Test-retest reliability, validation, and 'minimal detectable change' scores for frequently reported tests of objective physical function in patients with non-dialysis chronic kidney disease

Thomas J. Wilkinson, PhD [1]; Soteris Xenophontos, MSc[1]; Douglas W. Gould, MSc[1]; Barbara P. Vogt, MSc [2]; João L. Viana, PhD [3, 4]; Alice C. Smith, PhD [1, 5]; Emma L. Watson, PhD [1]

[1] Leicester Kidney Exercise Team, Department of Infection, Immunity and Inflammation, University of Leicester, United Kingdom; [2] Department of Clinical Medicine, Faculdade de Medicina de Botucatu, Univ Estadual Paulista, UNESP, Brazil; [3] School of Sport, Exercise and Health Sciences, Loughborough University, United Kingdom; [4] Research Center in Sports Sciences, Health Sciences and Human Development, CIDESD, University Institute of Maia, ISMAI, Portugal; [5] John Walls Renal Unit, Leicester General Hospital, University Hospitals of Leicester Trust, United Kingdom.

Disclosure of interest(s)

24   This analysis forms part of a larger body of work being completed by our group (ISRCTN

25   registration number: 36489137).

26

27   **<u>Corresponding author</u>**

28   Dr. Thomas J Wilkinson, Leicester Kidney Exercise Team, Department of Infection,

29   Immunity and Inflammation, University of Leicester, United Kingdom, LE5 4PW. Email:

30   drthomaswilkinson@gmail.com.

31

32

ABSTRACT

Physical function is an important outcome in chronic kidney disease (CKD). We aimed to establish the reliability, validity, and the 'minimal detectable change' (MDC) of several common tests used in renal rehabilitation and research. In a repeated measures design, 41 patients with CKD not requiring dialysis (stage 3b to 5) were assessed at an interval of 6 weeks. The tests were: the incremental shuttle walk test (ISWT), 'sit-to-stand' (STS) test, estimated 1 repetition maximum for quadriceps strength (e1RM), and VO_{2peak} by cardiopulmonary exercise testing (CPET). Reliability was assessed using intraclass correlation co-efficient (ICC) and Bland-Altman analysis, and absolute reliability by standard error of measurement and MDC. The ISWT, STS-60, e1RM, and CPET had 'good' to 'excellent' reliability (.973, .927, .927, and .866). STS-5 reliability was poor (.676). The MDC are: ISWT, 20 m; STS-5, 7.5 secs; STS-60, 4 reps; e1RM, 6.4 kg; VO_{2peak}, 2.8 ml/kg/min. There was strong correlation between the ISWT and VO_{2peak} ($r = 0.73$ and 0.74). Whilst there was poor correlation between the STS-5 and e1RM ($r = 0.14$ and 0.47), better correlation was seen between STS-5 and ISWT ($r = 0.55$ and 0.74). In conclusion, the ISWT, STS-60, e1RM, and CPET are reliable tests of function in CKD. The ISWT is a valid means of exercise capacity. The MDC can help researchers and rehabilitation professionals interpret changes following an intervention.

INTRODUCTION

Chronic kidney disease (CKD) is associated with adverse clinical outcome and reduced quality of life (Levey et al., 2005). CKD patients have reduced physical functioning (Hiraki et al., 2013; Kuo et al., 2015; Segura-Ortí, Gordon, Doyle, and Johansen, 2017), which is partly attributable to fatigue and muscle wasting characteristic of the condition (Wang and Mitch, 2014). Reductions in physical function begin early in the disease process (Hiraki et al., 2013), and are independently associated with mortality (Roshanravan et al., 2013). Consequently, physical function is an important target for research and rehabilitative intervention.

Physical function can be assessed using a range of field tests. The incremental shuttle walk test (ISWT) is a popular measure of exercise capacity, and has been used extensively in patient populations such as chronic obstructive pulmonary disease (COPD) (Singh, Jones, Evans, and Morgan, 2008), haemodialysis (HD) (Wilund et al., 2010), and non-dialysis CKD (Greenwood et al., 2012; Watson et al., 2015). Another well-established measure is the multi-variant 'sit-to-stand' (STS) test. The STS-5 repetition test is an assessment of lower body muscle strength, dynamic balance (Mong, Teo, and Ng, 2010), and exercise capacity (Jones et al., 2013), whilst the STS-60 second test measures lower body muscle endurance (Jones et al., 2013; Mong et al., 2010; Puthoff and Saskowski, 2013; Rikli and Jones, 2013; Segura-Ortí and Martínez-Olmos, 2011). These tests reflect a common activity of daily living (i.e. getting up from a chair), and are widely used in clinical and CKD research (Greenwood et al., 2012; McIntyre et al., 2006; Segura-Ortí et al., 2017; Segura-Ortí and Martínez-Olmos, 2011).

82   Reliability indicates the degree to which scores of a test are free of measurement errors.

83   Recognising the error inherent in outcome measures is imperative to the understanding of

84   changes and interpretation of research or rehabilitative interventions. Reliability can be

85   expressed in both relative and absolute terms. Relative (or test-retest) reliability can be

86   measured using intraclass correlation co-efficient (ICC). Absolute reliability refers to

87   individual performance variation and measurement error, and is quantified as standard error

88   of measurement (SE_M) (Ries, Echternach, Nof, and Blodgett, 2009; Stratford, 2004).

89

90   Using the SE_M, a more clinically useful means of interpreting reliability is the 'minimal

91   detectable change' (MDC) (Haley and Fragala-Pinkham, 2006), defined as the smallest

92   amount of reliable change in a measurement necessary to conclude that the difference is not

93   attributable to error (Segura-Ortí and Martínez-Olmos, 2011). Change exceeding the MDC is

94   considered 'true' change (Haley and Fragala-Pinkham, 2006). The MDC can be calculated at

95   an individual (MDC_{indv}) and group (MDC_{group}) level. The MDC_{indv} shows whether observed

96   changes in the individual's status are greater than variations of chance (Lee et al., 2013),

97   whereas the MDC_{group} is required to determine the relevance of changes across samples

98   (Busija et al., 2008; De Vet, Bouter, Bezemer, and Beurskens, 2001).

99

100   Alongside reliability, validity is also an important construct of physical performance tests.

101   Whilst the 'gold standard' measure of exercise capacity is cardiopulmonary exercise testing

102   (CPET) of VO_{2peak}, it is often impractical in rehabilitative settings, particularly in vulnerable

103   clinical patients. As such, the ISWT is frequently used as its surrogate measure (Holland et

104   al., 2014; Singh et al., 1994). Although the ISWT has been validated against CPET (via cycle

105   ergometer and treadmill modality) in other clinical populations (Arnardóttir et al., 2006;

106   Green et al., 2001; Holland et al., 2014; MacSween, Johnson, Armstrong, and Bonn, 2001;

107   Moloney et al., 2003), it has *not* yet been validated in CKD.

108

109   Lower body strength is important in CKD as the muscle of the legs are typically atrophied

110   (Wang and Mitch, 2014). Whilst the STS-5 may act as a surrogate measure of strength

111   (McCarthy, Horvat, Holtsberg, and Wisenbaker, 2004; Mong et al., 2010; Rikli and Jones,

112   2013), calculating 1 rep maximum (1RM) strength using resistance machines may be more

113   applicable in non-laboratory settings (i.e. a gymnasium) where dynamometry is not available

114   (Gail and Künzell, 2014).

