Archives of Rehabilitation Research and Clinical Translation (2020) 2, 100066

ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Rehabilitation Research and Clinical Translation

Archives of Rehabilitation Research and Clinical Translation 2020;2:100066

Available online at www.sciencedirect.com

Original Research

Agreement and Reliability of Clinician-in-Clinic Versus Patient-at-Home Clinical and Functional Assessments: Implications for Telehealth Services

Shelley E. Keating, PhD [a], Amandine Barnett, M Nut & Diet Practice [b,c], Ilaria Croci, PhD [a,d], Amy Hannigan, BHSci (Nut & Diet) [c], Louise Elvin-Walsh, BHSci (Nut & Diet) [c], Jeff S. Coombes, PhD [a], Katrina L. Campbell, PhD [b,c], Graeme A. Macdonald, PhD, MBBS [e,f,g], Ingrid J. Hickman, PhD [c,h]

[a] *School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, Queensland, Australia*
[b] *Bond Institute of Health and Sport, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia*
[c] *Department of Nutrition and Dietetics, Princess Alexandra Hospital, Brisbane, Queensland, Australia*
[d] *K.G. Jebsen Center of Exercise in Medicine, Department of Circulation and Medical Imaging, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway*
[e] *Queensland Liver Transplant Service, Princess Alexandra Hospital, Brisbane, Queensland, Australia*
[f] *Department of Gastroenterology and Hepatology, Princess Alexandra Hospital, Brisbane, Queensland, Australia*
[g] *School of Medicine, The University of Queensland, Brisbane, Queensland, Australia*
[h] *Mater Research Institute, The University of Queensland, Brisbane, Queensland, Australia*

List of abbreviations: 6MWT, 6-minute walk test; DBP, diastolic blood pressure; ICC, intraclass correlation coefficient; LoA, limit of agreement; LTR, liver transplant recipient; MCID, minimal clinically important difference; SBP, systolic blood pressure; STST, sit-to-stand test.

Supported by the Princess Alexandra Hospital Research Support Scheme project grant. Shelley E. Keating is supported by the National Health and Medical Research Council of Australia via an Early Career Research Fellowship (grant no. 122190). Ilaria Croci is supported by the Swiss National Science Foundation with a Postdoctoral Fellowship.
Clinical Trial Registration No.: ACTRN12617001260314.
Disclosures: none.
Presented as a poster to the 2018 Exercise and Sports Science Australia Research to Practice, March 27-29, 2018, Brisbane, Queensland, Australia.
Cite this article as: Arch Rehabil Res Clin Transl. 2020;2:100066.

https://doi.org/10.1016/j.arrct.2020.100066

KEYWORDS
Chronic disease;
Rehabilitation;
Self-assessment;
Technology;
Telemedicine

Abstract *Objective:* To compare agreement and reliability between clinician-measured and patient self-measured clinical and functional assessments for use in remote monitoring, in a home-based setting, using telehealth.
Design: Reliability study: repeated-measure, within-subject design.
Setting: Trained clinicians measured standard clinical and functional parameters at a face-to-face clinic appointment. Participants were instructed on how to perform the measures at home and to repeat self-assessments within 1 week.
Participants: Liver transplant recipients (LTRs) (N=18) (52±14y, 56% men, 5.4±4.3y posttransplant] completed the home self-assessments.
Interventions: Not applicable.
Main Outcome Measures: The outcomes assessed were body weight, systolic and diastolic blood pressure (SBP and DBP), waist circumference, repeated chair sit-to-stand (STST), maximal push-ups, and the 6-minute walk test (6MWT). Intertester reliability and agreement between face-to-face clinician and self-reported home-based participant measures were determined by intraclass-correlation coefficients (ICCs) and Bland-Altman plots, which were compared with minimal clinically important differences (MCID) (determined *a priori*).
Results: The mean difference (95% confidence interval) and [limits of agreement] for measures (where positive values indicate lower participant value) were weight, 0.7 (0.01-1.4) kg [−2.2 to 3.6kg]; waist 0.4 (−1.2 to 2.0) cm [−5.9 to 6.8cm]; SBP 7.7 (0.6-14.7) mmHg [−19.4 to 34.9mmHg]; DBP 2.4 (−1.4 to 6.2) mmHg [−12.2 to 17.0mmHg]; 6MWT, 7.5 (−29.1 to 44.1) m [−127.3 to 142.4m]; STST 0.5 (−0.8 to 1.7) seconds [−4.3 to 5.3s]; maximal push-ups −2.2 (−4.4 to −0.1) [−10.5 to 6.0]. ICCs were all >0.75 except for STST (ICC=0.73). Mean differences indicated good agreement than MCIDs; however, wide limits of agreement indicated large individual variability in agreement.
Conclusions: Overall, LTRs can reliably self-assess clinical and functional measures at home. However, there was wide individual variability in accuracy and agreement, with no functional assessment being performed within acceptable limits relative to MCIDs >80% of the time.

The use of telehealth strategies for chronic disease management favors equitable access to health services for patients across wide geographical dispersion, increases patient's self-care management, and reduces the patient's time away from daily life as well as travel costs.[1] Although a variety of telehealth strategies have been used successfully to deliver lifestyle interventions for weight reduction and improvements in metabolic risk factors,[2,3] their translation into clinical practice is complex due to the need to monitor outcomes remotely.

