The 'minimum clinically important difference' in frequently reported objective physical function tests following a 12-week renal rehabilitation exercise intervention in non-dialysis chronic kidney disease

Thomas J. Wilkinson, PhD[1]; Emma L. Watson, PhD[1]; Soteris Xenophontos, MSc[1]; Douglas W. Gould, PhD[1,2]; Alice C. Smith, PhD[1, 3]

[1]Leicester Kidney Lifestyle Team, Department of Infection, Immunity and Inflammation, University of Leicester, United Kingdom; [2]Instensive Care National Audit & Research Centre (ICNARC), London, United Kingdom; [3]John Walls Renal Unit, Leicester General Hospital, University Hospitals of Leicester Trust, United Kingdom.

Corresponding author

Dr. Thomas J Wilkinson, Leicester Kidney Exercise Team, Department of Infection, Immunity and Inflammation, University of Leicester, United Kingdom, LE5 4PW. Email: t.j.wilkinson@leicester.ac.uk

Author disclosures

The authors report no conflicts of interest. Preliminary results from this analysis have been presented at the American Society of Nephrology Kidney Week 2016 (Chicago).

Funding statement

This work was gratefully part-funded by the Stoneygate Trust. The research was supported by the National Institute for Health Research (NIHR) Leicester Biomedical Research Centre (BRC).

Abstract

Objective

Chronic kidney disease (CKD) patients are characterized by impaired physical function. The goal of exercise-based interventions is an improvement in functional performance. However, improvements are often determined by 'statistically significant' changes. We investigated the 'minimum clinically important difference' (MCID), 'the smallest change that is important to the patient', for commonly reported physical function tests.

Design

Non-dialysis CKD patients completed 12-weeks of a combined aerobic (plus resistance training). The incremental shuttle walking test (ISWT), sit-to-stand-5 (STS-5) and 60 (STS-60), estimated 1 repetition maximum (e1RM) for the knee extensors, and VO_{2peak} were assessed. After the intervention, patients rated their perceived change in health. Both anchor- and distribution-based MCID approaches were calculated.

Results

The MCID was calculated as follows: ISWT, +45m; STS-5, -4.2 seconds; VO_{2peak}, +1.5 ml/kg/min. Due to comparable increases in 'anchor' groups, no MCID was estimated for the STS-60 or e1RM.

Conclusion

We have established the MCID in CKD for common tests of physical function. These values represent the minimum change required for patients to perceive noticeable and beneficial change to their health. These scores will help interpret changes following exercise interventions where these tests are employed. These MCIDs can be used to power future studies to detect clinically important changes.

MeSH key words

Exercise; Outcome Assessment; Kidney Diseases; Rehabilitation

Introduction

Chronic kidney disease (CKD) is the progressive loss of renal function and is associated with premature mortality, multi-morbidity, and reduced quality of life [1,2]. CKD is a global health burden estimated to affect between 11 to 14% of people worldwide [2]. With an economic cost of over US$60 billion per year [3], 14% of the USA and Canada population have moderate to severe CKD stage 3 to 5, with this rising to 25% in those aged ≥60 years [2,3]. Patients with CKD are characterised by impaired physical functioning and limited exercise capacity, which can further result in reduced health-related quality of life and poor clinical outcome [4]. The pathological mechanisms behind this reduction in physical function and exercise capacity are complex and multifactorial. Potential factors include wasting of the skeletal muscle, inflammation, endothelial dysfunction, anaemia, and inactivity [5-7]. Many of these factors are 'modifiable' and, as such, may be improved or corrected through successful intervention. Exercise rehabilitation is becoming an increasing important tool to help improve exercise capacity and functional performance in CKD [6,8], and a number of research studies have found favorable effects of exercise in this patient group [5,7,9-11].

A frequently reported outcome measure in rehabilitation research is VO_{2peak} [11,12], a measure of maximal aerobic capacity. VO_{2peak} is obtained during a cardiopulmonary exercise test (CPET) and is regarded as the 'gold standard' measure of exercise capacity. Another key outcome is muscular strength. In CKD, where lower limb muscles are often atrophied [10], strength interventions and testing are particularly important. Consequently, estimating maximum strength using resistance machines has previously been used as a valid measure of strength, and may be more applicable in non-laboratory settings where dynamometry is not available [13].

While more accurate, graded maximal laboratory exercise tests and progressive strength testing may be impractical in some rehabilitation or clinical research settings. Consequently, pragmatic 'field' tests of physical function and performance are also employed. These include the 'incremental shuttle walking test' (ISWT) [14], to assess exercise capacity, and various chair 'sit-to-stand' tests (e.g., 5-repetiton,

repetitions in 60 seconds; STS-5/60) to assess dynamic balance, lower body strength, and in the case of the STS-60, muscle endurance [15,16]. All of these tests are extensively used in a wide range of clinical populations including older adults [17,18], chronic obstructive pulmonary disease (COPD) [15,19]; heart transplant patients [20]; Parkinson's disease [21], and CKD [10,16,22,23] to assess outcomes following various interventions.

