Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

Full length article

How do patients improve their timed up and go test? Responsiveness to rehabilitation of the TUG test in elderly neurological patients

Antonio Caronni[a,*], Michela Picardi[b], Evdoxia Aristidou[b], Paola Antoniotti[b], Giuseppe Pintavalle[b], Valentina Redaelli[b], Irma Sterpi[b], Massimo Corbo[b]

[a] IRCCS Fondazione Don Carlo Gnocchi Onlus, Milan, Italy
[b] Department of Neurorehabilitation Sciences, Casa di Cura del Policlinico, Milan, Italy

ARTICLE INFO

Keywords:
Instrumental timed up and go test
Rehabilitation
Responsiveness
Elderly
ITUG
Balance and locomotor impairment

ABSTRACT

Background: The timed up and go (TUG) test is widely used for assessing treatments effectiveness on elderly mobility. Although the TUG test consists of different tasks (e.g. walking and turning), the total TUG duration (TTD) is usually the only outcome measure, with TTD shortening indicating the patient's improvement.

Research question: Does TTD shortening reflect the improvement of each TUG tasks or does it reflect the improvement of only some of them?

Methods: This retrospective study recruited 120 elderly patients (mean, SD: 76.9, 6.6 years) admitted to in-patient rehabilitation because of an acute or chronic neurological disease (acute patients, AP; chronic patients, CP). TTD and TUG tasks duration was measured on admission and discharge (five trials/session) by means of the instrumental TUG test (ITUG). Likelihood ratios (LRs) were used for inferring TUG tasks improvement from TTD improvement. TTD and TUG tasks have improved if at least four measurements on discharge were shorter than the shortest measurement on admission.

Results: TTD improvement *per se* is not enough to claim that all the TUG tasks have improved ($LR^{+}_{AP} = 1.32$; $LR^{+}_{CP} = 1.85$). Conversely, if TTD has not improved, not even a single TUG task has improved ($LR^{-}_{AP} = 0.13$; $LR^{-}_{CP} = 0.19$). If TTD has improved, there is at least one TUG task that actually improved ($LR^{+}_{AP} = 3.17$; $LR^{+}_{CP} = 9.54$). The improvement of all TUG tasks can be only inferred in the (unusual) event of a large TTD shortening (AP: > 39%, $LR^{+}_{AP} = 6.26$; CP: > 30%, $LR^{+}_{CP} = 9.0$).

Significance: In most cases, TTD improvement is not associated with the improvement of all TUG tasks. Moreover, when TTD has improved there is at least a TUG task that has improved, but that remains unknown. To actually understand how treatments ameliorate patients' mobility, ITUG with TUG task duration measurement should be preferred to TTD.

1. Introduction

The timed up and go (TUG) test [1,2] is widely used in elderly patients with a range of mobility impairments for evaluating basic locomotor activities (i.e. standing up, walking, turning around and sitting down). In this patients' population, the TUG test is a common outcome measure for assessing the effectiveness of treatments, including drugs [3] and rehabilitation [4]. As an outcome measure, the TUG test is employed not only in clinical trials [5], but also at a single subject level in routine clinical practice.

Inertial sensors (e.g. accelerometers and gyroscopes) are more and more used in health assessment [6]. In the instrumental timed up and go (ITUG) test, participants wear inertial sensors (commonly secured to the trunk) while completing the conventional TUG test. Acceleration and angular velocity recorded by the inertial sensor during the ITUG test are used to split the test into its different phases (e.g. sit to walk, walking), which are eventually measured (e.g. phase duration).

When used to measure treatment effectiveness, the shortening of the total TUG duration (TTD) marks the improvement of the patients' performance. However, when a patient shortens his/her TTD the clinician wonders whether this modification reflects the homogeneous improvement of all the TUG phases (i.e. locomotor tasks) or the improvement of only some of these. Put it simply: what does *"the patient's TUG is improved"* mean? Thanks to the ITUG it becomes possible to explore this issue. For a clinician, this is an important question, both when the TUG test is used as an outcome measure in clinical trials and

[*] Corresponding author.
E-mail address: acaronni@dongnocchi.it (A. Caronni).

https://doi.org/10.1016/j.gaitpost.2019.02.010
Received 25 April 2018; Received in revised form 12 January 2019; Accepted 14 February 2019
0966-6362/

in everyday practice.

In fact, it is obvious that the tasks assessed by the TUG test (such as moving from sitting to standing, walking in a straight line or along a curved trajectory) represent fundamental tasks for patients' daily lives. Difficulty in turning is a risk factor for falls in patients with a neurological disease [7] and difficulty or slowness in moving from sitting to standing is a known falls risk in the elderly [8]. Clinical trials showed that the improvement of the sit to stand movement (e.g. in mild Parkinson disease patients [9]) or the increase in walking speed (e.g. in the elderly [10]) are associated to a reduction of the risk of falling.

If there were evidence that the TTD shortening reflects the uniform shortening of all the TUG phases, the clinician could reasonably infer the improvement of all the tasks making up the TUG test from the improvement of the TTD. On the contrary, if this evidence does not exist, it would be erroneous to claim the improvement of the patients' ability to move from sitting to standing, to walk and turn just because of the TTD reduction *(a tempting unwarranted conditional)*. In this case, the clinician can only conclude that the overall patient's mobility is changed [11].