115

116   No previous estimates of MDC for physical function tests exist for non-dialysis CKD

117   patients, and neither the ISWT nor STS-5 have not been validated in this group. Despite only

118   ~5% of CKD patients progressing to end-stage renal disease (and requiring renal replacement

119   therapy e.g., HD) (Dalrymple et al., 2011), research, particularly in regard to rehabilitation,

120   into non-dialysis patients often falls behind that of HD (Heiwe and Jacobson, 2014). With

121   reductions in physical function, an independent measure of mortality (Roshanravan et al.,

122   2013), evident in the early stages of CKD (Hiraki et al., 2013), it appears fundamental that

123   interventions designed to improve functional status (e.g., exercise (Heiwe and Jacobson,

124   2014)) are initiated promptly in the disease process and before it can progress. Furthermore,

125   with non-dialysis patients often experiencing differing functional capacities to other disease

126   populations (e.g., COPD), and indeed other CKD groups (e.g., those on HD (Hiraki et al.,

127   2013; Segura-Ortí et al., 2017)), the identification of the reliability and validity of physical

128   function tests specific to the non-dialysis population is vital in the correct interpretation of

129   functional changes.

130

131   The aims of the current study were to: 1) determine the test-retest reliability and estimated

132   MDC of the common physical function tests in patients with non-dialysis CKD; and 2)

133   confirm the validity of the ISWT as a measure of exercise capacity (VO_{2peak}), and the STS as

134   a measure of lower body strength.

All assessments took place between December 2013 and June 2016 at Leicester General

Hospital, Leicester, UK. Patients were recruited from nephrology outpatient clinics at the

University Hospitals of Leicester NHS Trust. Patients gave written informed consent in

accordance with the Declaration of Helsinki and local Research Ethics Committee approval

was obtained. This analysis forms part of a larger body of work completed by our group

(ISRCTN registration 36489137). To ensure accurate reporting of study measurement

properties and analysis parameters, the 'COnsensus-based Standards for the selection of

health Measurement Instruments' (COSMIN) checklist was adhered to (Mokkink et al.,

2010).

Participants

Patients were recruited if they had: CKD stages 3b-5 (i.e. an estimated glomerular filtration

rate (eGFR) of $\leq$44 ml/min/1.73^2) *not* requiring renal replacement therapy (e.g., HD); were

aged $\geq$18; no significant co-morbidity (e.g., unstable hypertension, potentially lethal

arrhythmia, myocardial infarction within the previous six months) contraindicative to

physical exercise; no significant physical impairment; and sufficient ability to give informed

consent. Prior exercise and physical activity level was not a pre-requisite for inclusion.

The severity of comorbidity was recorded and scored according to Charlson Comorbidity

Index (CCI). A higher CCI score indicated greater comorbidity, and CCI scores of 1–2 were

classified as mild; scores of 3–4 as moderate; and scores $\geq$ 5 as severe (Huang et al., 2014).

Physical function assessments

160   Assessments were performed at two time points (test 1 and 2) separated by a six week period

161   in which patients were instructed to maintain their habitual lifestyle. The six week interval

162   forms part of a control period in the main trial, thus we used pre- and post-data as an

163   opportunity to assess the reliability of the tests employed. All researchers performing the tests

164   followed strict operating procedures for each of the tests performed to reduce investigator

165   bias. As different researchers were used, we do not present *intra*-rater reliability, but *inter*-

166   rater reliability. Each test was performed in the same order (STS tests, ISWT, and e1RM). As

167   CPET was used to screen patients for cardio-pulmonary contraindications, and due to the

168   relative exhaustive aspect of the test, CPET was performed on a separate visit several days

169   before. All tests, apart from CPET due to logistic and cost factors, had a familiarisation test

170   several days prior to their first test to minimise learning effect (Holland et al., 2014). The

171   familiarisation testing session involved an explanation of the test to the patient, short

172   demonstration of the basic movement(s) by the researcher present, a practice repetition (if

173   appropriate), and then the full test. On the day of testing, following a re-explanation of the

174   test, patients performed each test once.

175

176   <u>ISWT</u>

177   The ISWT is maximally progressive test that involves walking at a pace externally dictated

178   by an auditory tone (Holland et al., 2014; Singh et al., 1994). During the test, the patient

179   walked a total of 10m back and forth and around two cones. The walking pace was increased

180   by a rate of 0.17m/sec every minute for twelve stages until the patient could no longer keep

181   up with the pace because of breathlessness, pain, or other symptoms. Only completed shuttles

182   were counted. The outcome was distance walked (m).

183

184 The ISWT was preferred over other measures of exercise capacity, such as the 6-minute walk

185 test (6MWT), as this test requires at least 30m of walking space (Holland et al., 2014);

186 provisions not available at our research facility. However, along with reducing limitations

187 associated with being self-paced, the ISWT may also provide superiority when prescribing

188 exercise intensity as a % of peak performance (Holland et al., 2014), and thus may be more

189 beneficial in a renal rehabilitation setting when tailoring individualised exercise

190 interventions.

191

192 STS tests

193 The STS-5 and -60 tests were employed as measures of lower body strength and muscle

194 endurance (Mong et al., 2010; Rikli and Jones, 2013; Segura-Ortí and Martínez-Olmos,

195 2011). The patient sat on a seat (43.2 cm from the ground). With their hands across their

196 chest, patients were asked to: 1) perform five complete STS cycles as fast as possible (STS-

197 5); and 2) perform as many complete STS cycles in 60 secs (STS-60). The STS-5 time was

198 stopped when the patient was seated following their fifth repetition. If the patient was half-

199 way through a stand when STS-60 time had expired, this was counted as one repetition (Rikli

200 and Jones, 2013). The STS-5 test was preferred over other STS versions (e.g., STS-10) as its

201 short duration reduces patient burden during testing sessions involving multiple other

202 outcomes (Nilsagård, Andreasson, Carling, and Vesterlin, 2017).

203

204 Lower limb strength

205 The maximal strength (kg) of the quadriceps muscle was measured using a leg extension

206 machine (TechnoGym, Italy). Performing a true 1RM test is associated with an increased

207 injury risk and stress on the muscles and joints, particularly in untrained (Gail and Künzell,

208 2014) and clinical groups (Abdul-Hameed, Rangra, Shareef, and Hussain, 2012), therefore

209 we estimated 1RM (e1RM) for the leg extension exercise from a 5-rep maximum (5RM)

210 (Brzycki, 1993; Dudgeon et al., 2010). During the test, weight was progressively increased by

211 a minimum of 2.5 kg depending on participant feedback and ease of the previous 5

212 repetitions. To reduce cumulative fatigue and to ensure an accurate 5RM, patients were given

213 a minimum 60 second rest in-between each attempt, although researcher discretion was also

214 used. The 5RM was determined as the maximal weight the patient could lift five times with

215 correct technique.

216

217 CPET

218 Peak exercise capacity (VO_{2peak}) was assessed using CPET. Patients were asked to cycle for

219 as long as possible at a revolutions per minute (RPM) $\geq$60. Following a 3 minute warm up,

220 the resistance on the static ergometer (Lode Excalibur, Netherlands) increased from 30 Watts

221 by 1 Watt every 3 secs in a ramp protocol. Throughout the test, an echocardiogram was

222 performed and reviewed by an experienced exercise cardiac nurse or doctor. The test was

223 stopped if: RPM <60; the patient reached volitional exhaustion; or at the discretion of the

224 medical professional. Using online direct breath-by-breath measurement (Cortex Metalyzer,

225 Cranlea, UK) of oxygen consumption (VO_2), we calculated relative VO_{2peak} (peak

226 ml/kg/min).

227

228 **Statistical methods**

229 Test-retest relative reliability of data was assessed using the ICC (*r*). An ICC between .600-

230 .749 is considered 'fair', $\geq$.750 'good', whilst a value $\geq$.900 is considered 'excellent' for

231 clinical measures (Cicchetti, 1994). Data is also represented graphically as Bland-Altman

232 plots with mean bias and limits of agreement set at 95% confidence intervals (95CI) (Bland

11

233 and Altman, 1999). Here the difference of the two paired measurements is plotted against the

234 mean of the two measurements.

235

236 The SE_M and MDC were calculated as a measure of absolute reliability (Haley and Fragala-

237 Pinkham, 2006; Stratford, 2004). The SE_M was calculated as:

238

239 $SE_M = SD * \sqrt{(1-r)}$.

