated he does not 'do much in the way of recreational activities.' Still, he emphasized having no problem getting in and out of bed, or getting dressed or bathing. He may therefore base his 'good' rating on the ease of basic mobility tasks like walking, while the golfer may have rated himself as only 'good' because of a gradually worsening handicap on the course.

Older adults in this cohort appeared to recognize distinctions between the domains of mobility and balance, yet overlap was found. In fact, a 'balance' code emerged for the domain of mobility, supporting the clinical impression that balance is one of many contributors to mobility.[85] Still, ratings sometimes differed between domains. One of the participants who rated her mobility as 'excellent' rated her balance as only 'fair,' noting that she can no longer participate in step aerobics or roller skating because of fear of

falling, and she has difficulty standing on one leg. When asked what she was thinking of while rating mobility, she responded only, 'Getting around. . . just being able to walk without falling.' Another participant rated her mobility as 'somewhat worse' than six months ago, but her balance as 'somewhat better.' Her balance was rated as better because previous symptoms of dizziness and tinnitus have improved, but increasing leg pain from degenerative changes of the spine have worsened her mobility.

Although 29 participants (88%) indicated that falling is a theme within the domain of balance, two who did not recognize this theme admitted to a history of falls. One participant explained, 'I've had a few falls, but I don't believe they were mostly balance-related. I was just clumsy.' Another states, 'I fell being careless.'

Open-ended responses provided insight beyond closed-ended, as a number of identified themes for both mobility and balance were not predicted and therefore not provided as closed-ended responses. For example, the IADL theme was not anticipated for mobility, particularly the specific activities of driving and traveling out of town as identified by three participants. The three item list of balance definitions reflected only one of the open-ended balance themes that emerged, falls. Additionally, many closed-ended response options were selected even though the corresponding theme had not emerged during the open-ended portion of the interview. Comments made by a few participants indicate that their closed-ended selections were meant to validate our mobility definitions as plausible, but often were not the activities or themes actually being considered when they rated their own status. Both of these observations highlight the importance of including open-ended questions when exploring perceptions of older adults.

Timeframe Domain

The wide range of responses generated when participants were asked to specify the timeframe used to provide ratings of current status was unexpected. While some questionnaires ask for ratings over a

specific period of time, others give broader instructions such as 'now.' For example, while most of the later questions in the MOS SF-36 questionnaire[28] specify a timeframe of 'the past 4 weeks,' the opening health status question after which our current state questions were structured asks participants only to rate their health 'in general,' and then specifies 'now' compared to one year ago for the transition question. We anticipated that without specifying a timeframe, older adults would rate their recent status over the past few days to a month. We did not anticipate the majority (62.9%) of the 27 participants who were able to specify a rating giving responses of six months or longer. In particular, 41% of these 27 individuals, all participants in a longitudinal study with visits every six months, revealed timeframes of greater than six months when rating current status. This question was posed shortly after participants rated not only their current status for each domain, but also any change in status over the past six months. While this may have influenced some to specify a timeframe of six months for the current state rating, it makes the ratings of greater than one year even more interesting. This means that 11 participants considered the past year when rating current status, but then rated change relative to their status six months earlier. This could be one reason why we have found discordance in change in self-ratings when measured directly (using a six-month transition question) versus indirectly (by comparing answers to state questions provided now and six months ago).

While some report perceived fluctuations in their mobility and balance from day to day that would change their rating by at least one level, others indicate that their rating reflects a status that is more stable over time. A 72 year old participant who responded 'about the same' to both transition questions, rated her mobility as 'excellent' and her balance as 'very good,' noted that she has not changed much in at least the past 15 years. 'Everything seems to be like the status quo. I don't recall having any major ups or downs. . . so I can say it's about the same, because to me, it basically is.' However, some older adults with less stable mobility may take a broader view when providing a self-rating by considering their average or usual abilities and excluding daily fluctuations that they view as temporary or infrequent. [71] One participant who missed the six month follow-up experienced a decline in SPPB from 11/12 at baseline to

8/12 one year later, when she became unable to stand from a chair unassisted. At both timepoints, she rated her mobility as 'very good' and her balance as 'good,' explaining, 'A problem with my knee is very recent and has affected my abilities in the past three days, but I'm really answering the questions based on the time before the past three days. . . so I look at before these past three days because this will go away. I'm sure that whatever it is, when the doctor diagnoses it, I'm sure there will be something done for it.' As a result, the difference in her physical performance at two timepoints is not accompanied by a corresponding change in self-reported mobility and balance. This situation provides one explanation for discordance between self-report and performance. While performance may represent status on a single day, or even at a specific time of day, self-report may be more general, reflecting a longer timeframe.

Results related to the transition rating of any change over the past six months were consistent with the literature. Our finding that older adults tend to consider their current status more than their prior status when rating transition in balance or mobility supports earlier findings,[31, 34] but those results were not specific to the domains of mobility and balance, or to older adults. Recall bias is a well-established disadvantage of retrospective measures like transition-based ratings,[32, 38, 151] and may be expected to be even worse in an older adult population,[87, 88] however, our results indicate that recognized recall bias or memory transience is low in this population of relatively high-functioning and highly educated community-dwelling elders. While unrecognized recall bias could still exist, recognized recall bias may have been minimized by our study design, as the timepoint for past comparison was marked by a study visit during which participants performed multiple tests of balance and mobility. This event may have been memorable enough to minimize recall bias, but we have identified a situation we term 'Timeframe Incongruency' or TFI that poses another threat to the validity of transition-based ratings. We have speculated that intervening events like hospitalizations over the six month period could contribute to TFI's. People may think back to the event and feel that they are now much better, even though they continue to lag behind their status six months ago, prior to the event. Here we found that only one of those with TFI also identified an intervening event, however, conclusions based on these 'TFI' statements

are limited by two main factors. First, the statements were usually made long after the transition ratings had been given instead of during or shortly after the rating, and the incongruency was never recognized by the participant. This raises the possibility that statements coded as TFI actually represent a 'slip of the tongue' in which a participant referred to some prior timepoint other than six months, but had made a six-month comparison when answering the transition questions. In addition, the interview was not designed to elicit TFI statements, so this theme was identified only when participants volunteered such a statement at some point during the interview. Because all participants were not 'screened' for TFI's, any of the others could also fit this theme.

Frame Of Reference

Three frames of reference appear to be used by community-dwelling older adults when gauging their balance and mobility status, each based on a different comparison: 1) A comparison to one's own abilities in the past ('Past-Self'), 2) A general comparison to an expectation held for people of a specific age group, usually the age group to which the participant belongs ('Age Expectation'), and 3) A comparison to specific other individuals with whom the participant has a relationship. These were the same three frames anticipated when initially developing our closed-ended response options, however, we did not anticipate that most in our cohort would use more than one of these frames. This was revealed primarily through comments made about frames of reference at various points during the interview, most revealed before the closed-ended options were ever presented. In fact, a few participants began to reveal frames of reference while rating their own current state at the start of the interview, long before any prompts were given.

Open- and closed-ended responses were merged into a summary theme for frame of reference. A participant who selected the 'Other Comparison' option may have noted while rating his mobility that he was referencing his status in the past, and those revelations seemed as or more important to an understanding of the self-rating process than the closed-ended response selections. As a result, only

summary themes were used in all further analyses. Participants were grouped dichotomously according to the themes of 'Past-Self' and 'Other Comparison,' but the theme of 'Age Expectation' was not examined separately, as the distinction from 'Other Comparison' was challenging in many cases. We distinguished the groups based on whether a comparison was made to specific people ('Other Comparison') or a more general, theoretical statement was made about a certain age group ('Age Expectation'). However, we recognize that phrases like 'for my age. . .' imply a comparison to an internal view of the 'norm' for people of that age group, and this norm is likely based on previous personal interactions.

The most commonly identified frame of reference involved a comparison to one's self at some point in the past, and this was also the most common frame of reference for those identified as using only a single frame. While sample size limitations allow conclusions to be drawn from the Table 3.3 comparisons only with caution, hypotheses may be generated for further testing. While statistical significance was rarely found and not anticipated in a pilot study of this nature, the mean and frequency differences between frame of reference groups may have clinical relevance. In addition to the statistically significant mean difference in BADL/ IADL difficulty between those who do and do not use a 'Past-Self'' frame of reference at all, comparisons of means for age and gait speed may also have clinical meaning. The results may suggest that in addition to having difficulty with one additional BADL or IADL, those who use a 'past-self' frame of reference may be somewhat older, walk more slowly, and be less active than those who do not reference their past ability. The findings for age are no surprise, as individuals who have lived longer have more of a past to use for comparison, and a greater chance of having declined from their previous state. Lower gait speed and greater BADL/IADL difficulty could make a comparison to one's self in the past hard to avoid.[89]

While thirteen participants used both the 'Past-Self' and 'Other Comparison' frames of reference, eleven used only the 'Past-Self' frame of reference, and a few of these individuals made comments in which they

voiced strong opposition to drawing a comparison to others. This led us to further split the 'Past-Self' group into a 'Past-Self only' group and a 'Past-Self plus,' and these comparisons were even more informative. Based on Table 3.4 comparisons, the distribution of SPPB balance scores varies significantly between these two groups, with a greater frequency of higher balance scores for those who use other frames of reference in addition to 'Past-Self.' These statistically significant findings support the clinically significant findings otherwise noted from Table 3. The 'Past-Self only' group walks 0.6 m/s more slowly, a difference greater than the 0.5 m/s threshold reported to be small but meaningful change.[35] Additionally, their total SPPB scores are almost one full point lower. According to Perera, a difference of 0.5 points on the SPPB is small but meaningful change and a full point is substantial. [35] Taken together, these results suggest lower mobility for the group who restricts to a single 'Past-Self' frame of reference. While not statistically significant, lower balance and mobility may explain why they appear to be more likely to fear falling, and Table 3.3 suggests a trend toward fewer falls that may be explained by greater activity restriction and lesser activity. People who limit themselves to considering only their own past abilities with no comparison to others or age-based expectation appear to be more highly educated, and this knowledge may help them take action (possibly by restricting activity) to prevent falls. Alternatively, it is possible that lower mobility leads people to focus only on their own prior status and avoid any comparison to others, many of whom appear to be doing better.

The hypothesis above is supported by the distribution of self-report ratings in this cohort (Table 3.6). All ratings of 'Poor' or 'Fair' for both domains were made by participants who use a 'Past Self' frame of reference with no comparison to others, while nearly all 'excellent' ratings for both domains (1 out of 1 for balance, 4 of 5 for mobility) were given by people who use an 'Other Comparison' frame of reference. This is further supported by descriptive results found within the 'Other Comparison' group. Specifically, participants drew a comparison to people perceived as less mobile than themselves far more often than they compared to others believed to have higher mobility. Of 16 total participants in the 'Other Comparison' group, 10 felt they were better than those to whom they compare and 3 others did not

specify the direction of comparison. This left only 3 who compared themselves to people of higher mobility, and one of those used his grandchildren as a benchmark for mobility and balance.

Another example of the relationship between self-rating and frame of reference was provided by a participant with very different ratings for the domains of mobility and balance. She rated her mobility as excellent, noting that she compares herself to other people, but when rating balance, she drew comparisons to her own balance when she was younger. As a result, she rated her balance as only fair, detailing high-level recreation activities that she can no longer perform without fear of falling. Overall, we found that older adults who consider their status at some point in the past when rating their current balance and mobility are more likely to apply ratings of 'Fair' and 'Poor' than those who do not use a 'Past-Self' approach (Table 3.6). They may also be more likely to report worsening over a six month period using both a direct approach (transition rating of perceived change in mobility, Table 3.7) and an indirect approach by which successive state ratings of balance are compared to determine any change (Table 3.8). We have observed discordance when change in self-report of mobility and balance is measured using these same two approaches, however, no initial hypotheses based on frame of reference could be identified from the results of this exploratory pilot work (Table 3.9).

Our between group findings for the 'Past-Self' theme are more informative than those between the 'Other Comparison' groups. Few of the latter comparisons even approached statistical significance. Whether or not an older adult compares herself to others when providing self-ratings does not appear to depend on individual characteristics like gait speed and BADL/IADL ability, however, it should be noted that these comparisons are confounded by the use of other frames of reference. In all but one case, our participants who compared themselves to another person or group with whom they have had some form of contact use an additional frame of reference when rating their balance and mobility, either a comparison to their own abilities in the past, or an age-based expectation. Fear of falling was about twice as common in the group who do not compare to others, but the proportion who fell over the past year was nearly the same. This

may be due to differences in activity level, as it appears that the group who do not compare to others are less active, with a SAFFE Activity score nearly one full activity lower. Obviously, 10 of the 17 participants in the 'No Other Comparison' group also populate the 'Past-Self only' group, which would explain why some of these findings approximate those described previously for the comparison of 'Past-Self only' and 'Past-Self plus.'

Any study of self-report must address the influence of cognition and mood, and this is even more important in a study of older adults, as declines become more common with advancing age.[87, 88] With a mean DSST score of 48.0 $\pm$ 9.3 (range 27 to 64) and mean GDS of 1.75 $\pm$ 2.0 (range 0 to 8), our cohort did not appear particularly depressed or cognitively impaired with respect to speed of perceptual processing. Moreover, no significant differences in cognition or mood were found for any of the two group comparisons made based on frame of reference. We hypothesize that personality characteristics may play a larger role than cognition and mood in the frame(s) of reference used when rating one's own status, and our protocol did not include appropriate measures to test this theory.

Interpersonal variability in definitions of the self-rating response options, particularly the central option of 'good' as illustrated in Table 3.5, may help to explain some of the discordance reported between self – ratings and performance on a measure thought to represent the same domain, or between self- and observer-ratings. This may be especially true when those observers are healthcare providers who are basing their ratings on a continuum of performance, often with published norms or meaningful thresholds as with gait speed.

The generalizability of these interview findings may be limited to highly educated older adults. Nearly 80% of our interview cohort attended at least one year of college, and one-third of the sample reported some post-graduate training. Because our focus is on the personal perceptions of these older adults, our conclusions may not generalize to a less educated or more racially diverse cohort. Additionally, we would classify these older adults as having moderate to high mobility function, so the results may differ

in a cohort with greater mobility disability. Still, one of our primary interests is recognizing decline in its earliest stages, and promoting resolution of even mild mobility deficits, so this sample is appropriate to meet those aims.

Another limitation that cannot be overlooked is the potential for participation in one to three prior study visits to influence the perspectives shared during the interviews. Participation in balance and mobility testing during earlier study visits may have shifted responses,[93, 120] resulting in follow-up ratings that are either higher or lower than at baseline, depending on perceived performance on our tests. Descriptive review of the frequency of ratings across all timepoints does not support such a shift for the group as a whole (data not shown), however, exposure to our tests of balance and mobility, generally the more challenging ones, was referenced without prompting by seven participants (21%) over the course of the interviews. Quotation 2 in Table 3.5 is an example of reference to ratings based on high-level balance activities, and the participant specifies that they selected 'very good' instead of 'excellent' because of their earlier performance.

As mentioned previously, our between-groups comparisons based on frame of reference are underpowered, however, we aim only to generate hypotheses for further testing in a larger cohort. We recognize the need to correct for multiple comparisons, but given the exploratory nature of this pilot study, the resultant increase in type II error is of more concern than the risk of Type I error. Any conclusions we have drawn are meant to provoke further discussion rather than propose a true difference between the groups.

