Supplemental Information #1: Laboratory Measures of Physical Function

Physical fitness is defined as " A set of attributes that people have or achieve that relates to the ability to perform physical activity" [1]. Components of physical fitness are cardiorespiratory fitness, which is the ability of the circulatory and respiratory systems to supply oxygen during sustained physical activity [1]; muscle strength, which is the ability of the muscle to exert force; muscle endurance, which is the ability of a muscle to continue to perform without fatigue; flexibility, which is the range of motion available at a joint. The "gold standard" measures of each of these attributes requires testing in the laboratory. The following is a brief description of the principles of these laboratory tests.

Cardiorespiratory Fitness: Exercise testing is used to measure cardiorespiratory fitness, the gold-standard measurement being maximal oxygen uptake (VO_{2max}) or VO_{2peak}. The mode of exercise is typically a treadmill or calibrated cycle ergometer. Analysis of respiratory gases using open circuit indirect calorimetry (measurement of expired oxygen, carbon dioxide and ventilation) provides a direct measure of oxygen uptake. The protocol typically is a gradual increase in external work (i.e., increasing grade on the treadmill at a constant speed, or a combination of increasing speed and grade; or gradually increasing resistance on a cycle ergometer), which progresses until the subject is unable to keep up with the speed. The criteria for achieving maximal levels are 1) achievement of near age-predicted maximal heart rate, 2) leveling off of oxygen uptake despite increasing external work, 3) respiratory exchange ratio (expired CO_2/ expired O_2 ratio) > 1.0 and, 4) if measured, a blood lactate above 8 mm/L. Individuals who are less fit and/or compromised by chronic conditions typically do not achieve these criteria. In such cases, the protocol continues with increasing

work until the subject is unable to continue, making the test a "symptom-limited' test, and the measurement would be referred to as VO_{2peak}.

Peak exercise capacity can be estimated from the external work achieved on the treadmill or cycle ergometer if respiratory gas analysis is not possible. Estimates of oxygen requirements for a given level of work are derived from metabolic equations [2] that are based on a required oxygen requirement for a given speed and grade on the treadmill or a given wattage of work on a cycle ergometer. This estimate is often converted to MET (Metabolic unit) levels (1 Met is a unit of resting oxygen consumption that is estimated to be 3.5 ml/kg body weight/minute). Thus, a subject who stops exercise at a treadmill speed of 3.0 mph/10% grade would have an estimated oxygen uptake of 26 ml/kg/min or 7.4 METs based on metabolic equations. This estimate requires use of calibrated equipment and no use of support during treadmill walking.

The protocol used for assessing maximal or peak exercise capacity should start at a low level (i.e. 2 METs) and the increments should be gradual (i.e. 0.5 to 1.0 METs/ stage). Given the low exercise capacity that characterizes the CKD population (average 4-7 METs), this will allow for several stages, so the pattern of rise in heart rate, blood pressure and ventilation can be assessed. During maximal exercise testing electrocardiogram should be monitored as well as blood pressure, symptoms and rating of perceived exertion. Since most patients stop exercise tests because of leg fatigue, the interpretation of exercise testing may be difficult in patients with CKD, (see discussion on this topic by Copley and Lindberg [3].

Peak exercise capacity can also be estimated from submaximal exercise on a calibrated cycle ergometer, based on the assumption that heart rate increases linearly with increasing energy expenditure (an assumption that may or may not be appropriate in patients treated with dialysis). Heart rate is measured after 3 minutes (steady state) of 3 or 4 submaximal exercise levels and plotted against the external work performed . The plotted heart rate is extrapolated to the age-predicted max heart rate, and the oxygen uptake at the corresponding workload at this max heart rate is the estimated VO_{2max}. Submaximal exercise may be useful in comparing responses to exercise training, in that heart rate response to a standard exercise level should decrease with exercise training. Likewise the measurement of respiratory gases during submaximal work to determine the ventilatory threshold may be important in assessment of the metabolic responses to exercise training. For a review of physiologic measures that have been used during exercise in CKD studies and for more specific recommendations for protocols please refer to the excellent review by Koufaki and Kouidi [4].