240

241 This method of calculating SE_M has been used previously (Chiu et al., 2016; Ries et al., 2009;

242 Segura-Ortí and Martínez-Olmos, 2011). The MDC at an individual level (MDC_{indv}) was

243 calculated at the 95CI. The equation used was:

244

245 $MDC_{indv} = SE_M * 1.96 * \sqrt{2}$.

246

247 The 1.96 represents the z-score at the 95CI. The 'SE_M * 1.96' is multiplied by the $\sqrt{2}$ to

248 account for errors associated with repeated measures (De Vet et al., 2001; Haley and Fragala-

249 Pinkham, 2006; Segura-Ortí and Martínez-Olmos, 2011; Stratford, 2004). The MDC at group

250 level (Busija et al., 2008; De Vet et al., 2001) (MDC_{group}) was calculated as:

251

252 $MDC_{group} = MDC_{indv} / \sqrt{n}$.

253

254 Construct validity between the ISWT and CPET derived VO_{2peak}, and between the STS-5 and

255 e1RM was assessed using simple linear regression and data are represented as scatterplot

256 graphs with a trend line showing r and 95CI interval bands.

257

258   A minimum sample size of 39 patients was needed to estimate an ICC r of .600 (the minimal

259   acceptable ICC in clinical investigations (Shoukri, Asyali, and Donner, 2004)) with a β of

260   0.80 at a significance level of $P = 0.050$ (Walter, Eliasziw, and Donner, 1998). Data was

261   assessed using SPSS v24. Data are reported as mean (SD), if normally distributed, or as

262   median (interquartile range, IQR). Distribution was assessed using the Kolmogorov-Smirnov

263   test. Paired comparisons were tested using paired t-tests, or non-parametric Wilcoxon signed

264   rank test as appropriate. Difference (with 95CI) is reported as mean (if data were normally

265   distributed), or median (if data were non-normally distributed). Data for some individuals

266   were not collected for different assessments due to missed measures or an inability to

267   complete (see footnote of **Table 2**); consequently, with no comparable data, this patient was

268   excluded (listwise approach) from analysis in that test.

269

RESULTS

Forty-one patients (23 females, 56%) with non-dialysis CKD were recruited. Mean patient age was 62 ($SD:$ 11) years old with a body mass index of 30.1 ($SD:$ 5.7). The majority of patients were Caucasian (66%). The mean eGFR was 25 ($SD:$ 8) ml/min/1.73^2. Over half of patients (63%) had previously diagnosed hypertension and 22% of patients had diabetes mellitus type II. The mean CCI score was 2.5 ($SD:$ 0.7), with the majority of patients (59%) classified as having a mild CCI score. Full patient clinical and demographic characteristics are shown in **Table 1.** Apart from anticipated exercise-induced fatigue during the tests, no adverse events of complaints were recorded.

Test-retest relative and absolute reliability

There was minimal difference between the two tests for the ISWT, STS-60, e1RM, and VO$_{2peak}$ by CPET with ICC *r* values of .973, .927, .927, (all rated as 'excellent') and .866 (rated 'good'), respectively. Conversely, the ICC for the STS-5 displayed only 'fair' agreement (*r* = .676). Data analysis revealed no statistically significant differences between the two tests for each measure of function **(Table 2)**. **Figure 1A–E** shows Bland-Altman plots with bias and limits of agreement (at 95CI) for each of these tests. Absolute reliability data (both SE$_M$ and MDC) are shown in **Table 3**.

The proportion of patients whom performed best in test 1 was as follows: ISWT (n = 16, 39%), STS-5 (n = 18, 44%), STS-60 (n = 14, 35%), e1RM (n = 9, 23%), and VO$_{2peak}$ (n = 20, 54%). No change between test 1 and 2 was observed in n = 3 (7%) for the ISWT; n = 2 (5%) for the STS-60; n = 9 (23%) for the e1RM; and n = 1 (3%) for VO$_{2peak}$ by CPET.

Validation of the ISWT and STS-5

We found a 'strong' correlation between the ISWT and VO_{2peak} test with an r of 0.73 at time

point 1 ($P < 0.001$), and an r of 0.74 at time point 2 ($P < 0.001$) (**Figure 2A–B**). The

correlation between the STS-5 and the e1RM test was 'poor' at time point 1 ($r = 0.14$, $P =$

0.371), and (albeit better) 'weak' at time point 2 ($r = 0.47$, $P = 0.003$) (**Figure 2C–D**). When

we compared the STS-5 with ISWT, we found 'moderate' correlation at time point 1 ($r =$

0.55, $P < 0.001$), and 'strong' correlation at time point 2 ($r = 0.74$, $P < 0.001$) (**Figure 2E–F**).

DISCUSSION

This study is the first investigation into the test-retest reliability and identification of the

MDC for tests of physical function in a non-dialysis CKD population. Our results have large

clinical relevance in CKD research and rehabilitation, and demonstrate that the ISWT, STS-

60, e1RM, and VO_{2peak} via CPET have 'good' to 'excellent' reliability in non-dialysis CKD.

Conversely, the STS-5 test performed poorly with only 'fair' agreement. Using SEM data, we

were able to calculate the MDC at an individual and group level. We also confirmed the

ISWT as a valid measure of VO_{2peak}, but found the STS-5 only weakly associated with lower

limb strength.

The ISWT in our trial showed excellent test-retest reliability with an ICC of .973. This

corresponds well with previous estimates in cardiac rehabilitation patients (.990) (Hanson,

Taylor, and McBurney, 2015), patients with peripheral vascular disease (.990) (Cunha-Filho

et al., 2008), and non-cystic fibrosis (.950) (Lee et al., 2015). The ISWT is used extensively

in research and rehabilitation as a measure of exercise capacity, and as stated previously, was

chosen above the 6MWT, another common field test of exercise capacity, for several reasons

including space considerations and the limitations associated with self-paced tests.

Furthermore, given its progressive nature, the ISWT is superior when prescribing exercise

intensity (e.g., % of maximum) (Holland et al. 2014), and thus potentially more useful in a

renal rehabilitation setting. Studies validating the ISWT against CPET-measured VO_{2peak}

have found good agreement ($r = 0.74$ to 0.88) regardless of exercise modality (i.e. treadmill

versus cycle ergometer) (Arnardóttir et al., 2006; Green et al., 2001; Holland et al., 2014;

MacSween et al., 2001; Moloney et al., 2003). Our r values of 0.73 and 0.74 represent a

327 strong agreement, and confirm the ISWT as a valid and easy means to assess exercise

328 capacity in a CKD population.

329

330 The STS-60 had excellent reliability with an ICC of .927. Similarly, Segura-Ortí and

331 Martínez-Olmos (2011) found an ICC of .970 in HD patients; however, other clinical

332 research into the reliability of the STS-60 is limited. Conversely, the STS-5 displayed

333 relatively poor reliability with an ICC of .676 ('fair'). Whilst this contrasts with previous

334 estimates in clinical populations (ICC values between .940 and .990), (Jones et al., 2013;

335 Mong et al., 2010; Thomas and Hageman, 2002), similar reliability (ICC of .640) was

336 reported by Ostchega et al (Ostchega et al., 2000) in older adults. Unfortunately, the authors

337 in this trial offer no explanation for their moderate reliability score. The lack of reliability in

338 our study may be due to the large within-patient variability seen in our sample. Indeed,

339 several patients took considerably longer on their first attempt than on their second: one

340 patient performed 25.8 seconds quicker in the STS-5 at visit 2. When we removed this

341 patient, the ICC for the STS-5 increased to .752 (.577 - .860), and although defined as now

342 'good', this still lacked the reliability observed relative to our other tests. As the aim of this

343 study was to assess natural variation, and with no clinical justification to remove this patient,

344 we retained this data point in the analysis

345

346 As recommended by Nilsagård et al. (2017), the STS-5 was preferred over the STS-10,

347 another commonly reported STS variant, to: 1) ease patient burden (i.e. STS-5 takes ~50%

348 less time); and 2) to isolate the assessment of strength. Given the requirement to complete

349 double the repetitions, the STS-10 has been described as a measure of muscular endurance

350 (Nilsagård et al. 2017), whereas the STS-5 is purported as a measure of lower limb strength.