Aim of the current work is to unveil how rehabilitation modifies the TUG phases duration and thus the TTD. To this aim, we recorded the ITUG in elderly patients with a locomotor impairment caused by a neurological disease, both at the beginning and end of a rehabilitation program consisting in physiotherapy and occupational therapy. These results are then analysed and discussed in the responsiveness framework [12].

2. Methods

2.1. Participants

We conducted a retrospective cohort study, in which we recruited 120 consecutive patients attending the rehabilitation clinic of Casa di Cura del Policlinico in Milano (May 2016 – January 2018).

Patients were included if i) older than 65 years, ii) admitted to rehabilitation because of a neurological disease, iii) able to complete the TUG test without touching assistance. Patients were excluded because of i) an acute medical condition, ii) a condition causing *per se* a locomotor impairment (e.g. lower limb amputation), iii) an admission TUG test longer than 40 s or iv) shorter than 10 s.

All patients participated in a rehabilitation program consisting in inpatient physiotherapy and occupational therapy one-on-one sessions (physiotherapy: two sessions/day, 45 min each, five days/week; occupational therapy: one session/day, 45 min each, three days/week).

Physiotherapy and occupational therapy interventions were developed in accordance with international consensus (e.g. [13]). However, therapists were free to modulate the exact exercise program according to the patients' needs and compliance.

For the current analysis, we split the patients' sample in acute and chronic patients. Acute patients are those admitted to rehabilitation because of a neurological disease occurred within the month before the rehabilitation admission. As an example, stroke patients who had their cerebrovascular accident one week before the rehabilitation admission were classified as acute patients. By contrast, Parkinson disease patients are typical chronic patients. There is no *a priori* reason to assume that the TUG changes the same in acute and chronic patients. On the contrary, it is a common clinical finding that the improvement expected after rehabilitation is larger in acute than chronic patients and so the pattern of improvement could also be different. For these reasons we analysed the acute and chronic conditions separately. Moreover, precisely because the amount of improvement is expected to be smaller in chronic than acute patients, we decided to recruit more chronic (70) than acute (50) patients.

At our knowledge, there are no guidelines for choosing the right sample size in responsiveness studies. Previous works assessing the responsiveness of instrumental movement measures recruited about 20 participants [14,15] and thus the current sample size seems appropriate.

Table 1
Patients' demographics. Stroke: patients with hemiparetic gait because of a stroke; PNP: peripheral neuropathy; VP: vascular parkinsonism; PD: Parkinson disease; LSS: lumbar spinal stenosis with polyradiculopathy; NMD: neuromuscular disorders; TBI: traumatic brain injury. Acute PNP patients were affected by Guillain–Barré syndrome. Both acute and chronic LSS patients were affected by lumbar spinal stenosis with polyradiculopathy, but acute LSS only had spinal surgery (e.g. laminectomy plus fusion) within one month before the admission evaluation. Chronic NMD patients were affected by early amyotrophic lateral sclerosis or Kennedy disease. The single acute NMD patient was affected by statin-induced myopathy.

		Acute patients (n = 50)	Chronic patients (n = 70)
Mean age (SD), years		77.8 (6.6)	76.4 (6.5)
Number of females (%)		23 (46%)	31 (44.3%)
Diagnosis, number of patients	Stroke	36	7
	PNP	2	21
	VP	0	15
	PD	0	13
	LSS	5	6
	NMD	1	6
	TBI	5	0
	Myelopathy	1	2
FIM motor (IQR), score	Admission	40 (32.3)	73.5 (10.3)
	Discharge	70 (14.5)	78 (13)
Number of patients using a walking aid (%)	Admission	35 (50.0%)	19 (27.1%)
	Discharge	23 (32.9%)	15 (21.4%)

When possible, patients were tested without gait aid. A gait aid was used (commonly a walker), if the measuring clinician judged that the risk of falling during the test was too high.

Table 1 reports patients' demographics.

All participants gave their written informed consent to participate in the study, which received internal ethical approval (PRO.05.M.03).

2.2. Mobility assessment: the TUG test

For each patient, the ITUG test was recorded at the beginning and at the end of the rehabilitation program. The ITUG was recorded according to the very same protocol detailed in our previous work [16]. Briefly, the conventional three-meters TUG test was performed [17]. Participants were asked to get out of the chair, walk three meters, turn around, walk back to the chair and sit down, prompted by a go signal. During the TUG test, subjects wore a commercial inertial measurement unit (mHT-Mhealth technologies, Bologna, Italy) secured to their lower back.

In each measuring session, the TUG test was repeated five times each (five trials/session). To avoid falls during the tests, patients were instructed to use a comfortable walking speed.

2.3. Data analysis and statistics

Signals recorded by the inertial measurement unit were used to measure the total TUG duration (TTD) and to subdivide the TUG test into five phases (sit to stand, STS; walk 1, W1; turn 1, T1; walk 2, W2; turn and sit, TAS). Details on the ITUG splitting procedure can be found elsewhere [16] and are also given as Supplementary data file 1.