Muscle strength can be assessed with a variety of methods including a manual muscle test [5], 1 repetition maximum [6], or a dynamometer [7]. Though it is beyond the scope of this paper to review in detail all of these methods, manual muscle testing uses standard test positions, gravity and manual resistance to assign a strength grade from 0 (no contraction) to 5 (strong, normal contraction) for each muscle or muscle group tested. The 1 repetition maximum (1RM) method attempts to attain the patient's maximum load capability that can be lifted (concentrically), or lowered (eccentrically) one time. Because this can be difficult particularly in patient populations, protocols for predicting the 1 RM from up to 10 repetitions have been

advocated [8]. While manual muscle testing and 1 RM testing typically test isometric, and isotonic or fixed resistance (both concentric and eccentric) contractions respectively, instrumented dynamometers allow the addition of isokinetic or fixed speed muscle testing in both shortening (concentric) or lengthening (eccentric) contractions. Test positions, and hence joint angles should be standardized in all of these methods. Common muscle groups associated with mobility include knee, hip, and back extensors. Handgrip strength is also a useful measurement in older and impaired populations as it is significantly correlated with lower extremity strength [9]. Handgrip strength may be affected by the presence of an arterio-venous fistula in hemodialysis patients [10] , thus it is unknown whether it is reflective of overall strength in these patients.

Like muscle strength, **muscle endurance** can be assessed in a number of ways, but is commonly tested using functional activities like a repeated push up, curl-up or bench press test (Canadian Physical Activity, Fitness & Lifestyle Approach Protocol, 2003, webpage: http://www.csep.ca/english/view.asp?x=609), comparing the number of repetitions correctly performed to a set of age and sex specific norms. Muscle endurance can be assessed dynamically (repetitions) or statically (isometric holding time) as has been done with the lumbar paraspinal muscles [11]. Muscle endurance can also be tested isokinetically where muscle contractions (concentric or eccentric) are repeated at a set velocity (120-180°/sec) until the individual can no longer produce at least 50% of a maximal voluntary isometric contraction force. The number of repetitions performed serves as the metric for comparison and norms are often provided by the manufacturer of the equipment or published online (Wimpenny P. 2000. *Interpretation Endurance / Fatigue)*. (online: Isokinetics Explained. http://

www.isokinetics.net/isokinetics/interpretation/endurance--fatigue.html). Other ways to assess muscle endurance include using free-weights and counting the number of repetitions that can be successfully completed at a pre-determined percentage of the 1RM (at i.e., 70% maximal contraction), or at a pre-determined percentage of bodyweight. Norms for these measurements vary.

Although **muscle power** can be assessed by measuring pedaling or arm cranking for 30 seconds at maximal speed against a constant force (Wingate Anaerobic Power Test, [12]), or by a vertical jump test [13], a more commonly used measure of muscle power in older or impaired populations is leg extension power measured with the Nottingham Power Rig [14]. Peak concentric power output of unilateral leg extensors (knee and hip), recorded in watts is measured with this standard test that requires a specialized piece of equipment.

Alternatively, the stair climb power test (SCPT) has been advocated as a clinically relevant measure of leg power in mobility limited older adults and requires only a set of stairs with a known vertical distance and the individual's mass in kilograms [15]. The SCPT is associated with more complex modes of power testing and with mobility performance. Finally, a recent paper suggests that leg power in older adults may be accurately assessed using the initial 20 seconds of a 30 second chair rise test [16].

Table 1 presents values for muscle power and muscle strength measured in knee extension and hand grip in healthy populations.

The diagnosis and treatment of sarcopenia includes muscle strength. The European Working Group on Sarcopenia in Older People published a consensus report on the definition and diagnosis of sarcopenia [17], which provides recommendations for use of muscle strength and physical performance measures in the diagnosis and treatment of sarcopenia in both research and clinical practice.

For a review of muscle function assessment that have been used in CKD studies and for more specific recommendations for protocols please refer to the excellent review by Koufaki and Kouidi [4].

Table 1: Mean values and 95% confidence intervals for knee-extension torque, handgrip, and muscle power, in the InCHIANTI study participants, according to gender and age strata (reprinted with permission [18])