351 Therefore, to avoid multiple assessments of muscular endurance we used the STS-5.

352

Nonetheless, we observed a weak relationship between the STS-5 and lower limb strength

(by e1RM), suggesting that, in our patients, the STS-5 may not be a comprehensive proxy

measure of strength. Whilst this supports previous work which found the STS-5 had only a

modest relationships with lower limb strength in COPD (Jones et al., 2013; Roig et al., 2011),

it conflicts with research that suggests in both elderly and stroke patients, the STS-5 is a good

marker of lower body strength (Mong et al., 2010). Interestingly, although we were unable to

demonstrate any strong relationship between STS-5 and lower limb muscle strength, we did

observe a strong to moderate association between the STS-5 and the ISWT. Similar

observations have been reported by Jones et al (2013) in COPD patients. In CKD clinical

practice and research, other methods to measure lower limb strength should be employed

over the STS-5.

364

Limited research exists on the reliability of the e1RM estimated by a 5RM test. Our ICC

value of .927 compares well with previous estimates in healthy recreational athletes (.900)

(Gail and Künzell, 2014), and in untrained diabetic patients (.990) (Abdul-Hameed et al.,

2012). Although dynamometry is considered the 'gold standard' for the assessment of

strength, e1RM testing is commonly applied for ease and simplicity, and represents a valid

means to assess leg muscle strength (Verdijk, Van Loon, Meijer, and Savelberg, 2009). We

experienced no serious difficulties performing a 5RM test, and therefore in gymnasium-based

rehabilitative setting, where dynamometers are unavailable, a 5RM test may be a safe and

reliable tool in strength measurement in CKD.

374

As expected, the 'gold standard' of assessing exercise capacity (VO_{2peak}) by CPET showed

good (bordering on 'excellent') test-retest reliability. Our ICC of .866 compares well with

377 previous testing in clinical groups, including patients with multiple sclerosis (.933) (Heine et

378 al., 2015) and coronary heart disease (.970) (Coeckelberghs et al., 2016).

379

380 The MDC is the smallest change that falls outside the expected range of error (Segura-Ortí

381 and Martínez-Olmos, 2011), and is invaluable to know as any change from a clinical

382 intervention exceeding the MDC can be considered a 'true' change (Haley and Fragala-

383 Pinkham, 2006). We were able to calculate the MDC at both an individual and group level.

384 An MDC at individual level is should be used to determine whether a single patient has made

385 an improvement, not attributable to error, following an intervention (Haley and Fragala-

386 Pinkham, 2006; Lee et al., 2013). This may be useful to physical therapists, or other allied

387 professionals, whom work with, and interpret changes of, individual patients. Conversely, as

388 the majority of clinical trials compare changes between groups of patients, the MDC_{indv}

389 should be adapted to quantify the MDC of an outcome measure in a group of patients (De Vet

390 et al., 2001). Often the MDC_{group} is smaller than the corresponding MDC_{indv} implying

391 superiority at detecting change at group level (De Vet et al., 2001).

392

393 At an individual level, the MDC for the ISWT was 20m. The small MDC_{group} (3m) found

394 suggests that, at group level, small changes in ISWT scores can be reliably identified. To our

395 knowledge, this is the *first* estimate of the MDC for this test in *any* clinical population.

396

397 For the STS-60, we found an MDC_{indv} of 4 reps. Similar findings have been reported in HD

398 patients (4 reps) (Segura-Ortí and Martínez-Olmos, 2011). This is perhaps due to the similar

399 mean number of repetitions (26) between their HD patients and our non-dialysis CKD

400 sample. Whilst such comparable functional capacity between HD and non-dialysis CKD

401 patients may be unexpected, the majority of patients in the Segura-Ortí and Martínez-Olmos

402   trial were almost exclusively male (82%). Male CKD patients (including those on HD)

403   display superior functional capacity than females (Hiraki et al., 2013); as our sample was

404   majority female (56%), this may explain the comparable results observed. At a group level,

405   the high reliability and low measurement error of the STS-60 means the MDC_{group} is just 1

406   rep.

407

408   Whilst an MDC of 3.1 secs has been reported for the STS-5 in patients undergoing cardiac

409   rehabilitation (Puthoff and Saskowski, 2013) and 5.0 secs in those with stroke (Pardo et al.,

410   2013), our data revealed an MDC_{indv} of 7.5 secs and an MDC_{group} of 1.2 secs. Our larger

411   MDC values could be due to the poor reliability for this test as discussed previously.

412

413   Whilst no MDC exists for 1RM testing, using dynamometry, in patients with knee

414   osteoarthritis, the MDC using peak isokinetic knee extension torque was estimated at 3.5 kg

415   (Kean et al., 2010). This value is smaller than our estimated MDC_{indv} of 6.4 kg. A possible

416   explanation for the larger value in our trial may be due to equipment limitations. The machine

417   used only increased in increments of 2.5 kg and above, thus limiting the ability to increase

418   weight in smaller increments. Whilst the e1RM has been shown to represent a valid measure

419   of strength assessed by dynamometer (Verdijk et al., 2009), further research is needed to

420   assess the validity in a CKD group.

421

422   The MDC_{indv} for VO_{2peak} was 2.8 ml/kg/min, and 0.5 ml/kg/min at group level. Brehm et al.

423   (Brehm, Balemans, Becher, and Dallmeijer, 2014) previously estimated an MDC_{indv} of 5.7

424   ml/kg/min in patients with cerebral palsy, whilst in patients with multiple sclerosis, Heine et

425   al. (Heine et al., 2015) calculated the MDC_{indv} at 4.6 ml/kg/min and the MDC_{group} at 0.8

426   ml/kg/min. Like the ISWT, the small MDC_{group} identified suggests that small changes in

427 VO_{2peak} can be identified. This is perhaps unsurprising given the highly technical analysis of

428 CPET.

429

430 We have confirmed the ISWT as a valid assessment of exercise capacity, although

431 disappointingly the STS-5 was not associated with lower limb strength. The low *r* at time

432 point 1 between the STS-5 and e1RM may be attributed to the outlier situated to the far right

433 of the Bland-Altman plot. However, as stated above, with no clinical justification to remove

434 this patient, we retained this data point in the analysis. Nonetheless, even at time point 2, the

435 *r* value (0.47) was still poor. The STS test as a proxy measure of lower limb strength (Mong

436 et al., 2010; Rikli and Jones, 2013) is increasingly being questioned by researchers

437 (McCarthy et al., 2004), and study has shown that only a moderate proportion of STS

438 performance is attributable to strength, with factors such as balance and sensorimotor ability

439 also contributing (McCarthy et al., 2004). Indeed, our data, and that of others (Jones et al.,

440 2013), suggests that the STS-5 may actually be a better measure of exercise capacity than

441 strength. As such, additional physiological factors may be contributing, and further research

442 is required into the identifying these elements and assessing the validity of the STS-5.

443

444 **Study limitations**

445 Our patient population was an opportunistic sample taken from a larger body of work

446 completed by our group. Nonetheless, the 41 patients included in the current analysis was

447 above the minimum sample size of 39 patients required to estimate an ICC r of .600 (the

448 minimal acceptable ICC in clinical investigations). Consequently, our analysis was

449 sufficiently powered to detect any differences between time points.

450

451 Another limitation in our trial was the six week test-retest interval. However, as our data is

452 derived from the control period the larger trial, this period was unavoidable. Nevertheless,

453 our trial was suitably powered and our data shows even over a six week period, the reliability

454 of our function tests was good. Furthermore, previous test-retest estimates of the reliability of

455 objective function such as the STS (Schaubert and Bohannon, 2005a; Schaubert and

456 Bohannon, 2005b) and strength (Schaubert and Bohannon, 2005b) over a similar six week

457 interval have also established good agreement.

