Gait & Posture 73 (2019) 246–250

Contents lists available at ScienceDirect

Gait & Posture

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

Assessing a novel way to measure three common rehabilitation outcome measures using a custom mobile phone application

Donald H. Lein Jr.[a,*], James H. Willig[b], Christian R. Smith[c], Jeffrey R. Curtis[d], Andrew O. Westfall[e], Christopher P. Hurt[a]

[a] Department of Physical Therapy, University of Alabama at Birmingham, Birmingham, AL, United States
[b] Division of Infectious Disease, University of Alabama at Birmingham, Birmingham, AL, United States
[c] School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States
[d] Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, United States
[e] Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, United States

ARTICLE INFO

Keywords:
Phone application
Clinical functional measures
Sit-to-stand
Timed-up-and-go
Validity
Mobility

ABSTRACT

Background: Clinicians often use thirty-second-sit **(chair)**-to-stand (30CST), timed-up-and-go (TUG), and the five-times-sit-to-stand (5xSTS) since these outcome measures (OMs) are sensitive for strength, balance and mobility. Research Question: The purpose of this study was to validate a custom smart phone application (App) that can remotely assess the 30CST, TUG, and 5xSTS.

Methods: Thirty-one healthy adults (range: 22–55 y; 54.6–106.8 kg; 160–185 cm; 19 females) participated in this cross-sectional study. Each participant performed the 30CST, TUG, and 5xSTS at a slow and normal speed. They performed each OMs twice while the App collected their performance data using both an iOS and Android phone. The gold standard of each test was the average of the silent count of two investigators for the 30CST and the time recorded by two investigators using stopwatches for the TUG and 5xSTS. Investigators analyzed the data using Intraclass Correlation coefficients (ICC), Pearson R coefficients, Signed Rank Tests, and Wilcoxon Rank-Sum Tests.

Results and Significance: A significant correlation was observed between the performances recorded by the phones and the direct observation gold standard for all three OMs ($r > 0.97$). For 30CST, no significant mean count differences were found for the following comparisons: between phones, within phone types, or within phone-by-speed levels. (P-values range 0.06–1.00). While a statistically significant difference was found in all of the time comparisons when performing TUG and 5xSTS ($p < 0.0001$) except for the between phone comparison with TUG ($p = 0.27$). For TUG and 5xSTS, the time difference was less than a second when compared to the gold standard and ICCs showed moderate to strong agreement when comparing the phone application to the gold standard (ICCs range 0.60–0.99). These data suggested that the App could validly measure performance of these OMs.

1. Introduction

Clinical measures of muscle function and mobility can be sensitive prognostic and diagnostic tests for individuals with strength, gait, and balance dysfunction [1–5]. Clinicians often use the thirty-second sit **(chair)**-to-stand ([30CST], strength), timed-up-and-go ([TUG], mobility, balance and fall risk), and five-times-sit-to-stand ([5xSTS], strength and power) to assess aspects of mobility because these outcome measures (OMs) are easy and cost effective to perform [6]. These measures are underused outside clinical and research settings. Given the focus on increased efficiency and lowering costs in health care, the ability to accurately monitor patients' function at home may mean that more observation and ultimately care can be performed there, at less cost. Furthermore, longitudinally monitoring performance measures of patients with acute or chronic conditions could also lead to earlier detection of functional decline and prompt timely intervention.

Smart phones provide an attractive platform to develop functional assessments for home monitoring of patients due to their increasing ubiquity worldwide [7] and their ease of use [8]. Recently, several groups of investigators found that smartphone-based digital biomarkers were reliable, feasible, and valid [9–11]. No researchers have explored the ability of smartphones to measure the 30CST, TUG, and 5xSTS, tests

* Corresponding author at: Department of Physical Therapy University of Alabama at Birmingham, 1720 2nd Avenue South, Birmingham, AL 35294, United States.
E-mail address: dlein@uab.edu (D.H. Lein).

https://doi.org/10.1016/j.gaitpost.2019.07.251
Received 6 November 2018; Received in revised form 11 July 2019; Accepted 18 July 2019
0966-6362/

Table 1
Summary statistics.