Age yr	n	Knee-Extension Torque (N/dm)	Handgrip , Kg	Muscle Power, W
		Men		
20-29	25	802.0 (722.5-881.4)	61.1 (57.0-65.2)	279.5 (256.4-302.6)
30-39	25	766.9 (677.2-856.6)	56,4 (52,2-60.7)	255,8 (237.2-274,3)
40-40	27	643.4 (598.1-688.6)	53.2 (48.7-57.6)	240.7 (221.0-260.4)
50-64	43	656.5 (603.3-713.6)	49.1 (45.3-52.9)	196.3 (179.5-213.2)
65-745	230	524.5 (505.7-543.2)	39.2 (37.9-40.5)	150.6 (144.6-156.6)
75-85	97	453.8 (423.7-484.0)	31.8 (29.7-33.9)	111.8 (103.6-120.0)
85+	22	320.4 (270.9-370.0)	27.1 (22.8-31.3)	71.8 (55.2-88.4)
ANOVA		p<0.0001	P<0.0001	p<0.0001
		Women		
20-29	22	552.0 (500.5-603.5)	35.6 (32.0-39.1)	233.5 (217.8-249.1)
30-39	31	455.9 (413.6-498.2)	34.3 (32.3-36.3)	180.3 (164.6-196.1)
40-49	26	427.9 (387.8-467.9)	31,8 (29.5-34.1)	146.4 (127.0-165.8)
50-64	58	386.6 (360.9-412.3)	27.1 (25.3-29.0)	107.0 (98.1-115.9)
65-74	255	327.4 (315.2-339.6)	22.2 (21.2-23.2)	83.0 (78.7-87.2)
75-85	134	269.7 (254.90284.6)	19.3 (17.9-20.7)	59.9 (54.7-65.0)
85+	35	237.0 (211.1-263.0)	14.5 (12.9-16.2)	55.2 (47.7-62.7)
ANOVA		p<0.0001	P<0.0001	p<0.0001

Supplemental Materials #2: Meaningful change and Indicators of Impaired Performance

Meaningful change of Physical Performance Measures:

When selecting a measure to evaluate changes in a dimension of physical function as a result of a treatment or intervention, it is important to understand what magnitude of change is important. An intervention may result in a statistically significant change, however the magnitude of change may not be considered to be important either clinically or to a patient (in self-reported measures). The concept of minimal clinically important difference (MCID) has been proposed to refer to the smallest difference in a score that is considered to be worthwhile or important [19]. For patient reported outcomes, the MCID has been defined as the "smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management"[19]. Clinically it has been defined as "the smallest effect size that would lead them to recommend a therapy to their patients" [19]. Thus choosing a specific measure of physical function should include consideration of whether expected changes with an intervention or over time will result in changes that are clinically important, so it can determined if there are clinically important differences between groups or over time. Likewise, when using a measure to assess change, the interpretation of results should include some consideration of MCID.

The two most common approaches to determining clinical meaningful differences are distribution-based assessment and anchor-based assessment [19,20]. The most typical interpretation of change based on the distribution-based approach is in comparing the difference between two groups at one point or the change over time in one group to the

standard deviation at baseline, which is best know as the effect size (ES) statistic. A small ES would be 0.20 (0.20 of the baseline SD), a medium ES would be 0.50 and a large ES would be 0.80 [19]. Hays and Wooley [19] suggest that the threshold for an MCID would correspond to a small ES. Some have suggested that a MCID would be a difference or a change of 1/2 of a SD [21].

Anchor-based approaches link changes in the outcome measure with changes in a clinical parameter (prospective change), a global change such as better or worse (retrospective report), or linked with antecedent causes (life events, treatment, passage of time) or subsequent consequences such as utilization or mortality. Anchor-based approaches can be used in combination with a distribution-based statistic such as ES [19].

The MCID may be affected by the direction of change, not just the magnitude of change [19]. It may also depend on the baseline levels, such that less change may result for those who are close to the upper end of the measure and more for those who start at the lower end. Other factors that may affect the MCID, include the clinical and demographic characteristics of the population of interest and the trajectory of the measure over time in the population of interest [20].

Perera et al [22] used both distribution- and anchor-based methods to estimate the magnitude of meaningful change in the gait speed test, the SPPB and the 6 minute walk test. They present the estimates of small meaningful change and substantial change using several different data sources that included different populations of older adults both observational and clinical trial designs. Segura-Orti et al [23] used the distribution-based method to estimate

minimal detectable change in several performance tests in hemodialysis patients. The meaningful change in these performance measures are shown in table 1.