458

459 In order to ensure accurate baseline data and to reduce learning effects (Holland et al., 2014;

460 Johnson-Warrington, Sewell, Morgan, and Singh, 2015), patients completing the ISWT at

461 visit 1 did so after a familiarisation test several days before. However, due to time constraints

462 at the second visit, it was feasible that the ISWT, and other outcomes, were repeated only

463 once. Whilst we observed good reliability between the two visits for the ISWT without re-

464 familiarisation at visit 2, it is unclear from our data if lack of re-familiarisation has effected

465 outcome measure scores at the second visit and whether one test is sufficient if the test has

466 previously been conducted is still unclear (Holland et al., 2014). In intervention trials,

467 repeating, or re-familiarising patients with physical function tests may prevent

468 underestimation of the impact of an intervention (Holland et al., 2014). Interestingly

469 however, due to impracticality and costing, patients did not undergo familiarisation CPET at

470 baseline; however, our results show even without this familiarisation test, reliability for

471 CPET is high, perhaps due to the technical accuracy of the method, and a familiarisation is

472 perhaps unnecessary for this test.

473

474 In order to further improve the methodology of our study, further research should investigate

475 the identification of the MDC in earlier stages of CKD (1 and 2), as well as potentially

476   identifying differences in gender or ethnicity. In our sample, a large proportion (46%) of our

477   patients were obese (defined by BMI), with 34% 'overweight'. Whilst more current obesity

478   prevalence statistics are needed to fully determine the demographic of our sample, the

479   external validity of the results may be reduced. Dynamometry, over 1- 5RM testing, should

480   be used to accurately measure strength; as such, associations with the STS-5 should be

481   reassessed. Additional physiological processes that contribute to function, such as balance,

482   should also be considered in future trials assessing performance. The minimally clinical

483   important difference, which is the minimal meaningful important difference from a patient

484   perspective, should also be investigated.

485

CONCLUSIONS

In conclusion, the ISWT, STS-60, e1RM, and CPET are reliable and valid tests of function in non-dialysis CKD patients. The MDC's we have calculated in this trial are imperative in the interpretation of changes as a result of intervention trials or rehabilitation programmes. In order for a 'true' change to occur, i.e. that above individual performance variability or inherent measurement error, the value should exceed the MDC's presented here. The MDC at an individual level is useful in evaluating changes in a single patient, whilst the MDC at group level should be employed in research comparing changes between groups. The MDC should be used to help clinicians, researchers, and rehabilitation professionals interpret reliable changes in their patients and clinical trials.

REFERENCES

Abdul-Hameed U, Rangra P, Shareef MY, Hussain ME 2012 Reliability of 1-repetition maximum estimation for upper and lower body muscular strength measurement in untrained middle aged type 2 diabetic patients. Asian Journal of Sports Medicine 3: 267-73. No DOI available.

Arnardóttir RH, Emtner M, Hedenström H, Larsson K, Boman G 2006 Peak exercise capacity estimated from incremental shuttle walking test in patients with COPD: a methodological study. Respiratory Research 7:1. DOI: 10.1186/1465-9921-7-127

Bland JM, Altman DG 1999 Measuring agreement in method comparison studies. Statistical Methods in Medical Research 8: 135-60. No DOI available.

Brehm M-A, Balemans AC, Becher JG, Dallmeijer AJ 2014 Reliability of a progressive maximal cycle ergometer test to assess peak oxygen uptake in children with mild to moderate cerebral palsy. Physical Therapy 94: 121-8. DOI: 10.2522/ptj.20130197

Brzycki M 1993 Strength testing—predicting a one-rep max from reps-to-fatigue. Journal of Physical Education, Recreation & Dance 64: 88-90. No DOI available.

Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM 2008 Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. Health and Quality of Life Outcomes 6: 1. DOI: 10.1186/1477-7525-6-55

533 Chiu E-C, Wu W-C, Chou C-X, Yu M-Y, Hung J-W 2016 Test-retest reliability and minimal

534 detectable change of the Test of Visual Perceptual Skills-in patients with stroke. Archives of

535 Physical Medicine and Rehabilitation. DOI: 10.1016/j.apmr.2016.04.023

536

537 Cicchetti DV 1994 Guidelines, criteria, and rules of thumb for evaluating normed and

538 standardized assessment instruments in psychology. Psychological Assessment 6: 284. No

539 DOI available.

540

541 Coeckelberghs E, Buys R, Goetschalckx K, Pattyn N, Vanhees L, Cornelissen V 2016 Test-

542 Retest Reliability of Maximal and Submaximal Gas Exchange Variables in Patients With

543 Coronary Artery Disease. Journal of Cardiopulmonary Rehabilitation and Prevention. DOI:

544 10.1097/HCR.0000000000000158

545

546 Cunha-Filho ITd, Pereira DAG, Carvalho AMBd, Campedeli L, Soares M, Freitas JdS 2008

547 Reliability of walking tests in claudicating patients: a pilot study. Jornal Vascular Brasileiro

548 7: 106-11. No DOI available.

549

550 Dalrymple, LS, Katz R, Kestenbaum B, Shlipak M, Sarnak MJ, Stehman-Breen C, Seliger S,

551 Siscovick D, Newman AB, Fried L 2011 Chronic kidney disease and the risk of end-stage

552 renal disease versus death. Journal of General Internal Medicine 26: 379-85. DOI:

553 10.1007/s11606-010-1511-x.

554

555 de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Beurskens 2001 Reproducibility and

556 responsiveness of evaluative outcome measures. International Journal of Technology

557 Assessment in Health Care 17: 479-87. No DOI available.

558

Dudgeon WD, Phillips KD, Durstine JL, Burgess SE, Lyerly GW, Davis JM, Hand GA 2010

Individual exercise sessions alter circulating hormones and cytokines in HIV-infected men.

Applied Physiology, Nutrition, and Metabolism 35: 560-8. DOI: 10.1139/H10-045

562

Gail S, Künzell S 2014 Reliability of a 5-repetition maximum strength test in recreational

athletes. Deutsche Zeitschrift für Sportmedizin 65: 314-7. No DOI available.

565

Green D, Watts K, Rankin S, Wong P, O'Driscoll J 2001 A comparison of the shuttle and 6

minute walking tests with measured peak oxygen consumption in patients with heart failure.

Journal of Science and Medicine in Sport 4: 292-300. No DOI available.

569

 Greenwood SA, Lindup H, Taylor K, Koufaki P, Rush R, Macdougall IC, Mercer TH 2012

Evaluation of a pragmatic exercise rehabilitation programme in chronic kidney disease.

Nephrology Dialysis Transplantation 27: iii126-iii34. DOI: 10.1093/ndt/gfs272

573

Haley SM, Fragala-Pinkham MA 2006 Interpreting change scores of tests and measures used

in physical therapy. Physical Therapy 86: 735-43. No DOI available.

576

Hanson LC, Taylor NF, McBurney H 2015 The 10m incremental shuttle walk test is a highly

reliable field exercise test for patients referred to cardiac rehabilitation: a retest reliability

study. Physiotherapy. DOI: 10.1016/j.physio.2015.08.004

580

Heine M, van den Akker LE, Verschuren O, Visser-Meily A, Kwakkel G, Group T-AS 2015

Reliability and responsiveness of cardiopulmonary exercise testing in fatigued persons with

583 multiple sclerosis and low to mild disability. PloS One 10: e0122260. DOI:

584 10.1371/journal.pone.0122260

585

586 Heiwe S, Jacobson SH 2014 Exercise training in adults with CKD: a systematic review and

587 meta-analysis. American Journal of Kidney Diseases 64: 383-93. DOI:

588 10.1053/j.ajkd.2014.03.020

589

590 Hiraki K, Yasuda T, Hotta C, Izawa KP, Morio Y, Watanabe S, Sakurada T, Shibagaki Y,

591 Kimura K 2013 Decreased physical function in pre-dialysis patients with chronic kidney

592 disease. Clinical and Experimental Nephrology 17: 225-231. DOI: 10.1007/s10157-012-