We used two complementary approaches (*population analysis* vs *single-subject analysis*), which both adopt statistics from the responsiveness theory [12], to assess the way rehabilitation modifies the duration of the TUG phases and the TTD.

In the *population analysis,* paired t-tests were used for comparing the sample TUG measures on admission and discharge. With this analysis, we examined whether rehabilitation improves TTD or the duration of the ITUG phases in the population of patients *considered as a whole*. The

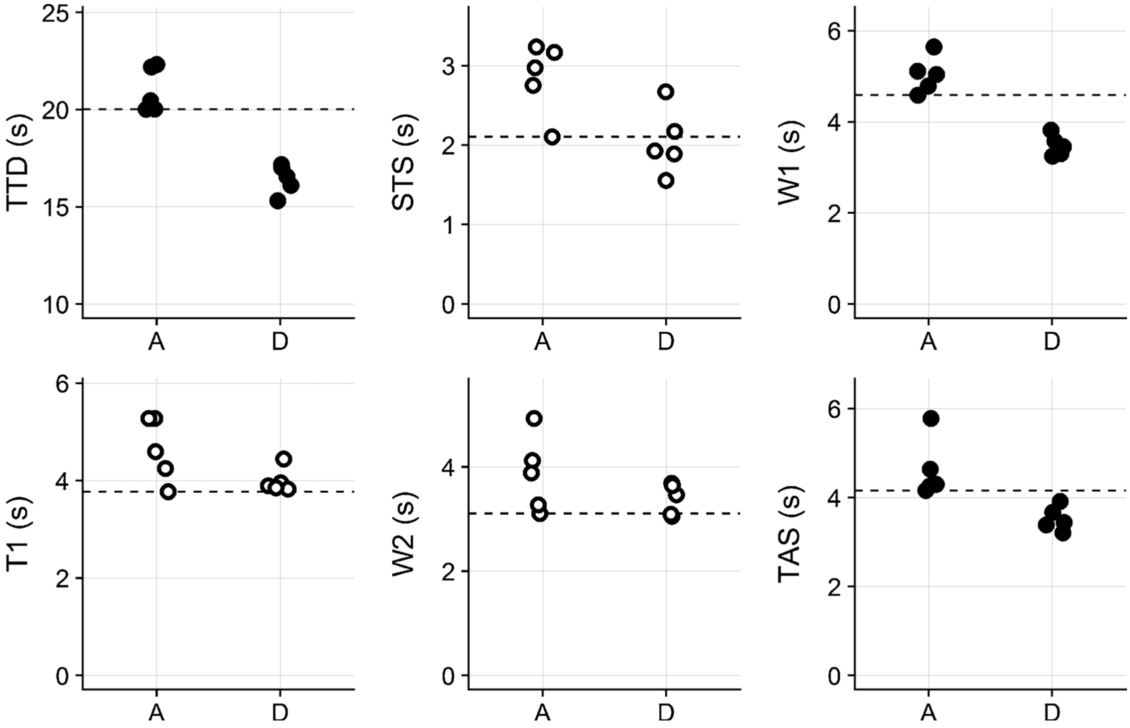

Fig. 1. TTD and TUG tasks in a single representative patient. Dots: TTD or TUG tasks duration in a single trial on admission (A) and discharge (D). Horizontal dashed line: the best trial recorded on admission. TTD, W1 and TAS got better (filled dots), while STS, T1 and W2 did not (empty dots).

Cohen's d was chosen as an effect size measure to quantify the difference between discharge and admission mean measures. Cohen's d was classified according to Cohen [18] so as to indicate a large ($d \geq 0.8$), medium ($0.5 \leq d < 0.8$), small ($0.2 \leq d < 0.5$) or negligible ($d < 0.2$) improvement. The mean value of the five trials was used in the population analysis.

In the *single-subject analysis,* we examined in *each single patient* who took part to the study whether rehabilitation improved TTD or the duration of the ITUG phases. For this analysis, a patient was considered actually improved if at least four out of five measurements on discharge were better than the best of the five measurements collected on admission (at least four non-overlapping data). Fig. 1 reports the TTD and TUG tasks duration in a single representative subject, before (admission) and after (discharge) completing the rehabilitation program. According to the single-subject analysis, this patient improved her W1 and TAS while no substantial modification was observed for STS, T1 and W2. The TTD was also shorter on discharge than admission. The definition of improvement we applied here to single subjects is in accordance with research on Single-Subject Experimental Design [19].

Positive ($^+$) and negative ($^-$) likelihood (LR) ratios were calculated to evaluate if, in the case of TTD improvement, it can be claimed the improvement of all the tasks making up the TUG test. Two approaches were used, both using the modification of TTD as a diagnostic test for detecting the improvement of all the TUG tasks.

The first one evaluated the diagnostic accuracy of the TTD improvement independently of the amount of the improvement. More specifically, the test was considered positive (thus indicating the improvement of all the TUG tasks) if at least four TTD on discharge were shorter than the shortest TTD recorded on admission. A conventional sensitivity and specificity analysis was thus run and LRs eventually calculated.