Effectiveness of telehealth exercise interventions often rely on monitoring change in functional tests such as the 6-minute walk test (6MWT) and the repeated sit-to-stand test (STST). These measures have shown good validity and patient acceptability when conducted by health care professionals.[4,5] However, these clinical studies have generally required participants to return to the clinical center for repeat testing or outcome assessment and therefore do not overcome the burden of face-to-face attendance at the health care facility. Limited data are available on the agreement and reliability of home-based self-assessment of common clinical outcome measures such as waist circumference[6-8] and blood pressure.[9-11]

This study aimed to determine the level of agreement and reliability of clinician-measured versus participant self-measured (ie, in an unsupervised, home-based setting) clinical and functional assessments commonly used for monitoring effectiveness of telehealth-delivered lifestyle interventions in liver transplant recipients (LTRs). This patient cohort has previously indicated a preference and motivation for flexible access to health care options including telehealth monitoring.[12] We hypothesized that LTRs would reliably self-assess clinical and functional outcomes with an acceptable level of agreement (ie, participant measures would be below a predetermined minimal clinically important difference at least 80% of the time for each outcome measure).

Materials and methods

This is a substudy from a larger randomized controlled feasibility study (35 participants randomized, with 27 participants completing the study) investigating telehealth-to-home delivered lifestyle intervention to enhance cardiometabolic health (ACTRN12617001260314). The study conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Metro South Human Research Ethics Committee (HREC/17/QPAH/208).

Participants were recruited from August to December 2017. Participants were adults (aged 18-70) under the care of the Queensland Liver Transplant Service, >6 months posttransplant, expected survival >1 year, living within 100 km of the hospital (primary care center). Participants were required to have current access to a mobile phone or computer hardware with internet access and capabilities for webcam attachment. Volunteers were excluded if they reported having a dietary restriction that would make the dietary component of the parent study inappropriate, had a physical disability that would impair participation in physical activity, were deemed unsafe to participate in a lifestyle intervention by a hepatologist or transplant surgeon, or did not have sufficient English literacy. All participants provided written informed consent prior to participation.

Procedure

Trained health professionals (exercise physiologists [n=2] and dietitians [n=4]) performed clinical and functional assessments at a baseline face-to-face clinic appointment. Baseline study appointments were attended for the purpose of research and not due to clinical follow-up. Assessments were performed by different health professionals from the multidisciplinary research team to replicate a real-world clinical setting. Participants were directed to undertake self-measured clinical and functional assessments at home, unsupervised, within 1 week of baseline clinic assessment. The 1-week timeframe was designed to minimize the effect of time on differences in repeat measures. Participants were instructed on how to perform each assessment at home and were provided with written instructions and links to video-recorded tutorials to view online. Equipment for participants to conduct all functional measures and waist circumference at home were provided (listed below). Results were recorded and sent to investigators via email, or verbally transcribed at the next telehealth appointment. Participants received up to 3 phone call reminders to complete the assessments. The clinician-measured face-to-face assessments were compared with the patient-at-home (unsupervised) assessments to determine agreement and reliability.

Outcome measures were chosen to reflect the pragmatic needs of real-world clinical telehealth practice where reliable, accessible, inexpensive measures of metabolic risk need to be longitudinally monitored to assess the effectiveness of telehealth-delivered lifestyle interventions.

Clinical measures

Body mass

Clinician-assessed body weight was recorded to the nearest 0.5 kg (Robusta 813[a]). Participants used their own personal scales to record weight (various brands, not recorded).

Waist circumference

Clinician-assessed waist circumference was measured midway between the lower rib margin and iliac crest with stomach muscles relaxed, to the nearest 1 cm. Identical tape measures were provided to all participants for the repeat home measure with written and pictorial instructions to take the measure at the same site.

Blood pressure

Clinician-assessed systolic and diastolic blood pressure (SBP and DBP) were performed seated using Welch Allyn Cerner Vital Signs Monitor[b] with an appropriate size cuff. An average of 3 measures was taken to the nearest 1 mmHg. Participants were instructed to either use home blood pressure machines or attend a local pharmacy to have blood pressure performed by a pharmacist (various brands, not recorded). Home- or pharmacy-based blood pressure was taken twice on the same arm, with at least 1 minute between readings after sitting quietly for a minimum of 5 minutes. If a >5 mmHg difference between the first and second readings was observed, a third measure was taken. Participants reported all BP results and investigators calculated the average. Participants were instructed to avoid caffeine, exercise, and smoking for 30 minutes prior to measurement.

Functional tests

6MWT

The 6MWT was conducted on a 10-m track within an indoor corridor after a standardized protocol.[13] A premeasured piece of string marked the 10-m track and turn around points. The trained clinician used a lap counter to count the number of laps that the participant completed within 6 minutes and calculated the number of meters walked to the nearest 10 m. Participants practiced using the lap counter during the baseline test and were provided with the string, lap counter, and stopwatch for home testing.