In summary, the findings from this mixed-methods approach with semi-structured interviews provided new insight into the process by which older adults rate their own balance and mobility. A variety of activity and non-activity themes emerge when community-dwellers are asked to rate their own balance and mobility, and the level of difficulty of the activities considered seems to impact the rating selected. Older adults also hold personal definitions of the 'poor' to 'excellent' response options that vary widely

between different people and may help to explain why self-ratings can seem so disparate from performance. The timeframe considered when rating 'current status' is more broad than anticipated for many older adults, and this may explain some of the discordance found when comparing transition ratings to the change in state ratings obtained at the same timepoints. Finally, the frame of reference used, particularly the use of a comparison to one's own performance in the past, may impact the rating selected, or underlying balance and mobility deficits may lead the individual to avoid the use of certain frames, specifically a comparison to others.

Given the above findings, we recommend some revisions to the global questions as they were asked of participants in this study (Appendix). First, stems should be added to specify a timeframe for the ratings of current status. To avoid a focus on daily fluctuations by individuals with more transient status or chronic conditions such as dizziness, we could follow the 4 week timeframe used for many of the questions in the MOS SF-36.[28] To standardize responses, we could provide 1) sample activities for each of the domains of interest, 2) brief definitions of the response options and/or 3) the frame of reference to be used when selecting a response. Unfortunately, specifying activities to be rated may induce floor or ceiling effects (depending on the mobility status of the sample), thereby decreasing responsiveness, and if a range of activities of varying levels of difficulty were provided, individuals would probably continue to focus on one end of the spectrum based on their own abilities. Providing definitions of the response options is likely to also involve specifying a frame of reference to be used.

Asking participants to use a certain frame of reference presents a few challenges. A 'Past-Self' frame of reference was used most commonly in this cohort, but we would then need to specify the timepoint for past comparison, and the current status question essentially becomes a transition question with a longer timeframe. Requesting a 'Past-Self' frame of reference may bias older adults to report greater change, particularly worsening of status, and any reported change may not be functionally significant from a healthcare perspective. Asking for an 'Other Comparison' may result in missing data from those who

insist on only a ‘Past-Self’ frame. In addition, an ‘Other Comparison’ frame is harder to standardize, as the individuals who serve as the basis for comparison will vary for each person. Future work should establish whether those who use more than one frame of reference have a ‘primary’ frame on which they rely most. Comparisons by primary frame of reference may provide greater insight into differences between groups.

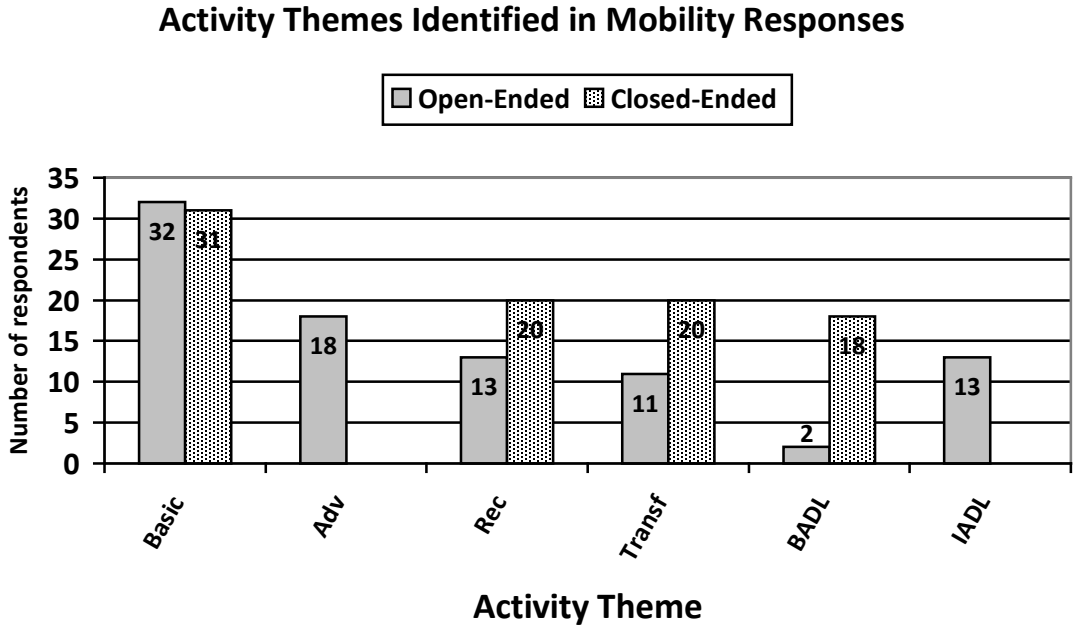

FIGURE 3.1: Themes Identified in Open-Ended and Closed-Ended Definitions of Mobility

KEY:

- **Basic = Basic Mobility:** walking on level surfaces, general statements like 'moving around'
- **Adv = Advanced Mobility**: walking on steps or curbs, in crowds, longer distances, uneven surfaces, in unfamiliar environments
- **Rec** = **Recreation**: includes seated exercise, swimming, golfing, aerobics, tennis, rollercoasters
- **Transf** = **Transfers**: Bed, chair or floor transfers, car transfers, bending or stooping
- **BADL** = Basic Activities of Daily Living (includes personal hygiene, dressing, eating, toileting)
- **IADL** = Instrumental Activities of Daily Living (includes shopping, housework/ meal prep, working outside the home, and traveling out of the community or driving)

Note *that the themes of Advanced Mobility and IADL were not presented to participants as closed-ended response options during the interview. The absence of respondents for these themes reflects that they were not available rather than not selected.*

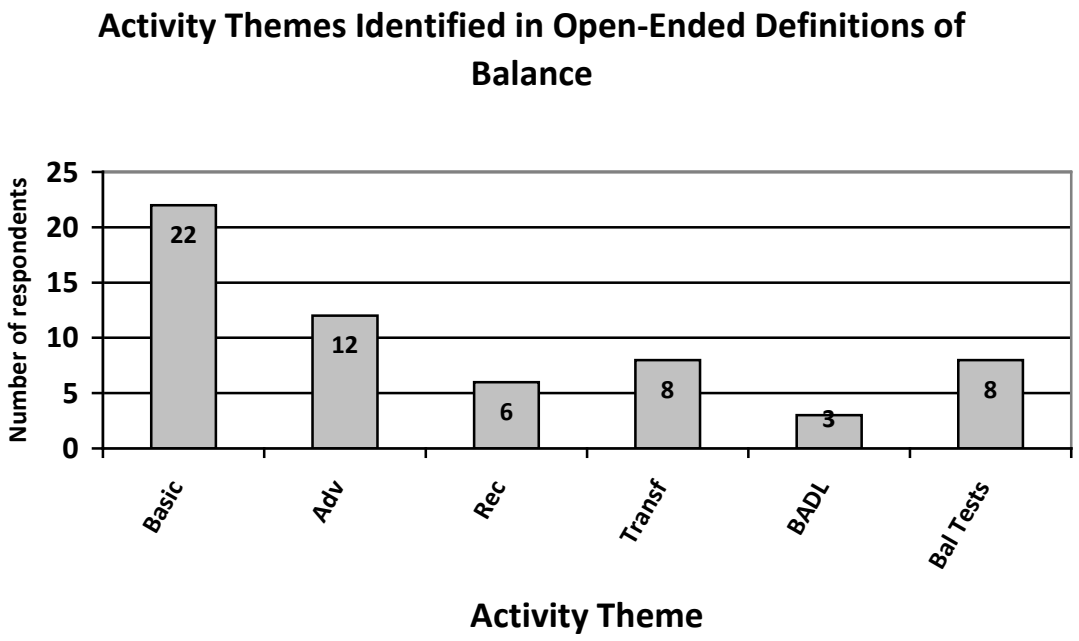

FIGURE 3.2: Themes Identified in Open-Ended Definitions of Balance

KEY:

- **Basic = Basic Balance:** Standing still, standing straight, walking on level surfaces
- **Adv = Advanced Balance Tasks**: Climbing steps or ladders, fast movements, head turns, stepping in different directions such as backwards or to the side, lifting, carrying, reaching on tip-toes, moving through crowds, being pushed or bumped
- **Rec** = **Recreation**: includes bowling, dancing, Tai Chi, step aerobics, tennis
- **Transf** = **Transfers**: Bed, chair, and car transfers, bending
- **BADL** = Basic Activities of Daily Living (includes bathing/showering, dressing)
- **Bal Tests** = Tests of balance: Standing on one leg or in tandem, walking 'tape ladders' on the floor, tandem walking

TABLE 3.1: Selection of Closed-ended Responses for Definition of Balance

Response Option	**# of Respondents (% of total) n=33**
Being steady on my feet	29 (88%)
Not being shaky	12 (36%)
Being confident that I will not fall	30 (91%)

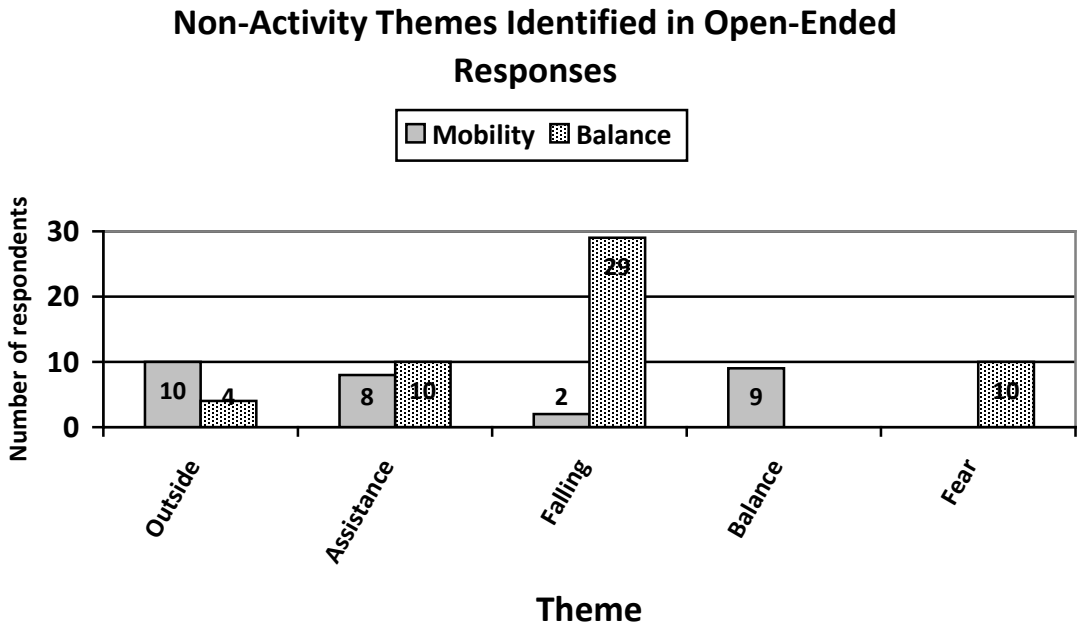

FIGURE 3.3: Non-Activity Themes Identified in Open-Ended Definitions of Balance and Mobility

KEY:

- **Outside:** Walking outside, negotiating curbs in the community, reference to garden or woods
- **Assistance**: Mentions either needing or NOT needing assistance from another person or the use of assistive devices like canes or handrails
- **Falling:** Specifically refers to falling
- **Balance**: Uses the term 'balance,' or any related term (e.g. steadiness, staggering); *This theme was identified only for the mobility domain*
- **Fear:** = Expresses fear by using the term 'fear,' or any related term such as 'worry' or 'concern'

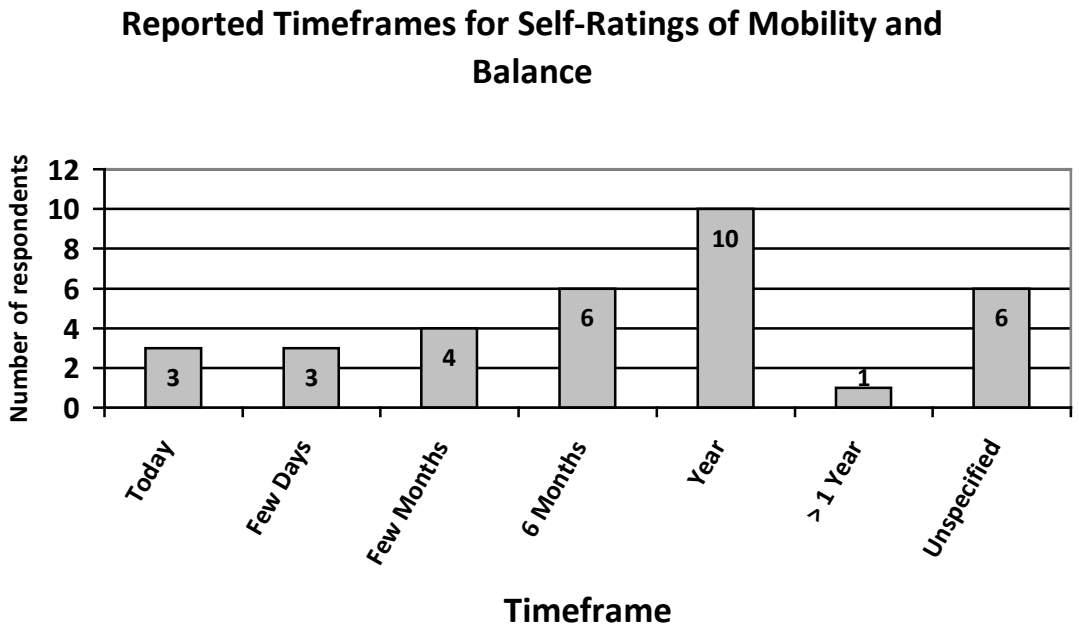

FIGURE 3.4: Timeframes of Global Mobility and Balance Self-Ratings

TABLE 3.2: Selection of Closed-ended Responses for Frames of Reference

FRAME OF REFERENCE	No. of Repondents (% of Total) n=33		
	Closed-Ended	**Open-Ended**	**Summary Theme**
THEME 1: PAST-SELF			
Overall Results: (Closed: '*How I did at some point in the past'*)	25 (75.8%)	16 (48.5%)	27 (81.8%)
Sample Open-Ended Quotations: • 'My answers represent me in comparison to my own life, what I used to be like.' • 'Last year I went roller skating and I realized that my balance has disintegrated because I haven't done that for years, and I thought I was going to get out there and skate. Well, I realized how much worse my balance was.' • 'I'm not as good at physical things as I was, but I'm comparing it to myself.'			
THEME 2: AGE EXPECTATION			
Overall Results: (Closed: '*How I think I should be doing at my age'*)	12 (36.4%)	7 (21.2%)	15 (45.5%)
Sample Open-Ended Quotations: • 'There's no way at my age I could feel excellent. Even very good is a stretch.' • 'I feel as though I do well for my age, and I do everything at home, and so far, so good.' • 'At 77 years old, you can't do what you did at 20, so I'm better than a lot of people my age.'			
THEME 3: OTHER COMPARISON			
Overall Results: (Closed: '*A specific person I know'*)	4 (12.1%)	16 (48.5%)	16 (48.5%)
Sample Open-Ended Quotations: • 'Compared to you I'm probably poor, but when I think of my friends, I would say good. A lot of them are walking with canes or not walking.' • 'I compare myself to my Dad. He was 94 and still doing well, so I sort of judge myself with him really. I'm trying to keep up with him.' • 'I'm good to excellent compared to them [friends 10 years younger]. Some of them have heart problems and can't walk real well, and can't breathe real well, and they have knee problems or hip replacements. . . I feel blessed because my problems [cancer] have not interfered with my balance and mobility.'			