Thirty-Second Sit-to-Stand Test Count	Phone			Gold Standard			Difference between Phone and Gold Standard			p-value#
	N	Mean (Std)	Median (q1, q3)	N	Mean (Std)	Median (q1, q3)	N	Mean (Std)	Median (q1, q3)	
Overall										
Total	113	11.41 (5.65)	10.00 (7.00, 15.00)	118	11.69 (5.81)	11.00 (7.00, 16.00)	113	−0.24 (1.35)	0.00 (0.00, 0.00)	0.13
Phone										
iOS	60	11.70 (5.32)	11.00 (7.00, 16.00)	60	11.78 (5.48)	11.00 (7.00, 15.50)	60	−0.08 (1.09)	0.00 (0.00, 0.00)	1.00
Android	53	11.08 (6.03)	10.00 (7.00, 15.00)	58	11.60 (6.18)	10.00 (7.00, 16.00)	53	−0.42 (1.59)	0.00 (0.00, 0.00)	0.04
Phone and Activity Speed										
iOS slow	30	7.80 (2.38)	7.00 (6.00, 10.00)	30	7.80 (2.43)	7.00 (6.00, 10.00)	30	0.00 (1.44)	0.00 (0.00, 0.00)	0.31
Android slow	27	6.85 (2.84)	7.00 (5.00, 8.00)	29	7.03 (2.78)	7.00 (5.00. 9.00)	27	−0.07 (0.27)	0.00 (0.00, 0.00)	0.50
iOS normal	30	15.60 (4.52)	16.00 (13.00, 18.00)	30	15.77 (4.73)	15.50 (13.00, 18.00)	30	−0.17 (0.59)	0.00 (-1.00, 0.00)	0.23
Android normal	26	15.46 (5.29)	15.00 (11.00, 19.00)	29	16.17 (5.17)	16.00 (13.00, 19.00)	26	−0.77 (2.21)	0.00 (-1.00,0.00)	0.09
Five-times Sit-to-Stand Test Time										
Overall										
Total	112	13.30 (5.01)	13.20 (9.73, 16.40)	118	12.59 (5.11)	12.29 (8.69, 15.32)	112	0.75 (0.54)	0.77 (0.57, 0.96)	< 0.0001
Phone										
iOS	60	13.23 (4.37)	13.26 (9.74, 16.82)	60	12.42 (4.36)	12.40 (8.82, 15.58)	60	0.81 (0.66)	0.88 (0.70, 1.05)	< 0.0001
Android	52	13.39 (5.70)	13.11 (9.73, 16.17)	58	12.77 (5.81)	11.94 (8.69, 15.32)	52	0.68 (0.35)	0.66 (0.48, 0.85)	< 0.0001
Phone and Activity Speed										
iOS slow	30	16.34 (3.54)	16.80 (13.55, 18.59)	30	15.53 (3.53)	15.16 (12.73, 17.54)	30	0.81 (0.75)	0.88 (0.67, 1.02)	< 0.0001
Android slow	26	17.30 (5.28)	16.17 (13.52, 20.30)	29	16.78 (5.45)	15.32 (13.33, 19.61)	26	0.67 (0.36)	0.66 (0.47, 0.79)	< 0.0001
iOS normal	30	10.12 (2.51)	10.01 (8.17, 11.47)	30	9.31 (2.50)	9.17 (7.42, 10.76)	30	0.81 (0.56)	0.90 (0.71, 1.09)	< 0.0001
Android normal	26	9.48 (2.57)	9.73 (7.68, 10.80)	29	8.75 (2.39)	8.69 (6.99, 10.08)	26	0.69 (0.34)	0.65 (0.49, 0.89)	< 0.0001
Timed –Up-and–Go Test										
Overall										
Total	113	10.25 (3.16)	9.60 (7.79, 11.92)	118	9.34 (3.14)	8.76 (6.86, 11.31)	113	0.90 (0.39)	0.87 (0.71, 1.08)	< 0.0001
Phone										
iOS	59	10.10 (2.99)	9.35 (7.97, 11.70)	60	9.08 (2.98)	8.41 (6.83, 10.87)	59	1.01 (0.44)	1.00 (0.82, 1.13)	< 0.0001
Android	54	10.41 (3.36)	9.94 (7.62, 12.59)	58	9.61 (3.30)	9.11 (6.97, 11.84)	54	0.78 (0.28)	0.76 (0.60, 0.96)	< 0.0001
Phone and Activity Speed										
iOS slow	30	12.35 (2.46)	11.54 (10.48, 14.65)	30	11.35 (2.51)	10.87 (9.14, 13.53)	30	1.01 (0.53)	1.04 (0.82, 1.15)	< 0.0001
Android slow	27	12.88 (2.92)	12.59 (10.44, 14.44)	29	12.09 (2.78)	11.84 (9.84, 13.36)	27	0.71 (0.28)	0.66 (0.50, 0.87)	< 0.0001
iOS normal	29	7.77 (1.08)	7.97 (7.06, 8.42)	30	6.82 (1.11)	6.86 (6.07, 7.48)	29	1.01 (0.33)	0.94 (0.83, 1.13)	< 0.0001
Android normal	27	7.94 (1.36)	7.62 (7.16, 8.46)	29	7.12 (1.31)	6.97 (6.44, 7.63)	27	0.84 (0.26)	0.81 (0.71, 0.99)	< 0.0001