Repeated STST

The repeated STST provides a measure of functional leg power and strength and was conducted according to a standardized protocol.[13] The test required a straight-back armless chair of standard height (42cm, recommended for both clinic- and home-based measures) placed firmly against a wall. After performing an initial single chair stand with arms folded across the chest and feet flat on the floor, the time to complete 5 repeat chair stands was recorded in seconds. A stopwatch was provided to participants who used a suitable chair in their home for all home-based tests.

Maximal push-up test

The maximal push-up test is a measure of upper body strength and endurance. For men, the test was conducted in the full push-up position (hands a comfortable distance apart, back straight, and using the toes as the pivot point) and for women the test was conducted in the modified

Table 1 Baseline clinical and functional characteristics of participants who did and did not complete self-assessments at home

Characteristic	Completed Home Retest (n=18)	Did Not Complete Home Retest (n=17)	*P* Value
Age (y)	51.7±13.9	49.7±15.6	.7
Sex (n)			
Women	8	2	.03*
Men	10	15	
BMI (kg/m^2)	27.5±7.7	27.4±5.2	>.99
Weight (kg)	78.6±20.3	87.7±17.5	.17
Waist circumference (cm)	96±15	101±16	.35
SBP (mmHg)	137±17	132±17	.42
DBP (mmHg)	84±9	81±9	.43
6MWT (m)†	466±80	471±93	.87
STST (s)‡	10 (8-16)	12 (9-16)	.04*
Push-ups (n)‡	8±6	10±11	.48

NOTE. Data are presented as means ± SD or medians (range, min to max). All data presented are assessments completed by the clinician at the baseline face-to-face appointment. An independent sample *t* test was used to compare the 2 groups.
Abbreviation: BMI, body mass index.
* *P*<.05.
† n=16 completers and n=19 noncompleters.
‡ n=17 completers and n=18 noncompleters.

push-up posture (as per men, but with the lower legs together in contact with the ground and ankles plantar flexed), according to a standardized protocol.[13] A marker (standard sized can of food) was placed on its end below the head of the participant to mark the range of motion required for each push-up. Prior to the test, participants completed up to 3 repetitions to ensure correct technique and then rested for 1 minute before completing the test. Participants then completed the maximum number of push-ups possible with good technique, consecutively without rest. Participants were given an exercise mat[c] and instructed to use a can of equal size to complete their home-based test.

Statistical analysis

Data analyses was performed using SPSS 25.[d] Data were verified for normal distribution via the Shapiro-Wilks test, the Kolmogorov-Smirnov test, and visualization of Q-Q plots. Patient characteristic data were compared between those who did and did not complete the home assessments using an independent *t* test (numerical data) or chi-square test (categorical data). A 1-sample *t* test was conducted to determine whether there were statistically significant differences between the clinician-measured assessments and participant home assessment for each clinical and functional test variable, with a test value of 0. Agreement between clinician and study participant assessments was determined according to Bland-Altman plots. The Bland-Altman plots show mean difference and limits of agreement between the 2 methods, with limits of agreement calculated as mean difference ± 1.96 standard deviation. Bland-Altman analyses are reported as mean difference (95% confidence interval) and limits of agreement (lower limit of agreement [LoA]-upper LoA). The mean difference was calculated as clinician measurement minus participant measurement, and thus a positive mean difference indicated that the participant-measured value was lower than the clinician-measured value. Systematic bias was assessed by the 1-sample *t* test to determine if the mean difference was significantly different from 0 and/or if the line of equality (y=0) was outside the 95% confidence interval of the mean difference. Proportional bias was assessed by linear regression to determine whether the average of the measures (independent variable) was significantly related to the difference between the clinician and participant measures (dependent variable).

The extent to which measurements can be replicated between the clinician and the participant was further assessed via intraclass correlation coefficients (ICCs) using a 2-way mixed model with absolute agreement. ICCs were interpreted as poor (<0.5), moderate (0.5-0.74), good (0.75-0.9), and excellent (>0.9).[14] With 2 observations per participant (ie, the clinician assessment and the participant assessment), a minimum sample size of n=13 would be required to achieve the statistical significance, based on ICCs of 0.70, for an alpha value set at 0.05 and a minimum power of 90%.[15]

Determining clinically meaningful agreement

In addition to the statistical approaches, minimal clinically important differences (MCIDs) were used to interpret the level of clinically meaningful reliability and agreement for each assessment. Differences between clinician-measured and participant-measured scores below the MCID were judged as representing an acceptable level of agreement. As recommended by Revicki et al,[16] multiple methods (anchor and distribution methods) were used to determine the MCID for each measure, in particular where existing literature was sparse. Accordingly, for this study, a clinically meaningful agreement (ie, a difference lower than the *a priori* MCID) for each variable was defined as a difference less than weight ≤3%, waist ≤2%, SBP ≤10 mmHg,

Table 2 Statistical, anchoring, and distribution methods to determine agreement between clinician-measured and participant-measured outcome assessments