TABLE 3.3: Comparison of Means Based on Frame of Reference

	Past-Self		Past-Self Only	
	Past-Self (n=27)	No Past-Self (n=6)	Past-Self Only (n=11)	Past-Self Plus (n=16)
CHARACTERISTIC				
Age (years)	78.19	73.00	80.18	76.81
Gender (% female)	70.4	83.3	72.7	68.8
Race (% black)	11.1	16.7	9.1	12.5
Educational Level (% college educated)	77.8	83.3	100	62.5
Comorbidities*	2.81	3.0	2.7	2.9
Geriatric Depression Scale (0-15)*	1.8	1.5	1.8	1.8
DSST	47.0	52.3	45.8	47.8
ERA-12 (0-12)	53.1	57.4	55.8	51.2
Gait Speed (m/sec)	0.97	1.09	.93	.99
SPPB Summary Score (0-12)	9.89	9.83	9.36	10.25
BADL/IADL difficulty (0-14)	1.60**	0.33**	1.80	1.47
ABC (0-100%)	78.5	85.7	79.0	78.2
Fear of Falling (% fearful)	40.7	66.7	54.5	31.3
Falls (% Fell Past Year)	37.0	33.3	27.3	43.8
SAFFE Activity (0-11)	8.9**	10.2**	8.6	9.2
SAFFE Restriction (11-0)	2.7	2.2	3.0	2.5

*These measures represent data collected at the baseline study visit. All other data were collected at the timepoint of the interview, either 12 or 18 months from baseline.
**Significant at the 0.05 level by independent samples t-test.

TABLE 3.4: Comparison of SPPB Balance Sub-Scale Scores Between 'Past-Self Only' and 'Past Self Plus' groups

SPPB BALANCE SCORE **(Clinical Meaning)**	**NUMBER (PERCENT) WITH EACH SCORE**	
	PAST-SELF ONLY N=11	***PAST-SELF PLUS*** N=16
1 ***(Holds side-by-side for 10 sec)***	1 (9%)	0 (0%)
2 ***(Holds semi-tandem for 10 sec)***	3 (27%)	2 (12.5%)
3 ***(Holds full tandem for 3 sec)***	4 (36%)	2 (12.5%)
4 ***(Holds full tandem for 10 sec)***	3 (27%)	12 (75%)

TABLE 3.5: Personal Meaning of Current-State Response Options

Sample Open-Ended Quotation	Current State Rating	
	Mobility	***Balance***
Quote 1: Excellent would mean no problems whatsoever, having the body, will, and vigor of a 25-year old. Very good – I would look at it as good for my age. **Good starts to imply a little bit of negativity, specific problems.** If I stop to think, yeah – I worry about my balance, especially if I try to stand on one foot or close my eyes.	Excellent	Very Good
Quote 2: . . . I can't stand for 30 seconds on one leg, and I say that precludes me from having excellent balance because I can't do special exercises other than just regular walking. I recall that standing on one leg was one of the activities that I was judged on for this particular test . . .	Excellent	Very Good
Quote 3: I don't have excellent balance or mobility. Excellent would mean like if you walk and have no problems. Everything is working good and you go about your business. I think mine is like very good – when you don't have to worry about what you're doing. You feel confidence that you would do it. **Good is that you're just getting by**.	Very Good	Good
Quote 4: Excellent I think is consistent. . . I'm always up to doing something. Very good might mean some days I don't feel like doing it.	Very Good	Good
Quote 5: . . . I think I'm better than fair because I'm able to do things that I want to do.	Good	Good
Quote 6: **I'm doing very well when I say good**, and a few things wrong, I say fair.	Good	Fair
Quote 7: Well, I say mine is fair because I sort of remember what I used to be, and what I was able to do before, and now I can't. . . like I can't run, and I used to be able to run. (*Interviewer: 'If you could run now, would you say that was excellent?*') Not excellent, but good.	Fair	Fair
Quote 8: **Good means that I'm able to do what I want to do when I want to do it.** Fair means that I have to think twice before I can do something, if I'm going to do it at all. Poor means I can't do it. . . '	Fair	Poor

TABLE 3.6: Distribution of State Ratings by Frame of Reference

STATE RATING	FRAME OF REFERENCE			
	Other Comparison n=16	No Other Comparison n=17	Past-Self* n=27	No Past-Self* n=6
MOBILITY				
Excellent	4 (25%)	1 (6%)	**2 (7%)**	**3 (50%)**
Very Good	4 (25%)	6 (35%)	8 (30%)	2 (33%)
Good	4 (25%)	2 (12%)	5 (18.5%)	1 (17%)
Fair	4 (25%)	6 (35%)	**10 (37%)**	**0 (0%)**
Poor	0 (0%)	2 (12%)	**2 (7%)**	**0 (0%)**
BALANCE				
Excellent	1 (6%)	0 (0%)	**0 (0%)**	**1 (17%)**
Very Good	2 (12.5%)	3 (18%)	**3 (11%)**	**2 (33%)**
Good	8 (50%)	8 (47%)	13 (48%)	3 (50%)
Fair	5 (31%)	3 (18%)	**8 (30%)**	**0 (0%)**
Poor	0 (0%)	3 (18%)	**3 (11%)**	**0 (0%)**

*Significant at the .05 level using Mann-Whitney U Test (***Mobility p = .010***, ***Balance p = .013***)

Comparison of state rating distributions across 'Other Comparison' groups is **not** significant at the .05 level *(Mobility p = .157, Balance p = .575)*

TABLE 3.7: Distribution of Transition Ratings by Frame of Reference

TRANSITION RATING	FRAME OF REFERENCE			
	Other Comparison N=16	No Other Comparison N=17	Past-Self N=27	No Past-Self N=6
MOBILITY				
Better	1 (6%)	3 (18%)	4 (15%)	0 (0%)
About the Same	10 (62.5%)	3 (18%)	8 (30%)	5 (83%)
Worse	5 (31%)	11 (65%)	15 (56%)	1 (17%)
BALANCE				
Better	1 (6%)	2 (12%)	2 (7%)	1 (17%)
About the Same	11 (69%)	9 (53%)	16 (59%)	4 (67%)
Worse	4 (25%)	6 (35%)	9 (33%)	1 (17%)

Comparisons of transition rating distributions across both of the 'Other Comparison' and 'Past-Self' groups were **not** significant

TABLE 3.8: Distribution of Six-Month Change in State Ratings by Frame of Reference

STATE RATING	FRAME OF REFERENCE			
	Other Comparison n=16	No Other Comparison n=15	Past-Self n=26	No Past-Self n=5
MOBILITY*				
1 Level Better	0 (0%)	1 (7%)	1 (4%)	0 (0%)
No Change	13 (81%)	8 (53%)	16 (61.5%)	5 (100%)
1 Level Worse	3 (19%)	3 (20%)	**6 (23%)**	**0 (0%)**
2 Levels Worse	0 (0%)	3 (20%)	**3 (11.5%)**	**0 (0%)**
BALANCE*				
1 Level Better	3 (19%)	1 (7%)	2 (8%)	2 (40%)
No Change	7 (44%)	10 (67%)	14 (54%	3 (60%)
1 Level Worse	6 (37.5%)	3 (20%)	**9 (35%)**	**0 (0%)**
2 Levels Worse	0 (0%)	1 (7%)	**1 (4%)**	**0 (0%)**

* Mann-Whitney U Test for Past-Self compared to No Past-Self, for mobility p=.193, for balance p=.032 . Other vs No Other Comparison groups not significant.

TABLE 3.9: Concordance/Discordance Between Change in State and Transition Self-Ratings by Frame of Reference

		CONCORDANT n (%)	**DISCORDANT** n (%)
MOBILITY	**Other Comparison,** n =16	11 (69%)	5 (31%)
	No Other Comparison, n=15	10 (67%)	5 (33%)
	Past-Self, n=26	16 (61.5%)	10 (38.5%)
	No Past-Self, n=5	5 (100%)	0 (0%)
	Past-Self ONLY, n=10	8 (80%)	2 (20%)
	Past-Self PLUS, n=16	8 (50%)	8 (50%)
BALANCE	**Other Comparison,** n=16	7 (44%)	9 (56%)
	No Other Comparison, n=15	9 (60%)	6 (40%)
	Past-Self, n=26	14 (54%)	12 (46%)
	No Past-Self, n=5	2 (40%)	3 (60%)
	Past-Self ONLY, n=10	6 (60%)	4 (40%)
	Past-Self PLUS, n=16	8 (50%)	8 (50%)

4. RELATIONSHIP BETWEEN CHANGE IN SELF-REPORT AND PERFORMANCE-BASED CHANGE FOR MOBILITY AND BALANCE

4.1 INTRODUCTION

Discordance between self-report of abilities and performance has been recognized for decades. Although much work has been published comparing self-report of status to results of performance-based measures thought to reflect the same domain, much of this work has been cross-sectional rather than prospective. [96, 99, 100, 103]

One study found discordance between self-report measures of walking and actual walking performance, with 13.6% of people unable to complete an 8 foot walk reporting that they could walk one-half mile without help. [105] A global question about mobility or balance may be more responsive than specific questions in picking up more subtle change in abilities, as people may interpret the global questions to encompass any areas in which they are having difficulty, even in tasks not typically explored in our questionnaires. Screening for mobility is important because mobility disability (sometimes defined as walking one-half mile and climbing stairs) has been shown to precede disability with activities of daily living.[74-76]

Ultimately, the debate about relative contributions of the two methods for measuring functional ability rages on, and the role that each should play in drawing conclusions in both clinical and research settings is unclear. Additionally, there is no definite answer in the literature to the question of which performance-based measures best reflect the domains of balance and mobility as defined by community-dwelling older adults.

Measuring mobility in older adults can be complicated.[99, 109] While measures such as gait speed and the Short Physical Performance Battery[105] require little time, they do require contact with an individual who may have difficulty finding transportation to the clinic. Additionally, they represent performance at a moment in time, possibly better or worse than the patient's 'average' daily ability. Self-report measures of daily activities are useful, but have their own disadvantages: people may over- or under-report ability,[112] and we hypothesize a number of reasons including preferences for a certain living environment or social interaction with caregivers. Anecdotally, patients are sometimes reluctant to complete lengthy self-report measures, may require assistance to complete them in a clinic setting, and do not always return them when taken home for completion.

While questionnaires with more specific items are advantageous in some situations,[64-66] it can be difficult to find a questionnaire that spans a wide enough range of items to fit people of all abilities, and to cover a multi-faceted domain such as mobility. In order to catch the area in which an individual is beginning to decline, a number of questions may be required, and this increases burden for both the patient and provider. Even then, interpretation of an item could cause the older adult to answer in a manner that does not reveal their new difficulty.

Self-report may also provide more information about overall mobility than commonly used performance measures such as gait speed, or even a multi-pronged test battery like the Short Physical Performance Battery.[105] Fried and colleagues found self-report of mobility to provide additional predictive information about incident difficulty and disability not provided by walking speed.[20, 124] Moreover, according to Bandura's social cognitive and self-efficacy theories,[82, 83] actual performance is determined more by beliefs than actual capacity, so a decline in perception of mobility ability may have adverse effects on social participation that ultimately lead to a decline in performance.[76, 117, 118] We are interested in the ability of older adults to detect change in their own mobility, particularly decline as an age-based

phenomenon, which may be very different than detecting decline after a major event (hospitalization) or improvement in response to an intervention.

Our exploration into the measurement of change in perceived mobility and balance in an older adult cohort would not be complete without a comparison between any change in self-ratings and performance-based change over the same period. Earlier when we compared state and transition-based global ratings, we showed that change assessed using self-ratings often does not agree when measured directly (using transition questions to gauge the individual's perceived change over a period of time) or indirectly (by comparison of successive state ratings obtained at those same times). In order to use this information to improve the measurement of mobility in older adults, we seek to determine which of these methods (state or transition-based) most closely agrees with any change in performance.

Purpose:

The purpose of this investigation is to quantify the relationship between change measured using global self-ratings of mobility and balance and measured change in performance-based tasks thought to represent the same domains. Specifically, we seek to determine which method of measuring change in self-ratings (state or transition-based) relates more closely to measured change in performance-based tasks over the same 6-month period. For the domain of mobility, we will then examine whether individual characteristics such as age, aging expectations, or mobility-related functional status determine the method of self-rating that more closely agrees with performance.

4.2 METHODS

4.2.1 Participants

One hundred nineteen community-dwelling adults age 65 and older participated in an observational study to refine protocols and develop measures for the assessment of balance and mobility in older adults.

Participants completed a battery of self-report and performance-based measures upon enrollment, and then approximately six and twelve months later. In addition, nine participants were invited to participate in an 18-month addendum study visit conducted in their own homes. During this visit, an abbreviated battery of self-report and performance-based measures of mobility, balance, and function was administered. Addendum home visits were offered based on time since the previous study visit. The cohort represented a range of functional abilities, but all were independently ambulatory for at least household distances using an assistive device if needed.

Participants were recruited from the University of Pittsburgh Claude D. Pepper Older American's Independence Center's (OAIC) Research Registry. The Registry is comprised of community-dwelling adults who agree to be contacted for participation in studies of balance and mobility. Effort was made to sample a wide range of functional mobility and balance by pre-screening over the phone for self-reported mobility and perceived balance. Individuals were excluded for unstable medical conditions (cancer with current treatment, angina), or for progressive or persistent neuromuscular conditions (stroke, Parkinson's) or pain that restrict movement.

4.2.2 Measures

Primary Self-Report Measures

Self-Rated Mobility and Balance: Global state and transition self-ratings of balance and mobility were obtained every 6 months for one year, along with a four hour battery of other self-report and performance-based measures felt to represent primarily the domains of balance and mobility. Participants were not reminded of the self-report ratings they had given six months earlier, and global self-ratings were obtained before any performance-based tests of balance or mobility were administered in that session. The latter was done out of concern that first attempting high-level balance tests such as unsupported

tandem and unilateral stance could influence balance and mobility self-ratings to reflect test performance rather than real-world functional ability. State questions were modeled after the Medical Outcomes Study SF-36 question of current health status[28] and simply asked participants to rate their 'level of *mobility* (alternatively *balance*) in general' with no further description of each domain or the specific period of time. Response options were provided on a 5-point scale of 1) Excellent, 2) Very Good, 3) Good, 4) Fair, and 5) Poor (Figure 4.1). Participants were then asked the transition question, 'In the past six months, has there been any change in your overall mobility (alternatively 'balance')? Response options were provided on a 7-point Likert-based scale of 3) Much Better, 2) Somewhat Better, 1) A Little Better, 0) About the Same, -1) A Little Worse, -2) Somewhat Worse, and -3) Much Worse.