Undeniably, the goal of many therapeutic interventions, such as exercise programs, is an improvement in functional performance. Unfortunately, 'real' improvements in clinical research trials are often decided by 'statistically significant' changes. Describing the limitations of significance testing is beyond the scope of this paper (for further reading see the American Statistical Association 2016 position statement [24]), however, a statistically significant change merely indicates a result did not occur by chance, and may have no translation into a useful recognized benefit to the patient [25-27].

The concept of the 'minimum clinically important difference' (MCID) was first developed in 1989 [28], and is defined as 'the smallest change that is important to the patient'. Described as 'fundamental to all clinical trials' [29], the MCID constitutes a threshold for outcome scores over which a patient would consider meaningful [25,30,31], and is important in determining the 'clinical effectiveness' of an intervention.

The MCID is preferably estimated using an 'anchor' based approach. Here, an external criterion, or 'anchor', is matched with an outcome of interest (e.g., outcome measure score). Often this 'anchor' is a patient-report subjective Likert scale with varying degrees of changes (e.g., 'good' through to 'bad') [25,26]. Alternatively, statistical driven 'distribution' methodology can be employed. These methods do not require patient input, but use estimations of outcome variability (e.g., standard deviation (SD) or effect size) to estimate an MCID.

No MCIDs currently exist for the commonly reported physical function tests mentioned above in a non-dialysis CKD population. However, as exercise rehabilitation and interventions are increasingly being

used to overcome functional deficits in CKD [8], establishment of the MCID for such tests would aid healthcare professionals and researchers to interpret improvements from a patient perspective after the implementation of a particular intervention. This may be particularly beneficial in determining the clinical effectiveness in individual patients or in studies with small sample sizes. In addition, the MCID has implications for the design of clinical trials, in terms of sample size calculation and the selection of endpoints [31].

Aims

The aim of this study was to calculate an MCID for the ISWT, STS-5/60, e1RM, and VO_{2peak}, by both anchor- and distribution-based methods, using a 12-week exercise program as the therapeutic intervention in non-dialysis CKD. A secondary aim was to observe the agreement between the patient-reported (anchor) and statistical (distribution) methods.

Methods

Design, setting, and patients

This was a retrospective secondary analysis of a study investigating the effects of a 12-week combined aerobic and resistance training intervention in non-dialysis CKD patients conducted at the Leicester General Hospital, UK between December 2013 and October 2016. Patients gave written informed consent and national Research Ethics Committee approval was obtained.

To be eligible for this study participants: (a) were diagnosed with moderately severe CKD (stages 3b-5); (b) were aged ≥18 years; (c) had no physical impairment or significant co-morbidities that were a contraindication to exercise (unstable hypertension, potentially lethal arrhythmia, uncontrolled diabetes mellitus (HbA1c >9%)); and had sufficient command of English to give informed consent. Patients were excluded if they began any form of renal replacement therapy (e.g., dialysis) or developed any physical impairment or significant co-morbidities that were a contraindication to exercise (as described above). The patient's clinician had the final decision on their inclusion to the study. The inclusion/exclusion criteria was kept broad to maximize recruitment rates and allow as many patients as possible the opportunity to take part in the intervention.

Renal exercise rehabilitation intervention

A full description of the methods for this study can be found in Watson et al.[32]. Patients attended Leicester Diabetes Centre, UK exercise gym three times a week for 12-weeks. Based around current cardiac and pulmonary rehabilitation principles, the exercise consisted of either moderate to vigorous aerobic only (70-80% of heart rate maximum, 30 minutes duration) performed on standard cardiovascular training equipment (e.g., treadmill, cycle ergometer, rowing machine), or a combined aerobic (as above but for 20 minutes duration) plus resistance training (70% 1-repetition maximum 3 sets of 10-12 repetitions) on a leg extension and leg press machine. Lower limb muscle exercises were chosen due to their prominent role in functional ability (e.g., walking, getting out of a chair) and proneness to muscle atrophy [33]. Patients gradually built up to the required intensities if they could not

attain them at the start of the program. For the purpose of this analysis, both groups (aerobic and aerobic plus resistance) are combined together to increase sample number.

Outcome measures

On initial assessment, basic anthropometric variables (height and weight) and medical history were taken. Current kidney function (eGFR) (ml/min/1.73m^2) was taken from the patient's latest routine blood samples.

The ISWT was used as a measure of exercise capacity and cardiorespiratory fitness [14,34]. For this test, the patient was required to walk between two cones 10m apart. Patients maintained a speed regulated by an external auditory beep. The walking speed was initially very slow (0.5m/s), but for every minute stage the required walking speed increased by 0.2m/s. The patient maintained cadence with the beeps until volitional exhaustion or until they could no longer keep up with the required pace. The total number of shuttles and distance walked (m) was calculated. Patients completed a familiarization of the ISWT prior to their assessment visit.

