

RESEARCH ARTICLE

Reproducibility of measurements on physical performance in head and neck cancer survivors; measurements on maximum mouth opening, shoulder and neck function, upper and lower body strength, level of physical mobility, and walking ability

Gerben van Hinte^{1*}, Ruud A. Leijendekkers^{1,2,3}, Bram te Molder⁴, Lizzy Jansen⁴, Corinda Bol⁴, Matthias A. W. Merkx⁵, Robert Takes⁶, Maria W. G. Nijhuis-van der Sanden³, Caroline M. Speksnijder^{5,7,8}

1 Department of Rehabilitation, Radboud University Medical Center, Nijmegen, The Netherlands, **2** Orthopaedic Research Laboratory, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands, **3** Radboud Institute for Health Sciences, IQ Healthcare, Radboud University Medical Center, Nijmegen, The Netherlands, **4** Research Group Musculoskeletal Rehabilitation, HAN University of Applied Sciences, Nijmegen, The Netherlands, **5** Department of Oral and Maxillofacial Surgery, Radboud University Medical Center, Nijmegen, The Netherlands, **6** Department of Otorhinolaryngology, Head and Neck Surgery, Radboud University Medical Center, Nijmegen, The Netherlands, **7** Department of Head and Neck Surgical Oncology, University Medical Center Utrecht Cancer Center, Utrecht University, Utrecht, The Netherlands, **8** Department of Oral and Maxillofacial Surgery and Special Dental Care, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands

* gerben.vanhinte@radboudumc.nl

Abstract

Background

Survivors of Head and Neck Cancer experience specific problems in functional performance. The aim of this study was to obtain the test-retest reliability of measurements on Maximal Mouth Opening (MMO), shoulder and neck function, lower and upper body strength, level of mobility and walking ability.

Materials and methods

Test-retest study design. Measurements on MMO (intra- and extra orally), Active range of motion of shoulders and neck, 30 Seconds Chair Stand Test, Grip Strength, Timed Up and Go test, and Six Minute Walk test.

Results

In total 50 participants were included. The mean age was 68.6. ± 9.9 years and median time since end of treatment was 3.0 years (Q1–Q3: 1.0–5.25 years). We found good to excellent test-retest reliability on the core set of measurements (Intraclass Correlation Coefficient (ICC) 0.77 to 0.98). Measurement of MMO with cardboard card, forward flexion shoulder and Six Minute Walk test had a relatively small measurement error (Smallest Detectable



OPEN ACCESS

Citation: van Hinte G, Leijendekkers RA, te Molder B, Jansen L, Bol C, Merkx MAW, et al. (2020) Reproducibility of measurements on physical performance in head and neck cancer survivors; measurements on maximum mouth opening, shoulder and neck function, upper and lower body strength, level of physical mobility, and walking ability. PLoS ONE 15(9): e0233271. <https://doi.org/10.1371/journal.pone.0233271>

Editor: Peter Dziegielewski, University of Florida, UNITED STATES

Received: February 13, 2020

Accepted: July 26, 2020

Published: September 3, 2020

Copyright: © 2020 van Hinte et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: GvH: This study was funded by the “Michel Keijzer Fonds” of the Dutch patient federation for head and neck cancer survivors. There is no grant number available for this grant. Website fund: <https://pvhh.nl/informatie/>

[organisatie/michel-keijzer-fonds/](#). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Change (SDC) % 5.4% - 15.1%). Measurement of MMO with a caliper, shoulder abduction, shoulder external rotation, later flexion and rotation of the neck, grip strength, 30 Seconds Chair Stand Test, and Timed up and Go test had a relatively large measurement error (SDC % 19.8% - 44.7%).

Conclusion

This core set of measurements on physical performance is found reliable and therefore able to differentiate in physical performance. The reported measurement errors should be taken into consideration when interpreting the results of repeated measurements.

Implications for cancer survivors

A core set of physical measurements can be used to measure physical performance in survivors of Head and Neck Cancer.

Introduction

Curative treatment of Head and Neck Cancer (HNC) may consist of surgery, radiotherapy, chemotherapy, or a combination of these treatments. The choice and extent of treatment is influenced by tumor size and cervical lymph node involvement expressed in TNM-status [1]. Survivors of HNC (sHNC) commonly experience treatment-related morbidity that impairs their physical, social, emotional, and psychological performance [2, 3].

Local morbidity can be related to alterations in the functional anatomy and physiology of the head and neck. Local limitations can occur in the Maximum Mouth Opening (MMO) and other oral functions (speech, swallowing) [4–7]. Consequently, a decrease in oral function is associated with malnutrition, which is an important outcome factor for recovery and survival rate [8–10]. Semi-regional morbidity can be found in decreased active range of motion (AROM) of shoulders and neck, as well as a decrease in upper body strength [11–13]. The etiology of limitations in neck and shoulder function is multifactorial and lies in a combination of nerve and soft tissue damage and a change in movement patterns due to pain and shoulder disuse mostly related to surgery and radiotherapy [11, 13]. Neck and shoulder problems in sHNC have a high incidence and can pose severe problems during activities in daily life and participation [11].

General morbidity can concern cancer-related fatigue [14], a lower level of physical mobility, and decreased walking ability which limits return to work and daily activities [15]. Treatment related morbidity may be caused by surgery due to resection, reconstruction, neck dissection (ND), by radiotherapy causing fibrosis, skin problems, mucositis, or by systemic responses of chemotherapy [16]. Local, semi-regional and general morbidity lead to a decrease in functional performance in sHNC resulting in limitations in daily activities and difficulty returning to work, which subsequently negatively influences Health-Related Quality of Life (HRQoL) [17, 18]. These findings, together with an increasing number of sHNC, reveal a clear need for rehabilitation interventions focusing on problems in the physical domain. In contrast to this, research shows that sHNC are mostly sedentary (> 50%) and very few participate in moderate or vigorous exercise [15]. However, during treatment 73% of the patients indicated the need for physiotherapy. After 8–11 years, 23% still indicate a need [19].

Several measurements provide insight into the limitations within the physical performance, such as MMO, shoulder and neck mobility, upper and lower body muscle strength, level of

mobility, and walking ability. Measurement methods on MMO vary and are performed both intra- and extra-orally [7, 20]. In cancer rehabilitation, a frequently used core set of measurements to objectify physical performance, consists of the measurement of AROM with goniometer or digital inclinometers, grip strength (GS) as proxy for upper body strength, the 30-second chair-to-stand test (30SCTS) for lower body strength, the Timed Up and Go test (TUG) for level of mobility, and the 6-Minute Walk Test (6MWT) for walking ability. This core set of physical performance measurements can be used in addition to Patient Reported Outcome Measurements (PROM's) on physical status. Insight in test-retest reproducibility of these instruments is important as it illustrates if measurements have the capacity to differentiate between sHNC when measured twice under the same conditions [21]. Insight into agreement parameters is important because it provides information on the Standard Error of Measurement (SEM) and the Smallest Detectable Change (SDC) which are essential for clinical interpretation of the (re)assessment of sHNC. Up to now, this core set of measurements was primarily studied on reliability in other patient populations or included in a case mix of sHNC and HNC patients still undergoing treatment [22–25]. Therefore, this study aims to examine the reliability, by investigating test-retest reproducibility, SEM, and SDC, of a core set of measurements on physical performance in sHNC.