Non-parametric Signed Rank tests for paired data were used to test if the observed differences were different from 0. A separate test was used for each row in the table.

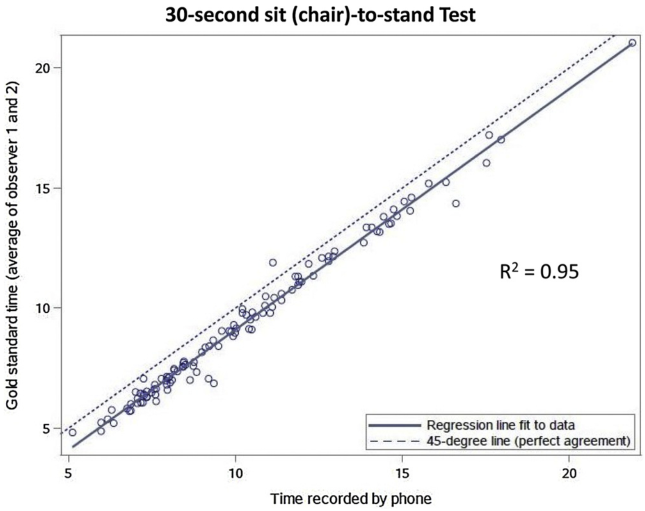

Fig. 1. Correlations between the time-based metrics of tasks in this investigation. Data is plotted with respect to a line of identity which represents perfect agreement in observer and phone metrics. Correlation of TUG was R = 0.99(95% CI 0.96 – 0.98; R^2 = 0.95.

commonly used by clinicians in various patient settings. The purpose of this research study is to validate a smart phone application that monitors the before-mentioned OMs.

2. Methods

2.1. Study Population

Before participating, all enrolled adults provided informed consent for this institutional review board approved study. Investigators selected a convenience sample of graduate students. They recruited participants of any gender, race or ethnicity, 18–60 years of age, and reported no health issues that would preclude them completing the three OMs.

2.2. Procedure

This application (CGI Group Inc., Montreal, Quebec) uses the angular rotation measured about the z-axis of the phone's coordinate system that counted an event when the phone passed 56-degrees and would not count another event until the phone decreased to less than 40-degrees. They assumed that the thigh moves 90-degrees from sitting (the horizontal axis) to standing (the vertical axis). The chosen values ensured that the thigh moves over halfway to full extension (i.e. > 45-degrees), needed for standing, and another event would not be recorded by the phone until the thigh moves to less than 45-degrees, more than halfway to sitting. Participants performed all measures with two investigators present. Participants performed the three OMs at two speeds, normal and slow. Investigators instructed the participants to perform the slow-paced test at less speed than the normal-paced test. The same two phones were used to assess these OMs; both sixth generation smartphones for iOS (iPhone 6) and an Android product (Samsung Galaxy S6). The three OMs were performed separately for each phone. Participants first familiarized themselves with using the application. The user initialized the software and placed the phone in their front pants/shorts pocket. Investigators randomized test sequence, speed selection, and sequence of phone used for measurement to avoid an order effect. The investigators told participants to wear pants with front pockets on the day of their data collection and the investigators made note of how the phones fit in their pocket (tight, normal or lose). Participants rested 30 to 60 s between OMs. Participants used the same armless chair (seat height = 45 cm) for all tests.

Table 2
Intraclass coefficient (ICC) values with 95% confidence interval (95% CI) for within group mean differences between observers and phone application measured performance in thirty-second sit-to-stand count (30CST), five-time sit-to-stand time (5xSTS), and timed-up-and-go (TUG) performance measures using iOS and Android smartphones tested at slow and normal walking speeds.