The STS-5 and STS-60 tests were employed as measures of lower body strength, balance, and muscle endurance. For these tests the patient sat on a seat (17 inches (43.2cm) from the ground) with their feet slightly apart. With their hands across their chest, patients were asked to: 1) perform 5 STS cycles as fast as possible [19]; and 2) perform as many STS cycles in 60 seconds [16].

The maximal strength (kg) of the quadriceps muscle was measured using a leg extension machine (TechnoGym, Italy). Performing a true 1RM test is associated with an increased injury risk and stress on the muscles and joints, particularly in untrained [13] and clinical groups [35], therefore we estimated 1RM (e1RM) of the knee extensors from a 5-rep maximum (5RM) [36]. During the test, weight was progressively increased in 2.5 kg increments, and the 5RM was determined as the maximal weight the patient could lift five times with correct technique.

Peak exercise capacity (VO_{2peak}) was assessed using CPET. Patients were asked to cycle for as long as possible at ≥60 revolutions per minute (RPM). Following a 3 minute warm up, the resistance on the static ergometer (Lode Excalibur, Netherlands) increased from 30 Watts by 1 Watt every 3 seconds in a ramp protocol. Throughout the test, an electrocardiogram was reviewed by an exercise cardiac nurse or doctor. The test was terminated if: RPM <60 and was unable to be increased with encouragement; the patient reached volitional exhaustion; or at the discretion of the medical professional. Oxygen consumption was measured using directly using an online breath-by-breath system (Cortex Metalyzer, Cranlea, UK) and relative VO_{2peak} (peak ml/kg/min) calculated.

All outcome measures had a familiarization test to reduce learning effects [37,38] and were completed pre- and post-exercise intervention.

Anchor-based MCID

Immediately after completion of the 12-week exercise intervention, patients were asked to rate how much they felt their health had changed using the Medical Outcomes Short Form-36 (SF-36) questionnaire: 'My health is...than 1 year ago' (i.e. prior to commencement of exercise intervention). Responses were categorized on a five-point Likert scale as: (a) 'much better'; (b) 'somewhat better'; (c) 'about the same'; (d) 'somewhat worse'; and (e) 'much worse'.

As described previously, response (b) (i.e. for patients to perceive their health as 'somewhat better') was used as the minimum MCID threshold [39], and for ease of interpretation, patient responses were grouped into two categories: 'better' (responses a + b), and 'the same or worse' (responses c + d + e). Similar grouping has been performed previously [40].

Distribution-based MCID

Two distribution-based methods were employed:

1) SD method = 0.5 x the SD of the change score (pre-post-exercise) [25];

2) Effect size method = change in scores corresponding to a small effect size (0.2) (0.2 x the mean change score) [26].

Statistical methods

As this was a secondary analysis, no a-priori sample size was calculated for the outcomes reported here. Only data for patients (n = 26) who completed the post-intervention questionnaire was included. The baseline variables were normally distributed. The change in physical function scores achieved by patients for each response category of the anchor-based approach is reported as the mean (with 95% confidence intervals, 95CI). As baseline performance may impact upon the change in physical function achieved [39,40]; mean percentage change was calculated to adjust for baseline values. Parametric independent sample t-tests were used to investigate the differences between groups, while within-group changes were assessed using paired sample t-tests. Data were analyzed using SPSS 24 (IBM UK Ltd, UK). Statistical significance was set at $P < .050$.

Results

Patients

Of the 54 patients consented to the main trial, 41 completed the 12-week exercise program. Of these, 26 patients completed the SF-36 questionnaire post-exercise and were eligible for analysis. Baseline characteristics for these 26 patients are found in **Table 1**. The majority (96%) of patients were CKD stage 3b to 4. One patient had an eGFR of 8 ml/min/1.732 (i.e. CKD stage 5), although she did not require any modality of renal replacement therapy.

Summary of main effects of renal rehabilitation program

Adherence to the exercise program was good with an average of 31/36 (86%) sessions attended. The majority of patients achieved the target intensity in the first session. For the aerobic component, 85% of patients reached their heart rate % target and goal duration (minutes) in the first session. All patients had reached the required intensity for the aerobic component by the end of week 2 (i.e. by session 6). For the resistance component, 60% reached the required 70% 1RM goal in the first session. Apart from 3 patients (whom all achieved desired intensity by week 6), the rest of the patients reached 3 sets of 10-12 repetitions 70% 1RM by week 2/session 6).

In this sub-group (n = 26), 12-weeks exercise training had favorable effects on physical function. Mean distance walked on the ISWT increased by +32m [95CI: 5 to 58] (P= .021); the number of seconds taken to complete the STS-5 was reduced by -3.1 seconds [95CI: 1.1 to 5.0] (P = .004); and the number of repetitions on the STS-60 increased by +4 [95CI: 2 to 7] (P= .001). e1RM increased by +18.7kg [95CI: 13.1 to 24.4] (P< .001)). VO_{2peak} increased by +1.3 ml/kg/min [95CI: -0.2 to 2.8], although this was not significant (P= .112).