Outcome Measures (units)	Clinician Measures	Participant Measures	Mean Diff (95% CI) [Lower LoA-Upper LoA]	*P* Value*	ICC (95% CI)	Anchor Method for MCID	Distribution Method for MCID	Acceptable Limit for MCID	n (%) Within Acceptable MCID
Weight (kg)	78.6±20.3	77.9±20.3	0.7 (0.01-1.4) [−2.17 to 3.57]	.59	0.998 (0.995-0.999)	<3% considered not clinically meaningful change[17]	$^1/_2$ SD: <0.75 kg SEM: <0.07 kg	≤3%	16/18 (89)
Waist (cm)	96±15	96±14	0.4 (−1.2 to 2.0) [−5.86 to 6.57]	.56	0.989 (0.970-0.996)	Intertester reliability should be <2%.[13,18] MCID=5% and WC maintenance=3%.[19] Mean coefficient of variation in measure between health professionals=1.5%[20]	$^1/_2$ SD: <1.6 cm SEM: <0.33 cm	≤2%	12/18 (67)
SBP (mmHg)	137 (102-164.5)	129 (100-156)	7.7 (0.6-14.7) [−19.35 to 34.85]	.04	0.768 (0.348-0.917)	SBP should be retested if readings differ >10 mmHg.[13]	$^1/_2$ SD: <6.85 mmHg SEM: <6.60 mmHg	≤10 mmHg	11/17 (65)
DBP (mmHg)	84±9	81±13	2.4 (−1.4 to 6.2) [−12.16 to 16.97]	.20	0.872 (0.657-0.953)	DBP should be retested if readings differ >6mmHg.[13]	$^1/_2$ SD: 3.7 mmHg SEM: <2.65 mmHg	≤5 mmHg	8/17 (47)
STST (s)	10 (8-16)	10 (5-16)	0.5 (−0.8 to 1.7) [−4.31 to 5.29]	.42	0.733 (0.269-0.903)	MCID=1.7 s in patients with COPD.[21] MCID=2.5 s (95% LoA = −2.6 to 2.6s) in older female adults.[22]	$^1/_2$ SD: <1.2 s SEM: <1.24 s	≤1.5 s	6/17 (35)
6MWT (m)	466±80	459±98	7.5 (−29.1 to 44.1) [−127.3 to 142.4]	.67	0.934 (0.521-0.942)	MCID=36 m in patients with chronic heart failure.[23] MCID=43.1 (16.8) m and 95% CI, 31.8-54.4 m in patients with chronic heart failure.[24] MCID=25 m in patients with coronary artery disease.[25]	$^1/_2$ SD: < 34.45 m SEM: <17.7 m	≤30 m	6/16 (38)

(*continued on next page*)

Table 2 (*continued*)

Outcome Measures (units)	Clinician Measures	Participant Measures	Mean Diff (95% CI) [Lower LoA-Upper LoA]	*P* Value*	ICC (95% CI)	Anchor Method for MCID	Distribution Method for MCID	Acceptable Limit for MCID	n (%) Within Acceptable MCID
Push-ups	10 (11)	13 (11)	−2.2 (−4.4 to −0.1) [−10.48 to 6.01]	.04	0.953 (0.845-0.985)	MCID = 2 in healthy recreationally active adults aged 18-45, with SEM = 1.[26] Mean difference (95% CI) = 3 (2.1-3.9) in healthy adults aged 37-57 with SEM = 2.6.[27]	$^1/_2$ SD: <3.0 push-ups SEM: <1.3 push-ups	≤2 push-ups	12/17 (71)

NOTE. $SEM = SD \times \sqrt{1-r}$, where r = ICC for the sample.[28]
Abbreviation: 95% CI, 95% confidence interval.
* *P* value from 1-sample *t* test.

DBP ≤5 mmHg, 6MWT ≤30 m, push-ups ≤2, and STST ≤1.5 seconds. An outcome measure was considered acceptable overall if the limits of agreement fell below the predetermined MCID. A self-assessed unsupervised outcome measure was also deemed acceptable overall if study participants were within the limits of clinically meaningful agreement at least 80% of the time.

Results

Thirty-five participants (71% men, mean age: 50.7±14.6y, mean body mass index: 27.4±6.5kg/m^2) were recruited to the parent study, with all participants consenting to undertake the repeated home measures. Of the 35 participants who had baseline face-to-face clinical and functional assessments with the clinician, 18 (51%) performed repeat baseline self-assessments at home. The mean time between clinician measures and participant home retest assessments was 12.6±9.5 days. There were no significant differences in baseline characteristics and physical function for those who did and did not complete self-assessments at home, except for sex and STST. A greater proportion of women completed the home retest than men (80% vs 40%; $P = .03$). Higher STST completion time, indicating reduced function, was recorded in those completing home tests than those who did not (10s [8-16s] vs 12s [9-16s]; $P = .04$) (table 1).