The second approach evaluated if the TTD improvement above a certain threshold can be used to claim the improvement of all the TUG tasks. The amount of TTD improvement (TTD improvement ratio, $TTD_{i.ratio}$) was calculated as follow:

$$TTD_{i.ratio} = (TTD_{admission} - TTD_{discharge}) \ / \ TTD_{admission},$$

with both $TTD_{admission}$ and $TTD_{discharge}$ mean of five trials. We preferred ($TTD_{admission} - TTD_{discharge}$) so that larger the $TTD_{i.ratio}$, larger the patient's improvement (the higher, the healthier). The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to evaluate the ability of $TTD_{i.ratio}$ to detect the improvement of all five TUG tasks and LRs were calculated for choosing the optimal threshold of the $TTD_{i.ratio}$.

$LR^+ > 5$ and $LR^- < 0.2$ can be considered *clinically useful* for confirming and excluding a diagnosis, respectively, while $LR^+ > 10$ and $LR^- < 0.1$ are commonly considered *clinically conclusive* [17]. For significance tests and confidence intervals (subscript in square brackets for better readability), the conventional 0.05 type I error probability was chosen.

All analyses were done in R 3.3.0 [20] (packages: effsize [21], epiR [22], OptimalCutpoints [23],plotROC [24], ggplot2 [25], cowplot [26]).

3. Results

The population analysis showed that, on discharge, acute patients significantly shortened their TTD and the duration of each individual TUG task. Chronic patients improved their TTD and all the TUG phases but STS duration (Fig. 2 and Table 2).

Acute patients' improvement was much larger than that of chronic ones, as highlighted by the effect size measure (Fig. 3A). In acute patients, Cohen's d indicated a large improvement of TTD, W1, T1, W2 and TAS and a medium improvement of STS. In chronic patients, Cohen's d indicated a medium improvement of TTD, W1 and TAS, a small improvement of T1 and W2 and a negligible improvement of STS.

At a single-subject level, the probability of a single patient of getting better was larger in acute than chronic patients and, both acute and chronic patients, more likely improved their TTD rather one of the TUG tasks (Fig. 3B).

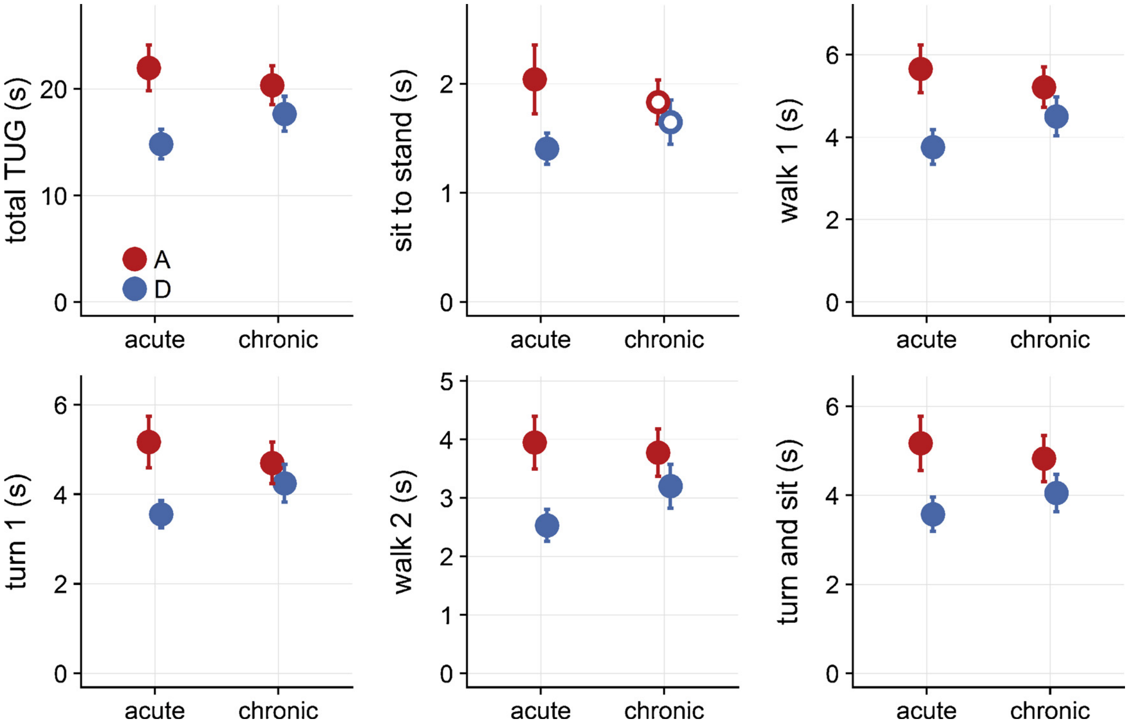

Fig. 2. Population data. Mean (and 0.95 CI for the mean) of the TTD and TUG tasks' duration in acute and chronic patients, on admission (A, red dots) and discharge (D, blue dots). All measures but sit to stand in chronic patients (empty dots) were significantly larger on A than D (filled dots) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

3.1. How do TUG tasks change when TTD shortens? LR analysis of single subjects' data

For this analysis, we used the TTD modification as a test (improved TTD: positive test; unimproved TTD: negative test) and calculated positive and negative LRs to evaluate if this test can be used to claim the improvement of all the TUG tasks. With this regard, among acute patients, only 13 out of 50 (i.e. 26%) improved all five TUG tasks and this percentage is even lower in chronic patients (7 out of 70, i.e. 10%).