The majority of patients (17 out of 26 (65%)) rated their health as ‘better’ on the SF-36 questionnaire, whilst 7 (27%) patients felt that it was the ‘same’ and 2 (8%) rated it as ‘worse’. As such, 9 (35%) patients felt their health was ‘same or worse’.

Anchor-based MCID

The average age of the 'better' group was 62.4 (±14.6) whilst the 'same or worse' group was 59.4 (±13.1) (*P*= .616). There was no difference in the number of females in each group (9 versus 5, *P*= .899), or in eGFR (26.4 (±8.8) versus 24.3 (±7.2) ml/min/1.73m^2, *P*= .550) and body mass index (29.2 (±5.8) versus 29.2 (±6.5), *P*= .988). No significant differences were also observed in any of the physical function tests pre-intervention (P's = .123 to .969).

Using anchor-based analysis, there were small yet distinct differences in outcome measure changes between those who rated their health as 'better' versus those who rated their health as the 'same or worse' following the exercise programme (**Table 2**).

In the ISWT, patients that rated themselves as 'better' significantly increased their distance walked by 45m (*P*= .033). In comparison, those who rated themselves as the 'same or worse' increased by 26m (*P*= .339). The time taken to complete the STS-5 was reduced by 4.2 seconds in the 'better' patients (*P*= .007) compared to -0.9 seconds (*P*= .221) in the 'same or worse' group. This difference of 3.3 seconds was significant (*P*= .041). VO_{2peak} was increased in both groups, although the greater improvement was seen in the 'better' patients (1.5 ml/kg/min, *P*= .073), compared to just 0.5 ml/kg/min (*P*= .788) in the 'same or worse' patients. Improvements in the STS-60 and e1RM were comparable between the groups.

Distribution-based MCID

Table 3 shows MCID estimates using distribution-based methodology. The SD method yielded values closest to those estimated using the anchor-based approach. For VO_{2peak}, the MCID estimated was greater than that calculated using the anchor method (1.5 versus 1.8 ml/kg/min). The MCID's calculated using the effect size method were considerably lower than using an anchor-based method.

Discussion

Following 12-weeks of supervised exercise, we have been able to identify the MCID for several common measures of physical function. The MCID estimated in the current trial for the ISWT is 45m (11%) and is the first reported in a non-dialysis CKD population. This value compares well with previous estimates in COPD patients following 7-weeks pulmonary rehabilitation (PR) (48m/28%) [34] and patients with idiopathic pulmonary fibrosis following PR (46m/17%) [40]. The MCID in patients following 6-weeks cardiac rehabilitation was slightly greater at 70m (18%) [39]; however, while in absolute terms our improvement of 45m is 25m less than that described in this study, the baseline ISWT distance in that trial (391m) was less than that seen in our cohort (420m). As such, patients with a lower baseline score may 'have more to gain' [39] and when normalized to a percentage change, the disparity is somewhat reduced (11% versus 18%). Further, disease-specific limitations may also contribute to the disparities seen.

A valid MCID should be *at least* as large as the observed 'minimal detectable change' (MDC) [26]. The MDC is the smallest amount of reliable change in a measurement necessary to conclude that the difference is not attributable to error. Change exceeding the MDC is considered 'true' change [41]. Our MCID change value of 45m seen in the 'better' group (*and* the absolute difference between the 'better' and 'same or worse' group: 19m) exceeds the MDC for the ISWT in CKD [42]. As such, this change cannot be attributed to the inherent error of the ISWT.

The mean change in our 'same or worse' group for the ISWT (26m) was similar to that in seen in COPD [39] where the 'about the same' group increased by 29m. This demonstrates that patients failed to rate, positive yet, small changes in exercise capacity. This may be a consequence of a 'response shift' in which contact with a health professional shifts a patient perspective on how they assess current state [39]. It may also be that an increase of 26m is simply insufficient to have a noticeable effect on self-perceived health in CKD, and that greater improvements (i.e. 45m) are required to have noticeable benefits to the patients.

We estimated the MCID for the STS-5 test at -4.2 seconds (-33%). Notably, this was significantly different than the -0.9 seconds (-7%) in the 'same or worse' group. Our MCID value, and the difference between groups, exceeds the MDC for the STS-5 [42]. Previous estimates of an MCID in the STS-5 test are scarce. Jones et al. previously valued the MCID as a reduction of -1.7 seconds (-11%) following PR [15], whilst Meretta et al. estimated it at -2.3 seconds in patients with peripheral, central, or mixed vestibular dysfunction following an individualized outpatient vestibular rehabilitation programme [43]. Using a distribution-method approach, the MCID was estimated at -3.7 seconds (-27%) in patients with multiple sclerosis [44].