Reliability for assessments between clinician and participant measurement was excellent for weight and waist circumference (ICC>0.9), good for DBP (ICC = 0.87) and SBP (ICC = 0.77) (table 2). There were no differences between the mean clinician-measured scores and mean participant-measured scores for weight ($P = .59$), waist ($P = .56$), and DBP ($P = .20$); however, SBP was significantly lower ($P = .04$) for the participant-measured score than that of the clinician. For all clinical outcomes, participants measured lower scores for their clinical measures on average when compared with clinician-measured values (see table 2). Moreover, the limits of agreement for all clinical measures fell outside the MCID, demonstrating wide individual variability for self-assessed reliability (figs 1A-D). SBP demonstrated the greatest variability (lower LoA (−19.35 mmHg) and upper LoA [34.58 mmHg]) with wide 95% confidence intervals for interrater reliability (ICC = 0.35-0.92) indicating large variability (see table 2) (fig 1C). Bland-Altman analysis also demonstrated systematic bias for SBP whereby participant measures were systematically lower than clinician measures (7.7mmHg [0.6-14.7]; $P = .04$). Linear regression of the Bland-Altman analysis showed proportional bias in DBP ($r = 0.56$, $P = .02$) whereby the higher the DBP reading, the poorer the agreement in the measures between the clinician and the participant. No other proportional bias was observed ($r = 0.31$, $r = 0.30$, $r = 0.23$, $P > .05$ for weight, waist, and SBP, respectively). Weight was the only clinical measure that was measured by the participant within the acceptable level of agreement at least 80% of the time (89%, 16/18 participants) (see table 2 and fig 1A).

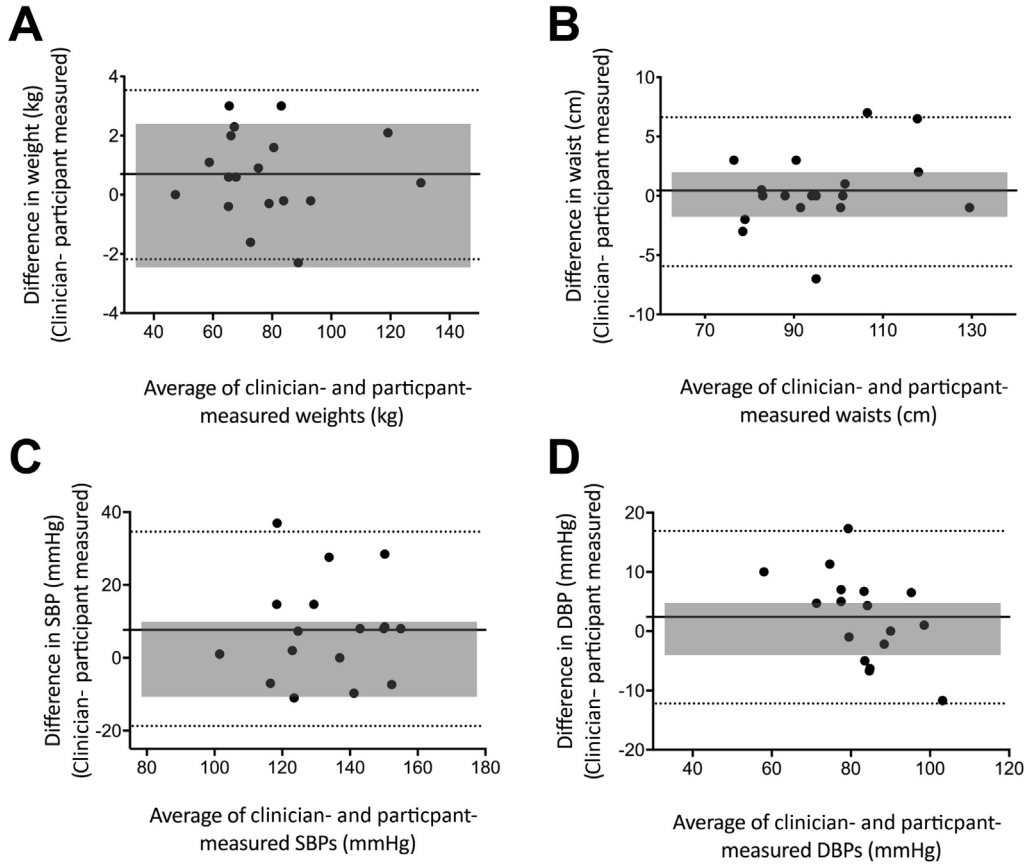

Fig 1 Bland-Altman plots of the difference between face-to-face clinician-measured and participant self-assessments performed at home versus the mean of clinician-measured and participant-measured assessments for (A) weight, n=18; (B) waist, n=18; (C) SBP, n=17; and (D) DBP, n=17. Solid line indicates the mean difference, and dashed lines indicate the upper and lower limits of agreement. Shaded regions represent the clinically acceptable limits of agreement.