The LRs analysis showed that there is no guarantee that all the TUG tasks are shorter (i.e. improved) when TTD is improved, both in acute (LR$^+$: 1.32 $_{[1.10–1.59]}$) and chronic (LR$^+$: 1.85 $_{[1.48–\ 2.33]}$) patients.

However, in chronic patients, if TTD is improved, it can be concluded that there is at least one TUG task that is actually improved (LR$^+$: 9.54 $_{[2.52–36.10]}$). In acute patients, the conclusion that at least one TUG task is improved if TTD is improved is much weaker (LR$^+$: 3.17 $_{[0.98–10.28]}$). On the contrary, if TTD is not improved, there is good evidence that not even a single TUG task is improved, both in chronic (LR$^-$: 0.19 $_{[0.10–0.35]}$) and acute patients (LR$^-$: 0.13 $_{[0.05–0.37]}$).

Fig. 4 shows the ROC curves for the different cutpoints of TTD$_{i.ratio}$. The AUC of the ROC curves was significantly larger than 0.5, both for acute and chronic patients (0.89 $_{[0.80–0.99]}$ and 0.95 $_{[0.90–1.00]}$, respectively). The LRs analysis showed that, in chronic patients, there is strong evidence that patients whose TTD$_{i.ratio}$ is larger than 0.30 actually get better in all five TUG task (LR$^+$: 9.0 $_{[3.97–20.41]}$), while patients whose TTD$_{i.ratio}$ is smaller than 0.30 do not (LR$^-$: 0.16 $_{[0.03–0.97]}$). Similarly, acute patients whose TTD$_{i.ratio}$ is larger than 0.39 likely get better in all five TUG tasks (LR$^+$: 6.26 $_{[2.68–14.61]}$) and acute patients whose TTD$_{i.ratio}$ is smaller than 0.39 likely do not (LR$^-$: 0.18 $_{[0.05–0.64]}$).

At last, among the 59 chronic patients whose TTD$_{i.ratio}$ was smaller than 0.30 (and thus flagged as *unimproved*), there were 36 patients (i.e. 61.0%) who improved at least one of the five TUG tasks. More

Table 2
Total TUG duration and TUG tasks duration, population analysis. Data plotted in Fig. 2 are also given here in tabular form. TTD and TUG tasks are significantly shorter on discharge than admission (Student's t-test, paired sample, $p < 0.001$) with the only exception of STS in chronic patients (Student's t-test, paired sample, $p = 0.14$; lowercase italics). TTD and TUG tasks duration is given in seconds (s). Cohen's d: paired Cohen's d statistics. Probability of getting better: proportion of improved patients according to single subjects analysis.

	TTD and TUG tasks	Admission, mean duration (s) [SD]	Discharge, mean duration (s) [SD]	Cohen's d [0.95 CI]	Cohen's d classification	Probability of getting better [0.95 CI]
Acute patients	TTD	22.0 [7.61]	14.8 [4.90]	1.22 [0.78 - 1.66]	Large	0.82 [0.68 - 0.91]
	STS	2.04 [1.1]	1.40 [0.51]	0.78 [0.36 - 1.19]	Medium	0.58 [0.43 - 0.72]
	W1	5.66 [2.05]	3.76 [1.49]	1.07 [0.64 - 1.50]	Large	0.66 [0.51 - 0.78]
	T1	5.17 [2.03]	3.55 [1.08]	0.91 [0.49 - 1.33]	Large	0.56 [0.41 - 0.70]
	W2	3.95 [1.59]	2.53 [0.96]	1.03 [0.61 - 1.46]	Large	0.54 [0.39 - 0.68]
	TAS	5.17 [2.14]	3.57 [1.35]	0.98 [0.55 - 1.40]	Large	0.64 [0.49 - 0.77]
Chronic patients	TTD	20.3 [7.69]	17.7 [6.89]	0.56 [0.22 - 0.91]	Medium	0.59 [0.46 - 0.70]
	sts	1.83 [0.84]	1.65 [0.86]	0.18 [-0.16 - 0.52]	Negligible	0.33 [0.22 - 0.45]
	W1	5.21 [2.07]	4.51 [1.99]	0.54 [0.19 - 0.88]	Medium	0.40 [0.29 - 0.52]
	T1	4.70 [1.94]	4.25 [1.78]	0.41 [0.07 - 0.75]	Small	0.29 [0.19 - 0.41]
	W2	3.77 [1.70]	3.20 [1.57]	0.48 [0.14 - 0.82]	Small	0.36 [0.25 - 0.48]
	TAS	4.83 [2.16]	4.05 [1.75]	0.59 [0.25 - 0.94]	Medium	0.39 [0.27 - 0.51]