We found that an improvement of 1.5 ml/kg/min (7%) in VO_{2peak} was sufficient for patients to detect a positive change in their health. This value exceeds the MDC for the VO_{2peak} [42]. Despite being a frequently reported outcome in clinical research, there is limited analysis into the MCID for maximal oxygen uptake (measured as either VO_{2peak} or VO_{2max}). In 17 elderly patients with abdominal aortic aneurysms undergoing 6-weeks of supervised exercise (2x/week), the MCID was defined as an improvement of 2.0 ml/kg/min (19%) [45]. While MCID data is limited for changes in maximal aerobic capacity, any improvements should not be understated. In a large meta-analysis, Kodama et al. found an improvement of 1-MET (3.5 ml/kg/min) resulted in a 13% reduction in all-cause mortality and 15% reduction in cardiovascular disease related events (in healthy men and women) [46]. Further, Myers et al. reported each 1-MET increase in exercise capacity conferred a 12% improvement in survival in elderly men referred for clinical treadmill exercise testing [47]. Our overall improvement of 1.3 ml/kg/min equates to 0.37-MET (and extrapolated could mean a ~5% reduction in all-cause mortality based on previous research).

Following 12-weeks of exercise, we found that performance in the STS-60 and e1RM tests were both increased. When separating patients into our MCID 'anchors' (i.e. 'better' or 'same or worse'), we found that both groups increased at comparable rates. As such, we could not distinguish an obvious MCID value for these tests. Notably, the difference (1 repetition) between the 'better' and 'same or

worse' groups for the STS-60 did not exceed the MDC for this test, and thus the difference observed could be a result of natural test variation. While explanations for absence of a clear MCID for the STS-60 and e1RM are unknown, it may be that a greater improvement is needed for patients to perceive this as having a positive effect on their health. Tests such as the ISWT or STS-5 have more transferable application to everyday activities (e.g., walking, standing up from a chair several times). Conversely, performing ~30-70 STS repetitions (as seen in the STS-60) may not seem useful or relevant, and due to the duration of the test patients may be unable to comprehend the somewhat small improvements seen. For the e1RM, it is unlikely patients would repeat this movement in everyday life, and while a good measure of strength, it is localized to the quadriceps. More overall lower limb strength measures may have a greater influence on patients.

Anchor versus distribution methodology

We employed two statistically driven distribution approaches of estimating the MCID: the effect size and SD method. These methods do not require a subjective patient input, but use estimations of outcome variability. We observed that, overall, these methods markedly underestimated the MCID when compared to favored anchor-based estimates. The exception to this was the MCID for VO_{2peak} which was estimated by the SD approach at 1.8 ml/kg/min compared to the anchor-based method estimate of 1.5 ml/kg/min. Similar discordance between distribution- and anchor-based approaches have been observed previously [39], [48,49].

In particular, the effect size method severely underestimated MCID values. The change in scores corresponding to a small effect size (0.2) has previously been considered the MCID [26]. However, our patients may have higher expectations of what constitutes a 'better' quality of life following an intervention such as exercise. Consequently, estimating the MCID using a large effect size (>0.8) may be more appropriate. While potentially useful in estimating the MCID when an anchor-based method is not available (no subjective quality of life criterion measured) or unachievable (as in the case of the STS-60 and e1RM in our trial), this approach (i.e. using a statistically driven methodology) ignores the

very purpose of the MCID as it does not address the question of clinical importance and, importantly, fails to distinctly separate clinical importance from statistical significance [26].

Strengths and limitations

The strength of this study is that it a uses patient-centered approach to determine the MCID. However, as an anchor-based approach relies on retrospective estimates of health [34], this approach is at risk of interference by 'recall bias' as patients are required to remember the intrinsic nature of their condition prior to an intervention [26,27]. Our analysis is limited by a small opportunistic sample size. This restricts the analysis in two aspects. Firstly, while some of our data are comparable to previous estimates from other clinical populations, it should be further substantiated in larger cohorts. Unfortunately, complex exercise interventions in research are difficult to perform, especially in CKD groups, and thus our data do provide useful and important preliminary data in this regard. Secondly, due to a reduced number of patients, we combined both groups of exercise modalities. This prevents differentiation of MCIDs for these interventions. However, it should be noted that changes in VO_{2peak}, ISWT, STS-5, and STS-60 performance were similar (i.e. no significant differences) between the two groups in the full analysis. Thus it appears for these tests, 12-weeks exercise of any modality is sufficient to improve performance. Unsurprisingly, the only difference in performance observed between the two groups was seen for e1RM (the additional resistance arm increased by 13 kg more); interestingly, e1RM was one of the variables we were unable to ascertain a MCID for. A small sample size also meant that only 9 (35%) patients identified their health as 'the same or worse' (two of these felt it got 'worse'). While this clearly shows positive benefits of exercise in this group (i.e. improving health and/or preventing further decline), it limits the generalisability of the findings. Lastly, our resistance programme contained only two lower limb resistance exercises. Whilst the quadriceps muscles are functionally beneficial in CKD 33, future training interventions should also look to maintain the musculature of other lower limb muscles important for physical functioning (e.g., calf muscles).