Primary Performance Measures

Gait Speed: Gait speed was measured in meters per second (m/s) using an instrumented walkway (GaitMat II) of approximately 6 meters in length. To avoid including acceleration and deceleration, recordings were taken only from the central 4 meters of the walkway. Up to four 'passes' were performed according to participant tolerance, and gait speed was averaged over all passes. Participants used their own assistive device if needed and were closely spotted by research staff, but were not allowed physical assistance from another person. If GaitMat gait speed could not be determined (due to equipment malfunction or administration of study measures in the home), the mean time to complete the SPPB timed walk was converted to velocity and substituted for the GaitMat data. Because the SPPB 4-meter timed walk is measured from a standing start, a conversion factor was developed by our statistician using both the GaitMat and SPPB timed walk data available on all participants at baseline. Gait speed in older adults is highly predictive of adverse outcomes, including institutionalization and mortality.[138-141]

Short Physical Performance Battery (SPPB):[105] The SPPB quantifies functional mobility using three categories: standing balance, walking, and repeated chair stands. Performance in each category is timed, and the time is converted to an ordinal score ranging from 0 (unable to perform) to 4 (best performance),

used to calculate a summary score (maximum 12). The full protocol and training instructions can be downloaded from the NIH website. Test-retest and inter-rater reliability of each of the three categories range from 0.73 to 0.97,[104, 136, 137] and the reliability of the summary scale was established using internal consistency (Chronbach's alpha = 0.76). [105] Originally validated in over 5,000 adults age 71 and older, summary scores correlated strongly with self-reported ADL disability for tasks like walking across a room and transferring from bed to chair; and with self-reported ability to walk up and down steps or walk ½ mile without help.[75] Summary scores were also found to strongly predict mortality and nursing home admission.[105]

Timed Up and Go:[47] The time in seconds required to stand from a chair with armrests, walk 3 meters using any assistive device, turn, walk back, and sit down was measured and averaged over two trials. Normal, healthy elderly usually complete the task in under 10 seconds, while completion times greater than 30 seconds are consistent with functional dependence and use of an ambulatory assistive device.[47]

Figure of 8:[46] Subjects are asked to walk in a Figure of 8 pattern around two traffic cones placed approximately 5 feet apart on the floor. The total elapsed time is measured in seconds, along with the number of steps to complete the task and an observer-rated score of smoothness. For this investigation, only elapsed time was considered in analyses.

30-Second Timed Unilateral Stance:[152] This test was administered as part of a brief battery of static standing balance measures. Participants were asked to bend one leg at the knee and lift the foot behind them so they are standing only on the opposite foot. Those who were unable to attain a full unilateral stance position (opposite foot raised completed off of the floor) without any form of external support were classified as unable to perform the test. Participants were asked to hold the position for up to 30 seconds, and could move their arms or bend at the waist to maintain their balance, but could not touch any objects for external support. Timing began as soon as the participant appeared stable in the position, and ended after 30 seconds, or when the participant made contact with any objects for support, or touched the lifted

foot to the ground. They were considered 'Able to Hold' the unilateral position, and a time was recorded, only if they could demonstrate stability by holding the position for at least 1 second after any support was released. Unilateral stance as a measure of balance has predictive validity for fall risk in older adults.[153]

Explanatory Variables

15-item Geriatric Depression Scale (GDS):[128] The 15-item version of this basic screening measure for depression in older adults was used. For each item, participants are assigned one point if the 'Yes/No' response option selected corresponds to lower mood, for total scores ranging from 0 to 15. When used clinically, scores higher than 5 suggest depression and signal the need for further evaluation.

Digit Symbol Substitution Test:[129] This paper and pencil test from the Wechsler Adult Intelligence Scale measures perceptual processing speed. Participants are given a coding key with nine boxes. The upper half of each box contains a number from 1-9, and the bottom contains the symbol that corresponds to that number. The rest of the page is filled with rows of split boxes. The top of each box contains randomly ordered numbers (ranging from 1-9), and the coding key is used to fill in the bottom half of each box with the corresponding symbol. Participants fill in as many of the boxes as possible in 90 sec. Performance declines with age.[130] When combined with a gait speed of less than 1.0 m/s, DSST scores of less than 27 have been associated with increased mortality and incident disability in an older adult cohort of mean age 70 years.[131]

Expectations Regarding Aging (ERA-12)[60]- A twelve-item self-report tool designed to measure expectations regarding aging in the domains of physical health, mental health, and cognitive function. Responses have demonstrated acceptable levels of reliability and construct validity.[60]

Activities-Specific Balance Confidence Scale (ABC):[132] Participants were asked to rate their confidence that they will not fall or lose their balance during the performance of each of 16 mobility-related activities. A response of 0 indicates no confidence and 100 is complete confidence. The total

score is calculated as the average confidence reported for all items and is expressed as a percentage out of a possible 100. Scores of greater than 80% are common in high functioning, physically active older adults.[132, 133]

Survey of Basic and Instrumental Activities of Daily Living (BADL/ IADL Questionnaire): A self-report measure of basic and instrumental activities of daily living (BADL/ IADL) taken from the National Health Interview Survey (NHIS) was used to assess function for daily activities.[134] The BADL/ IADL scale measured self-reported ability to perform 16 tasks, including getting up from bed or chairs, walking, stair climbing, getting outside the home, and shopping. The total reported here indicates the number of tasks out of a possible 16 identified by the participant as either difficult or not performed for reasons related to health and physical function. Interrater and test-retest reliability ICC's above 0.9 have been reported in community-dwelling older adults.[25]

Survey of Activities and Fear of Falling in the Elderly (SAFFE):[135] This self-report measure assesses physical activity and fear of falling relative to eleven specific mobility-related activities, ranging from bending over and reaching overhead to walking for exercise and going out in slippery environments. SAFFE activity is scored as the number of activities performed out of 11, and SAFFE restriction is calculated as the number of activities reported as performed less frequently now than five years ago. Both activity and restriction scores range from 0 to 11, with higher scores indicating greater activity or greater restriction, respectively. To determine SAFFE fear score, participants are asked how worried they are about falling with each activity performed. Total fear is calculated as the mean response across only the activities that are currently performed, ranging from 0 (not at all worried) to 3 (very worried).

4.2.3 Statistical Analysis

PASW® Statistics version 18.0 (IBM SPSS, Inc., Chicago, Illinois) was used for all statistical analyses. Descriptive Statistics including mean, standard deviation, minimum value and maximum value were used to describe the distribution of timed performance measures across global self-rating state categories for both domains. Because nearly 25% of the cohort could not perform the unilateral stance test at baseline, we categorized performance on that test using our own ordinal scale. Categories were assigned as follows: 0 = Unable to perform, 1 = Held less than 10 seconds, 2 = Held between 10 and 29 seconds, 3 = Held for the full 30 seconds (ceiling). A contingency table was then used to provide the distribution of performance categories at baseline for each balance state rating.

Level of Agreement between Global Self-Ratings and Performance: Nonparametric Correlations (Spearman Rho) were used to determine the level of agreement between global self-ratings and performance-based measures thought to represent each of the domains, mobility and balance. This was done for both state and transition-based self-ratings, and for change over the time period of interest as well as cross-sectionally (at baseline). For analyses of change, the **time period of greatest change in transition rating** was determined for each individual and used for all subsequent analyses. This was done separately for each of the domains (mobility and balance). The default time period of 6 to 12 months was used for those participants whose greatest change was equivalent at two or more time periods. We chose this period because our results for discordance between state and transition-based change in self-ratings suggest that it was the period of greatest self-perceived change (by transition) for the cohort overall.

Change in State Rating over the time period of greatest change by transition report was expressed as the rating (From Excellent [1] to Poor [5]) reported at the earlier timepoint minus the rating selected at the later timepoint. As a result, state change could range from -4 to +4, with negative values indicating

selection of a 'worse' state rating at the later timepoint, while positive values indicated that a more favorable rating was selected. **Transition Rating** was expressed as the raw value (-3 to +3) selected by the participant at the later timepoint. Negative values therefore indicate a self-reported worsening of status.

Concordance Groups: Participants were categorized into four groups based on which method of determining change in global self-rating (state or transition) was more concordant with any change in their gait speed over the same period. In order to establish categories based on concordance between change in self-ratings and performance, change in gait speed was converted to a three-level ordinal scale based on the parameters for meaningful change established by Perera and colleagues.[35, 142] Change of less than 0.10 m/s in either direction was categorized as no change or 'same,' change of +0.10 m/s or greater was categorized as 'better,' and change of the same magnitude in the negative direction was categorized as 'worse.' To facilitate comparisons with this three-level scale of performance change, and because the clinical relevance of the qualifiers 'much,' 'somewhat,' and 'a little' worse or better is unknown, transition ratings at the later timepoint were then collapsed into the same categorical scale of 'better,' 'same,' and 'worse.' Thus, ratings of positive 1, 2, and 3 were all categorized as 'better,' and all negative ratings were categorized as 'worse.' Finally, the difference between the consecutive global state ratings selected at the two end points of the time period of interest was collapsed using the same three-level categorical scale. For each individual at the time period of greatest reported mobility change by transition, the change in state ratings and the transition rating were each compared to the change in gait speed to determine which method of self-rating change related more closely to performance change. **Each participant was categorized as into one of the following groups: 1) STATE ONLY concordant with performance change, 2) TRANSITION ONLY concordant with performance change, 3) BOTH concordant with performance change**, **4) NEITHER concordant with performance change.** For example, if the change in gait speed from 6 to 12 months had declined by at least 1.0 m/sec (classified as 'worse'), the transition rating at 12 months was 'same,' but the state rating had worsened from 'Good'

at 6 months to 'Fair' at 12 months (classified as 'worse'), the individual would be categorized as concordant by 'STATE ONLY.'

One-way ANOVA with Least Significant Difference post-hoc testing was performed to compare the four concordance groups on individual characteristics such as age, mood, cognition, expectations regarding aging, balance confidence, and functional level. A Type I error rate of 0.05 was used. No adjustment was made for multiple comparisons given the exploratory nature of these analyses, and resultant concern with inflation of the Type II error rate.

We then re-established concordance groups using the published guideline of $\pm$1 point (out of a possible 12 points) as substantial change in SPPB score[35, 142] and repeated all subsequent analyses. As a final check of the degree to which our findings are robust, we repeated all analyses again using $\pm$0.05 m/s as small but meaningful change in gait speed. [35]

4.3 RESULTS

Of the original 120 participants, 104 had data for at least two consecutive timepoints of the four possible (baseline, 6 months, 12 months, and 18 months). The most common reason for missing data at six months was travel out of the area. By 12 months, home visits minimized missing data to only cases of serious illness or death. Baseline characteristics for these 105 participants who had data for comparison across at least one of three possible time periods of change are described in Table 4.1

Cross-Sectional Baseline Findings: Based on distributions of performance across baseline self-ratings for mobility (Table 4.2) and balance (Table 4.3), better performance on select measures is associated with higher state ratings overall for both domains. All of the performance measures that we selected correlated more strongly with self-ratings for the domain of mobility (Spearman Rho range .489-.499) than for balance (range .273 - .340). Of the measures we selected for the domain of balance, the TUG and Figure

of 8 correlated as highly with self-ratings as unilateral stance, although the correlations were still only fair. Table 4.4 provides the distribution across state ratings for unilateral stance performance once converted to an ordinal variable with four levels as described in the methods section. As evidenced by the correlation of 0.300 (Spearman), there is a general trend toward better performance by those who selected higher balance state ratings, however, as with all of the mobility and balance comparisons made, individual outliers exist (Tables 2, 3, 4).

Comparison of Change in Self-Report and Performance: Correlations between change in selected performance measures and each of two methods of measuring change in self-ratings are listed in Table 4.5 for the domains of both mobility and balance. Change in gait speed correlated more strongly with self-ratings change for the domain of balance than for mobility, with only transition ratings reaching significance for the domain of mobility. By contrast, only change in state ratings achieved statistical significance when compared to change in SPPB scores over the same period, and only for the domain of mobility. Correlations between change in self-ratings and change in TUG and Figure of 8 were lower than for gait speed change for both domains and both types of self-ratings, and were statistically significant (although still modest correlations) only for state ratings for the domain of balance. The correlations between unilateral stance (as a continuous measure using the raw hold time with values of '0' seconds for unable to attempt) and self-ratings were lower than anticipated for the domain of balance. Correlations between the two methods of determining change in self-report (state and transition) were slightly higher for the domain of mobility (tau = .412) than for balance (tau = .372), but correlations between self-ratings change for both domains were modest.

Comparison of Concordance Groups for Change in Self-Ratings and Gait Speed Change: The categorical distribution of change in gait speed for each of the four concordance groups established based on agreement between change in self-rating and change in gait speed over the same period are presented in Figure 4.1. Overall, at least one of two methods of determining change in self-ratings (state or transition-

based) was concordant with the direction of change in gait speed for 65% of the sample, and 25% were concordant using both types of self-ratings. The majority (69%) of those who were concordant only by transition had declined in gait speed over the six-month period of interest, and only 2 of 42 who were concordant by transition had improved in gait speed. Overall, over 70% of those whose gait speed had improved by at least 0.10 m/s did not report improvement in mobility by either serial state or transition-based approaches. Conversely, almost 80% of those who declined in gait speed reported decline in mobility by transition, and no participant who declined in gait speed was concordant by state ratings only.

Mean gait speed and SPPB change by concordance group were calculated for only the individuals with significant gait speed decline (represented in blue in Figure 4.1). The six who declined in the 'BOTH' concordant group had a mean gait speed decline of 0.24 m/sec, compared to 0.17 m/sec for the 'TRANSITION ONLY' group and 0.21 m/sec for the 'NEITHER' concordant group. For change in SPPB scores among only those who declined in gait speed, 2.7 points was the average SPPB decline for those in the 'BOTH' concordant group, compared to decline of 1.1 points for the 'TRANSITION ONLY' group and 0.4 points for the NEITHER' concordant group. Results were similar for concordance groups established using $\pm$0.05 m/s as small significant change for gait speed and $\pm$1 point as substantial change for the Short Physical Performance Battery.

Between-group comparisons based on the degree of concordance for each method of assessing change in self-ratings and change in gait speed are presented in Table 4.6 for the domain of mobility. Older adults whose self-ratings were concordant in direction with gait speed change by 'TRANSITION ONLY' had lower ABC scores, gait speed, SPPB scores, TUG performance, and SAFFE activity levels at the later of the two timepoints than the 'BOTH' and 'NEITHER' concordant groups. Additionally, this 'TRANSITION ONLY' group had significantly greater SAFFE fear and greater activity restriction in the past five years than the 'BOTH' and 'NEITHER' groups, and greater comorbidity than the 'BOTH' group. Compared to the 'BOTH' and 'NEITHER' groups, GDS scores were significantly higher and

ERA scores were lower for the 'STATE ONLY' group. DSST scores were highest for the 'TRANSITION ONLY' and 'NEITHER' concordant groups.

4.4 DISCUSSION

These results add to our series investigating whether global self-ratings are useful in detecting changes, particularly decline, in mobility and balance in older adults with the passage of time. Change in each of the performance measures we selected correlated more modestly with change in global self-ratings than performance and self-ratings correlate cross-sectionally. However, the results of this investigational pilot work suggest that transition mobility ratings are better at communicating decline in gait speed and performance on the Short Physical Performance Battery than sequential state ratings, and neither method is useful to detect improvements in these performance measures without intervention. As healthcare providers seeking to detect, remediate, and when possible, prevent age-related changes, we are more interested in decline. Decline in mobility poses the greater threat to the individual's safety than improvement, and transition-based questions may make better screening tools for the detection of decline.