Methods

Study setting and participants

Two subgroups participated in this cross-sectional study. Between January and June 2018, the first group of sHNC was recruited by convenience sampling from three regional patient support groups of the Dutch Head and Neck Oncology patient federation (regional support groups: Nijmegen, West-Brabant and Centre of Holland). Between March and June 2019, the second group was recruited from sHNC scheduled for usual care follow-up appointments at the Radboud university medical center. Inclusion criteria were: sHNC, completed medical treatment, 18 years or older, and able to walk unaided.

sHNC that were not able to speak or understand Dutch, patients receiving palliative care, and patients at risk when performing physical measurements were excluded. The safety and possible risk when performing physical measurements was assessed before inclusion, using a modified Physical Activity Readiness Questionnaire (PARQ), leading to the exclusion of willing participants who answered both yes to one or more out of seven questions and were judged on these items by their general practitioner to be unfit or unsafe for exercise [26, 27].

Sample size calculation

An a-priori sample size calculation was conducted following the recommendations of Donner & Eliasziw [28]. With a more than acceptable intraclass correlation coefficient (ICC) of 0.80, an level of significance of 0.05, and power of 0.8 ($\beta = 0.2$) it was established that 45 participants were required in the final analysis. It was anticipated that approximately 10% would drop out for motivational or practical reasons. Thus, the goal became including at least 50 patients in total. This number is sufficient to achieve a score of good on adequate sample size conform the COSMIN checklist [29]. The COSMIN checklist can be used to evaluate the methodological quality of studies on measurement properties of health status measurement instruments.

Study procedure

Members of the Dutch Head and Neck Oncology patient federation attended a presentation about the research project during a regular federation meeting. If interested, they received the

patient information brochure. Before their follow-up appointment, the usual care follow-up group was contacted by telephone to inform about the study and send the patient information brochure. The week following the presentation or phone call, both groups were contacted by telephone to determine if there were any questions and acquire verbal informed consent. Participants then received the PARQ digital questionnaire using Castor (Ciwit BV, Amsterdam, The Netherlands) electronic data capture (EDC) program (<http://www.castoredc.com>). The measurements took place at the physical therapy department of the Radboud university medical center. Prior to the physical measurements written consent was obtained. The study was conducted according to the principles of the Declaration of Helsinki (64th version, October 19th, 2013). The protocol (NL2017-3508) was approved by the Ethics Committee of the Radboud university medical center. This study followed the COSMIN checklist to ensure methodological and statistical quality and reduce bias [29].

Measurements

The patient's demographic and clinical data including age, sex, body weight, body height, smoking status (yes/ no/ history of smoking, packyears), alcohol usage (yes/no, number of units daily), level of education (lower, middle, higher), social status (living alone, living with partner), years since completion of medical intervention, tumor location (oral cavity, nasopharynx, oropharynx, larynx, other), treatment modality (surgery, radiotherapy, chemotherapy, or combinations of these), and neck dissection status (yes, unilateral/bilateral, no) were obtained using a custom patient reported questionnaire send by the electronic data capture software program Castor (see also Table 1). Measurements were performed in a standardized order and according to a standardized measurement protocol. The MMO was measured using two methods. Method one measured intra-orally with a cardboard ruler (TheraBite[®] Range of Motion Scale, Atos Medical Inc., New Berlin, Wisconsin, United States). Method two measured MMO extra-orally with a calibrated caliper (Electronic Digital Caliper 150 mm/6", Somultishop, Echt, Holland) following a previously described protocol [7]. Shoulder abduction and forward flexion were measured with a digital inclinometer (Baseline[®] Digital Inclinometer, Fabrication Enterprises Inc., White Plains, New York, USA) [30]. External rotation of the shoulder was measured with a goniometer (Universal goniometer, Mathys Synthes, Bettlach, Switzerland). The CROM (Cervical Range of Motion Instrument, Performance Attainment Associates, Lindstrom, Minnesota, USA) was used to measure the lateral flexion and rotation of the neck [31]. Grip strength was measured with a hand-held dynamometer (JAMAR[®], Sammons Preston Rolyan, Warrenville, Illinois, USA) [32]. The 30SCTS was used to examine lower body strength [33]. The level of mobility was measured with the TUG [24]. Walking ability was evaluated using a self-paced 6MWT on a 20-meter circuit [24].

Measurements were performed by physical therapy students who received intensive training. Measurements were supervised by an experienced physical therapist. The time interval between the test and retest measurement was at least one hour and maximal two hours. Test and retest were performed by the same physical therapy student. After the first test session, the data collection form was collected by the researcher to limit bias. In accordance with guidelines, during both the test- and retest session the 30SCTS and 6MWT were measured once, MMO and neck and shoulder function were measured twice, and GS and TUG were measured three times. For both test and retest measurement, the best score of each participant was used.

Statistical analysis

The demographic, personal, and treatment characteristics of the participants were described. Categorical data were presented as exact numbers and percentages were calculated. For the

Table 1. Demographic, participant, and treatment characteristics.

Characteristic	Total (n = 50)	SD	PCTL 25 th centile; 75 th centile	%
Sex				
Male, n	28			56
Female, n	22			44
Age (years), mean	68.6	9.9		
Body Mass index, median	25.0		23.5–26.7	
Smoking				
Yes, n	4			8
Pack-years, median	19		4.0–34.0	
No, but used to, n	39			78
Pack-years in history, median	20		9.0–31.0	
Never, n	7			14
Alcohol usage (>1 daily)				
Yes, n	22			44
Glasses per day, median	2		0.5–3.5	
No, n	28			56
Level of education				
Lower, n (%)	20			40
Middle, n (%)	17			34
Higher, n (%)	13			26
Social status				
Living alone, n (%)	16			32
Living with a partner, n (%)	34			68
Years since cancer treatment, median	3.0		1.0–5.25	
Tumor location				
Oral cavity, n (%)	28			56
Nasopharynx, n (%)	1			2
Oropharynx, n (%)	2			4
Larynx, n (%)	12			24
Other, n (%)	7			14
Oncology treatment				
Surgery, n (%)	19			38
Surgery and radiotherapy, n (%)	18			36
Radiotherapy, n (%)	4			8
Surgery, radiotherapy and chemotherapy, n (%)	7			14
Radiotherapy + chemotherapy, n (%)	2			4
Neck dissection				
Unilateral, n (%)	22			44
Bilateral, n (%)	6			12
No, n (%)	22			44

SD: standard deviation; PCTL: Percentile

<https://doi.org/10.1371/journal.pone.0233271.t001>

continuous data, means and standard deviations (SD) were calculated. Differences in MMO between the two measurement methods were tested with paired samples t-test in case of normally distributed data or Wilcoxon signed rank test for not normally distributed data. Reliability was divided into test-retest reproducibility and agreement parameters [34]. Test-retest reproducibility was tested using the intraclass correlation coefficient (ICC). ICC's were

calculated using a two-way mixed effect model ($ICC_{3.1_{agreement}}$) with absolute agreement and 95% confidence intervals (CI). Cut-off points for the ICC were chosen as poor (0.01–0.20), slight (0.21–0.51), fair (0.41–0.60), good (0.61–0.80), very good (0.81–0.92), and excellent (0.93–1.00) [35]. Both were expressed in the unit of measurement. SEM was calculated as $SEM_{agreement} = \sqrt{\sigma_{error}^2} = \sqrt{(\sigma_{2o} + \sigma_{2residual})}$ [36]. The variance due to systematic differences between measurements (σ_{2o}) and the residual variance ($\sigma_{2residual}$) was obtained from the varcomp analysis [36]. The $SEM_{agreement}$ was used to calculate the $SDC_{agreement} = 1.96 * \sqrt{n} * SEM$. In this formula, 'n' refers to the number of measurements, which was two in this study. Additionally, the SDC% was calculated as agreement outcome independent of the unit of measurement. The SDC% was calculated by dividing the SDC by the mean of the summed test and retest score, then multiplied by 100. For SDC% a 20% difference was set as cut off value for measurement error being relatively small (<20%) or large (>20%). Bland-Altman plots visualize the relationship between the measurement error and the observed value including the presence of systematic bias and bias related to the magnitude of the test outcome [37]. These plots show the test-retest difference (y-axis) against the mean of the first and second test outcomes (x-axis). Mean differences between the test and retest measurements were calculated with their standard deviations to calculate the 95% limits of agreement (95% LoA). In the plot, 95% LoA are shown (mean difference \pm 1.96 * SD of the difference). All analyses were performed using IBM SPSS Statistics v25 (SPSS, Inc., Chicago, Illinois, United States). In all cases, two-sided p-values smaller than 0.05 were considered to be statistically significant.