Reliability for assessments between clinician- and participant-measured 6MWT and push-ups were excellent (ICC>0.9) and good for STST (ICC=0.73) with wide 95% confidence intervals indicating large variability (see table 2). There were no significant differences between the mean clinician-measured scores and the mean participant-measured scores for 6MWT or STST ($P>.05$); however, the mean number of push-ups measured by participants was systematically higher than that measured by clinicians (−2.2 [−4.4, −0.1]; $P=.04$). On average, participants measured lower distances for their 6MWT and STST (see table 2). The limits of agreement for all functional tests fell outside the MCID (figs 2A-C), demonstrating high individual variability in agreement for each functional measure, with the 6MWT demonstrating the greatest variability (lower LoA to upper LoA: −127.3 to 142.4m) (see fig 2A). Linear

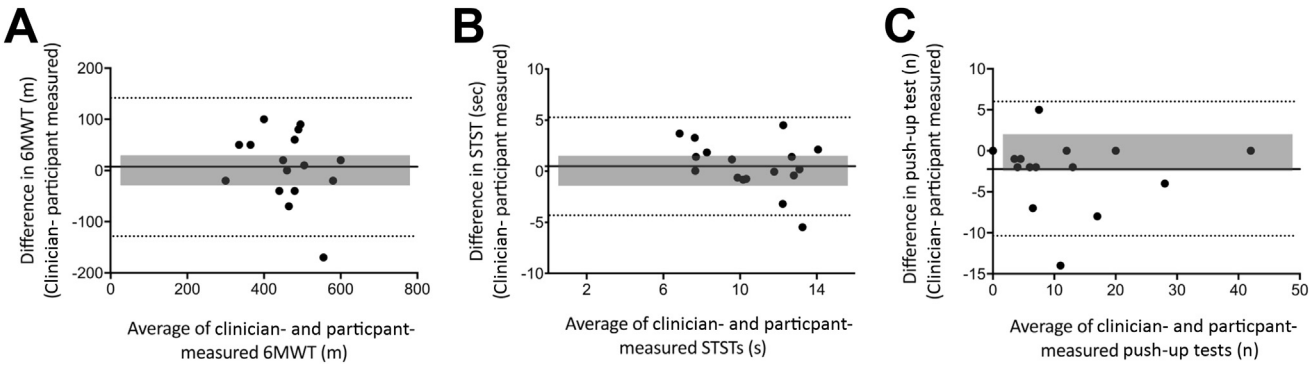

Fig 2 Bland-Altman plots of the difference between face-to-face clinician-measured and participant self-assessments performed at home versus the mean of clinician-measured and participant-measured assessments for (A) 6MWT, n=16; (B) STST, n=17; and (C) push-up test n=17. Solid line indicates the mean difference, and dashed lines indicate the upper and lower limits of agreement. Shaded regions represent the clinically acceptable limits of agreement.

regression of the Bland-Altman analysis did not find proportional bias in any functional test ($r=0.27$, $r=0.38$, $r=0.10$, $P>.05$, for 6MWT, STST, push-up test, respectively). No functional test was determined to be acceptable; that is, none were conducted with ≥80% of participants' scores within the MCID (see table 2).

Discussion

This study evaluated the level of agreement and reliability of participant-measured home-based (unsupervised) assessments for use as outcome measures for remote monitoring in telehealth lifestyle interventions. We employed a multipronged approach to establish test-retest reliability (using ICCs), levels of agreement (using Bland-Altman plots and 1-sample *t* test), and whether participants were able to measure clinical and functional outcomes within a clinically acceptable range (based on MCID determined via distribution and anchoring methods). Generally, based on ICCs and mean differences, the reliability between assessments performed by participants and clinicians was moderate to excellent; however, there was notable between-individual variability with wide limits of agreement. Bland-Altman plots define the limits of agreement but do not determine whether these limits are acceptable or not for the clinical purpose of the measure.[29] Therefore, *limits of acceptability* were determined *a priori* based on MCIDs that were derived from clinically and analytically relevant criteria. The self-assessments were further explored by comparing the limits of agreement of each measure with the MCID and the percentage of participants whose self-assessment measures were within the acceptable range.

Contrary to our hypothesis, despite the moderate to excellent intertester reliability, all limits of agreement were outside of the clinically acceptable range. Only body weight was retested by participants within clinically acceptable limits of agreement >80% of the time. Overall, the mean differences between clinician- and patient-measured outcomes seemed acceptable. However, for some participants, there were clinically significant differences between the results obtained by clinicians and participants that could affect clinical decision making for that participant. It is important to be attentive to the effect of this variability when making clinical decisions using outcomes obtained by patients. Our findings highlight the need for significant methodological consideration regarding the selection and implementation of appropriate, feasible, and reliable outcome measures for self-assessment in telehealth-to-home interventions and training of assessors.

Accurate and reliable blood pressure monitoring is notoriously difficult in both clinical and community settings,[30,31] and this study reinforces the numerous barriers to remote blood pressure monitoring including access to robust home BP monitoring equipment, biological variability, and patients understanding of, and compliance with, strict pretest criteria and instructions. Although SBP had the lowest intertester reliability with a significant difference between the average clinician- and participant-measured scores, clinician-measured blood pressure has also been to shown to be problematic.[32,33] Although clinician-measured cuff BP is the reference standard for managing high BP, it is widely acknowledged that it is not a criterion standard measure[31,32] and may be affected by circadian variability and *white coat effects*. Although 24-hour ambulatory monitoring is accepted as the criterion standard for clinical judgement regarding BP,[31] it is unlikely feasible for pragmatic study designs in large clinical studies or real-world practice. To reduce variability in home-based BP measures, researchers and clinicians should ensure standardization of methods (eg, measures taken at the same time of day, at the same location, with the same pretest instructions for known influences of BP including caffeine and physical activity).