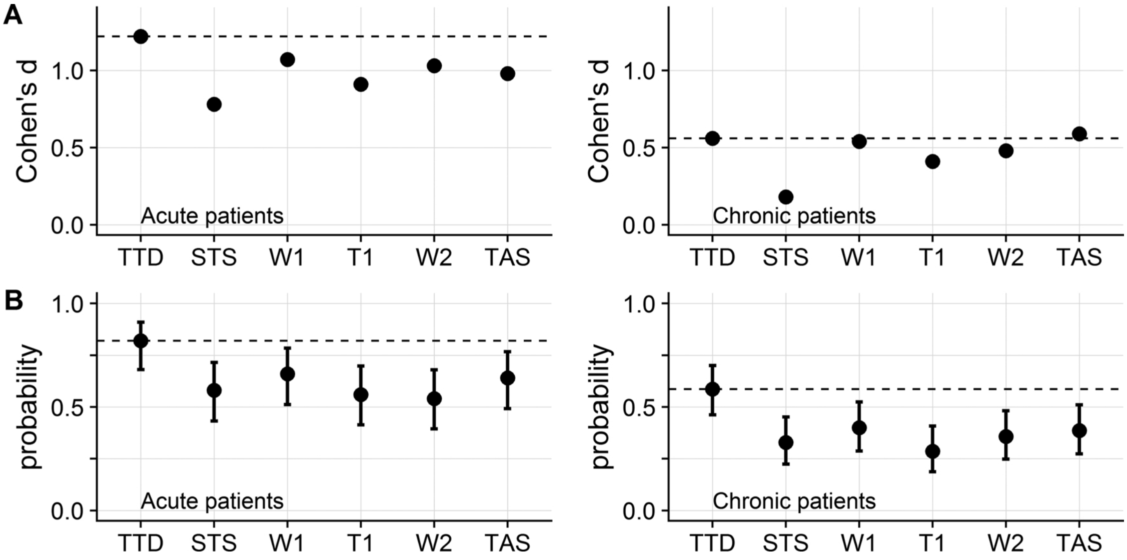

Fig. 3. ITUG responsiveness: population and single subjects analysis. A: Cohen's d. B: proportion (and 0.95 CI) of patients who improved their TTD and TUG tasks' duration.

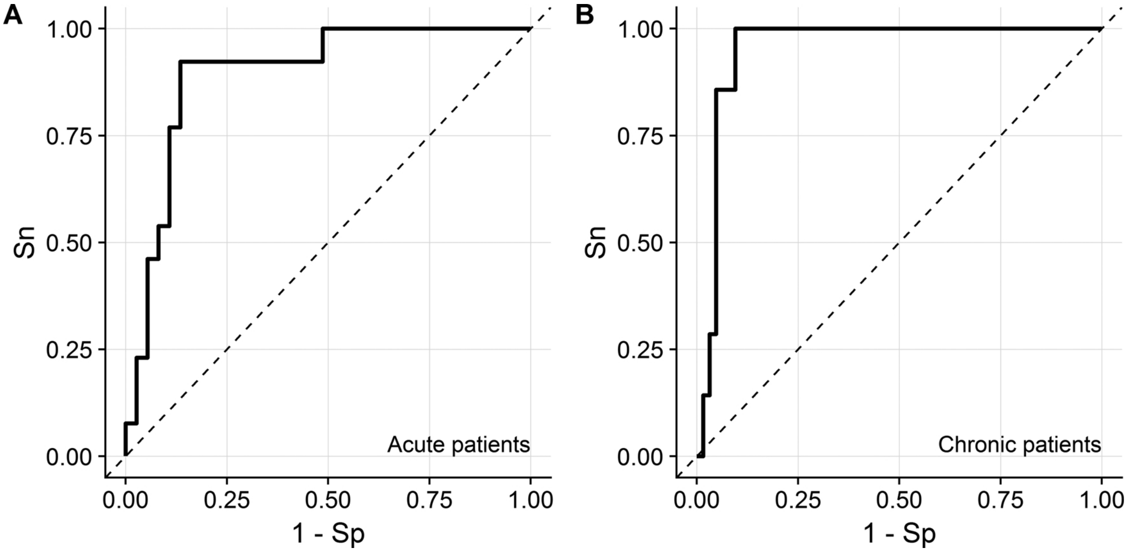

Fig. 4. $TTD_{i.ratio}$ ROC curves. Acute patients, AUC: 0.89 $_{[0.80–0.99]}$. Chronic patients, AUC: 0.95 $_{[0.90–1.00]}$.

strikingly, among the 35 acute patients whose $TTD_{i.ratio}$ was smaller than 0.39 there were 28 (i.e. 80.0%) improving at least one TUG task.

4. Discussion

Aim of the current work was to shed light on how elderly neurological patients with a mobility impairment modify their TUG test after rehabilitation. The TUG test is often used as an outcome measure to evaluate the effectiveness of different treatments (including rehabilitation). As clinicians, we believe that this issue is of essential importance. The TUG test consists of different tasks (i.e. moving from sitting to standing, walking, turning and sitting down) whose importance for patients' everyday life is obvious. If we had shown that the TTD improvement is invariably associated with the improvement of each TUG task, in the face of a patient who shortens his/her TTD the clinician could have concluded that the treatment ameliorated all the fundamental locomotor tasks making up the TUG test.