Results

In total 50 sHNC participated in the study, of which 29 were male and 21 were female. Fig 1 shows the flowchart of the recruitment and enrollment of participants. The mean age of participants was 69 years, with a standard deviation of 9.9. The median time of cancer survivorship was 3 years. All demographic, participant, and treatment characteristics are presented in Table 1.

MMO showed no significant difference between the cardboard ruler and the digital caliper at the test measurement ($p = 0.08$) but MMO measured using the digital caliper was significantly larger (10.1%) at the retest measurement compared to the card ruler ($p < 0.001$).

The calculated ICC values ranged from 0.77 to 0.98 (see Table 2). These values indicate good to excellent test-retest reproducibility [35]. Agreement expressed in SDC% ranged between 5.4% and 44.7% for the whole core set of physical measurements. MMO measured with cardboard card, forward flexion shoulder and 6MWT had an acceptable measurement error (SDC%: 5.4% - 15.1%) compared to caliper measured MMO, shoulder abduction, shoulder external rotation, later flexion and rotation of the neck, grip strength, 30SCST, and TUG (SDC%: 19.8% - 44.7%). The Limits of Agreement for all measurements are visualized in Figs 2, 3 and 4.

Discussion

This study establishes good to excellent test-retest reliability of a core set of measurements on physical performance for sHNC in two frequently used measurements on MMO (Therabite[®] cardboard card (intra orally) and a digital caliper (extra orally)), shoulder and neck AROM, upper body strength (GS), lower body strength (30SCTS), level of mobility (TUG), and walking ability (6MWT). It also provides clinically usable information on measurement error to interpret and evaluate physical performance in sHNC. The measurement error reported in caliper measured MMO, shoulder abduction, shoulder external rotation, lateral flexion of the neck, rotation of the neck, GS, 30SCST and TUG is large in relation to the mean scores of the

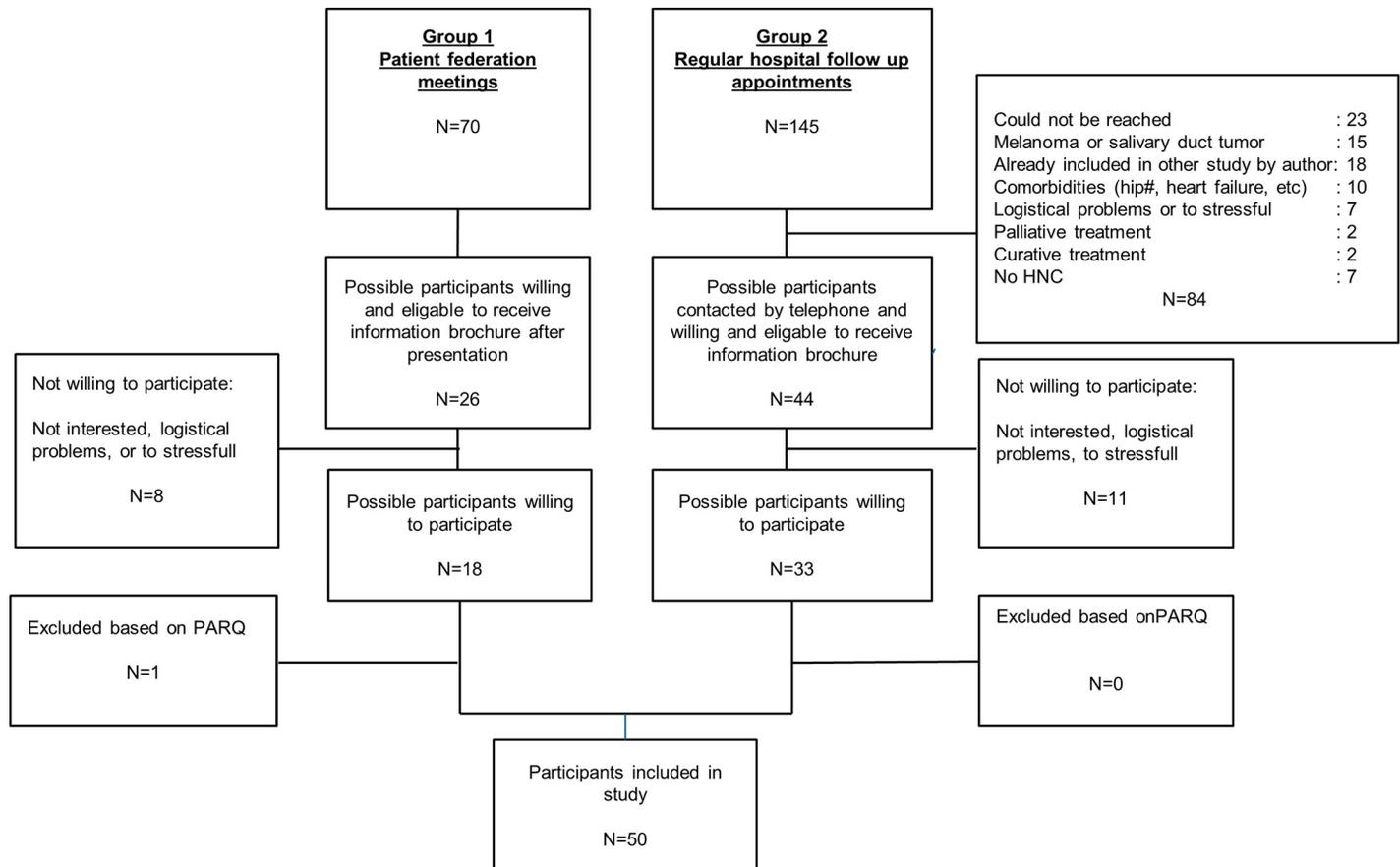


Fig 1. Recruitment and enrollment participants. HNC: Head and Neck Cancer, PARQ: Physical Activity Readiness Questionnaire.

<https://doi.org/10.1371/journal.pone.0233271.g001>

test and retest measurements. This leads to the question if this variance is related to the testers, the measurement procedure, or the participants. Although measurements were performed by physical therapy students, they received extensive training and supervision during measurements by experienced physical therapists. The measurement protocol was based on guidelines and training sessions were performed to solve possible uncertainties. This advocates that measurement error caused by variance in testers or the measurement procedure should be limited. Possible variance between measurements caused by participants will be discussed per measurement.

MMO measured with the Therabite[®] cardboard ruler (ICC 0.95) and the digital caliper (ICC 0.90) showed ICC's that are slightly lower in comparison with measurement of MMO using a normal ruler intra-orally (ICC 0.99) [20]. This however still indicates a good ability to differentiate in MMO between sHNC [20]. Digital caliper scores are systematically higher for the retest measurement compared to the cardboard ruler, indicating more variation in MMO with the digital caliper (Fig 3). One hypothesis for the higher MMO is related to observations made by the students performing the measurements. They observed sHNC experiencing fear of the digital caliper being directly in their field of view during the first test measurement. The participants might have experienced discomfort related to possible contact between the nose or chin and the metal digital caliper. This fear was less present during the retest measurement possibly resulting in a larger MMO. This variation is also illustrated by a higher SEM (3.81 to 2.38), SDC (10.57 to 6.60), and SDC% (22.7% to 15.1%) compared to the Therabite[®] cardboard

Table 2. Reliability of measurements on physical performance in sHNC.