Table 1: Recommendations for criteria for meaningful change for common physical performance tests (from Perera, et al [22] and Segura-Orti [23])

Performance Measure	**Recommended Criterion for Meaningful Change (22)**	**Minimal Detectable Change* in Hemodialysis Patients (23)**
10 foot gait speed		
small meaningful change	0.05 m/s	
substantial meaningful change	0.10 m/s	
10 meter gait speed		
small meaningful change	0.05 m/s	
substantial meaningful change	0.10 m/s	
4 meter gait speed		
small meaningful change	0.05 m/s	
substantial meaningful change	0.10 m/s	
SPPB score		
small meaningful change	0.5 points	
substantial meaningful change	1 point	
6 minute walk distance		66.3
small meaningful change	29 m	
substantial meaningful change	50 m	
Chair stand 10 (sec)		8.4
Chair stand 60sec (reps)		4.0

* Minimal Detectable Change Scores at 90% Confidence Intervals

Singh, et al [24] used a less rigorous analysis of the changes on the ISWT resulting from a 12 week pulmonary rehabilitation program to determine minimally important improvement in the intermittent shuttle walk distance using the anchor of patient perceived change in their exercise performance. They reported that minimal perceived improvement was associated with an increase in 47.5 meters, and additional benefit resulted at ISW distance of 78.7 meters.

Interpretation of results: Interpreting changes in physical function measures resulting from an intervention should be done thoughtfully, with changes put into perspective.

An excellent example of this in terms of physical function domains is the report of a change in VO_{2peak} with exercise training in hemodialysis patients from 18.9 ± 7.9 ml/kg/min at baseline to 21.4 ± 9.5 ml/kg/min following 5 months of exercise training [25]. The change was statistically significant (p=0.03), however, the magnitude of change (2.5 ml/kg/min) is less than the 1/2 standard deviation of the baseline value (3.6), suggesting the change is not clinically significant. Likewise, the post-training value of 21.4 ml/kg/min remains remarkably low (average 68% of age-predicted values), and within the VO_{2peak} characteristic of patients with mild congestive heart failure, another indication that the change, although statistically significant may not be considered clinically significant.

Another example that demonstrates the importance of careful interpretation of data is found in the gait speed data reported in the Renal Exercise Demonstration Project [25]. The change in gait speed in the intervention group was 4.6±1.7 m/sec, which is a clinically meaningful change (as per Perera, et al, 22), however the change was more important in the context of a negative change in the control group (-1.0±1.8 m/sec. Thus, the difference in trajectory of change over time between the groups was both statistically significant, and clinically important.

Indicators of Impaired Performance

When assessment of physical performance is done in the clinic as a routine part of the patient care as suggested (figure 4 of manuscript), it is important to have a guide that indicates impaired performance on standardized mobility assessments. The data found in

table 2 is extracted from several different sources that have measures healthy individuals of various ages.

Table 2: Indicators of impaired performance on standardized mobility assessments

Age	Gait speed m/sec [a]	6 minute walk (meters) [a]	Timed Up and Go test (seconds) [a]	SPPB [b]	Chair stand 5 (seconds) [c]
all				<7*	>13.7*
men 60-69	<1.2	<511	>8		
men 70-79	<1.2	<482	>11		
men 80-89	<0.83	<385	>11		
women 60-69	<1.1	<460	>9		
women 70-79	<1.0	<442	>10		
women 60-69	<0.8	<316	>12		

[a] values are the lower 95% CI for each age decade from a meta-analyses of gait speed, (ref [26]) and of 6 minute walk (ref [27]. Timed Up and Go test values are the lower 95% CI reported by Steffen, et al [28].

[b] SPPB score < 7 is associated with increased risk of mobility-related disability (RR 2.0-4.9 compared to a score of 10-12) in the Established Population for Epidemiologic Studies in the Elderly (EPESE) cohort (ref [29].

[c] 13.7 seconds is the 50th percentile of chair stand 5 values in the Established Population for Epidemiologic Studies in the Elderly (EPESE) cohort (ref[30].

References

1. Office of the U.S. Surgeon General. *Physical Activity and Health: A Report of the Surgeon General*: U.S. Department of Health and Human Services, Public Health Service; 1996.

2. American College of Sports Medicine. *Guidelines for Exercise Testing and Prescription 5th edition*. Philadelphia: Williams & Wilkins; 1995.

3. Copley JB, Lindberg JS: The Risks of Exercise. *Adv Renal Repl Ther.*6(2):165-171, 1999

4. Koufaki P, Kouidi E: Current Best Evidence Recommendations on Measurement and Interpretation of Physical Function in Patients with Chronic Kidney Disease. *Sports Med.*40(12):1055-1074, 2010

5. Kendall FP, McCreary EK. *Muscles: Testing and Function*. Baltimore: Williams and Wilkins; 1983.

6. Nieman DC: Fitness and Sports Medicine. *MedSciSports Exer.* 23(5), 1991

7. Osternig LR: Isokinetic dynamometry: implications for muscle testing and rehabilitation. *Exerc Sport Sci Rev.*14:45-80, 1986