Based on results of sub-analyses for only those individuals who declined in gait speed over the six-month period of interest, the majority of individuals who experience decline of at least 0.10 m/sec in gait speed do report perceived decline in mobility by transition. A minority of these individuals also select a lower mobility state rating six months later, and these may be the individuals who have undergone the greatest performance change based on both gait speed and SPPB scores. The consistency of these results when repeated using small significant change ($\pm$0.05 m/s) in gait speed strengthens the conclusion that serial use of state ratings could be more specific to large declines in mobility performance, while transition ratings are more sensitive to smaller decline. If this is true, a report of worsening mobility by global transition report may be a useful first-level screening tool to determine the need for further evaluation of

balance and mobility. More detailed self-report measures and performance-based screening tools could then be used to decide whether a referral to a geriatric mobility specialist is in warranted.[17, 25, 125, 140, 154-156] Our results indicate that about 25% of individuals who report worsening mobility will be 'false positives' based on a lack of significant decline in gait speed, however, we recognize that mobility encompasses far more than merely gait speed, and that these individuals may have indeed experienced decline in aspects of mobility that do not affect gait speed (e.g. transfers out of a chair or bed). These findings oppose the 'Intentional Censoring' theory that we have proposed in the past in an attempt to explain why discordance may occur. Although censoring of perceived decline may still have occurred at the individual level, our cohort as a whole readily reported decline. However, our research team members did not function as healthcare providers for these older adults, so intentional censoring was not truly tested. Even participants who may be prone to censoring may have identified no need to censor during study visits.

Instead of censoring decline, we found that older adults are far less likely to report improvement in mobility, even the 25% of the sample whose gait speed improved by at least 0.10 m/sec. When improvement was reported, the serial state technique was generally used, however, the majority (72%) did not report improvement at all. One possibility is that the improvement occurred gradually over the six month period, so was not recognized by the individual, however, changes of more than 0.20 m/s and 4 points on the SPPB are large,[35, 142] particularly over a period as short as six months. It is also possible that older adults do not anticipate improvement in this domain without intervention, so a phenomenon opposite the 'placebo effect'[157] was responsible. Other studies have found transition ratings to be more responsive to improvement after intervention,[44, 94, 95, 114, 115] but older adults may be more reluctant to report being 'better' in the absence of intervention, or even an intervening event such as an illness from which they subsequently recovered.

Most of these conclusions emerged from examination of the results in Figure 4.1 and Table 4.6, between group comparisons of concordance groups for change in self-rating and performance. Older adults whose

transition self-ratings, but not serial state ratings, were concordant in direction with gait speed change had the lowest mobility-related function of all concordance groups. They had slower self-selected walking speed, lower SPPB scores, and worst performance on the TUG. In addition, they had lowest balance confidence, SAFFE activity levels, new activity restriction, and fear. The 'NEITHER' concordant group, who predominately experienced improvement or no significant change in gait speed, were the highest functioning of all four groups based on mean gait speed, SPPB score, and TUG performance. A possible explanation is that gait speed improvement in higher-functioning older adults with no mobility difficulty or disability is not reported because it has no recognizable impact on function for the individual.

Cognition, depression, and aging expectations did not explain discordance between self-ratings change and change in performance. For cognition, the 'NEITHER' and 'TRANSITION ONLY' concordance groups both demonstrated higher performance on a test of processing speed than the 'BOTH' concordant group, however, means for all groups were within normal ranges based on published age-adjusted norms.[129] Scores on a geriatric depression screen were significantly higher for the 'STATE ONLY' group than for the 'BOTH' and 'NEITHER' concordant groups, but even the highest group mean (2.3) fell well below the minimum score of 6.0 for referral for possible depressed mood. [128] The 'STATE ONLY' group also had the lowest aging expectations, however, to our knowledge, no clinical meaning has been established for the difference between scores of 45 and 56. These scores are close to the sample means for the ERA-12 validation studies, and are well above mean scores for subsamples of individuals with ADL disability.[60]

Mean change in SPPB scores for the sub-group in each concordance group who declined in gait speed suggests that a decline in serial state ratings may reflect more substantial performance decline than detected by a report of decline by transition alone. Thus, while serial state ratings do not detect performance decline missed by transition, when accompanying decline in transition, decline by state rating may justify immediate referral without further screening. The most obvious explanation for this

phenomenon is the difference in global state and transition question response scales. Perceived decline that warrants a transition rating of 'a little,' or even 'somewhat,' worse may not be enough to lower state rating by one full level (e.g. from 'good' to 'fair').

While the transition mobility question may be more responsive than repeated assessments of mobility state rating when monitoring for gait speed and SPPB decline, certainly mobility is more than just gait speed. We decided to use these two measures to establish concordance between performance and self-ratings because they are well-researched and standardized measures for which estimates of meaningful change are available, but we are not implying that change in gait speed or SPPB score is the 'gold standard' for mobility decline without intervention. New impairments could affect mobility even if gait speed remains constant, and SPPB scores are ordinal scores established based on quartiles of performance,[105] so the same degree of change in gait speed or timed chair stand may or may not alter the ordinal score. For example, the ability to walk long distances in the community could be impaired by endurance, the ability to swim could be impacted by a shoulder injury, or transfers out of bed may be disrupted by new orthostasis, but in all cases, gait speed over four meters and time to rise from a chair may not change significantly. Therefore, our results are limited to a discussion of the ability of our global self-ratings to detect changes in mobility as defined by gait speed and SPPB score.

In addition to the choice of performance measures, validity of the concordance group comparison results depend on thresholds selected to define 'change' in the performance-based measures. We use meaningful change thresholds originally established by Perera et al.,[35] but these initial findings have subsequently held true using other large cohorts[142] The findings for small but meaningful change of $\pm$0.05 m/s in gait speed may be relevant in detecting pre-clinical disability, defined as a state in which older adults change their technique for completing daily tasks but do not report true difficulty or disability. [145]

Although some authors have viewed change of two or more levels in a global rating as having 'substantial clinical meaning,'[25, 44] we have chosen to consider change of even a single level as substantial. We

believe that at least three factors support this stance. First are the documented tendencies of adults to avoid the ratings at each end of a Likert scale (central tendency bias),[122] and to reconstruct the past for consistency with the present (consistency bias).[38] Because our transition question includes a central option of 'about the same' instead of 'exactly the same,' we believe that those who choose to indicate improvement or decline of even one level perceive actual change and would choose the midline option if uncertain. We are also interested in the concept of preclinical disability,[145] which may be lost if the threshold for change is raised to two levels. Additionally, we are interested in the concepts of balance and mobility, and even small variations in these domains can have important public health implications for the individual, especially related to risk for falls and subsequent injury.[5, 54, 141, 158] Therefore, we believe it is important to investigate the impact of even a small perceived decline. Finally, an average change of 0.5 on a 7-point scale has been reported as important to patients, and this amount of change is equivalent to 'a little bit' better or worse.[33, 159]

Tables 4.2 and 4.3 provide some evidence for the discrepancy between global self-ratings and performance at the individual level. Neither of the two individuals with fastest gait speed rated their mobility as 'excellent.' One selected 'very good,' and the other selected only 'good.' Another participant with gait speed of 1.33 m/s rated himself as only 'fair' for mobility. While these individuals may have mobility limitations not evident from gait speed, geriatric mobility specialists such as physical therapists would likely rate all of these individuals higher than 'good' if presented only with their performance scores.

Similar individual results were seen for the domain of balance. Of those who rated their balance as 'excellent,' one could not perform the unilateral stance test, and only 50% could hold the position for more than 10 seconds. Only seven participants could hold the position for the full 30 seconds, yet 71% of these individuals at the ceiling of unilateral stance performance rated their balance as only 'good' or 'fair.' Of the four measures selected for comparison in Table 4.3, clinicians and participants would

probably agree that unilateral stance is the one recognized as most specific to the domain of balance,[85] however, clinicians recognize that as a static test, it may be least reflective of daily function, and therefore perceived balance.[81] While participants provided the global ratings at the start of each study visit, prior to the administration of actual performance tests like unilateral stance, this baseline timepoint represents the only study visit at which they did not have prior experience with the test. Participants may not have been considering such high-level balance and mobility tasks when selecting a global state response, but at 6, 12, and 18 months, all had prior experience with the test, and it may have shifted their ratings as part of a phenomenon called 'Response Shift.'[120, 160]

The largest group of participants (n=36) at the timepoint of greatest change in mobility transition rating was discordant by both transition and state report. Rather than being 'wrong,' these individuals could be detecting decline or improvement in areas of mobility not impacting gait speed and SPPB scores. Alternatively, changes in gait speed and SPPB scores that do not impact ability to perform daily activities may not be perceived as significant, or even recognized by the older adult, so many could report being 'about the same' and continue to select the same state rating. This is supported by our between-groups comparisons for self-ratings and performance change concordance groups (Table 4.6). The 'TRANSITION ONLY' group was lower functioning than the other concordance groups, and while the majority of this group had experienced performance decline in the past six months, they also appear to have started out with lower function (results not shown), so decline in these individuals may have impacted daily function more than the same degree of decline in individuals who started out with the highest levels of function.

Based on the correlations in Tables 4.2, 4.3, and 4.4, we have yet to find the best performance measures to reflect self-perceived ability for the domains of mobility and balance in an older adult cohort. It is likely that a battery of measures is needed, but constraints include space, equipment, manpower, and

patient tolerance. If this is true, use of global ratings may be even more useful, as they may capture aspects of mobility and balance that our current longer questionnaires and performance measures do not.

Our results support those published previously to suggest that while self-report and performance overlap, they each bring distinct and valuable information to the table. Studies of mobility in mid-life to older adults with knee osteoarthritis have demonstrated only moderate (0.44) correlation between self-report and performance-based measures, and the authors have concluded that performance results alone do not adequately reflect functional mobility.[103] More recently, self-report measures in a population with arthritis were found to relate most strongly to reported pain, while performance measures related to self-efficacy for mobility tasks.[107] Suggestions have been made not only to include both self-report and performance measures, but also to include both disease-specific and more general health-related quality of life self-reports.[107, 108]

Reuben and colleagues[100] urge healthcare providers to 'apply caution when using a single measure of physical function' to screen, diagnose, or monitor individual patients. These recommendations came after finding only weak to moderate associations between multiple self-report measures and a single performance-based measure of function that they designed to measure the same construct of physical function.[100] For older women with gait speeds above 0.6 m/sec, self-report of mobility was found to provide additional predictive information about incident difficulty and disability not provided by walking speed alone. (Fried 2001)

In closing, the study of self-report (in our case, global self-ratings) and performance change in older adults is complicated, and it appears that no one 'rule' can be instituted that will apply to all conditions. Optimal recommendations likely vary according to the individual (based on characteristics including level of function) and the situation (monitoring natural history of decline or response to a rehabilitative intervention). Although transition-based global ratings of mobility appear to be more sensitive for screening decline in mobility without intervention, changes in state ratings may be more sensitive for

larger decline. As for the domain of health, we hypothesize that global state ratings of mobility may play other roles as well, such as predicting future decline in cases of high-functioning individuals who rate their abilities on the lower end of the response scale.[48, 51, 161] While the roles of both state and transition-based global ratings may be based on initial function and the direction of change to be detected, the relative value of each type of rating can be more fully elucidated only with further work in larger cohorts with a greater range of mobility function.

TABLE 4.1: Baseline Description of the Sample (n=104)

CHARACTERISTIC	Mean (SD)	Range
Age (yrs)	77.4 (6.0)	64 - 92
Gender (% Female)	74%	__
Race (% Black)	12.5%	__
Education (% Who Attended College)	70%	__
Comorbidities (0-17)	2.9 (1.4)	0 - 6
15-Item GDS	1.48 (1.7)	0 - 8
DSST (# completed in 90 sec)	48.1	16-70
ERA (0-100)	53.1	11.1-97.2
Gait Speed (m/sec)	0.97 (0.24)	0.45 – 1.52
TUG (sec)	10.5 (4.1)	6.0-26.0
SPPB Score (0 -12)	9.5 (2.3)	3 - 12
BADL/IADL difficulty (0 - 16)	2.5 (2.9)	0 - 14
SAFFE Activity (0 – 11)	8.7 (1.5)	3 - 11
SAFFE Fear (0-3)	0.13 (0.19)	0-0.78
SAFFE Restriction (0 -11)	2.9 (2.8)	0 - 9

TABLE 4.2: Baseline Performance and Global Self-Rated State Correlations for the Domain of MOBILITY

MOBILITY RATING	**PERFORMANCE MEASURE Mean (SD) Min -- Max**		
	Gait Speed (m/sec)	**TUG (s)**	**Figure of 8 (s)**
Excellent **N=13**	1.24 (0.15) 1.02-1.42	7.8 (1.11) 6.04-10.00	7.62 (1.28) 5.99 – 10.10
Very Good **N=38**	1.16 (0.22) 0.69-**1.59**	9.13 (1.88) 6.34-16.15	8.89 (1.75) 6.39-14.86
Good **N=42**	1.08 (0.22) 0.54-1.53	10.15 (3.23) 7.15-23.07	9.84 (2.67) 6.87-18.31
Fair **N=25**	0.84 (0.24) 0.44-1.33	13.75 (5.11) 6.97-25.92	13.05 (5.07) 6.60-26.88
Poor **N=1**	0.30	23.46	23.84
CORRELATIONS	.493*	.499*	.489*

*$p \leq 0.05$

TABLE 4.3: Baseline Performance and Global Self-Rated State Correlations for the Domain of BALANCE

BALANCE RATING	**PERFORMANCE MEASURE Mean (SD) Min -- Max**			
	Gait Speed (m/s)	**TUG (s)**	**Unilateral Stance (s)**	**Fig of 8 (s)**
Excellent	1.13 (0.25)	8.85 (2.75)	14.10 (11.79)	8.33 (2.26)
N=8	0.72-1.42	6.04-13.22	2.72 – 30.00	5.99 – 12.05
Very Good	1.16 (0.21)	9.19 (2.25)	8.99 (6.78)	8.53 (1.68)
N=19	0.71-1.57	6.69-14.72	1.36 – 20.95	6.63-13.04
Good	1.11 (0.25)	9.73 (3.12)	9.70 (9.15)	9.48 (2.41)
N=48	0.54-**1.59**	6.34-23.07	1.01 – 30.00	6.74-18.31
Fair	0.99 (0.27)	11.91 (5.01)	8.49 (8.73)	11.62 (4.93)
N=34	0.44-1.45	6.97-25.92	1.03 – 30.00	6.47-26.88
Poor	0.91 (0.28)	12.47 (4.48)	4.41 (1.93)	12.07 (4.87)
N=10	0.30-**1.59**	7.94-23.46	1.84 – 6.38	6.60-23.84
CORRELATIONS	.273*	.327*	.161	.340*

*$p \leq 0.05$

TABLE 4.4: Distribution of Performance on Unilateral Stance Test by Global BALANCE State Rating at Baseline

BALANCE RATING	**Performance on Unilateral Stance Test** n (% within Unilateral Stance Performance Category)			
	Unable N=28	<10 sec N=60	10-29.5 sec N=24	30 sec (Ceiling) N=7
Excellent N=8	1 (4%)	3 (5%)	2 (8%)	2 (29%)
Very Good N=19	2 (7%)	9 (15%)	8 (33%)	0 (0%)
Good N=48	9 (33%)	26 (43%)	8 (33%)	5 (62.5%)
Fair N=34	10 (36%)	17 (28%)	6 (25%)	1 (14%)
Poor N=10	5 (18%)	5 (8%)	0 (0%)	0 (0%)