	30CST ICC (95% CI)	5xSTS ICC (95% CI)	TUG ICC (95% CI)
Overall			
Total (pooling results from both phones)	0.97 (0.97, 0.98)	0.98 (0.98, 0.99)	0.95 (0.94, 0.96)
Phone			
iOS	0.98 (0.97, 0.98	0.97 (0.96, 0.98)	0.93 (0.92, 0.95)
Android	0.96 (0.95, 0.97)	0.99 (0.99, 0.99)	0.97 (0.96, 0.98)
Phone, and Activity Speed			
iOS slow	0.83 (0.77, 0.88)	0.95 (0.94, 0.97)	0.90 (0.87, 0.93)
Android slow	1.00 (0.99, 1.00)	0.99 (0.99, 0.99)	0.96 (0.95, 0.98)
iOS normal	0.99 (0.99, 0.99)	0.93 (0.90, 0.95)	0.60 (0.51, 0.70)
Android normal	0.90 (0.86, 0.94)	0.95 (0.93, 0.97)	0.80 (0.73, 0.86)

When performing 30CST, participants performed as many sit-to-stand-to-sit cycles as possible in 30-seconds from the study chair. For the 30CST, both investigators silently counted the number of full sit-to-stand-to-sit sequence in 30-seconds and recorded our counts before looking at the phone measure.

For the TUG test, participants started from the chair in a seated position. An individual initiated movement when they heard the audible beep of the phone, walked around a cone placed 3 m from the chair, returned to the chair and sat down. Both investigators started stopwatches on participant movements and stopped the time when participants contacted the seat at the end of the test. The software started the timer on an audible beep and stopped the timer when the person sat back down.

For the 5xSTS, participants were instructed to sit-to-stand-to-sit five times from the chair without using their hands. Participants started and ended the test in a seated position. Both investigators monitored time the same way as the TUG except investigators stopped their watches when participants contacted the seat during the fifth repetition and the software stopped the timer when the person completed five sit-to-stand cycles.

2.3. Analysis

Intraclass correlation coefficients (ICC) were calculated to evaluate the interrater reliability between the 2 observers as well as for agreement between gold standard and phone application within all phones, within each phone type and at normal and slow speeds within each phone. Because the ICCs were very high between the observers, the average of the 2 observers' results was used as the gold standard. The investigators also calculated Pearson R correlations between the gold standard and time or count recorded by the phones. For each trial, the difference between the phone app value and the gold standard average observer value was calculated. Non-parametric Signed Rank tests for paired data were used to test if the observed differences were different from 0 within phone types and within phone*speed levels. Two-sided non-parametric Wilcoxon Rank-Sum tests were used to test for a difference in the differences between phone types. All analyses were performed using SAS v. 9.4 (SAS, Cary, NC).