Currently, no universally agreed method exists to estimate MCID. While MCID values are dynamic and often context (and intervention)-specific [50], the most useful concept of MCID should indicate a

treatment is effective and alert the clinician or healthcare professional and the patient to its impact on patient's life [26]. While an anchor-based method which is patient-centered is perhaps the most pertinent approach to do this, it is important to remember that condemning an intervention because a group failed to meet the MCID may be misguided, as the weakness may be borne within the MCID rather than the intervention itself [27]. In an ideal world, adjunct metrics, such as the MDC along with statistical descriptive of variations (e.g., 95CI, effect size), should be used alongside the MCID to help interpret and evaluate whole changes from an intervention.

Summary

We have established the MCID in a non-dialysis CKD population for several commonly reported tests of physical function. These values represent the minimum change required in order for the patient to see noticeable and beneficial change on their overall health status. These scores will help rehabilitation professionals, clinicians, and researchers interpret 'meaningful' changes following interventions such as renal exercise rehabilitation programs where these tests are commonly employed. Further, these MCID's can be used to inform future trial methodology by estimating sample size to help power studies to detect clinically important changes over statistical ones.

References

1. Levey AS, Eckardt K-U, Tsukamoto Y, et al. Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney Int. 2005;67(6):2089-2100.

2. Hill NR, Fatoba ST, Oke JL, et al. Global prevalence of chronic kidney disease–a systematic review and meta-analysis. PloS One. 2016;11(7):e0158765.

3. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: global dimension and perspectives. The Lancet. 2013;382(9888):260-272.

4. Roshanravan B, Robinson-Cohen C, Patel KV, et al. Association between physical performance and all-cause mortality in CKD. J Am Soc Nephrol. 2013; 24(5):822-830.

5. Gould DW, Graham-Brown MP, Watson EL, Viana JL, Smith AC. Physiological benefits of exercise in pre-dialysis chronic kidney disease. Nephrology. 2014;19(9):519-527.

6. Wilkinson T, Shur N, Smith A. "Exercise as medicine" in chronic kidney disease. Scand J Med Sci Sports. 2016;26(8):985-988.

7. Roshanravan B, Gamboa J, Wilund K. Exercise and CKD: Skeletal Muscle Dysfunction and Practical Application of Exercise to Prevent and Treat Physical Impairments in CKD. Am J Kidney Dis. 2017;69(6):837-852.

8. Greenwood SA, Lindup H, Taylor K, et al. Evaluation of a pragmatic exercise rehabilitation programme in chronic kidney disease. Nephrol Dial Transplant. 2012;27(3):126-134.

9. Viana JL, Kosmadakis GC, Watson EL, et al. Evidence for anti-inflammatory effects of exercise in CKD. J Am Soc Nephrol. 2014;25(9):2121-2130.

10. Watson EL, Greening NJ, Viana JL, et al. Progressive resistance exercise training in CKD: a feasibility study. Am J Kidney Dis. 2015;66(2):249-257.

11. Howden EJ, Coombes JS, Strand H, Douglas B, Campbell KL, Isbel NM. Exercise training in CKD: efficacy, adherence, and safety. Am J Kidney Dis. 2015;65(4):583-591.

12. Beck-da-Silva L, Piardi D, Soder S, et al. IRON-HF study: a randomized trial to assess the effects of iron in heart failure patients with anemia. Int J Cardiol. 2013;168(4):3439-3442.

13. Gail S, Künzell S. Reliability of a 5-repetition maximum strength test in recreational athletes. Dtsch Z Sportmed. 2014;65(11):314-317.

14. Singh S, Morgan M, Hardman A, Rowe C, Bardsley P. Comparison of oxygen uptake during a conventional treadmill test and the shuttle walking test in chronic airflow limitation. Eur Respir J. 1994;7(11):2016-2020.

15. Jones SE, Kon SS, Canavan JL, et al. The five-repetition sit-to-stand test as a functional outcome measure in COPD. Thorax. 2013;68(11):1015-1020.

16. Segura-Ortí E, Martínez-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. 2011;91(8):1244-1252.

17. Goldberg A, Chavis M, Watkins J, Wilson T. The five-times-sit-to-stand test: validity, reliability and detectable change in older females. Aging Clin Exp Res. 2012;24(4):339-344.

18. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist. 2013;53(2):255-267.

19. Crook S, Puhan MA, Frei A. The validation of the sit-to-stand test for COPD patients. Eur Respir J. 2017;50(3):1701506.

20. Lewis ME, Newall C, Townend JN, Hill SL, Bonser RS. Incremental shuttle walk test in the assessment of patients for heart transplantation. Heart. 2001;86(2):183-187.