Body weight and waist circumference were conducted with the highest level of agreement, which is consistent with the literature.[17] However, only body weight was determined to be acceptable based on comparisons with MCID. For the 6MWT, despite the high relative reliability determined by ICCs, which is consistently reported in the literature,[34,35] the limits of agreement were broad. Thus, although the 6MWT may demonstrate high statistical intertester reliability, participants were only within clincially acceptable agreement with clinician measures 38% of the time. Interestingly, a recent study by Hwang et al[36] examined the reliability of clinicians to measure supervised face-to-face 6MWT assessment versus supervised telehealth 6MWT assessment and found that, although ICCs were high (ICC=0.90), the limits of agreement (−84 to 76m) were above the predetermined clinically acceptable range (±36m) in patients with chronic heart failure.[36] The authors also observed that clinician test-retest measures were outside the clinically acceptable range 29% of the time. Together, our studies suggest that the high variability in reliability of the 6MWT may not be limited to participant skill alone. The source of variation in both clinic- and home-based measurements may be influenced by numerous factors including biological variation, measurement variation, skill or technique, or psychological factors such as motivation and self-efficacy. LTRs have a lifelong relation with their specialist care center and have expressed a desire for remote monitoring via telehealth services.[12] Thus, there is bidirectional motivation for clinicians and patients to explore enablers and barriers to reliable home self-assessments to monitor longitudinal outcomes remotely from the clinic using telehealth-to-home innovations. Although the 6MWT may be a pragmatic choice of outcome for lifestyle interventions, there is a need for more training and facilitation for patients to improve self-assessment feasibility for telehealth use. Although adding to resource burden, supervision of the 6MWT via telehealth may improve reliability.[36] Promoting to the patients the value of self-assessment and remote monitoring of health outcomes may also facilitate improved telehealth interactions between stakeholders.

Study limitations

Despite all necessary equipment and resources provided, only half of the participants performed home assessments, and the mean time difference between clinician assessment and participant measures in those who did complete

the home assessments (12±9.5d) exceeded the requested 1-week timeframe. This indicates that there are additional barriers to completing home self-assessment. Although it is unlikely that any clinically relevant changes to outcomes would have occurred in this timeframe, real-world clinical feasibility is lowered, irrespective of the reliability of those who did perform the measures. Patient activation and engagement with remote monitoring technology is critical and remains an ongoing challenge for telehealth monitoring services. Further methodological strategies to enhance compliance with self-assessment, as well as exploration of barriers and enablers to self-assessment, are required to increase the uptake of unsupervised home assessment within telehealth interventions. Our interpretation of agreement and reliability for all outcome measures was also limited by the lack of evidence to inform MCID for outcome measure, specifically in LTRs.

Finally, given the known sensitivity of blood pressure measures to day-to-day environmental factors, participants were asked to avoid caffeine, exercise, and smoking for 30 minutes prior to the measure and to sit quietly for 5 minutes prior to measures. However, because of the pragmatic design of the study, we were unable to ensure that participant blood pressure measures were undertaken at the same time of day and were unable to detect possible white coat effects. Identifying the source of variation for each measure was beyond the scope of this study.

Conclusion

This study determined that overall, patient self-assessed clinical and functional outcome measures for metabolic health and fitness had good agreement and reliability on average with face-to-face clinician-assessed outcome measures in LTRs. However, at an individual level, there was considerable variation. With the current methodologies, aside from body weight, no clinical or functional outcome was deemed acceptable when compared with MCIDs. Moreover, only 51% of LTRs in the study undertook self-assessed measures at home. Clinicians wishing to remotely monitor health outcomes from lifestyle-based interventions in chronic disease via telehealth may need to consider greater standardization, and to provide training and facilitation to patients to improve the uptake and reliability of patient home-based health assessments.

Suppliers

a. Robusta 813; Seca.
b. Welch Allyn Cerner Vital Signs Monitor; Hillron.
c. Exercise mat; Kmart Australia Limited.
d. SPSS 25; IBM.

Corresponding author

Shelley E. Keating, PhD, School of Human Movement and Nutrition Sciences, The University of Queensland, St Lucia, QLD 4072, Australia. *E-mail address:* s.keating@uq.edu.au.

Acknowledgment

We thank Metro South Health Biostatistics, Princess Alexandra Hospital, Brisbane, for their statistical support.