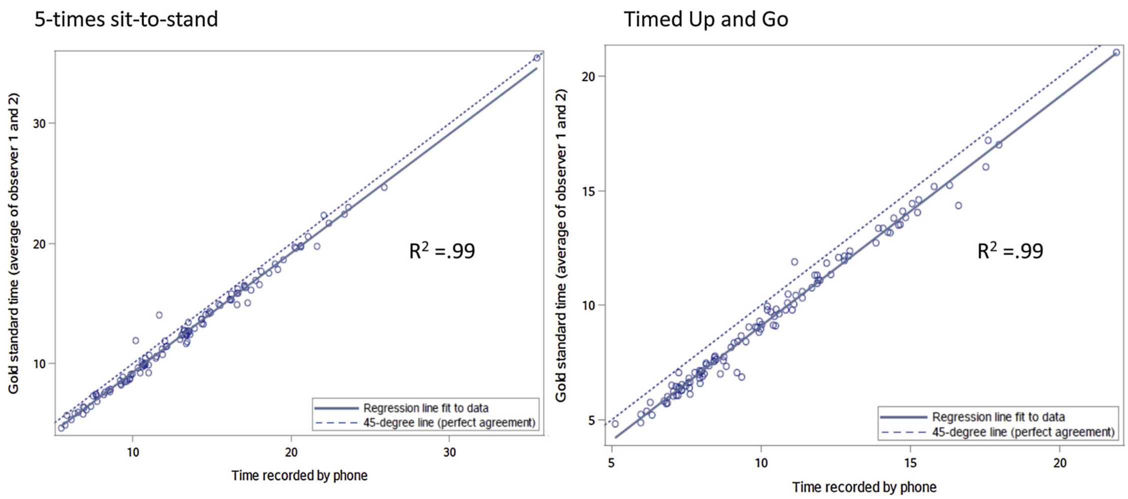

Fig. 2. Correlations between the time-based metrics of tasks in this investigation. Data is plotted with respect to a line of identity which represents perfect agreement in observer and phone metrics. A strong correlation suggests the presence of a constant offset between these measurements for most observations: 5xSTS (r = 0.99, 95% CI 0.991 – 0.996) (R2 = 0.99); TUG (r = 0.99, 95% CI 0.990 – 0.994) (R2 = 0.99).

3. Results

Thirty-one healthy adults (19 females and 12 males), aged 22–55 years, weight ranging from 54.55 to 106.82 kg, and height extending 160.02–185.42 cm were enrolled. Between observer ICC values were greater than 0.98 for all 3 OMs. Mean differences of 1 count or second or less were observed when comparing the mean of both phone recordings to direct observation gold standard for all OMs (Table 1).

For 30CST, investigators found no significant difference within both phone's and each phone's mean counts and the gold standard or the mean count difference within each phone when compared to the observer's counts at both slow and normal paced performances (Table 1). In addition, the investigators found a strong and positive correlation between the gold standard and phones (r = 0.97, 95% CI 0.96 – 0.98) (R^2 = 0.95) (Fig. 1). The investigators also found ICC values ranging from 0.83 to 1.00 with the majority of the values over 0.90 (6/7, 85.7%) (Table 2) These results suggested strong agreement within both phone's and each phone's mean times and the gold standard or the mean time difference within each phone when compared to the observer's time at both slow and normal paced performances and gold standard. **Bland-Altman** plots supported the ICC findings (Appendices 1 and 2). A Wilcoxon signed-rank test showed a non-significant mean count difference between phones for 30CST when combining data for both speeds (Z = -1.36, p = 0.18).

For both TUG and 5xSTS, there were significant differences when comparing the phone application's mean time to the gold standards mean time for all within group analyses (Table 1). However, ICC values showed moderate to strong agreement for both phone's and each phone's mean times and the gold standard or the mean time difference within each phone when compared to the observer's time at both slow and normal paced performances and gold standard. ICC values ranged from 0.60 to 0.99 with the majority of values equal to or above 0.90 (12/14, 85.7%) (Table 2). In addition, strong and positive correlation was observed between the gold standard and time or count recorded by the phones: TUG (r = 0.99, 95% CI 0.990 – 0.994) (R2 = 0.99); 5xSTS (r = 0.99, 95% CI 0.991 – 0.996) (R^2 = 0.99). (Fig. 2). Fig. 2 shows that a consistent distance exists between the observed linear regression line and perfect-fit line in both tests suggesting a systematic error. Bland-Altman plots showed that most if not all of the points plotted above the 0-difference line for all the within comparisons performed for these two outcome measures (Appendices 3 to 6). This would represent a systematic bias suggesting that the phone always overestimated the time. Finally, Wilcoxon signed rank tests showed no significant mean time difference between the phones for TUG (Z = 1.10, p = 0.27) but with the 5xSTS a significant mean time difference was found between the phones (Z = -3.50, p = 0.0007) when combing data for both speeds in both analyses. The Android phone performed better than the IOS when measuring the 5xSTS.

4. Discussion

The study was the first step in the validation of a novel smart phone application that can assess the 30CST, TUG, and 5xSTS in non-clinical (e.g. home) settings. For the 30CST, the phone application performed well under all conditions. For the two time-based OMS, the results were statistically different from the investigators' measurements. However, the observed data varied in a consistent manner (R2 = .99, See Fig. 1, and Appendix 1) and moderate to strong agreement was observed between the gold standard and the phone application for all within comparisons. We believe this was due to a delay of when timing of the task began, between the phone (on the tone) and when the participant initiated movement. Modifying the application to begin on movement instead of a tone could account for the mean difference between the gold standard and the application observed between these two OMS improving the accuracy and precision of the phone when compared to the gold standard.