593 0681-8

594

595 Holland AE, Spruit MA, Troosters T, Puhan MA, Pepin V, Saey D, McCormack MC, Carlin

596 BW, Sciurba FC, Pitta F, Wanger J, MacIntyre N, Kaminsky DA, Culver BH, Revill SM,

597 Hernandes NA, Andrianopoulos V, Camillo CA, Mitchell KE, Lee AL, Hill CJ, Singh SJ.

598 2014 An official European Respiratory Society/American Thoracic Society technical

599 standard: field walking tests in chronic respiratory disease. European Respiratory Society 44:

600 1428-46. DOI: 10.1183/09031936.00150314

601

602 Huang YQ, Gou R, Diao YS, Yin QH, Fan WX, Liang YP, Chen Y, Wu M, Zang L, Li L,

603 Zang J, Cheng L, Fu P, Liu F 2014 Charlson comorbidity index helps predict the risk of

604 mortality for patients with type 2 diabetic nephropathy. Journal of Zhejiang University

605 Science B 15: 58-66. DOI: 10.1631/jzus.B1300109

606

607   Johnson-Warrington V, Sewell L, Morgan M, Singh S 2015 Do we need a practice

608   incremental shuttle walk test for patients with interstitial lung disease referred for pulmonary

609   rehabilitation? Respirology 20: 434-8. DOI: 10.1111/resp.12469

610

611   Jones SE, Kon SS, Canavan JL, Patel MS, Clark AL, Nolan CM, Polkey MI, Man WDC

612   2013 The five-repetition sit-to-stand test as a functional outcome measure in COPD. Thorax

613   68: 1015-20. DOI: 10.1136/thoraxjnl-2013-203576

614

615   Kean CO, Birmingham TB, Garland SJ, Bryant DM, Giffin JR 2010 Minimal detectable

616   change in quadriceps strength and voluntary muscle activation in patients with knee

617   osteoarthritis. Archives of Physical Medicine and Rehabilitation 91: 1447-51. DOI:

618   10.1016/j.apmr.2010.06.002

619

620   Kuo Y-T, Chiu K-M, Tsang Y-M, Chiu C-M, Chien M-Y 2015 Influence of Chronic Kidney

621   Disease on Physical Function and Quality of Life in Patients after Coronary Artery Bypass

622   Grafting. Cardiorenal Medicine 5: 237-45. DOI: 10.1159/000433447

623

624   Lee AL, Cecins N, Holland AE, Hill CJ, McDonald CF, Burge AT, Rautela L, Thompson PJ,

625   Stirling RG, Jenkins S 2015 Field Walking Tests Are Reliable and Responsive to Exercise

626   Training in People With Non–Cystic Fibrosis Bronchiectasis. Journal of Cardiopulmonary

627   Rehabilitation and Prevention 35: 439-45. DOI: 10.1097/HCR.0000000000000130

628

629   Lee P, Liu CH, Fan CW, Lu CP, Lu WS, Hsieh CL 2013 The test–retest reliability and the

630   minimal detectable change of the Purdue Pegboard Test in schizophrenia. Journal of the

631   Formosan Medical Association 112: 332-37. DOI: 10.1016/j.jfma.2012.02.023

632

633   Levey AS, Eckardt K-U, Tsukamoto Y, Levin A, Coresh J, Rossert J, Zeeuw DD, Hostetter

634   TH, Lameire N, Eknoyan G 2005 Definition and classification of chronic kidney disease: a

635   position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney

636   International 67: 2089-100. DOI: 10.1038/sj.ki.5002343

637

638   MacSween A, Johnson NJ, Armstrong G, Bonn J 2001 A validation of the 10-meter

639   incremental shuttle walk test as a measure of aerobic power in cardiac and rheumatoid

640   arthritis patients. Archives of Physical Medicine and Rehabilitation 82: 807-10. DOI:

641   10.1053/apmr.2001.23185

642

643   McCarthy EK, Horvat MA, Holtsberg PA, Wisenbaker JM 2004 Repeated chair stands as a

644   measure of lower limb strength in sexagenarian women. The Journals of Gerontology Series

645   A: Biological Sciences and Medical Sciences 59: 1207-12. No DOI available.

646

647   McIntyre CW, Selby NM, Sigrist M, Pearce LE, Mercer TH, Naish PF 2006 Patients

648   receiving maintenance dialysis have more severe functionally significant skeletal muscle

649   wasting than patients with dialysis-independent chronic kidney disease. Nephrology Dialysis

650   Transplantation 21: 2210-6. DOI: 10.1093/ndt/gfl064

651

652   Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet

653   HCW 2010 The COSMIN checklist for evaluating the methodological quality of studies on

654   measurement properties: A clarification of its content. BMC Medical Research Methodology

655   6: 2. DOI: 10.1186/1471-2288-6-2

656

657 Moloney E, Clayton N, Mukherjee D, Gallagher C, Egan J 2003 The shuttle walk exercise

658 test in idiopathic pulmonary fibrosis. Respiratory Medicine 97: 682-7. No DOI available.

659

660 Mong Y, Teo TW, Ng SS 2010 5-repetition sit-to-stand test in subjects with chronic stroke:

661 reliability and validity. Archives of Physical Medicine and Rehabilitation 91: 407-13. DOI:

662 10.1016/j.apmr.2009.10.030

663

664 Nilsagård Y, Andreasson M, Carling A, Vesterlin H 2017 Examining the validity and

665 sensitivity to change of the 5 and 10 sit-to-stand tests in people with multiple sclerosis.

666 Physiotherapy Research International. [Epub ahead of print]. DOI: 10.1002/pri.1681

667

668 Ostchega Y, Harris TB, Hirsch R, Parsons VL, Kington R, Katzoff M 2000 Reliability and

669 prevalence of physical performance examination assessing mobility and balance in older

670 persons in the US: data from the Third National Health and Nutrition Examination Survey.

671 Journal of the American Geriatrics Society 48: 1136-41. No DOI available.

672

673 Pardo V, Knuth D, McDermott B, Powell J, Goldberg A 2013 Validity, reliability and

674 minimum detectable change of the maximum step length test in people with stroke. Journal of

675 the Neurological Sciences 325: 74-8. DOI: 10.1016/j.jns.2012.11.021

676

677 Puthoff ML, Saskowski D 2013 Reliability and responsiveness of gait speed, five times sit to

678 stand, and hand grip strength for patients in cardiac rehabilitation. Cardiopulmonary Physical

679 Therapy Journal 24: 31. No DOI available.

680

681   Ries JD, Echternach JL, Nof L, Blodgett MG 2009 Test-retest reliability and minimal

682   detectable change scores for the timed "up & go" test, the six-minute walk test, and gait

683   speed in people with Alzheimer disease. Physical Therapy 89: 569-79. DOI:

684   10.2522/ptj.20080258

685

686   Rikli RE, Jones CJ 2013 Development and validation of criterion-referenced clinically

687   relevant fitness standards for maintaining physical independence in later years. The

688   Gerontologist 53: 255-67. DOI: 10.1093/geront/gns071

689

690   Roshanravan B, Robinson-Cohen C, Patel KV, Ayers E, Littman AJ, De Boer IH, Ikizler TA,

691   Himmelfarb J, Katzel LI, Kestenbaum B, Seliger S 2013 Association between physical

692   performance and all-cause mortality in CKD. Journal of the American Society of Nephrology

693   2012070702. DOI: 10.1681/ASN.2012070702.

694

695   Schaubert K, Bohannon RW 2005a Reliability of the sit-to-stand test over dispersed test

696   sessions. Isokinetics and Exercise Science 13: 119-22. No DOI available.