Unfortunately, this is not the case. Indeed, we have shown that in most cases, TTD improvement is not associated with the improvement of all the TUG tasks. For the clinician this indicates that some of the tasks making up the TUG test did not respond to treatment. In addition, the LRs analysis showed that when TTD has improved it can be said that there is at least one TUG task that has improved. However, which task got better (and which did not) remains unknown.

Knowledge of which task did not ameliorated at treatment end is an important information for the clinician, indicating that the risk associated with the impaired task is still there. In addition, this information can be used in the future for tailoring treatments.

In this study, we used statistics commonly used in the responsiveness theory. Responsiveness is the ability of a measure to detect change over time in the construct to be measured [27]. Simply put, a measure is responsive if able to show the patient's modification when the patient is actually changed (i.e. improved or worsened) [28]. The results presented here can also be discussed from the responsiveness perspective. In a responsiveness study, responsiveness indices (e.g. Cohen's d) are calculated for different measures and the measure with the best responsiveness index (e.g. the largest Cohen's d) is the best measure for showing and measuring patients' modification. According to this framework, our population analysis shows that TTD (i.e. the TUG parameter with the largest Cohen's d) is the most responsive TUG test measure. Accordingly, the single-subject analysis showed that it is highly unlikely that a TUG task has improved if TTD did not get shorter. Note that this last aspect is not a foregone conclusion. In fact it could have happened that the improvement of a short task (e.g. STS) was not apparent as TTD reduction because of the variability of the remaining tasks. The fact that TTD is the most responsive index is of clear interest for the clinician. For a clinician just interested to show that his/her patient *got better* (with no interest on understanding *how* that patient got better), TTD is the most sensitive parameter to show the change.

Responsiveness indices are fundamental for choosing the appropriate sample size when planning clinical trials. For example, consider a researcher interested in testing the effectiveness of a new rehabilitative intervention for elderly patients with a neurological disease. She plans a randomised controlled trial in which the new intervention and the conventional treatment are compared and chooses T1 duration as the main outcome. She expects that the new intervention is twice as effective as the conventional treatment. From the current work, the researcher knows that conventional rehabilitation produces a large improvement of T1 in acute patients and a small improvement in chronic patients. Based on Table 2 data, about 10 patients per group should be recruited if the trial is run on acute patients, while about 250 chronic patients per group should be recruited (type I error probability: 0.05, power: 0.8, independent groups, two tails).

The current results can also be used to propose clinically important differences for the TUG test [29]. We show here that chronic patients whose $TTD_{i.ratio}$ is larger than 0.30 actually get better in all five TUG tasks and that acute patients whose $TTD_{i.ratio}$ is larger than 0.39 likely get better in all five TUG tasks. Given that such improvements are associated with the improvement of all the fundamental tasks making up the TUG test, an improvement ratio larger than 0.30 and 0.39 can thus be proposed as clinically important in chronic and acute patients, respectively.

We are aware that our study has some limitations. The patients' sample is heterogeneous, with patients with different diseases and different mobility impairment grouped together. In addition, we focused our analysis on just TTD (s) and the duration (s) of the different TUG tasks. However, thanks to the ITUG, different movement parameters (e.g. angular velocity) can be obtained from the TUG test [30]. It is obvious that, at the present moment, we cannot exclude that some of these measures have higher responsiveness than that of the TUG tasks duration or even larger responsiveness than that of TTD.

In conclusion, we showed that TTD is the most responsive parameter to show the patient's improvement after rehabilitation. However, in most instances, TTD is not enough to understand how the patient got better. The clinician actually interested in understanding how treatments ameliorate patients' mobility (and thus the TUG test) should measure the duration of the different TUG tasks, for example by using the ITUG.

Contributors and authorship

All authors materially participated in the research or article preparation. Antonio Caronni conceived and designed the study, contributed to data acquisition, analysed and interpreted data, wrote the first version of the manuscript and updated it according to the other Authors suggestions. Irma Sterpi and Massimo Corbo contributed to the interpretation of data and critically revised the manuscript. Paola Antoniotti, Evdoxia Aristidou, Michela Picardi, Giuseppe Pintavalle and Valentina Redaelli collected data, participated to the interpretation of data and revised the first version of the manuscript. All Authors gave their final approval of the submitted manuscript.