8. Brzycki M: Strength Testing -- Predicting a One-Rep Max from Reps-to-Fatigue. *J Phys Ed, Rec & Dance.*64(1):88, 1993

9. Viitasalo JT, Leskinen AL, Heikkinen E: Muscular strength profiles and anthropometry in random samples of men aged 31-35, 51-55 and 71-75 years. *Ergonomics.*28:1563-1574, 1985

10. Tawney KW, Tawney PJW, Hladik G, Hogan S, Falk R, Weaver C. Moore D, Lee M: The Life Readiness Program: A Physical Rehabilitation Program for Patients on Hemodialysis. *Am J Kidney Dis.*36:581-591, 2000

11. Udermann BE, Mayer JM, Graves JE, Murray SR: Quantitative Assessment of Lumbar Paraspinal Muscle Endurance. *J Athl Training.*38(3), 2003

12. Patton JF, Murphy MM, Frederick FA: Maximal power outputs during the Wingate anaerobic test. *Int J Sports Med.*6(2):82-85, 1985

13. Vandewalle H, Peres G, Monod H: Standard anaerobic exercise tests. *Sports Med (Auckland, N.Z.).*4(4), 1987

14. Bassey EJ, Short AH: A new method for measuring power output in a single leg extension: feasibility, reliability and validity. *Eur J Appl Physiol Occ Physiol.*60(5):385-390, 1990

15. Bean JF, Kiely DK, LaRose S, Alian J, Frontera WR: Is Stair Climb Power a Clinically Relevant Measure of Leg Power Impairments in At-Risk Older Adults? *Arch Phys Med Rehab.*88(5):604-609, 2007

16. Smith WN, Del Rossi G, Adams JB, Abderlarahman KZ, Asfour SA, Roos BA, Signorile JF: Simple equations to predict concentric lower-body muscle power in older adults using the 30-second chair-rise test: a pilot study. *Clin Interv Aging.*5:173-180, 2010

17. Cruz-Jentoft JF, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F, Martin FC, Michel JP, Rolland Y, Schneider SM, Topinkova E, Vandewoude M Zamboni M: Report of the European Working Group on Sarcopenia in Older People. Sarcopenia: European consensus on definition and diagnosis. *Age Ageing* 39:412-423, 2010

18. Lauretani F, Russo CR, Bandinelli S, et al.: Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. *J Appl Physiol.*95:1851-1860, 2003

19. Hays RD, Woolley JM: The Concept of Clinically Meaningful Difference in Health-Related Quality of LIfe Research. How Meaningful is it? *Qual Life Res.*3:329-338, 2000

20. Revicki D, Hays R.D., Cella D, Sloan JA: Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 61:102-109, 2008

21. Jaeschke R, Singer J, Guyatt G: Measurement of health status: ascertaining the minimal clinically important difference. *Con Clin Trials.*10:407-415, 1989

22. Perera S, Mody SH, Woodman RC, Studenski SA: Meaningful Change and Responsiveness in Common Performance Measures in Older Adults. *J Am Geriatrics Soc.*54:743-749, 2006

23. Segura-Ortı E, Martınez-Olmos FJ: Test-Retest Reliability and Minimal Detectable Change Scores for Sit-to- Stand-to-Sit Tests, the Six-Minute Walk Test, the One-Leg Heel-Rise Test, and Handgrip Strength in People Undergoing Hemodialysis. *Phys Ther.*91(8):1244-1252, 2011

24. Singh SJ, Jones PW, Evans R, Morgan MDL: Minimum clinically important improvement for the incremental shuttle walking test. *Thorax* 63:775-777, 2008

25. Painter PL, Carlson L, Carey S, Paul SM, Myll J: Physical Functioning and Health Related Quality of Life Changes with Exercise Training in Hemodialysis Patients. *Am J Kidney Dis.*35(3):482-492, 2000

26. Bohannon RW, Williams Andrews A: Normal walking speed: a descriptive meta-analysis. *Physiother.*97(3):182-189, 2011

27. Bohannon RW: Six-Minute Walk Test A Meta-Analysis of Data From Apparently Healthy Elders. *Topics Geriatric Rehab.*23(2):155-160, 2007

28. Steffen TM, Hacker TA, Mollinger L: Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. *Phys therapy.*82(2):128-137, 2002

29. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV, Studenski S, Berkman LF, Wallace RB: 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.*55A:M221-M231, 2000

30. Guralnik JM, Simonsick EM, Ferrucci L: A short physcial performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol.*49:M85-M94, 1994