Recently, investigators successfully examined remote monitoring by smart phones for other performance OMs providing evidence for the feasibility of this approach [8–10]. This same phone application successfully counted steps when compared to a gold standard [12]. While sample size, **lack of application public availability**, testing only healthy participants, pocket type and a single institutional setting limit our pilot study, we believe that this phone application, with further development and testing, could lend further support that smartphone-based digital biomarkers are a feasible and valid method to monitor patients. Further research is also needed to determine optimal implementation of this technology in a home-based setting.

Translation of this technology to routine patient care may decrease cost to the health system, by facilitating remote monitoring and early detection of negative trends in function. In addition to the timely detection of functional decline by this technology, clinician or algorithm aided monitoring of results could trigger early patient driven interventions, decreasing time to care and engaging patients in performing more of their own care.

Author statement

Author involved in the conception and design of the study— Donald H Lein Jr, James H Willig, Christian R Smith, Jeffrey R Curtis, Andrew O Westfall, Jonathan Cortis, Clayton Rice, Christopher P Hurt

Author involved in acquisition of data— Donald H Lein Jr, Christian R Smith, Christopher P Hurt

Author involved in analysis and interpretation of data— Donald H Lein Jr, James H Willig, Jeffrey R Curtis, Andrew O Westfall, Christopher P Hurt

Author involved in drafting of the article— Donald H Lein Jr, James H Willig, Jeffrey R Curtis, Christopher P Hurt

Author involved in revising it critically for important intellectual content— Donald H Lein Jr, James H Willig, Christian R Smith, Jeffrey R Curtis, Andrew O Westfall, Jonathan Cortis, Clayton Rice, Christopher P Hurt

Author involved in final approval of the version to be submitted— Donald H Lein Jr, James H Willig, Christian R Smith, Jeffrey R Curtis, Andrew O Westfall, Jonathan Cortis, Clayton Rice, Christopher P Hurt

Funding

No funding was provided for this work.

Declaration of Competing Interest

All of the authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.gaitpost.2019.07.251.

References

[1] E.V. Applebaum, D. Breton, Z.W. Feng, A.T. Ta, K. Walsh, K. Chasse, et al., Modified 30-second sit to stand test predicts falls in a cohort of institutionalized older veterans, PLoS One 12 (2017) e0176946.

[2] G. Spagnuolo, C. Faria, B.A. da Silva, A.C. Ovando, J. Gomes-Osman, A. Swarowsky, Are functional mobility tests responsive to group physical therapy intervention in individuals with Parkinson's disease? NeuroRehabilitation (2018).

[3] B. Pieper, T.N. Templin, A. Goldberg, A comparative study of the five-times-sit-to-stand and timed-up-and-go tests as measures of functional mobility in persons with and without injection-related venous ulcers, Adv. Skin Wound Care 27 (2014) 82–92 quiz 3-4.

[4] N. Millor, P. Lecumberri, M. Gomez, A. Martinez-Ramirez, M. Izquierdo, An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit, J. Neuroeng. Rehabil. 10 (2013) 86.

[5] T.B. Hafsteinsdottir, M. Rensink, M. Schuurmans, Clinimetric properties of the timed Up and Go test for patients with stroke: a systematic review, Top Stroke Rehabil. 21 (2014) 197–210.

[6] Academy of Neurologic Physical Therapy. Academy of Neurologic PT Outcome Measures Recommendations, (2018).

[7] https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. Accessed 29th April 2019.

[8] L. Rainie, A. Perrin, 10 Facts About Smartphones as the Iphone Turns 10, (2017).

[9] F. Lipsmeier, K.I. Taylor, T. Kilchenmann, D. Wolf, A. Scotland, J. Schjodt-Eriksen, et al., Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial, Mov. Disord. (2018).

[10] S. Arora, V. Venkataraman, A. Zhan, S. Donohue, K.M. Biglan, E.R. Dorsey, et al., Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study, Parkinsonism Relat. Disord. 21 (2015) 650–653.

[11] A. Zhan, S. Mohan, C. Tarolli, R.B. Schneider, J.L. Adams, S. Sharma, et al., Using smartphones and machine learning to quantify Parkinson disease severity: the Mobile Parkinson disease score, JAMA Neurol. (2018).

[12] C.P. Hurt, D.H. Lein Jr., C.R. Smith, J.R. Curtis, A.O. Westfall, J. Cortis, et al., Assessing a novel way to measure step count while walking using a custom mobile phone application, PLoS One 13 (2018) e0206828.