Conflict of interest statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] D. Podsiadlo, S. Richardson, The timed "Up & Go": a test of basic functional mobility for frail elderly persons, J. Am. Geriatr. Soc. 39 (1991) 142–148.
[2] S. Mathias, U.S. Nayak, B. Isaacs, Balance in elderly patients: the "get-up and go" test, Arch. Phys. Med. Rehabil. 67 (1986) 387–389.
[3] M.H. Emmelot-Vonk, H.J.J. Verhaar, H.R.N. Pour, A. Aleman, Lock TMTW, Bosch JLHR, et al., Effect of testosterone supplementation on functional mobility, cognition, and other parameters in older men: a randomized controlled trial, JAMA 299 (2008) 39–52.
[4] N. Kerse, K. Peri, E. Robinson, T. Wilkinson, M. von Randow, L. Kiata, et al., Does a functional activity programme improve function, quality of life, and falls for residents in long term care? Cluster randomised controlled trial, BMJ 337 (2008) a1445.
[5] F. Li, P. Harmer, K. Fitzgerald, E. Eckstrom, R. Stock, J. Galver, et al., Tai chi and postural stability in patients with Parkinson's disease, N. Engl. J. Med. 366 (2012) 511–519, https://doi.org/10.1056/NEJMoa1107911.
[6] S. Patel, H. Park, P. Bonato, L. Chan, M. Rodgers, A review of wearable sensors and systems with application in rehabilitation, J. Neuroeng. Rehabil. 9 (2012) 21.
[7] F.-Y. Cheng, Y.-R. Yang, C.-J. Wang, Y.-R. Wu, S.-J. Cheng, H.-C. Wang, et al., Factors influencing turning and its relationship with falls in individuals with Parkinson's disease, PLoS One 9 (2014) e93572.
[8] D.A. Ganz, Y. Bao, P.G. Shekelle, L.Z. Rubenstein, Will my patient fall? JAMA 297 (2007) 77, https://doi.org/10.1001/jama.297.1.77.
[9] C.G. Canning, C. Sherrington, S.R. Lord, J.C.T. Close, S. Heritier, G.Z. Heller, et al., Exercise for falls prevention in Parkinson disease A randomized controlled trial, Neurology 84 (2015) 304–312.
[10] R. Patil, K. Uusi-Rasi, K. Tokola, S. Karinkanta, P. Kannus, H. Sievänen, Effects of a multimodal exercise program on physical function, falls, and injuries in older women: a 2-year community-based, randomized controlled trial, J. Am. Geriatr. Soc. 63 (2015) 1306–1313.
[11] C. Schlenstedt, S. Paschen, A. Kruse, J. Raethjen, B. Weisser, G. Deuschl, Resistance versus balance training to improve postural control in Parkinson's Disease: a randomized rater blinded controlled study, PLoS One 10 (2015) e0140584.
[12] D.E. Beaton, C. Bombardier, J.N. Katz, J.G. Wright, A taxonomy for responsiveness, J. Clin. Epidemiol. 54 (2001) 1204–1217.
[13] L.D. Gillespie, M.C. Robertson, W.J. Gillespie, C. Sherrington, S. Gates, L.M. Clemson, et al., Interventions for preventing falls in older people living in the community, Cochrane Database Syst. Rev. 9 (2012).
[14] J.I. Hoff, A.A. v/d Plas, E.A.H. Wagemans, J.J. Van Hilten, Accelerometric assessment of levodopa-induced dyskinesias in Parkinson's disease, Mov. Disord. 16 (2001) 58–61.
[15] S. Bolink, B. Grimm, I.C. Heyligers, Patient-reported outcome measures versus inertial performance-based outcome measures: a prospective study in patients undergoing primary total knee arthroplasty, Knee 22 (2015) 618–623.
[16] A. Caronni, I. Sterpi, P. Antoniotti, E. Aristidou, F. Nicolaci, M. Picardi, et al., Criterion validity of the instrumented timed up and go test: a partial least square regression study, Gait Posture (2018).
[17] A. Caronni, C. Cattalini, A.M. Previtera, Balance and mobility assessment for ruling-out the peripheral neuropathy of the lower limbs in older adults, Gait Posture 50 (2016) 109–115.
[18] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd edn., Hillsdale, New Jersey: L, 1988.
[19] B.J. Byiers, J. Reichle, F.J. Symons, Single-subject experimental design for evidence-based practice, Am. J. Speech Lang. Pathol. 21 (2012) 397–414.
[20] R Core Team, R: A Language and Environment for Statistical Computing, (2017).
[21] M. Torchiano, effsize: Efficient Effect Size Computation, (2017).
[22] M.S. Telmo Nunes, C. Heuer, J. Marshall, J. Sanchez, R. Thornton, J. Reiczigel, et al., epiR: Tools for the Analysis of Epidemiological Data, (2017) with contributions from.
[23] M. López-Ratón, M.X. Rodríguez-Álvarez, C.C. Suárez, F.G. Sampedro, {OptimalCutpoints}: An {R} package for selecting optimal cutpoints in diagnostic tests, J. Stat. Softw. 61 (2014) 1–36.
[24] M.C. Sachs, {plotROC}: a tool for plotting roc curves, J. Stat Software, Code Snippets 79 (2017) 1–19, https://doi.org/10.18637/jss.v079.c02.
[25] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2009.
[26] Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2", (2017).
[27] L.B. Mokkink, C.B. Terwee, D.L. Patrick, J. Alonso, P.W. Stratford, D.L. Knol, et al., The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes, J. Clin. Epidemiol. 63 (2010) 737–745.
[28] H.C.W. De Vet, C.B. Terwee, L.B. Mokkink, D.L. Knol, Measurement in Medicine: a Practical Guide, Cambridge University Press, 2011.
[29] A. Caronni, L. Sciumè, Is my patient actually getting better? Application of the McNemar test for demonstrating the change at a single subject level, Disabil. Rehabil. 39 (2017), https://doi.org/10.1080/09638288.2016.1194486.
[30] D. Vervoort, N. Vuillerme, N. Kosse, T. Hortobágyi, C.J.C. Lamoth, Multivariate analyses and classification of inertial sensor data to identify aging effects on the timed-up-and-go test, PLoS One 11 (2016) e0155984.