	Test	Retest	Diff test-retest	95% LoA	ICC3.1 (95% CI)	SEM _{agreement} **	SDC _{agreement} **	SDC %
	Mean (SD)	Mean (SD)	Mean (SD)					
MMO (millimeter)								
Card	43.16 (10.57)	43.98 (10.51)	-0.82 (3.30)	-7.29; 5.65	0.95 (0.91–0.97)*	2.38	6.60	15.1%
Caliper	44.78 (11.90)	48.42 (11.07)	-3.63 (4.03)	-11.53; 4.27	0.90 (0.54–0.96)*	3.81	10.57	22.7%
AROM shoulder forward flexion (degrees)								
Left	162.14 (17.41)	160.16 (17.31)	2.86 (8.38)	-13.56; 19.28	0.95 (0.89–0.97)*	4.08	11.30	7.0%
Right	160.64 (19.48)	159.28 (19.86)	1.36 (4.20)	-6.87; 9.59	0.95 (0.96–0.99)*	3.09	8.57	5.4%
AROM shoulder abduction (degrees)								
Left	156.90 (25.19)	152.62 (29.32)	4.28 (18.40)	-31.79; 40.35	0.77 (0.62–0.86)*	13.23	36.68	23.7%
Right	158.36 (24.97)	154.68 (26.48)	3.68 (15.68)	-27.06; 34.42	0.81 (0.69–0.89)*	11.28	31.27	20.0%
AROM shoulder external rotation (degrees)								
Left	51.88 (14.46)	52.66 (14.48)	-0.78 (8.10)	-16.67; 15.11	0.85 (0.74–0.91)*	5.70	15.80	30.2%
Right	55.82 (13.07)	54.52 (14.83)	1.30 (7.45)	-13.27; 15.87	0.86 (0.76–0.92)*	5.28	14.65	26.6%
AROM neck lateral flexion (degrees)								
Left	31.68 (10.47)	31.50 (10.93)	0.18 (5.48)	-10.56; 10.92	0.87 (0.78–0.93)*	3.84	10.64	33.7%
Right	32.76 (8.41)	32.20 (8.92)	0.56 (5.63)	-10.46; 11.58	0.79 (0.66–0.88)*	3.96	10.97	33.8%
AROM neck rotation (degrees)								
Left	63.86 (11.89)	62.66 (14.48)	-1.52 (7.18)	-15.60; 12.56	0.80 (0.67–0.88)*	5.14	14.25	22.1%
Right	64.26 (12.05)	64.52 (14.83)	-0.66 (7.93)	-16.20; 14.88	0.79 (0.65–0.87)*	5.57	15.44	23.9%
GS (kilogram)								
Left	29.08 (12.73)	28.86 (14.60)	0.22 (6.67)	-12.86; 13.30	0.88 (0.80–0.93)*	4.67	12.96	44.7%
Right	30.20 (14.67)	29.68 (15.42)	0.52 (4.23)	-7.76; 8.80	0.96 (0.93–0.98)*	2.98	8.26	27.6%
30SCST (number of times)	10.56 (3.55)	11.04 (3.90)	-0.48 (1.47)	-3.31; 2.35	0.92 (0.85–0.95)*	1.07	2.96	27.4%
TUG (second)	7.79 (3.69)	7.73 (3.56)	0.05 (0.79)	-1.50; 1.60	0.98 (0.96–0.99)*	0.55	1.54	19.8%
6MWT (meters)	447.18 (117.04)	456.68 (120.12)	-9.50 (27.59)	-63.57; 44.57	0.97 (0.95–0.98)*	20.45	56.67	12.5%

*: p<0.001

** : expressed in unit of measurement; AROM: Active range of motion; CI: confidence interval; ICC: intraclass correlation coefficient; GS: grip strength; LoA: limits of agreement; MMO: maximal mouth opening; SDC: smallest detectable change; SD: standard deviation; SEM: standard error of measurement; TUG: timed up and go; 6MWT: 6-minute walking test; 30SCST: 30-second chair-to-stand test.

<https://doi.org/10.1371/journal.pone.0233271.t002>

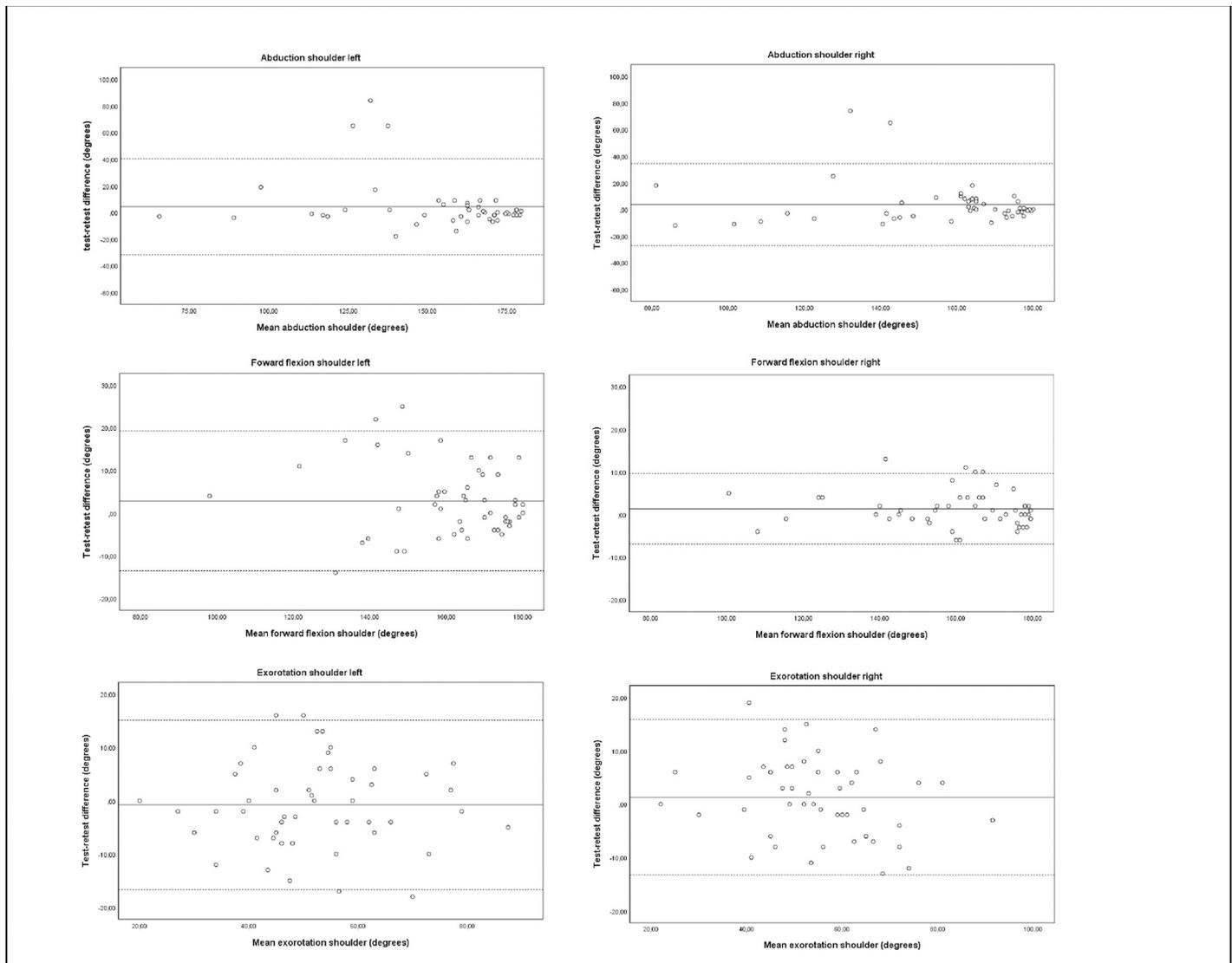


Fig 2. Bland–Altman plots for test-retest reproducibility of maximal mouth opening, shoulder abduction, forward flexion of the shoulder, external rotation of the shoulder. The solid line represents the mean difference (systematic bias) and the dashed lines illustrate the 95% limits of agreement (mean difference \pm 1.96 SD of the difference).