697

698   Schaubert KL, Bohannon RW 2005b Reliability and validity of three strength measures

699   obtained from community-dwelling elderly persons. The Journal of Strength & Conditioning

700   Research 19: 717-20. DOI: 10.1519/R-15954.1

701

702   Segura-Ortí E, PL Gordon, JW Doyle, KL Johansen 2017 Correlates of Physical Functioning

703   and Performance Across the Spectrum of Kidney Function. Clinical Nursing Research

704   1054773816689282. DOI: https://doi.org/10.1177/1054773816689282

705

706   Segura-Ortí E, Martínez-Olmos FJ 2011 Test-retest reliability and minimal detectable change

707   scores for sit-to-stand-to-sit tests, the six-minute walk test, the one-leg heel-rise test, and

708   handgrip strength in people undergoing hemodialysis. Physical Therapy 91: 1244-52. DOI:

709   10.2522/ptj.20100141

710

711   Shoukri M, Asyali M, Donner A 2004 Sample size requirements for the design of reliability

712   study: review and new results. Statistical Methods in Medical Research 13: 251-71. No DOI

713   available.

714

715   Singh SJ, Morgan M, Hardman A, Rowe C, Bardsley P 1994 Comparison of oxygen uptake

716   during a conventional treadmill test and the shuttle walking test in chronic airflow limitation.

717   European Respiratory Journal 7: 2016-20. No DOI available.

718

719   Singh SJ, Jones P, Evans R, Morgan M 2008 Minimum clinically important improvement for

720   the incremental shuttle walking test. Thorax 63: 775-7. DOI: 10.1136/thx.2007.081208

721

722   Stratford PW 2004 Getting more from the literature: estimating the standard error of

723   measurement from reliability studies. Physiotherapy Canada; 56: 27-30. No DOI available.

724

725   Thomas VS, Hageman PA 2002 A preliminary study on the reliability of physical

726   performance measures in older day-care center clients with dementia. International

727   Psychogeriatrics 14: 17-23. No DOI available.

728

729   Verdijk LB, Van Loon L, Meijer K, Savelberg HH 2009 One-repetition maximum strength

730   test represents a valid means to assess leg strength in vivo in humans. Journal of Sports

731   Sciences 27: 59-68. DOI: 10.1080/02640410802428089

732

733   Walter SD, Eliasziw M, Donner A 1998 Sample size and optimal designs for reliability

734   studies. Statistics in Medicine, 17: 101-10. No DOI available.

735

736   Wang XH, Mitch WE 2014 Mechanisms of muscle wasting in chronic kidney disease. Nature

737   Reviews Nephrology 10: 504-16. DOI: 10.1038/nrneph.2014.112

738

739   Watson EL, Greening NJ, Viana JL, Aulakh J, Bodicoat DH, Barratt J, Feehally J, Smith, AC

740   2015 Progressive resistance exercise training in CKD: a feasibility study. American Journal

741   of Kidney Diseases 66: 249-57. DOI: 10.1053/j.ajkd.2014.10.019

742

743   Wilund KR, Tomayko EJ, Wu P-T, Chung HR, Vallurupalli S, Lakshminarayanan B,

744   Fernhall B 2010 Intradialytic exercise training reduces oxidative stress and epicardial fat: a

745   pilot study. Nephrology Dialysis Transplantation 25: 2695-701. DOI: 10.1093/ndt/gfq106

746

747 Table 1. Patient clinical and demographic characteristics

748

	$n = 41$
Age (years) ($_{SD}$) [min-max]	62 (11) [27-80]
Females, n (%)	23 (56%)
BMI (kg/m^2) ($_{SD}$) [min-max]	30.1 (5.7) [16.4-41.5]
eGFR (ml/min/1.72^2) ($_{SD}$) [min-max]	25 (8) [9-41]
Haemoglobin (g/l) ($_{SD}$) [min-max]	118 (14) [90-161]
Albumin (mg/l) ($_{SD}$) [min-max]	41 (2.7) [34-47]
Ethnicity	
Caucasian, n (%)	27 (66%)
South Asian, n (%)	6 (15%)
Indian, n (%)	4 (10%)
Black Caribbean, n (%)	2 (5%)
Other, n (%)	2 (5%)
Cause of disease	
Diabetic nephropathy type II, n (%)	5 (12%)
Interstitial nephritis, n (%)	5 (12%)
IgA nephropathy, n (%)	3 (7%)
Polycystic kidney disease, n (%)	3 (7%)
Other, n (%)	5 (12%)
Unknown / aetiology uncertain, n (%)	20 (49%)

Comorbidities

Diabetes mellitus type II, *n* (%)	9 (22%)
Hypertension, *n* (%)	26 (63%)
CCI score (SD) [min-max]	2.5 (0.7) [2-4]
- Mild CCI score (1-2), *n* (%)	24 (59%)
- Moderate CCI score (3-4), *n* (%)	17 (41%)
- Severe CCI score ($\geq$ 5), *n* (%)	0 (0%)
Median *n* of morbidities [min-max]	3 [0-6]

749

750 Unless stated, data presented as mean and SD: standard deviation; BMI: body mass index;

751 CCI: Charlson Comorbidity Index; eGFR: estimated glomerular filtration rate

752

Table 2. Relative reliability results of physical function tests in patients with CKD

Test	n	Test 1 score	Test 2 score	Difference (95CI)	ICC (r)[†]	95CI for ICC (r)	P[†]
ISWT (m)	41[‡]	400 (SD: 186)	409 (SD: 185)	9 (-4 – 23)	.973	.950 - .986	0.164
STS-5 (secs)	41[§]	11.0 (IQR: 8.8 – 13.4)	10.5 (IQR: 8.7 – 12.4)	-0.1 (-0.9 – 0.4)	.676	.468 - .813	0.248
STS-60 (repetitions)	40[‖]	26 (IQR: 20 – 29)	28 (IQR: 22 – 30)	2 (0 – 2)	.927	.866 - .961	0.093
e1RM (kg)	40[¶]	50.6 (IQR: 34.5 – 64.7)	51.8 (IQR: 36.6 – 64.7)	2.7 (0.0 – 4.3)	.927	.866 - .961	0.053
VO_{2peak} (ml/kg/min)	37[#]	19.6 (SD: 5.4)	19.6 (SD: 5.7)	0.0 (-0.9 – 1.0)	.866	.755 - .929	0.936