21. Petersen C, Steffen T, Paly E, Dvorak L, Nelson R. Reliability and minimal detectable change for sit-to-stand tests and the functional gait assessment for individuals with Parkinson disease. J Geriatr Phys Ther. 2017;40(4):223-226.

22. Rossi AP, Burris DD, Lucas FL, Crocker GA, Wasserman JC. Effects of a renal rehabilitation exercise program in patients with CKD: a randomized, controlled trial. Clinical J Am Soc Nephrol. 2014;9(12):2052-2058.

23. Segura-Ortí E, Gordon P, Doyle J, Johansen K. Correlates of Physical Functioning and Performance Across the Spectrum of Kidney Function. Clin Nurs Res. 2017:1054773816689282.

24. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. Am Stat. 2016; 70(2): 129-133.

25. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. J Rheumatol. 2001;28(2):406-412.

26. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007;7(5):541-546.

27. Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. J Man Manip Ther. 2008;16(4):82-83.

28. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Contemp Clin Trials. 1989;10(4):407-415.

29. Gibbs NM, Weightman WM. The minimum clinically important difference is fundamental to all clinical trials. Eur J Anaesthesiol. 2016;33(1):60.

30. Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. Arthritis Res Ther. 2015;17(1):143.

31. Rees J. Patients not P values. BJU Int. 2015;115(5):678-679.

32. Watson EL, Gould DW, Wilkinson TJ, et al. 12-weeks combined resistance and aerobic training confers greater benefits than aerobic alone in non-dialysis CKD. Am J Physiol Renal Physiol. 2018; 314(6):1188-1196.

33. McIntyre CW, Selby NM, Sigrist M, Pearce LE, Mercer TH, Naish PF. Patients receiving maintenance dialysis have more severe functionally significant skeletal muscle wasting than patients with dialysis-independent chronic kidney disease. Nephrol Dial Transplany. 2006;21(8):2210-2216.

34. Singh SJ, Jones P, Evans R, Morgan M. Minimum clinically important improvement for the incremental shuttle walking test. Thorax. 2008;63(9):775-777.

35. Abdul-Hameed U, Rangra P, Shareef MY, Hussain ME. Reliability of 1-repetition maximum estimation for upper and lower body muscular strength measurement in untrained middle aged type 2 diabetic patients. Asian J Sports Med. 2012;3(4):267-273.

36. Dudgeon WD, Phillips KD, Durstine JL, et al. Individual exercise sessions alter circulating hormones and cytokines in HIV-infected men. Appl Physiol Nutr Metab. 2010;35(4):560-568.

37. Young HML, March D, Churchward D, et al. Practice tests are required for physical function testing in a haemodialysis patients. Physiotherapy. 2016;102:261-262.

38. Johnson-Warrington V, Sewell L, Morgan M, Singh S. Do we need a practice incremental shuttle walk test for patients with interstitial lung disease referred for pulmonary rehabilitation? Respirology. 2015;20(3):434-438.

39. Houchen-Wolloff L, Boyce S, Singh S. The minimum clinically important improvement in the incremental shuttle walk test following cardiac rehabilitation. Eur J Prev Cardiol. 2015;22(8):972-978.

40. Nolan CM, Delogu V, Maddocks M, et al. Validity, responsiveness and minimum clinically important difference of the incremental shuttle walk in idiopathic pulmonary fibrosis: a prospective study. Thorax. 2017:thoraxjnl-2017.

41. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Phys Ther. 2006;86(5):735-743.

42. Wilkinson TJ, Xenophontos S, Gould DW, et al. 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. Phys Ther Prac. 2018; 30: 1-12.

43. Meretta BM, Whitney SL, Marchetti GF, Sparto PJ, Muirhead RJ. The five times sit to stand test: responsiveness to change and concurrent validity in adults undergoing vestibular rehabilitation. J Vestib Res. 2006;16(4, 5):233-243.

44. Jensen HB, Mamoei S, Ravnborg M, Dalgas U, Stenager E. Distribution-based estimates of minimum clinically important difference in cognition, arm function and lower body function after slow release-fampridine treatment of patients with multiple sclerosis. Mult Scler Relat Disord. 2016;7:58-60.

45. Kothmann E, Batterham AM, Owen SJ, et al. Effect of short-term exercise training on aerobic fitness in patients with abdominal aortic aneurysms: a pilot study. Br J Anaesth. 2009;103(4):505-510.

46. Kodama S, Saito K, Tanaka S, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA. 2009;301(19):2024-2035.

47. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. N Engl J Med. 2002;346(11):793-801.

48. Gremeaux V, Troisgros O, Benaïm S, et al. Determining the minimal clinically important difference for the six-minute walk test and the 200-meter fast-walk test during cardiac rehabilitation program in coronary artery disease patients after acute coronary syndrome. Arch Phys Med Rehabil. 2011;92(4):611-619.