<https://doi.org/10.1371/journal.pone.0233271.g002>

ruler. The SEM and the SDC of the Therabite[®] cardboard are comparable to measurements performed in a population with temporomandibular joint problems (SEM 2.9 and SDC 8.1 mm), providing evaluative values for clinical use. Based on these findings we would favor the use of the Therabite[®] cardboard ruler in sHNC.

The ICC's on shoulder abduction, forward flexion and external rotation (ICC 0.77 to 0.95) found in our study are slightly lower than ICC's measured in healthy subjects (ICC 0.95 to 0.99) With specific problems in shoulder problems to be expected in sHNC, these ICC's still demonstrate a good ability to differentiate in shoulder function between sHNC [30]. A remarkable finding is the high SDC and SDC% for shoulder abduction and external rotation. Shoulder abduction is an important indicator of accessory nerve damage, associated with a high risk of shoulder pain and limitations in activities in daily life [38, 39]. The high shoulder abduction SDC illustrates a large measurement error between test and retest scores. This

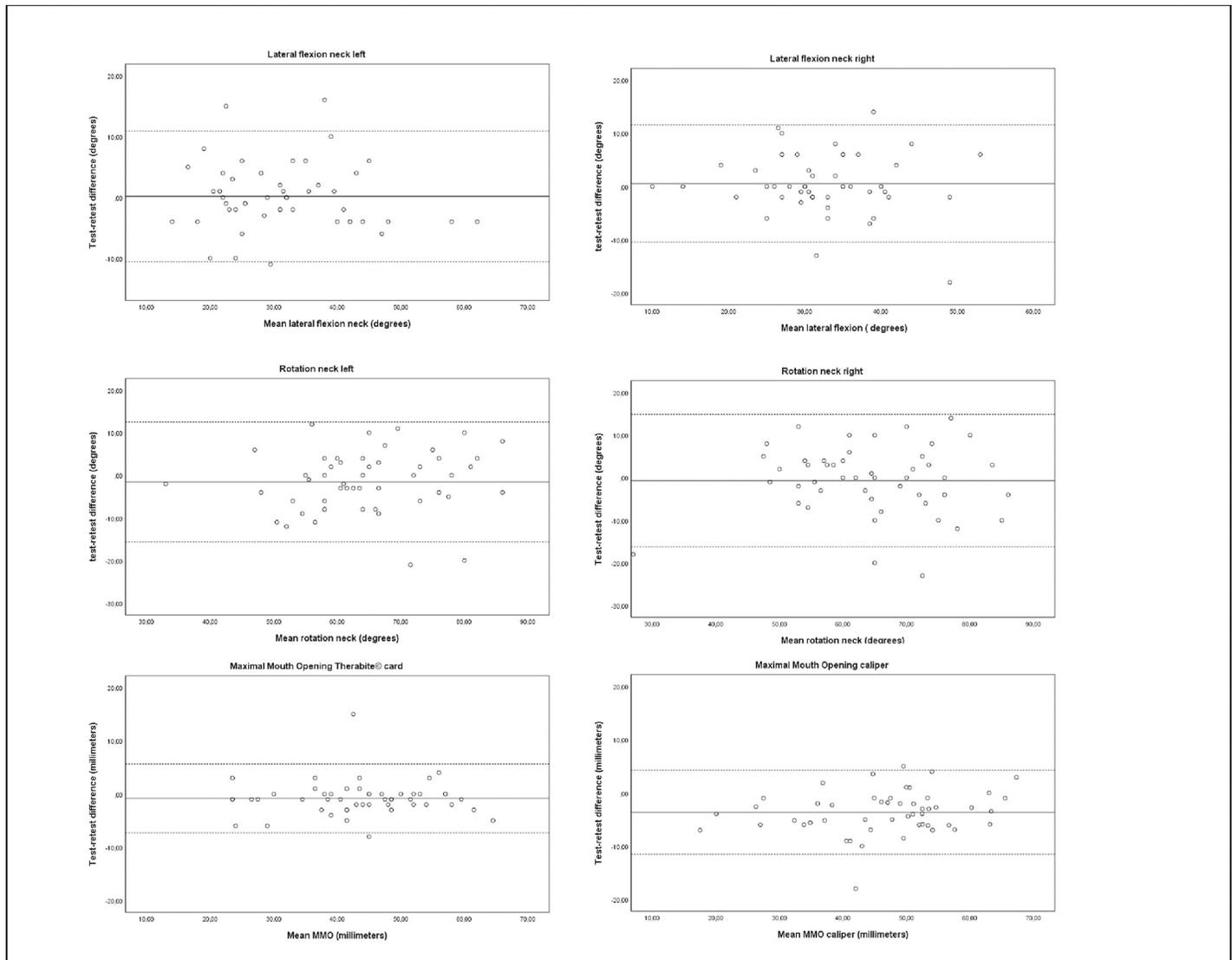


Fig 3. Bland–Altman plots for test-retest reproducibility of lateral flexion of the neck, rotation of the neck, maximal mouth opening. The solid line represents the mean difference (systematic bias) and the dashed lines illustrate the 95% limits of agreement (mean difference \pm 1.96 SD of the difference).

<https://doi.org/10.1371/journal.pone.0233271.g003>

measurement error is especially observed in scores on shoulder abduction smaller than 150 degrees (Fig 2). Pain, proprioceptive dysfunction, or decreased upper body strength may have contributed to the use of compensation strategies which could have resulted in confounded measurement results, increasing the measurement error. However, even with extensive training of the testers and the use of a strict measurement protocol, these compensation strategies could not be prevented. This supports clinical examination of the shoulder function by a physical therapist. Future research should take this into account when standardizing measurement protocols.