Spearman Rho Correlation = 0.300* ($p \leq 0.01$)

TABLE 4.5: Correlations between **Change** in Self-Ratings and Performance at Timepoint of Greatest Transition Change (n=103)

CHANGE IN PERFORMANCE MEASURE	**SELF-REPORT DOMAIN OF CHANGE**			
	Mobility		**Balance**	
	State	***Transition***	***State***	***Transition***
	(Mob Ratings Correlated at .412)*		*(Bal Ratings Correlated at .372*)*	
Gait Speed	.191 (p=.052)	**.267****	**.353****	**.312****
SPPB	**.218***	.183	.158	.096
TUG	.066	.146	**.228***	.035
Fig of 8	-.008	.117	**.243***	.127
Unilateral Stance (Continuous)	N/A	N/A	.025	-.142

**p≤ 0.01 level * p≤ 0.05 level

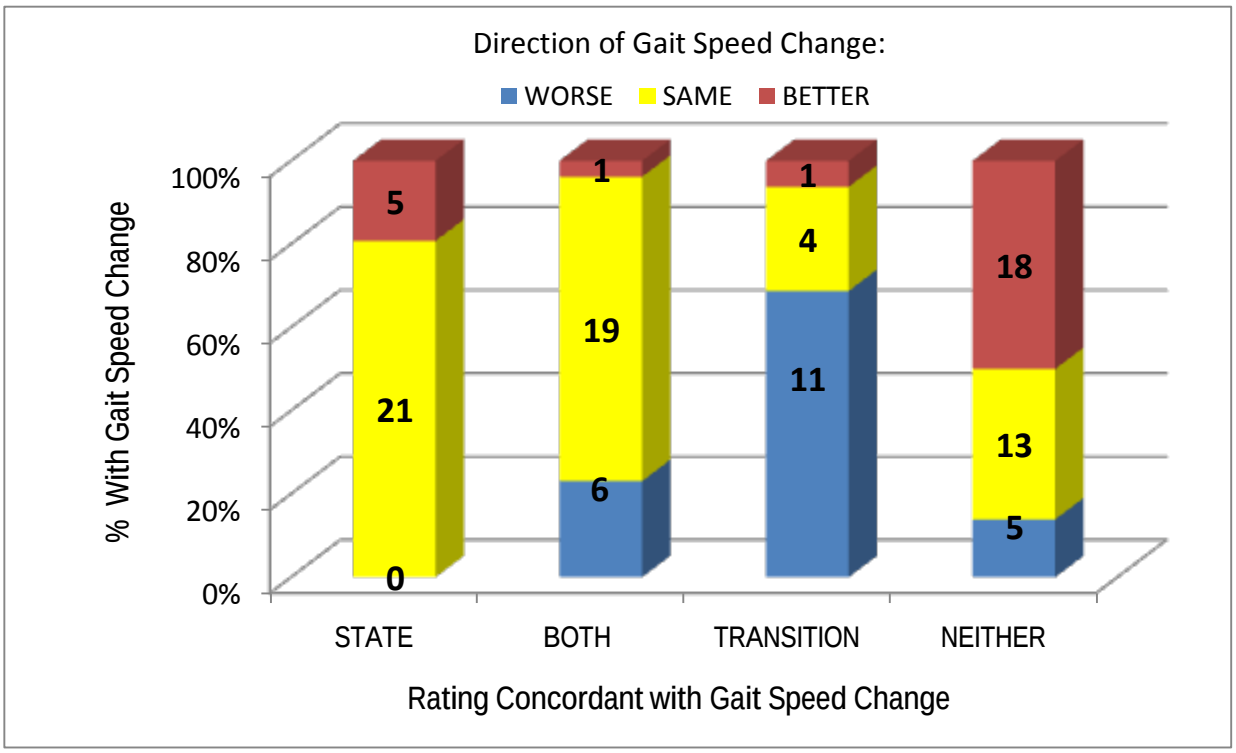

FIGURE 4.1: Comparison of Concordance Groups for Change in Global Mobility and Gait Speed

TABLE 4.6: Comparison of Concordance Groups for Change in Self-Ratings and Performance

EXPLANATORY MEASURE *Different than BOTH group # Different than NEITHER group p≤0.05	RELATIONSHIP BETWEEN CHANGE IN MOBILITY SELF-RATINGS and CHANGE IN GAIT SPEED (±0.10 m/s as sig. change) STATE ONLY CONCORDANT N=26	BOTH CONCORDANT N=26	TRANSITION ONLY CONCORDANT N=16	NEITHER CONCORDANT N=36
GDS (0-15)	**2.3*#**	0.92	1.6	1.2
DSST (# in 90 sec)	46	42	**51.5***	**52***
Comorbidities (0-17)	2.9	2.3	**3.6***	2.9
ERA Score (0-100)	**45.1*#**	56.9	55.6	54.4
BADL/IADL difficulty (0-16)	3.8	2.2	3.2	3.4
ABC (0-100%)	67.6	76.8	**60.1*#**	78.2
Gait Speed (m/sec)	**0.97#**	1.01	**0.84*#**	**1.11**
SPPB (0-12)	9.5	9.7	**8.2*#**	**10.3**
TUG (sec)	11.4	10.3	**12.8*#**	**9.4**
SAFFE Activity (0-11)	**8.3#**	8.9	**7.1*#**	9.1
SAFFE Fear (0-3)	**0.22***	0.08	**0.28*#**	0.13
SAFFE Restriction (0-11)	3.7	2.4	**4.1*#**	2.8

Bold text indicates values viewed as clinically meaningful.

5 SIGNIFICANCE AND FUTURE DIRECTIONS

The results of this work have provided much insight into the self-ratings made by community-dwelling older adults when presented with state- and transition-based global questions for the domains of mobility and balance. We have found discordance when comparing change in state and transition-based self-ratings over a six month period, and discordance when either type of self-rating change is compared to change in performance measures over the same period. While ceiling and floor effects were rarely a factor, some of this discordance is probably explained by the difference in the ratings scales used for the state and transition questions. Perceived change of 'a little' or even 'somewhat' worse by transition may not warrant decline by a full rating in state. In addition, our findings have provided evidence to support or refute a few of the discordance theories presented in the background. We emphasize that conclusions must be made with caution in an exploratory work of this nature, and we intend for these findings to be used only to generate hypotheses for further testing.

Intentional Censoring: Little support was found for the theory of intentional censoring. According to our theory, older adults may fail to communicate perceived decline in mobility function, out of fear that recommendations for lifestyle changes will be made that they do not wish to follow.[89] For the domain of mobility at the timepoint of greatest change by transition rating, less than 25% of the sample who demonstrated significant decline in gait speed selected transition ratings of 'about the same' or 'better' over the same time period. These five individuals did have the smallest mean SPPB score decline of all concordance sub- groups who declined in gait speed, so it is possible that rather than being intentionally censored, the decline in mobility was not recognized by the individual because it had little functional impact. Moreover, the entire older adult cohort was far more willing to communicate decline than we had anticipated, with about 50% of the sample reporting mobility decline by transition at some point over

study participation, over half of these with no corresponding decrease in gait speed. Still, our semi-structured interview was not designed to uncover intentional censoring, so it may have existed on an individual level in this cohort. It is also possible that adults who would normally censor for healthcare providers may not have felt it necessary for to censor for our research staff.

While reports of worsening mobility status were more common than anticipated, reports of improvement were rare, even though 25% of the sample experienced an increase of at least 0.10 m/s in gait speed over the time period of greatest change by transition rating. As a result, we now hypothesize that older adults do not anticipate improvement in mobility in the absence of an intervention.

Response Shift: There is some support for response shift as a result of participation in our battery of performance-based measures, but most of this is anecdotal evidence obtained from the interviews. A number of participants referred to their previous performance on our high-level tests of balance and mobility when answering questions about how they selected self-ratings, particularly for the domain of balance. Additionally, a comparison of the distributions of state ratings at baseline and six months suggest regression toward the mean for the domain of balance, but no such distribution-based findings emerged for the domain of mobility. We suspect that response shift contributed to discordance, particularly between the baseline and six month timepoints, however, we do not view it as the primary factor. These conclusions are limited to response shift induced by study participation, as our study was not designed to capture shift from most external sources.

The sub-theory of **Timeframe Incongruency** was supported by the interview results. While the transition question asked for a comparison to status six months earlier, most in our cohort indicated that their state rating actually reflected a more general rating of their status over as much as six months to one year. This would not create discordance for individuals of very stable mobility, whether high or low functioning, but one goal as healthcare providers for older adults is to detect change, particularly decline, in mobility. The selection of state ratings that reflect status 'in general' over the last few months to a year

greatly confounds any attempt to use such ratings to detect change. When performing calculations to determine change in perceived status over a period of time, we assume that these state ratings indicate status at a discrete point in time. If this assumption is not met, change values derived in this manner will likely be discordant with both transition-based ratings and change in performance measures collected at the same two points in time.[44] We suspect that timeframe incongruency due to the generalization of state ratings is one reason for the discordance found between serial state ratings and gait speed change, particularly decline, in our sample. The majority (77%) of those who were concordant by serial state ratings demonstrated no significant change in performance. Further research is needed to determine whether the individuals who specify longer timeframes for their state ratings are those who are indeed more stable in mobility over time, in which case, the resulting timeframe incongruency will result in little discordance. Conclusions from our data are limited by having timeframe results for only the 33 participants who completed the interviews.

Recall Bias: Minimal recognized recall bias was identified from our interviews, however, our study design did not facilitate complete investigation into this phenomenon. Our questionnaires were designed to detect a recognized inability to remember the prior timepoint of comparison (memory transience), but distortion of the memory was not explored. The participants in our study may not have been consciously aware that their recall of the past was inaccurate. Studies suggest that adults subconsciously reconstruct the past for consistency with the present ,[31, 32] and this could have contributed to the consistency bias observed in our sample as described below.

Consistency and Central Tendency Bias: The central option of 'about the same' was indeed the response selected most frequently for transition-based questions in both domains, and over 60% of the sample selected the same state rating for at least two of the consecutive timepoints. Still, distribution results support acceptable use of the entire rating scale for state-based questions, and willingness to communicate perceived change, particularly decline, by transition. About 40% of the sample selected the

'about the same' central option for at least two consecutive timepoints, however, a number of these people demonstrated no significant change in gait speed, and we believe that the domain of mobility encompasses far more than gait speed and SPPB scores, so it is difficult to distinguish between the accurate perception of 'no change' and a tendency to select the central option regardless of change.

Gradual Change: This theory is one of the most difficult to test using our protocol. We hypothesized that it would manifest as change in performance without a corresponding change in self-rating, and while this occurred in our sample, our protocol was not designed to determine if these observations were due to gradual change or to another of the proposed theories. Moreover, age-related change that is truly gradual may not reach functional significance over time periods as short as six-months, so these individuals were probably classified as having no performance change when we defined concordance groups for change in self-ratings and performance. Further exploration with time periods of one year or longer are recommended.

Low Expectations of Aging: Based on Expectations Regarding Aging[60] scores, there was little evidence for low expectations overall in our sample. In addition to low ERA scores, these individuals might be anticipated to consistently report being worse by transition, and about 12% of this sample rated their mobility as 'worse' by transition at every timepoint, without a corresponding decrease in gait speed over at least one of the time periods. The lowest aging expectations were found for the 'STATE ONLY' self-rating and gait speed change concordance group, but while the difference in ERA scores was statistically significant, the clinical significance is unclear. While we focused on transition ratings of decline, low expectations of aging may have played a bigger role in the reluctance of this sample to report improvement in mobility when gait speed had increased.

Pre-Clinical Disability: As noted above , about 12% of participants reported worsening by transition when no change was found in gait speed at the timepoint of greatest change by transition rating. However, only about 6% of these individuals reported no increase in difficulty or disability (based on

BADL/IADL score) over the same period, a requirement for preclinical disability.[125, 145] We believe that gait speed does not represent the entire domain of mobility,[103] and because we did not inquire into changes in the technique used by our participants to perform their daily activities, few conclusions can be drawn about the phenomenon of pre-clinical disability.

Disparity in Definitions of the Domains and Frames of Reference: A number of themes were identified to define the domains of mobility and balance. Individual variability does exist in our cohort, and appears to be based, at least in part, on functional status and level of social participation. While most participants used more than one theme when self-rating mobility and balance, some described a focus on higher-level skills or those that reflect greater social participation, while others considered only more 'standard' mobility activities such as walking around the house and rising from chairs. We conclude that this disparity in definitions is one of the more significant contributors to discordance between change in self-ratings and performance. We defined performance change as change in gait speed or SPPB scores, but older adults who perceive decline in their tennis or golf game may not show decline on our performance measures. Alternatively, those who have declined in gait speed may perceive no change in their ability to get around the house and get up from a chair.

Three frames of reference appear to be used to draw comparisons when selecting a current state response. In addition to discordance with performance, these frames may explain why older adults sometimes select a different rating than healthcare providers would have selected for them. Instead of a rating based on the norm for people of their own age group, this community-dwelling older adult cohort often drew comparisons to their own performance many years ago, or to the performance of friends who appear to be outliers for their age group. Use of either of these frames could interfere with the communication of change over a finite period of time. Comparison to the past may predispose to low ratings and to the perception of decline when no recent decline has occurred, but comparison to others could cause ratings to change with acquaintances, even with no change in performance. Standardizing frame of reference is

difficult because a certain frame (such as a comparison to others known by the older adult) is refused by some, and variation is inherent between individuals. Most of the individuals in our interview sample used more than one frame of reference, so the relative contributions of each frame to the self-rating chosen may be the key.

Future Directions

These results demonstrate that mobility and balance are multi-faceted domains, defined differently even within a relatively homogeneous community-dwelling older adult cohort. Given the individual variability in definitions provided by participants for each of these domains, it is clear that the performance-based measures we selected do not encompass every facet of a domain. Stems could be used to convert global questions to more specific inquiries into activities using an imposed frame of reference, and while this may standardize results for research purposes and improve the concordance between self-ratings and performance measures (and change in both), it will probably diminish the value of the self-ratings in the clinic setting. As previously discussed, a number of studies support self-report as providing meaningful information beyond that collected using performance measures. [25, 101, 106] Thus, modifying global questions to mirror results of our often artificial performance tests is likely to increase the correlation between the two, but this is not the ultimate goal in geriatric healthcare. If these questions are to be used to detect incident difficulty with life activities, a more useful option may be to collect information from the individual to qualify the ratings they have selected.

We selected performance measures of mobility and balance that are low-tech and low burden, therefore feasible in almost any clinic or home-based healthcare setting, however, our selections may have limited our findings. The cross-sectional correlations found between global ratings and performance measures are evidence that our performance tests have failed to capture the domains of mobility and balance

entirely as they are perceived by older adults. Because participation is thought to be based more on perceived ability than performance on timed measures,[82, 83] finding the optimal measure or battery of measures to capture the domain of mobility as perceived by older adults will continue to be a goal.