49. Jayadevappa R, Malkowicz SB, Wittink M, Wein AJ, Chhatre S. Comparison of Distribution- and Anchor-Based Approaches to Infer Changes in Health-Related Quality of Life of Prostate Cancer Survivors. BMC Health Serv Res. 2012;47(5):1902-1925.

50. Lang CE, Edwards DF, Birkenmeier RL, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. Arch Phys Med Rehabil. 2008;89(9):1693-1700.

Table 1. Demographics of patients at baseline

	n = 26
Age (years) [±SD]	61.4 [±13.7]
No. of males (%)	12 (44)
Body mass index (kg/m^2) [±SD]	29.5 [±6.0]
eGFR (ml/min/1.73m^2) [±SD]	25.5 [±8.1]§
Causes of disease	
IgA nephropathy, n (%)	2 (8%)
Diabetic nephropathy, n (%)	3 (12%)
Polycystic kidney disease, n (%)	1 (4%)
Chronic pyelonephritis (interstitial nephritis), n (%)	2 (8%)
Unknown etiology, n (%)	14 (53%)
Other etiology, n (%)	4 (15%)
Co-morbidities	
Hypertension, n (%)	13 (50%)
Type II diabetes mellitus, n (%)	7 (27%)
Baseline physical function score	
ISWT (m) [min-max]	420 [140-890]
STS-5 (seconds) [min-max]	12.7 [3.7-49.9]
STS-60 (repetitions) [min-max]	30 [8-73]
e1RM (kg) [min-max]	48.7 [15.0-121.4]
VO_{2peak} (ml/kg/min) [min-max]	20.7 [13.7-33.0]

Data presented as mean [±SD] or [min-max]. § = 96% of patients were CKD stage 3b to 4, only 1 person had an eGFR of <15 (i.e. CKD stage 5) however she was not treated using a renal replacement therapy.

eGFR = estimated glomerular filtration rate; ISWT = incremental shuttle walking test; STS-5 = sit-to-stand 5-repetiton test; STS-60 = sit-to-stand 60 second test; e1RM = estimated 1-repetition maximum; VO_{2peak} = peak oxygen uptake.

Table 2. Mean change [95CI] in physical function scores from baseline following a 12-week renal exercise intervention in each response category

	Perception of health					
	'Better' (*n* = 17)		'The same or worse' (*n* = 9)		Between-group	
	Mean difference [95CI], %	*P*§	Mean difference [95CI], %	*P*§	difference	*P*
Mean change in ISWT (m)	**+45 [3 to 66], +11%**	.033*	+26 [-32 to 83], +6%	.339	19 [-30 to 69]	.427
Mean change in STS-5 (seconds)	**-4.2 [-7.0 to -1.3], -33%**	.007*	-0.9 [-2.5 to 0.7], -7%	.221	**3.3 [-0.7 to 7.3]**	**.041***
Mean change in STS-60 (repetitions)	**+5 [1 to 8], +17%**	**.009***	**+4 [0 to 7], +13%**	**.037***	1 [-4 to 6]	.669
Mean change in e1RM (kg)	**+18.3 [10.9 to 25.7], +38%**	**<.001***	**+20.3 [8.6 to 32.1], +42%**	**.004***	2.0 [-10.5 to 14.5]	.744
Mean change in VO_{2peak} (ml/kg/min)	+1.5 [-0.2 to 3.2], +7% ¥	.073	+0.5 [-3.4 to 4.3], +2%	.788	1.1 [-2.2 to 4.4]	.509

95CI = 95% confidence intervals; ISWT = incremental shuttle walking test; STS-5 = sit-to-stand 5-repetiton test; STS-60 = sit-to-stand 60 second test; e1RM = estimated 1-repetition maximum; VO_{2peak} = peak oxygen uptake. * = significant ($P < .050$); § = *P* value donates significance of the within-group change.

Table 3. Differences between anchor- and distribution-based methods

	ISWT (m)	STS-5 (seconds)	STS-60 (repetitions)	e1RM (kg)	VO_{2peak} (ml/kg/min)
Anchor-based	*+45*	*-4.2*	*N/A* [a]	*N/A* [a]	*+1.5*
Distribution-based					
SD method	+29	-2.5	+3	+7.3	+1.8
Effect size method	+6	-0.6	+1	+3.7	+0.3

SD = standard deviation; ISWT = incremental shuttle walking test; STS-5 = sit-to-stand 5-repetiton test; STS-60 = sit-to-stand 60 second test; e1RM = estimated 1-repetition maximum; VO_{2peak} = peak oxygen uptake. SD method: MCID = 0.5 × the SD of the change score (pre-post-exercise); Effect size method: MCID = 0.2 x the mean change score (pre-post-exercise). [a] = as both groups significantly increased at comparable rates, estimates of the MCID (i.e. using the mean improvement for the 'better' group) are inconclusive.