The ICC's on neck function measured with CROM device are slightly lower, and the SEM's are higher compared to literature investigating healthy subjects [31]. This could advert to the CROM device being able to differentiate between sHNC. However, the measurement error is slightly higher compared to healthy subjects when it is used in an evaluative setting. The high

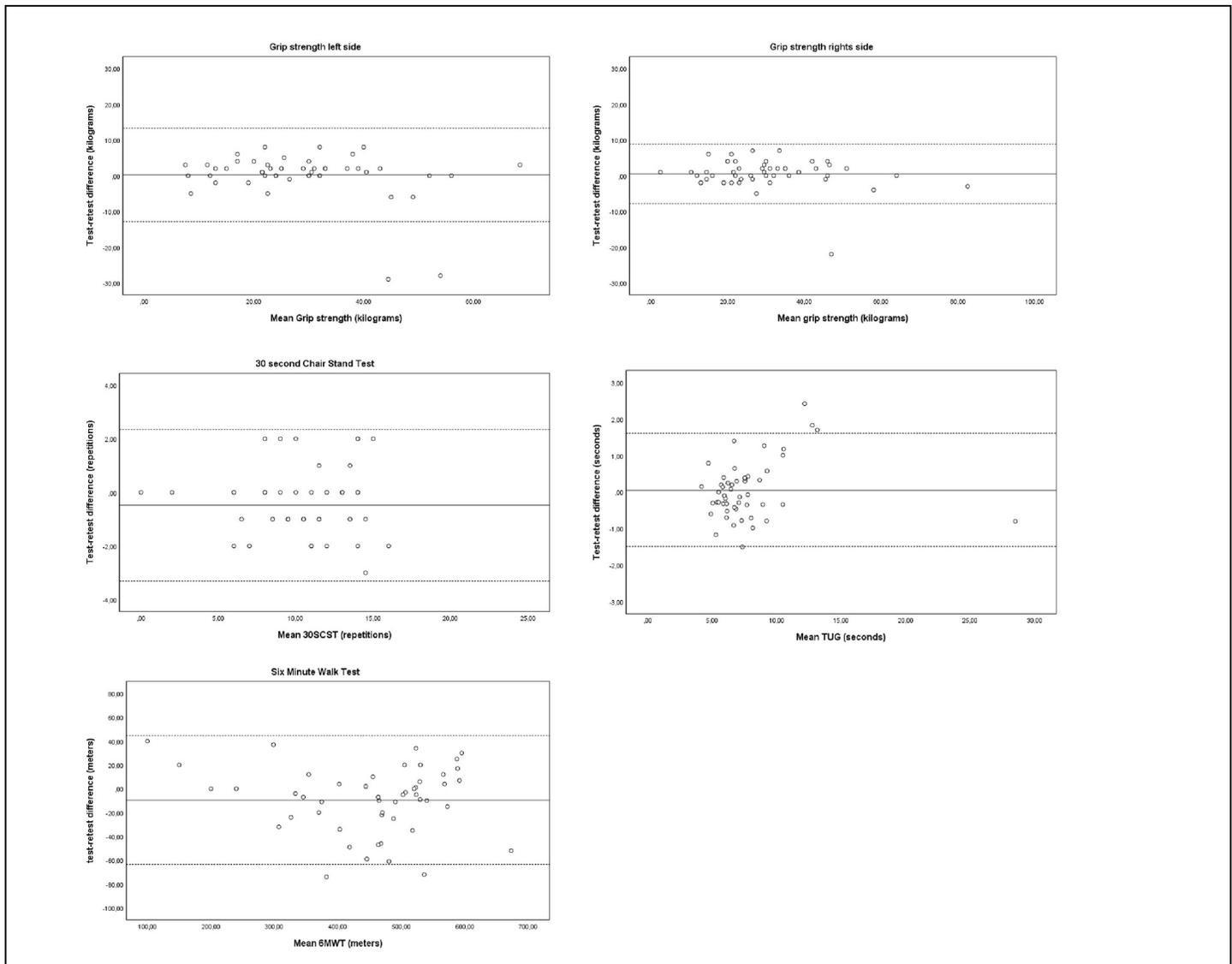


Fig 4. Bland–Altman plots for test-retest reproducibility of grip strength, 30 second chair stand test, Six Minute Walk Test, and Timed Up and Go test. The solid line represents the mean difference (systematic bias) and the dashed lines illustrate the 95% limits of agreement (mean difference \pm 1.96 SD of the difference).

<https://doi.org/10.1371/journal.pone.0233271.g004>

SDC% values confirm poor evaluative measurement properties. Variation in measurement outcomes could be related to sHNC undergoing ND surgery and radiotherapy, leading to local alterations in anatomy and physiology causing different compensation strategies [12, 40].

GS ICC scores of 0.96 for the right side and 0.88 for the left side are in line with community-dwelling elderly (right ICC 0.95 and left ICC 0.91), which demonstrates a good ability to differentiate in upper body strength between sHNC [41]. When compared to literature, the SEM for GS was higher (SEM left 4.67, right 2.98) in sHNC compared to healthy individuals (SEM scores for men 2.77, women 1.66). The high SDC% values (45% for the left side, 28% for the right side) illustrate that the measurement error for the GS is too large to be used in a clinical setting which limits evaluative usability [42].

The ICC found for the 30SCST (ICC 0.92) is in line with previous research investigating HNC patients (ICC 0.95) [25] and a study investigating community-dwelling adults (ICC 0.84

men, ICC 0.92 women) [33]. This indicates that the 30SCST can differentiate in functional lower body strength between sHNC. With a mean test-retest score of 11 repetitions and an SDC being nearly 3 repetitions, a sHNC must show an improvement of at least 3 repetitions (SDC% 27%) to be above the measurement error, which limits clinical evaluative usability.

The 6MWT demonstrated an ICC value of 0.97, which is in line with a study that included sHNC and patients with HNC receiving treatment (ICC 0.97). This indicates excellent capability to differentiate in walking ability between sHNC. The SEM of 20.5 meters is lower compared to patients undergoing hemodialysis (SEM 28.4) and comparable to patients with Alzheimer's, SEM 20.28. The SDC and SDC% indicate that in relation to mean 6MWT test- and retest scores a 13% change is above the measurement error.

Level of mobility was assessed by the TUG which showed a comparable ICC (ICC 0.98) to test-retest studies in people with chronic conditions as Parkinson or stroke [43, 44]. The Bland Altman plot showed homogenous scores for the TUG in our sample (Fig 4). This disputes whether the TUG should be a standard test to differentiate in the level of mobility in sHNC. The SDC score (1.54 sec.) seems relatively small but in percentage (SDC%: 20%) to the average scores (7.73 to 7.79 sec) it is quite large regarding evaluative purposes.

Strengths and limitations of this study

This study followed the COSMIN checklist to ensure methodological and statistical quality and reduce bias. Similar to other studies, the participants in this study represent a heterogeneous group of sHNC, displaying different characteristics [45–49]. Although specific subgroups in sHNC (for example, patients after laryngectomy) are known to have specific problems in physical performance [50]. The heterogeneity in this sample is likely to provide an adequate representation of the total group of sHNC as found in daily practice. Therefore, this study provides clinically useful information on reliability of a core set of measurements on physical performance.

The selection of participants came from two different groups resulting in a heterogeneous sample of sHNC that improves generalizability. The sHNC contacted through the patient federation had no treatment relationship with the researcher. For this reason, they were asked to report on treatment and tumor characteristics. This allows for mistakes and misinterpretations by the sHNC. The time interval between the test and retest measurement was at least one hour and maximal two hours. Even though intervals of one or two weeks are typically recommended by experts to allow recovery and limit recall bias [51]. The time between the test and retest measurement was chosen because of logistical reasons and was estimated to be long enough to recover from fatigue; the data showed no signs of fatigue. Higher retest measurements were found for both measurements on MMO and 6MWT. This indicates a possible learning effect for these outcomes. This initial learning effect has not been found in previous literature for measurements on MMO and is in line with literature for 6MWT [52]. For all three measurements it does not influence reliability. Another limitation is the absence of measurements on inter-rater reliability. An additional measurement to determine inter-rater reliability was deemed to be too exhausting and time consuming for participants.

Clinical relevance

More than half of sHNC are sedentary and experience specific problems in physical performance due to treatment of the head and neck area [15]. Insight into reliability of a core set of measurements on physical performance in sHNC is essential to improve supportive care and research on the physical performance of sHNC. To gain full insight into sHNC physical status

these measurements can be used in addition to Patient Reported Outcome Measurements (PROMs) that measure patients' perceptions and views on physical status and performance.

Conclusion

This study demonstrated good to excellent test- retest reliability of a core set of measurements on physical performance which illustrates that this coreset can be used to differentiate in physical performance between sHNC. The reported measurement errors should be taken into consideration when interpreting the results of repeated measurements.