Another direction for future research is the predictive validity of global mobility and balance state-based ratings. Just as Schoenfeld and colleagues[48] concluded for global health ratings, it is possible that self-rated mobility may be even more meaningful in individuals of moderate to high mobility function rather than their mobility-disabled counterparts. Individuals who perform well on standardized performance tests of mobility but rate their own mobility as only 'fair' or 'poor' may be more likely to develop mobility disability in the coming years. Even if deficits cannot be identified for rehabilitation, a low self-rating may signal the need for more frequent visits to a geriatric medical or rehabilitation professional.

The findings from our semi-quantitative interviews could be used to design more streamlined interviews feasible for administration to a larger older adult cohort. The focus would be on the primary frame of reference (one's self in the past or a comparison to others) used when providing ratings of global state for the domain of mobility, possibly asking each person to rate themselves twice, once using each frame of reference. Performance measures could be limited to only tasks representing daily activities (Timed Up and Go, gait speed) in an attempt to minimize response shift from study participation, or ratings could be repeated at the end of the session to investigate response shift. More data would be collected about potential intervening events since the previous study visit, events that could create response shift with or without timeframe incongruency. Collection of transition ratings relative to not only the past six months, but also to the past year, would allow more thorough investigation into the natural history of change in mobility, and theories such as gradual change. Administration in a larger cohort would provide more power to facilitate quantitative analyses based on the frame of reference.

When considered together, these results support global self-ratings of mobility and balance as valid and meaningful in a moderate to high-functioning community-dwelling cohort. While we have focused on

discordance, concordance was also found, particularly for transition ratings of mobility and decline in gait speed over a six month period without intervention. This suggests that transition ratings could be useful as an initial screening tool of almost no burden to monitor older adults for mobility decline with the passage of time. Ultimately, because these ratings are not without both 'false-positives' and 'false-negatives,' self-ratings would be used to supplement rather than replace longer self-report and performance-based measures. Older adults who report worse mobility might undergo subsequent assessment, with measures organized from lowest to highest burden, before referring to a mobility specialist such as a physical therapist for more detailed evaluation and intervention.

APPENDIX

Semi-Structured Interview Script

Domain Definitions:

OPEN-ENDED DEFINITIONS OF MOBILITY AND BALANCE:

'The investigators in this study would like to learn more about how older adults answer the questions that I have just asked you. We think that different people may answer these questions in different ways. For example, some of the terms we use may not mean the same thing to everyone. When you answered the previous questions about your ***Mobility*** *(alt:* ***Balance****), what did you mean by* ***Mobility****? What were you thinking of?'* The following prompts should be used by domain:

- ***MOBILITY***: *Do you think of any activities in particular? Do you have to be up on your feet, or do you consider other activities where you are not up on your feet?*
    - *Does it matter where you perform the activities/ do you think of certain environments?*
- ***BALANCE***: *Do you only consider activities where you are standing and walking?*
    - *Does it have anything to do with falling?*

CLOSED-ENDED MOBILITY DEFINITIONS*:*

If you had to choose from the following options, which best describe what ***mobility*** *means to you?*

CHECK ALL THAT APPLY or NONE, Indicate our own definition if desired. . .

__ *Walking in all environments that I encounter daily*

__ *Walking around my home*

__ *Getting outside of my home/ getting around in the community*

__ *Recreational activities in the community (such as golfing, swimming, bowling, dancing)*

__ *Transfers – moving in and out of bed, getting up and down from a chair, or on/off of a toilet*

__ *Completing basic daily activities like getting dressed or bathing*

__ *Getting around without an assistive device (cane/walker)*

__ *Other:* _____

CLOSED-ENDED BALANCE DEFINITIONS*:*

*If you had to describe **balance** using one or more of the options below, which would you choose?*

CHECK ALL THAT APPLY or NONE, Indicate our own definition if desired. . .

__ *Being steady on my feet*

__ *Not being shaky*

__ *Being confident that I will not fall*

__ *Other: _____*

Timeframe Domain:

***CURRENT STATE TIMEFRAME**: 'When you answered the question in each set about how you were doing in general, what **period of time** were you thinking of?'*

Prompt used as needed: '*Were you thinking of this hour? Today? Over the past week? The past month? Past year?'*

TRANSITION METHOD: '*When I asked you to compare how you are doing now to how you were doing six months ago, how did you do that? Which did you think of first and/or most*?

__ How I am doing now

__ How I was doing then (6 months ago)

RECALL BIAS: '*Was it hard to remember how you were doing 6 months ago?*'

__ Yes

__ No

Interviewer: Check here ___ if participant reports no difficulty, but verbalized difficulty while answering the global self-rating questions.

Frame of Reference Domain:

OPEN-ENDED FRAME OF REFERENCE QUESTION: '*When you choose an answer to describe how you are doing, how do you decide which answer to choose? What does each of those response categories mean to you?'*

Prompts to be asked of everyone:

- '*Do you compare yourself to someone when you choose to pick or not pick a certain answer?'*
- *'Do you think of specific people or imagine a type of person who would fit into each category?'*

CLOSED-ENDED FRAME OF REFERENCE QUESTION*:*

'Please finish this sentence: 'When asked to describe how I am doing, I compare myself to. . . ''

CHECK ALL THAT APPLY or NONE, Indicate our own definition if desired. . .

__ *How I did at some point in the past*

__ *How I think I should be doing at my age*

__ *A specific person I know*

BIBLIOGRAPHY

1 Haentjens P, Magaziner J, Colon-Emeric CS, et al. Meta-analysis: excess mortality after hip fracture among older women and men. *Ann Intern Med*.152(6):380-90.

2 Cauley JA, Thompson DE, Ensrud KC, et al. Risk of mortality following clinical fractures. *Osteoporos Int*. 2000;11(7):556-61.

3 Tinetti ME, Williams CS. Falls, injuries due to falls, and the risk of admission to a nursing home. *N Engl J Med*. 1997;337(18):1279-84.

4 Dellinger AM, Stevens JA. The injury problem among older adults: mortality, morbidity and costs. *J Safety Res*. 2006;37(5):519-22.

5 Sattin RW. Falls among older persons: a public health perspective. *Annu Rev Public Health*. 1992;13:489-508.

6 von Bonsdorff M, Rantanen T, Laukkanen P, et al. Mobility limitations and cognitive deficits as predictors of institutionalization among community-dwelling older people. *Gerontology*. 2006;52(6):359-65.

7 Guralnik JM, Simonsick EM. Physical disability in older Americans. *J Gerontol*. 1993;48 Spec No:3-10.

8 Brach JS, VanSwearingen JM, Newman AB, et al. Identifying early decline of physical function in community-dwelling older women: performance-based and self-report measures. *Phys Ther*. 2002;82(4):320-8.

9 Rubenstein LZ, Powers CM, MacLean CH. Quality indicators for the management and prevention of falls and mobility problems in vulnerable elders. *Ann Intern Med*. 2001;135(8 Pt 2):686-93.

10 Tinetti ME. Clinical practice. Preventing falls in elderly persons. *N Engl J Med*. 2003;348(1):42-9.

11 Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2009(2):CD007146.

12 Steinberg M, Cartwright C, Peel N, et al. A sustainable programme to prevent falls and near falls in community dwelling older people: results of a randomised trial. *J Epidemiol Community Health*. 2000;54(3):227-32.

13 Wolf SL, Barnhart HX, Kutner NG, et al. Reducing frailty and falls in older persons: an investigation of Tai Chi and computerized balance training. Atlanta FICSIT Group. Frailty and Injuries: Cooperative Studies of Intervention Techniques. *J Am Geriatr Soc*. 1996;44(5):489-97.

14 Li F, Harmer P, Fisher KJ, et al. Tai Chi and fall reductions in older adults: a randomized controlled trial. *J Gerontol A Biol Sci Med Sci*. 2005;60(2):187-94.

15 Protas EJ, Tissier S. Strength and speed training for elders with mobility disability. *J Aging Phys Act*. 2009;17(3):257-71.

16 VanSwearingen JM, Perera S, Brach JS, et al. A randomized trial of two forms of therapeutic activity to improve walking: effect on the energy cost of walking. *J Gerontol A Biol Sci Med Sci*. 2009;64(11):1190-8.

17 Fried LP, Bandeen-Roche K, Williamson JD, et al. Functional decline in older adults: expanding methods of ascertainment. *J Gerontol A Biol Sci Med Sci*. 1996;51(5):M206-14.

18 Weiss CO, Hoenig HM, Fried LP. Compensatory strategies used by older adults facing mobility disability. *Arch Phys Med Rehabil*. 2007;88(9):1217-20.

19 Tomey KM, Sowers MR. Assessment of physical functioning: a conceptual model encompassing environmental factors and individual compensation strategies. *Phys Ther*. 2009;89(7):705-14.

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

21 McNeely EA, Clements SD. Recruitment and retention of the older adult into research studies. *J Neurosurg Nurs*. 1994;26(1):57-61.

22 Mody L, Miller DK, McGloin JM, et al. Recruitment and retention of older adults in aging research. *J Am Geriatr Soc*. 2008;56(12):2340-8.

23 Sherbourne CD, Meredith LS. Quality of self-report data: a comparison of older and younger chronically ill patients. *J Gerontol*. 1992;47(4):S204-11.

24 Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11(3):193-205.

25 Studenski S, Perera S, Wallace D, et al. Physical performance measures in the clinical setting. *J Am Geriatr Soc*. 2003;51(3):314-22.

26 Montero-Odasso M, Schapira M, Varela C, et al. Gait velocity in senior people. An easy test for detecting mobility impairment in community elderly. *J Nutr Health Aging*. 2004;8(5):340-3.

27 DeSalvo KB, Fisher WP, Tran K, et al. Assessing measurement properties of two single-item general health measures. *Qual Life Res*. 2006;15(2):191-201.

28 Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473-83.

29 Read JL, Quinn RJ, Hoefer MA. Measuring overall health: an evaluation of three important approaches. *J Chronic Dis*. 1987;40 Suppl 1:7S-26S.

30 Bergner M. Measurement of health status. *Med Care*. 1985;23(5):696-704.

31 Wyrwich KW, Tardino VM. Understanding global transition assessments. *Qual Life Res*. 2006;15(6):995-1004.

32 Guyatt GH, Norman GR, Juniper EF, et al. A critical look at transition ratings. *J Clin Epidemiol*. 2002;55(9):900-8.

33 Juniper EF, Guyatt GH, Willan A, et al. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *J Clin Epidemiol*. 1994;47(1):81-7.

34 Guyatt GH. Making sense of quality-of-life data. *Med Care*. 2000;38(9 Suppl):II175-9.

35 Perera S, Mody SH, Woodman RC, et al. Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc*. 2006;54(5):743-9.

36 Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol*. 2002;14(2):109-14.

37 Guyatt G, Sackett D, Taylor DW, et al. Determining optimal therapy--randomized trials in individual patients. *N Engl J Med*. 1986;314(14):889-92.

38 Schacter DL. The seven sins of memory. Insights from psychology and cognitive neuroscience. *Am Psychol*. 1999;54(3):182-203.

39 Ross M. Relation of implicit theories to the construction of personal histories. *Psychol Rev*. 1989;96:341-7.

40 McFarland C, Ross M, Giltrow M. Biased recollections in older adults: the role of implicit theories of aging. *J Pers Soc Psychol*. 1992;62(5):837-50.

41 Samson MM, Meeuwsen IB, Crowe A, et al. Relationships between physical performance measures, age, height and body weight in healthy adults. *Age Ageing*. 2000;29(3):235-42.

42 Ferrucci L, Guralnik JM, Simonsick E, et al. Progressive versus catastrophic disability: a longitudinal view of the disablement process. *J Gerontol A Biol Sci Med Sci*. 1996;51(3):M123-30.

43 Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum*. 1993;36(6):729-40.

44 Fischer D, Stewart AL, Bloch DA, et al. Capturing the patient's view of change as a clinical outcome measure. *JAMA*. 1999;282(12):1157-62.

45 Herrmann D. Reporting current, past, and changed health status. What we know about distortion. *Med Care*. 1995;33(4 Suppl):AS89-94.

46 Hess RJ, Brach JS, Piva SR, et al. Walking skill can be assessed in older adults: validity of the Figure-of-8 Walk Test. *Phys Ther*.90(1):89-99.

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

48 Schoenfeld DE, Malmrose LC, Blazer DG, et al. Self-rated health and mortality in the high-functioning elderly--a closer look at healthy individuals: MacArthur field study of successful aging. *J Gerontol*. 1994;49(3):M109-15.

49 Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav*. 1997;38(1):21-37.

50 Grant MD, Piotrowski ZH, Chappell R. Self-reported health and survival in the Longitudinal Study of Aging, 1984-1986. *J Clin Epidemiol*. 1995;48(3):375-87.

51 DeSalvo KB, Bloser N, Reynolds K, et al. Mortality prediction with a single general self-rated health question. A meta-analysis. *J Gen Intern Med*. 2006;21(3):267-75.

52 Lee Y. The predictive value of self assessed general, physical, and mental health on functional decline and mortality in older adults. *J Epidemiol Community Health*. 2000;54(2):123-9.

53 Krause NM, Jay GM. What do global self-rated health items measure? *Med Care*. 1994;32(9):930-42.

54 Perera S, Studenski S, Chandler JM, et al. Magnitude and patterns of decline in health and function in 1 year affect subsequent 5-year survival. *J Gerontol A Biol Sci Med Sci*. 2005;60(7):894-900.

55 Davies AR, Ware JE, Jr., *Measuring health perceptions in the Health Insurance Experiment.* 1981: Santa Monica, CA.

56 Mavaddat N, Kinmonth AL, Sanderson S, et al. What determines Self-Rated Health (SRH)? A cross-sectional study of SF-36 health domains in the EPIC-Norfolk cohort. *J Epidemiol Community Health*.

57 Singh-Manoux A, Martikainen P, Ferrie J, et al. What does self rated health measure? Results from the British Whitehall II and French Gazel cohort studies. *J Epidemiol Community Health*. 2006;60(4):364-72.

58 Phelan EA, Anderson LA, LaCroix AZ, et al. Older adults' views of "successful aging"--how do they compare with researchers' definitions? *J Am Geriatr Soc*. 2004;52(2):211-6.

59 Zecevic AA, Salmoni AW, Speechley M, et al. Defining a fall and reasons for falling: comparisons among the views of seniors, health care providers, and the research literature. *Gerontologist*. 2006;46(3):367-76.

60 Sarkisian CA, Steers WN, Hays RD, et al. Development of the 12-item Expectations Regarding Aging Survey. *Gerontologist*. 2005;45(2):240-8.

61 Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-33.

62 Haley SM, Jette AM, Coster WJ, et al. Late Life Function and Disability Instrument: II. Development and evaluation of the function component. *J Gerontol A Biol Sci Med Sci*. 2002;57(4):M217-22.

63 Jette AM, Haley SM, Coster WJ, et al. Late life function and disability instrument: I. Development and evaluation of the disability component. *J Gerontol A Biol Sci Med Sci*. 2002;57(4):M209-16.

64 Guyatt GH, King DR, Feeny DH, et al. Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation. *J Clin Epidemiol*. 1999;52(3):187-92.

65 Liang MH, Larson MG, Cullen KE, et al. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum*. 1985;28(5):542-7.

66 Garratt AM, Brealey S, Robling M, et al. Development of the knee quality of life (KQoL-26) 26-item questionnaire: data quality, reliability, validity and responsiveness. *Health Qual Life Outcomes*. 2008;6:48.