References

1. Bashshur RL, Shannon GW, Smith BR, et al. The empirical foundations of telemedicine interventions for chronic disease management. Telemed J E Health 2014;20:769-800.
2. Hutchesson MJ, Rollo ME, Krukowski R, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015;16:376-92.
3. Williams ED, Bird D, Forbes AW, et al. Randomised controlled trial of an automated, interactive telephone intervention (TLC Diabetes) to improve type 2 diabetes management: baseline findings and six-month outcomes. BMC Public Health 2012;12:602.
4. Veloso-Guedes CA, Rosalen ST, Thobias CM, et al. Validation of 20-meter corridor for the 6-minute walk test in men on liver transplantation waiting list. Transplant Proc 2011;43:1322-4.
5. Bohannon EW, Bubela DJ, Magasi SR, Wang Y, Gershon RC. Sit-to-stand test: performance and determinants across the age-span. Isokinet Exerc Sci 2010;18:235-40.
6. Kushi LH, Kaye SA, Folsom AR, Soler JT, Prineas RJ. Accuracy and reliability of self-measurement of body girths. Am J Epidemiol 1988;128:740-8.
7. Bigaard J, Spanggaard I, Thomsen BL, Overvad K, Tjønneland A. Self-reported and technician-measured waist circumferences differ in middle-aged men and women. J Nutr 2005;135:2263-70.
8. Han TS, Lean ME. Self-reported waist circumference compared with the 'Waist Watcher' tape-measure to identify individuals at increased health risk through intra-abdominal fat accumulation. Br J Nutr 1998;80:81-8.
9. Warren RE, Marshall T, Padfield PL, Chrubasik S. Variability of office, 24-hour ambulatory, and self-monitored blood pressure measurements. Br J Gen Pract 2010;60:675-80.
10. Stergiou GS, Baibas NM, Gantzarous AP, et al. Reproducibility of home, ambulatory, and clinic blood pressure: implications for the design of trials for the assessment of antihypertensive drug efficacy. Am J Hypertens 2002;15(2 Pt 1):101-4.
11. Ragot S, Genès N, Vaur L, Herpin D. Comparison of three blood pressure measurement methods for the evaluation of two antihypertensive drugs: feasibility, agreement, and reproducibility of blood pressure response. Am J Hypertens 2000;13(6 Pt 1):632-9.
12. Hickman IJ, Coran D, Wallen MP, et al. 'Back to Life'- using knowledge exchange processes to enhance lifestyle interventions for liver transplant recipients: a qualitative study. Nutr Diet 2019;76:399-406.
13. Coombes J, Skinner T. ESSA's student manual for health, exercise and sport assessment. Chatswood, Australia: Elsevier; 2014.
14. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63.
15. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Arch Orofac Sci 2017;12:1-11.
16. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important

differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102-9.
17. Stevens J, Truesdale KP, McClain JE, Cai J. The definition of weight maintenance. Int J Obes 2006;30:391-9.
18. Stewart A, Marfell-Jones M. International Society for the Advancement of Kinanthropometry (ISAK) International Standards for Anthropometric Assessment. New Zealand: Lower Hutt; 2011.
19. Verweij LM, Terwee CB, Proper KI, Hulshof CTJ, van Mechelen W. Measurement error of waist circumference: gaps in knowledge. Public Health Nutr 2013;16:281-8.
20. Panoulas VF, Ahmad N, Kassamali RH, Nightingale P, Kitas GD. The inter-operator variability in measuring waist circumference and its potential impact on the diagnosis of the metabolic syndrome. Postgrad Med J 2008;84:344-7.
21. Jones SE, Kon SSC, Canavan JL, et al. The five-repetition sit-to-stand test as a functional outcome measure in COPD. Thorax 2013;68:1015-20.
22. 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;24:339-44.
23. Tager T, Hanholz W, Cebola R, et al. Minimal important difference for 6-minute walk test in individuals with chronic heart failure. Int J Cardiol 2014;176:94-8.
24. Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG. Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. Cardiopulm Phys Ther J 2013;24:21-9.
25. 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;92:611-9.
26. Negrete RJ, Hanney WJ, Kolber MJ, et al. Reliability, minimal detectable change, and normative values for tests of upper extremity function and power. J Strength Cond Res 2010;24:3318-25.
27. Suni JH, Oja P, Laukkanen RT, et al. Health-related fitness test battery for adults: aspects of reliability. Arch Phys Med Rehabil 1996;77:399-405.
28. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil 2005;84:719-23.
29. Giavarina D. Understanding Bland Altman analysis. Biochemia Med 2015;25:141-51.
30. Kallioinen N, Hill A, Horswill MS, Ward H, Watson MO. Sources of inaccuracy in the measurement of adult patients' resting blood pressure in clinical settings: a systematic review. J Hypertens 2017;35:421-41.
31. Staessen JA, Li Y, Hara A, Asayama K, Dolan E, O'Brien E. Blood pressure measurement anno 2016. Am J Hypertens 2017;30:453-63.
32. Sharman JE, Marwick TH. Accuracy of blood pressure monitoring devices: a critical need for improvement that could resolve discrepancy in hypertension guidelines. J Hum Hypertens 2018;33:89-93.
33. Moore MN, Schultz MG, Nelson MR, et al. Identification of the optimal protocol for automated office blood pressure measurement among patients with treated hypertension. Am J Hypertens 2018;31:299-304.
34. Uszko-Lencer NHMK, Mesquita R, Janssen E, et al. Reliability, construct validity and determinants of 6-minute walk test performance in patients with chronic heart failure. Int J Cardiol 2017;240:285-90.
35. Hanson LC, McBurney H, Taylor NF. The retest reliability of the six-minute walk test in patients referred to a cardiac rehabilitation programme. Physiother Res Int 2012;17:55-61.
36. Hwang R, Mandrusiak A, Morris NR, Peters R, Korczyk D, Russell T. Assessing functional exercise capacity using telehealth: is it valid and reliable in patients with chronic heart failure? J Telemed Telecare 2017;23:225-32.