Supporting information

S1 File. S1 File measurements reproducibility.
(SAV)

Acknowledgments

The authors want to thank I. van Baaijen and M. Arts for their help in organizing participant inclusion.

Author Contributions

Conceptualization: Gerben van Hinte, Matthias A. W. Merkx, Robert Takes, Maria W. G. Nijhuis-van der Sanden, Caroline M. Speksnijder.

Data curation: Gerben van Hinte, Ruud A. Leijendekkers.

Formal analysis: Gerben van Hinte, Ruud A. Leijendekkers.

Funding acquisition: Gerben van Hinte.

Investigation: Gerben van Hinte, Ruud A. Leijendekkers, Bram te Molder, Lizzy Jansen, Corinda Bol.

Methodology: Gerben van Hinte, Ruud A. Leijendekkers, Matthias A. W. Merkx, Robert Takes, Maria W. G. Nijhuis-van der Sanden, Caroline M. Speksnijder.

Project administration: Gerben van Hinte.

Supervision: Matthias A. W. Merkx, Robert Takes, Maria W. G. Nijhuis-van der Sanden, Caroline M. Speksnijder.

Visualization: Gerben van Hinte.

Writing – original draft: Gerben van Hinte.

Writing – review & editing: Gerben van Hinte, Ruud A. Leijendekkers, Bram te Molder, Lizzy Jansen, Corinda Bol, Matthias A. W. Merkx, Robert Takes, Maria W. G. Nijhuis-van der Sanden, Caroline M. Speksnijder.

References

1. Shah JP, Patel SG, Singh B. Jatin Shah's Head and Neck Surgery and Oncology E-Book: Elsevier Health Sciences; 2012.
2. Wells M, Swartzman S, Lang H, Cunningham M, Taylor L, Thomson J, et al. Predictors of quality of life in head and neck cancer survivors up to 5 years after end of treatment: a cross-sectional survey. Supportive care in cancer: official journal of the Multinational Association of Supportive Care in Cancer. 2016; 24(6):2463–72.

3. Eickmeyer SM, Walczak CK, Myers KB, Lindstrom DR, Layde P, Campbell BH. Quality of life, shoulder range of motion, and spinal accessory nerve status in 5-year survivors of head and neck cancer. *PM & R: the journal of injury, function, and rehabilitation*. 2014; 6(12):1073–80.
4. de Groot RJ, Wetzels JW, Merkx MAW, Rosenberg A, de Haan AFJ, van der Bilt A, et al. Masticatory function and related factors after oral oncological treatment: A 5-year prospective study. *Head & neck*. 2019; 41(1):216–24.
5. Kao SS, Peters MD, Krishnan SG, Ooi EH. Swallowing outcomes following primary surgical resection and primary free flap reconstruction for oral and oropharyngeal squamous cell carcinomas: A systematic review. *The Laryngoscope*. 2016; 126(7):1572–80. <https://doi.org/10.1002/lary.25894> PMID: 26865034
6. Kraaijenga SA, Oskam IM, van der Molen L, Hamming-Vrieze O, Hilgers FJ, van den Brekel MW. Evaluation of long term (10-years+) dysphagia and trismus in patients treated with concurrent chemo-radiotherapy for advanced head and neck cancer. *Oral oncology*. 2015; 51(8):787–94. <https://doi.org/10.1016/j.oraloncology.2015.05.003> PMID: 26027851
7. Wetzels JW, Merkx MA, de Haan AF, Koole R, Speksnijder CM. Maximum mouth opening and trismus in 143 patients treated for oral cancer: a 1-year prospective study. *Head & neck*. 2014; 36(12):1754–62.
8. van den Berg MG, Rasmussen-Conrad EL, van Nispen L, van Binsbergen JJ, Merkx MA. A prospective study on malnutrition and quality of life in patients with head and neck cancer. *Oral oncology*. 2008; 44(9):830–7. <https://doi.org/10.1016/j.oraloncology.2007.11.002> PMID: 18243771
9. Liu SA, Tsai WC, Wong YK, Lin JC, Poon CK, Chao SY, et al. Nutritional factors and survival of patients with oral cancer. *Head & neck*. 2006; 28(11):998–1007.
10. Jager-Wittenaar H, Dijkstra PU, Vissink A, Langendijk JA, van der Laan BF, Pruijm J, et al. Changes in nutritional status and dietary intake during and after head and neck cancer treatment. *Head & neck*. 2011; 33(6):863–70.
11. van Wilgen CP, Dijkstra PU, van der Laan BF, Plukker JT, Roodenburg JL. Shoulder and neck morbidity in quality of life after surgery for head and neck cancer. *Head & neck*. 2004; 26(10):839–44.
12. van Hinte G, Wetzels JGH, Merkx MAW, de Haan AFJ, Koole R, Speksnijder CM. Factors influencing neck and shoulder function after oral oncology treatment: a five-year prospective cohort study in 113 patients. *Supportive care in cancer: official journal of the Multinational Association of Supportive Care in Cancer*. 2019; 27(7):2553–60.
13. Gane EM, McPhail SM, Hatton AL, Panizza BJ, O'Leary SP. Neck and Shoulder Motor Function following Neck Dissection: A Comparison with Healthy Control Subjects. *Otolaryngology—head and neck surgery: official journal of American Academy of Otolaryngology-Head and Neck Surgery*. 2019; 160(6):1009–18.
14. Bossi P, Di Pede P, Guglielmo M, Granata R, Alfieri S, Iacovelli NA, et al. Prevalence of Fatigue in Head and Neck Cancer Survivors. *The Annals of otology, rhinology, and laryngology*. 2019; 128(5):413–9. <https://doi.org/10.1177/0003489419826138> PMID: 30700107
15. Rogers LQ, Courneya KS, Robbins KT, Malone J, Seiz A, Koch L, et al. Physical activity and quality of life in head and neck cancer survivors. *Supportive care in cancer: official journal of the Multinational Association of Supportive Care in Cancer*. 2006; 14(10):1012–9.
16. Nemeth D, Zaleczna L, Huremovic A, Engelmann J, Poeschl PW, Strasz M, et al. Importance of chewing, saliva, and swallowing function in patients with advanced oral cancer undergoing preoperative chemoradiotherapy: a prospective study of quality of life. *International journal of oral and maxillofacial surgery*. 2017; 46(10):1229–36. <https://doi.org/10.1016/j.ijom.2017.05.005> PMID: 28579265
17. Gane EM, McPhail SM, Hatton AL, Panizza BJ, O'Leary SP. The relationship between physical impairments, quality of life and disability of the neck and upper limb in patients following neck dissection. *Journal of cancer survivorship: research and practice*. 2018; 12(5):619–31.
18. Giuliani M, Papadakos J, Broadhurst M, Jones J, McQuestion M, Le LW, et al. The prevalence and determinants of return to work in head and neck cancer survivors. *Supportive care in cancer: official journal of the Multinational Association of Supportive Care in Cancer*. 2019; 27(2):539–46.
19. Oskam IM, Verdonck-de Leeuw IM, Aaronson NK, Witte BI, de Bree R, Doornaert P, et al. Prospective evaluation of health-related quality of life in long-term oral and oropharyngeal cancer survivors and the perceived need for supportive care. *Oral oncology*. 2013; 49(5):443–8. <https://doi.org/10.1016/j.oraloncology.2012.12.005> PMID: 23318122
20. Jager-Wittenaar H, Dijkstra PU, Vissink A, van Oort RP, Roodenburg JL. Variation in repeated mouth-opening measurements in head and neck cancer patients with and without trismus. *International journal of oral and maxillofacial surgery*. 2009; 38(1):26–30. <https://doi.org/10.1016/j.ijom.2008.10.001> PMID: 19041226