67 Crews JE, Campbell VA. Vision impairment and hearing loss among community-dwelling older Americans: implications for health and functioning. *Am J Public Health*. 2004;94(5):823-9.

68 Guralnik JM, Ferrucci L, Balfour JL, et al. Progressive versus catastrophic loss of the ability to walk: implications for the prevention of mobility loss. *J Am Geriatr Soc*. 2001;49(11):1463-70.

69 Fried LP, Guralnik JM. Disability in older adults: evidence regarding significance, etiology, and risk. *J Am Geriatr Soc*. 1997;45(1):92-100.

70 Tan TL, Ding YY, Lee A. Impaired mobility in older persons attending a geriatric assessment clinic: causes and management. *Singapore Med J*. 2001;42(2):68-72.

71 Gill TM, Allore HG, Hardy SE, et al. The dynamic nature of mobility disability in older persons. *J Am Geriatr Soc*. 2006;54(2):248-54.

72 Andersen-Ranberg K, Christensen K, Jeune B, et al. Declining physical abilities with age: a cross-sectional study of older twins and centenarians in Denmark. *Age Ageing*. 1999;28(4):373-7.

73 Weiss CO, Fried LP, Bandeen-Roche K. Exploring the hierarchy of mobility performance in high-functioning older women. *J Gerontol A Biol Sci Med Sci*. 2007;62(2):167-73.

74 Clark DO, Stump TE, Hui SL, et al. Predictors of mobility and basic ADL difficulty among adults aged 70 years and older. *J Aging Health*. 1998;10(4):422-40.

75 Guralnik JM, Ferrucci L, Simonsick EM, et al. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med*. 1995;332(9):556-61.

76 Hirvensalo M, Rantanen T, Heikkinen E. Mobility difficulties and physical activity as predictors of mortality and loss of independence in the community-living older population. *J Am Geriatr Soc*. 2000;48(5):493-8.

77 Onder G, Penninx BW, Lapuerta P, et al. Change in physical performance over time in older women: the Women's Health and Aging Study. *J Gerontol A Biol Sci Med Sci*. 2002;57(5):M289-93.

78 Liu-Ambrose T, Khan KM, Eng JJ, et al. Balance confidence improves with resistance or agility training. Increase is not correlated with objective changes in fall risk and physical abilities. *Gerontology*. 2004;50(6):373-82.

79 Frank JS, Patla AE. Balance and mobility challenges in older adults: implications for preserving community mobility. *Am J Prev Med*. 2003;25(3 Suppl 2):157-63.

80 VanSwearingen JM, Brach JS. Making geriatric assessment work: selecting useful measures. *Phys Ther*. 2001;81(6):1233-52.

81 Cyarto EV, Brown WJ, Marshall AL, et al. Comparative effects of home- and group-based exercise on balance confidence and balance ability in older adults: cluster randomized trial. *Gerontology*. 2008;54(5):272-80.

82 Bandura A, *Self-efficacy mechanism in physiological activation and health-promoting behavior*, in *Neurobiology of Learning, Emotion and Affect*, Madden J, Editor. 1991, Raven Press: New York.

83 Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev*. 1977;84(2):191-215.

84 Vellas BJ, Wayne SJ, Romero LJ, et al. Fear of falling and restriction of mobility in elderly fallers. *Age Ageing*. 1997;26(3):189-93.

85 Arnold CM, Busch AJ, Schachter CL, et al. A Randomized Clinical Trial of Aquatic versus Land Exercise to Improve Balance, Function, and Quality of Life in Older Women with Osteoporosis. *Physiother Can*. 2008;60(4):296-306.

86 Metz SM, Wyrwich KW, Babu AN, et al. Validity of patient-reported health-related quality of life global ratings of change using structural equation modeling. *Qual Life Res*. 2007;16(7):1193-202.

87 Luszcz MA, Bryan J. Toward understanding age-related memory loss in late adulthood. *Gerontology*. 1999;45(1):2-9.

88 Small GW. What we need to know about age related memory loss. *BMJ*. 2002;324(7352):1502-5.

89 Quine S, Morrell S. Fear of loss of independence and nursing home admission in older Australians. *Health Soc Care Community*. 2007;15(3):212-20.

90 Kamper SJ, Ostelo RW, Knol DL, et al. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol*.63(7):760-766 e1.

91 Boult C, Boult LB, Morishita L, et al. A randomized clinical trial of outpatient geriatric evaluation and management. *J Am Geriatr Soc*. 2001;49(4):351-9.

92 Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med*. 1999;48(11):1531-48.

93 Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med*. 1999;48(11):1507-15.

94 Mancuso CA, Charlson ME. Does recollection error threaten the validity of cross-sectional studies of effectiveness? *Med Care*. 1995;33(4 Suppl):AS77-88.

95 Aseltine RH, Jr., Carlson KJ, Fowler FJ, Jr., et al. Comparing prospective and retrospective measures of treatment outcomes. *Med Care*. 1995;33(4 Suppl):AS67-76.

96 Cress ME, Schechtman KB, Mulrow CD, et al. Relationship between physical performance and self-perceived physical function. *J Am Geriatr Soc*. 1995;43(2):93-101.

97 Guralnik JM, Winograd CH. Physical performance measures in the assessment of older persons. *Aging (Milano)*. 1994;6(5):303-5.

98 Kempen GI, Steverink N, Ormel J, et al. The assessment of ADL among frail elderly in an interview survey: self-report versus performance-based tests and determinants of discrepancies. *J Gerontol B Psychol Sci Soc Sci*. 1996;51(5):P254-60.

99 Myers AM, Holliday PJ, Harvey KA, et al. Functional performance measures: are they superior to self-assessments? *J Gerontol*. 1993;48(5):M196-206.

100 Reuben DB, Valle LA, Hays RD, et al. Measuring physical function in community-dwelling older persons: a comparison of self-administered, interviewer-administered, and performance-based measures. *J Am Geriatr Soc*. 1995;43(1):17-23.

101 Simonsick EM, Newman AB, Nevitt MC, et al. Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC study. *J Gerontol A Biol Sci Med Sci*. 2001;56(10):M644-9.

102 Alexander NB, Guire KE, Thelen DG, et al. Self-reported walking ability predicts functional mobility performance in frail older adults. *J Am Geriatr Soc*. 2000;48(11):1408-13.

103 Stratford PW, Kennedy D, Pagura SM, et al. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. *Arthritis Rheum*. 2003;49(4):535-40.

104 Seeman TE, Charpentier PA, Berkman LF, et al. Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies of successful aging. *J Gerontol*. 1994;49(3):M97-108.

105 Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;49(2):M85-94.

106 Reuben DB, Siu AL, Kimpau S. The predictive validity of self-report and performance-based measures of function and health. *J Gerontol*. 1992;47(4):M106-10.

107 Maly MR, Costigan PA, Olney SJ. Determinants of self-report outcome measures in people with knee osteoarthritis. *Arch Phys Med Rehabil*. 2006;87(1):96-104.

108 Hawker G, Melfi C, Paul J, et al. Comparison of a generic (SF-36) and a disease specific (WOMAC) (Western Ontario and McMaster Universities Osteoarthritis Index) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol*. 1995;22(6):1193-6.

109 Jette AM. The Functional Status Index: reliability and validity of a self-report functional disability measure. *J Rheumatol Suppl*. 1987;14 Suppl 15:15-21.

110 Dorevitch MI, Cossar RM, Bailey FJ, et al. The accuracy of self and informant ratings of physical functional capacity in the elderly. *J Clin Epidemiol*. 1992;45(7):791-8.

111 McGinnis GE, Seward ML, DeJong G, et al. Program evaluation of physical medicine and rehabilitation departments using self-report Barthel. *Arch Phys Med Rehabil*. 1986;67(2):123-5.

112 Magaziner J, Bassett SS, Hebel JR, et al. Use of proxies to measure health and functional status in epidemiologic studies of community-dwelling women aged 65 years and older. *Am J Epidemiol*. 1996;143(3):283-92.

113 Ramey DR, Raynauld JP, Fries JF. The health assessment questionnaire 1992: status and review. *Arthritis Care Res*. 1992;5(3):119-29.

114 Ward MM. Clinical measures in rheumatoid arthritis: which are most useful in assessing patients? *J Rheumatol*. 1994;21(1):17-27.

115 Ware JE, Jr. Monitoring and evaluating health services. *Med Care*. 1985;23(5):705-9.

116 Gill TM, Desai MM, Gahbauer EA, et al. Restricted activity among community-living older persons: incidence, precipitants, and health care utilization. *Ann Intern Med*. 2001;135(5):313-21.

117 Yardley L, Smith H. A prospective study of the relationship between feared consequences of falling and avoidance of activity in community-living older people. *Gerontologist*. 2002;42(1):17-23.

118 Brach JS, FitzGerald S, Newman AB, et al. Physical activity and functional status in community-dwelling older women: a 14-year prospective study. *Arch Intern Med*. 2003;163(21):2565-71.

119 Mor V, Murphy J, Masterson-Allen S, et al. Risk of functional decline among well elders. *J Clin Epidemiol*. 1989;42(9):895-904.

120 Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health Qual Life Outcomes*. 2004;2:14.

121 Ryff CD. Self-perceived personality change in adulthood and aging. *J Pers Soc Psychol*. 1982;42(1):108-15.

122 Streiner DL, Norman GR, *Health measurement scales: a practical guide to their development and use*. 2nd ed. 1995, New York: Oxford University Press.

123 Sarkisian CA, Hays RD, Berry SH, et al. Expectations regarding aging among older adults and physicians who care for older adults. *Med Care*. 2001;39(9):1025-36.

124 Fried LP, Bandeen-Roche K, Chaves PH, et al. Preclinical mobility disability predicts incident mobility disability in older women. *J Gerontol A Biol Sci Med Sci*. 2000;55(1):M43-52.

125 Gill TM, Robison JT, Tinetti ME. Difficulty and dependence: two components of the disability continuum among community-living older persons. *Ann Intern Med*. 1998;128(2):96-101.

126 Kane RA, Kane RL, *Assessing the elderly: a practical guide to measurement*. 1981, Lexington, MA: Lexington Books.

127 Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17(3):163-70.

128 Sheikh JI, Yesavage JA, *Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version*, in *Clinical Gerontology: A Guide to Assessment and Intervention*. 1986, The Haworth Press: NY. p. 165-173.

129 Weschler D, *Weschler Adult Intelligence Scale-Revised*. 1981: San Antonio, TX.

130 Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychol Rev*. 1996;103(3):403-28.

131 Rosano C, Newman AB, Katz R, et al. Association between lower digit symbol substitution test score and slower gait and greater risk of mortality and of developing incident disability in well-functioning older adults. *J Am Geriatr Soc*. 2008;56(9):1618-25.

132 Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. *J Gerontol A Biol Sci Med Sci*. 1995;50A(1):M28-34.

133 Myers AM, Fletcher PC, Myers AH, et al. Discriminative and evaluative properties of the activities-specific balance confidence (ABC) scale. *J Gerontol A Biol Sci Med Sci*. 1998;53(4):M287-94.

134 Fitti JE, Kovar MG. The Supplement on Aging to the 1984 National Health Interview Survey. *Vital Health Stat 1*. 1987(21):1-115.

135 Lachman ME, Howland J, Tennstedt S, et al. Fear of falling and activity restriction: the survey of activities and fear of falling in the elderly (SAFE). *J Gerontol B Psychol Sci Soc Sci*. 1998;53(1):P43-50.

136 Nevitt MC, Cummings SR, Kidd S, et al. Risk factors for recurrent nonsyncopal falls. A prospective study. *JAMA*. 1989;261(18):2663-8.

137 Winograd CH, Lemsky CM, Nevitt MC, et al. Development of a physical performance and mobility examination. *J Am Geriatr Soc*. 1994;42(7):743-9.

138 Abellan van Kan G, Rolland Y, Andrieu S, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. *J Nutr Health Aging*. 2009;13(10):881-9.

139 Cesari M, Kritchevsky SB, Penninx BW, et al. Prognostic value of usual gait speed in well-functioning older people--results from the Health, Aging and Body Composition Study. *J Am Geriatr Soc*. 2005;53(10):1675-80.

140 Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci*. 2000;55(4):M221-31.

141 Hardy SE, Perera S, Roumani YF, et al. Improvement in usual gait speed predicts better survival in older adults. *J Am Geriatr Soc*. 2007;55(11):1727-34.

142 Kwon S, Perera S, Pahor M, et al. What is a meaningful change in physical performance? Findings from a clinical trial in older adults (the LIFE-P study). *J Nutr Health Aging*. 2009;13(6):538-44.

143 Imms FJ, Edholm OG. Studies of gait and mobility in the elderly. *Age Ageing*. 1981;10(3):147-56.

144 Hoxie RE, Rubenstein LZ. Are older pedestrians allowed enough time to cross intersections safely? *J Am Geriatr Soc*. 1994;42(3):241-4.

145 Fried LP, Herdman SJ, Kuhn KE, et al. Preclinical Disability: Hypotheses about the bottom of the iceberg. *J Aging Health*. 1991;3(2):285-300.

146 Onder G, Penninx BW, Ferrucci L, et al. Measures of physical performance and risk for progressive and catastrophic disability: results from the Women's Health and Aging Study. *J Gerontol A Biol Sci Med Sci*. 2005;60(1):74-9.

147 Patton M, *Qualitative Evaluation and Research Methods*. 1990, Newbury Park, CA: Sage Publications.

148 Crabtree BF, Miller WL, eds. *Doing Qualitative Research.* 2nd ed. 1999, Sage: London. 406.

149 Strauss A, Corbin J, *Basics of qualitative research: Grounded theory procedures and techniques*. 1990, London: Sage.

150 Sandelowski M. Sample size in qualitative research. *Res Nurs Health*. 1995;18(2):179-183.

151 Feinstein AR, Josephy BR, Wells CK. Scientific and clinical problems in indexes of functional disability. *Ann Intern Med*. 1986;105(3):413-20.

152 Hile ES, Brach JS, Perera S, et al. Influence of Light Support on Tandem Stance in Community-Dwelling Older Adults *(Abstract) Combined Sections Meeting of the American Physical Therapy Association; San Diego, CA.* 2010.

153 Vellas BJ, Wayne SJ, Romero L, et al. One-leg balance is an important predictor of injurious falls in older persons. *J Am Geriatr Soc*. 1997;45(6):735-8.

154 Bischoff HA, Stahelin HB, Monsch AU, 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):315-20.

155 Guralnik JM, Ferrucci L. Assessing the building blocks of function: utilizing measures of functional limitation. *Am J Prev Med*. 2003;25(3 Suppl 2):112-21.

156 Jette AM, Cleary PD. Functional disability assessment. *Phys Ther*. 1987;67(12):1854-9.

157 Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. *Ann Intern Med*. 2002;136(6):471-6.

158 Nevitt MC, Cummings SR, Hudes ES. Risk factors for injurious falls: a prospective study. *J Gerontol*. 1991;46(5):M164-70.

159 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. *Control Clin Trials*. 1994;10:407-415.

160 Schwartz CE, Bode R, Repucci N, et al. The clinical significance of adaptation to changing health: a meta-analysis of response shift. *Qual Life Res*. 2006;15(9):1533-50.

161 Idler EL, Kasl S. Health perceptions and survival: do global evaluations of health status really predict mortality? *J Gerontol*. 1991;46(2):S55-65.