95CI: 95% confidence interval; e1RM: estimated 1 rep maximum-leg extension strength; ICC: intraclass correlation coefficient; ISWT: incremental shuttle walk test; IQR: interquartile range; SD: standard deviation; STS: sit-to-stand; [†] refers to test 1 versus test 2 analysis; [‡] 1 (2%) patient removed due to arthritic pain in hips and knees during test, 1 (2%) patient had incomplete data missing at time point 1 or 2; [§] 1 (2%) patient removed due to arthritic pain in hips and knees during test, 1 (2%) patient had incomplete data missing at time point 1 or 2; [‖] 2 (5%) patients removed due to arthritic pain in hips and knees during test, 1 (2%) patient had incomplete data missing at time point 1 or 2; [¶] 3 (7%) patients had incomplete data missing at time point 1 or 2; [#] 6 (14%) patients had incomplete data missing at time point 1 or 2
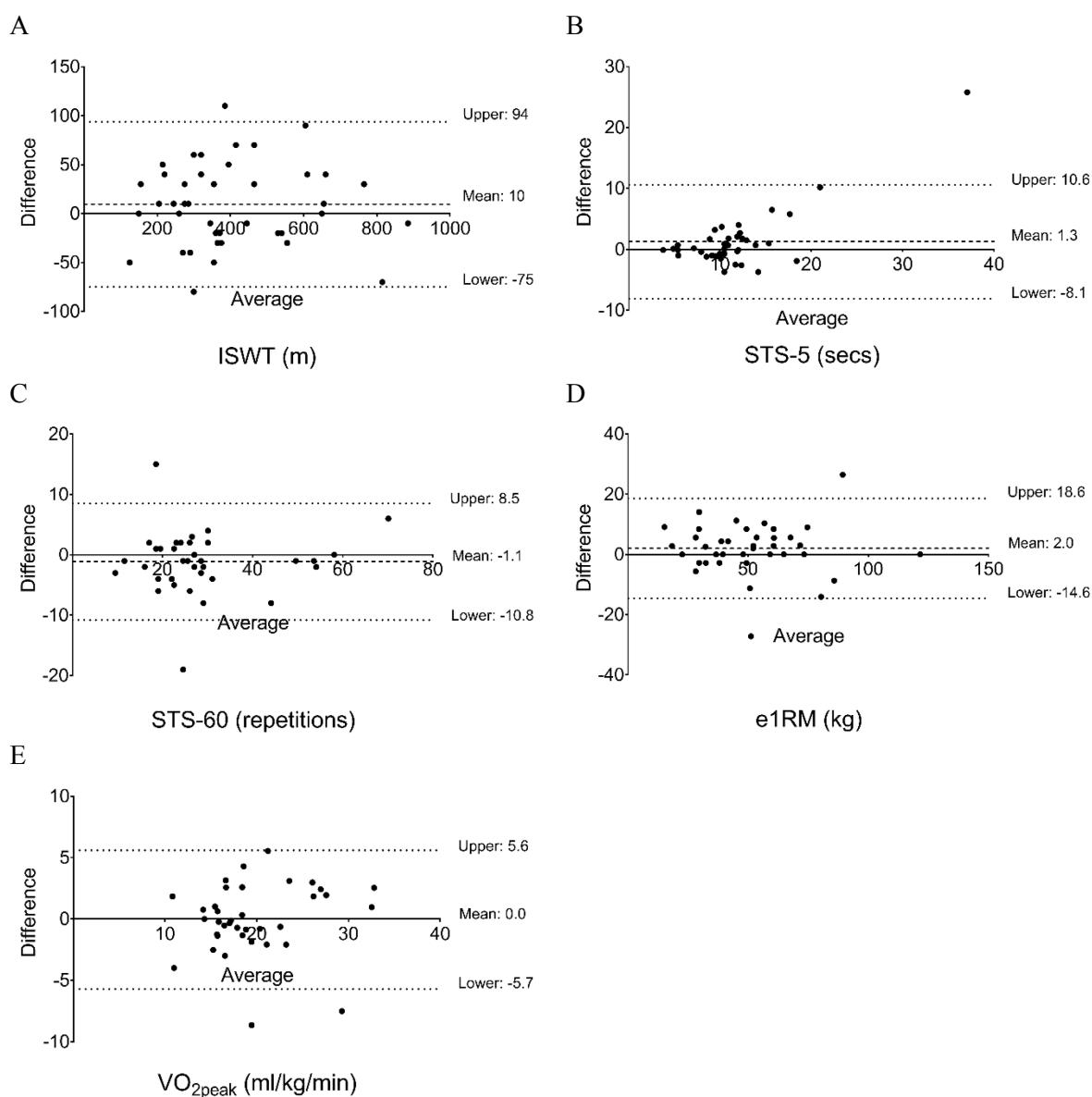
Table 3. SEM and MDC at both individual and group levels at 95CI for physical function tests in patients with CKD

Test	SE_M	MDC_{indv}	MDC_{group}
ISWT (m)	7.1	20	3
STS-5 (secs)	2.7	7.5	1.2
STS-60 (reps)	1.3	4	1
e1RM (kg)	2.3	6.4	1.0
VO_{2peak} (ml/kg/min)	1.0	2.8	0.5

e1RM: estimated 1 rep maximum-leg extension strength; ISWT: incremental shuttle walk test; MDC_{indv}: minimal detectable change at individual level; MDC_{group}: minimal detectable change at group level; SEM: standard error of measurement; STS: sit-to-stand.
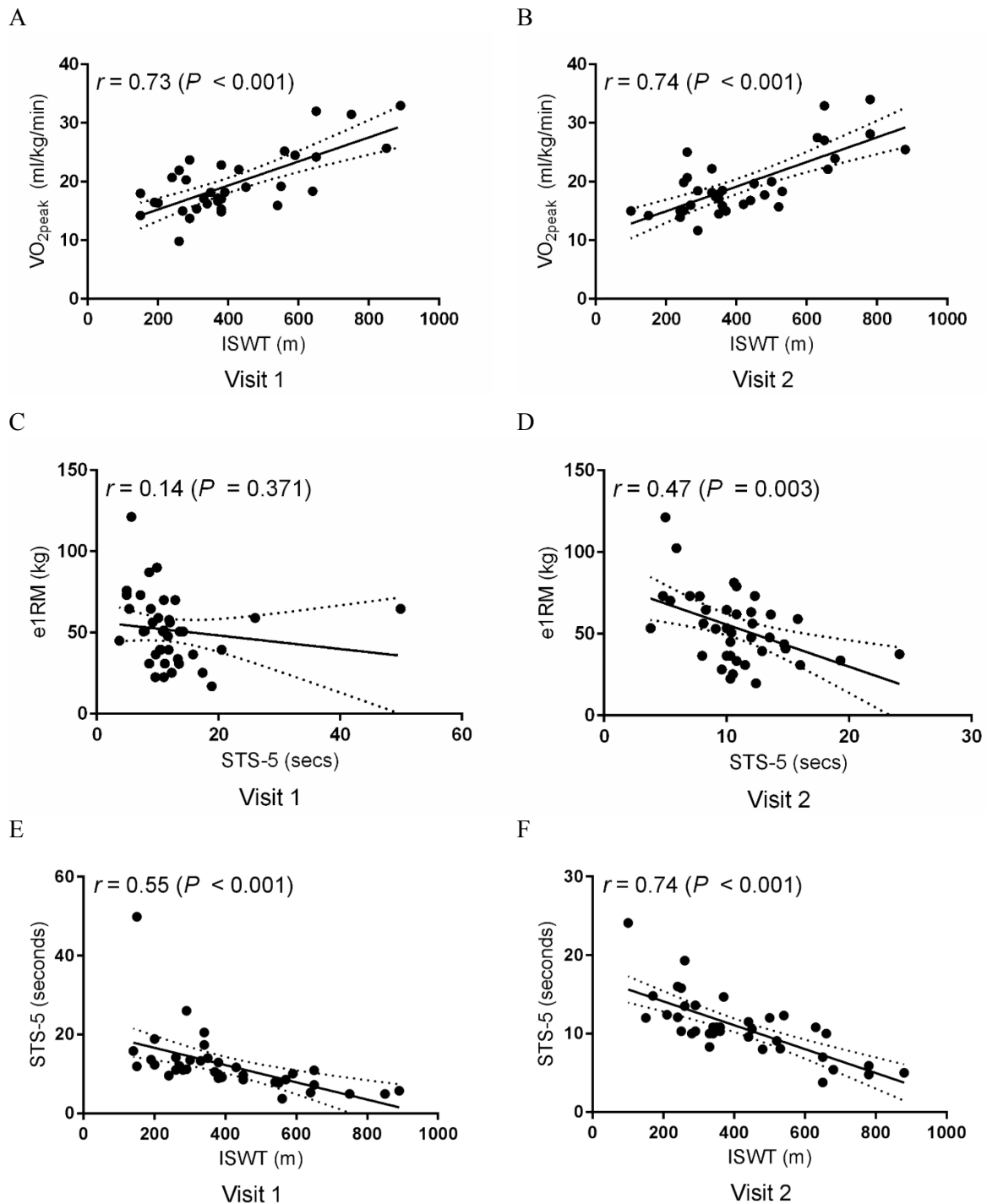
Figure 1. Bland-Altman plots showing test-retest reliability between test 1 and test 2

e1RM = estimated 1 repetition maximum-leg extension strength; ISWT = incremental shuttle walk test; STS = sit-to-stand. Top and bottom dotted lines represent upper and lower 95% confidence limits of agreement, respectively. Dashed line represents mean bias.

Figure 2. Linear regression line showing the ISWT against VO_{2peak}, and the STS-5 against e1RM and ISWT over the two testing visits

e1RM = estimated 1 repetition maximum-leg extension strength; ISWT = incremental shuttle walk test; STS-5 = sit-to-stand 5 repetition test. 95% confidence interval bands are shown as dotted lines that embrace the regression line in solid black fill.