21. Kottner J, Gajewski BJ, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *Int J Nurs Stud*. 2011; 48(6):659–60. <https://doi.org/10.1016/j.ijnurstu.2011.01.017> PMID: 21376316
22. Ries JD, Echternach JL, Nof L, Gagnon Blodgett M. Test-retest reliability and minimal detectable change scores for the timed "up & go" test, the six-minute walk test, and gait speed in people with Alzheimer disease. *Physical therapy*. 2009; 89(6):569–79. <https://doi.org/10.2522/ptj.20080258> PMID: 19389792
23. Segura-Orti E, Martinez-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. *Physical therapy*. 2011; 91(8):1244–52. <https://doi.org/10.2522/ptj.20100141> PMID: 21719637
24. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. *Physical therapy*. 2002; 82(2):128–37. <https://doi.org/10.1093/ptj/82.2.128> PMID: 11856064
25. Eden MM, Tompkins J, Verheijde JL. Reliability and a correlational analysis of the 6MWT, ten-meter walk test, thirty second sit to stand, and the linear analog scale of function in patients with head and neck cancer. *Physiotherapy theory and practice*. 2018; 34(3):202–11. <https://doi.org/10.1080/09593985.2017.1390803> PMID: 29068767
26. Shephard RJ. PAR-Q, Canadian Home Fitness Test and exercise screening alternatives. *Sports medicine (Auckland, NZ)*. 1988; 5(3):185–95.
27. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). *Canadian journal of sport sciences = Journal canadien des sciences du sport*. 1992; 17(4):338–45. PMID: 1330274
28. Donner A, Eliasziw M. Sample size requirements for reliability studies. *Statistics in medicine*. 1987; 6(4):441–8. <https://doi.org/10.1002/sim.4780060404> PMID: 3629046
29. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*. 2010; 19(4):539–49.
30. Kolber MJ, Vega F, Widmayer K, Cheng MS. The reliability and minimal detectable change of shoulder mobility measurements using a digital inclinometer. *Physiotherapy theory and practice*. 2011; 27(2):176–84. <https://doi.org/10.3109/09593985.2010.481011> PMID: 20690872
31. Audette I, Dumas JP, Cote JN, De Serres SJ. Validity and between-day reliability of the cervical range of motion (CROM) device. *J Orthop Sports Phys Ther*. 2010; 40(5):318–23. <https://doi.org/10.2519/jospt.2010.3180> PMID: 20436238
32. Dodds RM, Syddall HE, Cooper R, Kuh D, Cooper C, Sayer AA. Global variation in grip strength: a systematic review and meta-analysis of normative data. *Age and ageing*. 2016; 45(2):209–16. <https://doi.org/10.1093/ageing/afv192> PMID: 26790455
33. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Research quarterly for exercise and sport*. 1999; 70(2):113–9. <https://doi.org/10.1080/02701367.1999.10608028> PMID: 10380242
34. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *Journal of clinical epidemiology*. 2006; 59(10):1033–9. <https://doi.org/10.1016/j.jclinepi.2005.10.015> PMID: 16980142
35. Byrt T. How good is that agreement? *Epidemiology (Cambridge, Mass)*. 1996; 7(5):561.
36. De Vet HCW, Terwee BT, Mokkink LB, Knol DL. *Measurement in Medicine: A practical guide*. New-York: U.S.A.: Cambridge University Press; 2011
37. Portney LG WM. *Foundations of clinical research: applications to practice*. Harlow: Pearson. 2014.
38. McGarvey AC, Chiarelli PE, Osmotherly PG, Hoffman GR. Physiotherapy for accessory nerve shoulder dysfunction following neck dissection surgery: a literature review. *Head & neck*. 2011; 33(2):274–80.
39. van Wilgen CP, Dijkstra PU, van der Laan BF, Plukker JT, Roodenburg JL. Shoulder complaints after neck dissection; is the spinal accessory nerve involved? *The British journal of oral & maxillofacial surgery*. 2003; 41(1):7–11.
40. Speksnijder CM, van der Bilt A, Slappendel M, de Wijer A, Merkx MA, Koole R. Neck and shoulder function in patients treated for oral malignancies: a 1-year prospective cohort study. *Head & neck*. 2013; 35(9):1303–13.
41. Bohannon RW, Schaubert KL. Test-retest reliability of grip-strength measures obtained over a 12-week interval from community-dwelling elders. *Journal of hand therapy: official journal of the American Society of Hand Therapists*. 2005; 18(4):426–7, quiz 8.

42. King TI 2nd. Interinstrument reliability of the Jamar electronic dynamometer and pinch gauge compared with the Jamar hydraulic dynamometer and B&L Engineering mechanical pinch gauge. *The American journal of occupational therapy: official publication of the American Occupational Therapy Association*. 2013; 67(4):480–3.
43. Lyders Johansen K, Derby Stistrup R, Skibdal Schjott C, Madsen J, Vinther A. Absolute and Relative Reliability of the Timed 'Up & Go' Test and '30second Chair-Stand' Test in Hospitalised Patients with Stroke. *PloS one*. 2016; 11(10):e0165663. <https://doi.org/10.1371/journal.pone.0165663> PMID: [27798686](https://pubmed.ncbi.nlm.nih.gov/27798686/)
44. Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease. *Physical therapy*. 2001; 81(2):810–8. <https://doi.org/10.1093/ptj/81.2.810> PMID: [11175678](https://pubmed.ncbi.nlm.nih.gov/11175678/)
45. Boyes AW, Girgis A, D'Este C, Zucca AC. Prevalence and correlates of cancer survivors' supportive care needs 6 months after diagnosis: a population-based cross-sectional study. *BMC cancer*. 2012; 12:150. <https://doi.org/10.1186/1471-2407-12-150> PMID: [22510387](https://pubmed.ncbi.nlm.nih.gov/22510387/)
46. Giuliani M, McQuestion M, Jones J, Papadakos J, Le LW, Alkazaz N, et al. Prevalence and nature of survivorship needs in patients with head and neck cancer. *Head & neck*. 2016; 38(7):1097–103.
47. Jansen F, van Uden-Kraan CF, van Zwieten V, Witte BI, Verdonck-de Leeuw IM. Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. *Supportive care in cancer: official journal of the Multinational Association of Supportive Care in Cancer*. 2015; 23(6):1679–88.
48. Sammut L, Ward M, Patel N. Physical activity and quality of life in head and neck cancer survivors: a literature review. *International journal of sports medicine*. 2014; 35(9):794–9. <https://doi.org/10.1055/s-0033-1363984> PMID: [24554555](https://pubmed.ncbi.nlm.nih.gov/24554555/)
49. Wells M, Cunningham M, Lang H, Swartzman S, Philp J, Taylor L, et al. Distress, concerns and unmet needs in survivors of head and neck cancer: a cross-sectional survey. *European journal of cancer care*. 2015; 24(5):748–60. <https://doi.org/10.1111/ecc.12370> PMID: [26250705](https://pubmed.ncbi.nlm.nih.gov/26250705/)
50. Jansen F, Eerenstein SEJ, Lissenberg-Witte BI, van Uden-Kraan CF, Leemans CR, Leeuw IMV. Unmet supportive care needs in patients treated with total laryngectomy and its associated factors. *Head & neck*. 2018; 40(12):2633–41.
51. Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*. 2014; 23(6):1713–20.
52. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002; 166(1):111–7. <https://doi.org/10.1164/ajrccm.166.1.at1102> PMID: [12091180](https://pubmed.ncbi.nlm.nih.gov/12091180/)