Journal of Parkinson's Disease xx (20xx) x–xx
DOI 10.3233/JPD-202416
IOS Press

Review

Digital Health Technology to Measure Drug Efficacy in Clinical Trials for Parkinson's Disease: A Regulatory Perspective

Leonard Sacks* and Elizabeth Kunkoski*
Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA

Accepted 14 December 2020
Pre-press 11 January 2021

Abstract. Digital health technology (DHT), including wearable and environmental sensors, video cameras and other electronic tools, has provided new opportunities for the measurement of movement and functionality in Parkinson's disease. Compared to current standards for evaluation of the disease (MDS-UPDRS), DHT may offer new possibilities for more frequent objective measurements of the duration, severity and frequency of disease manifestations over time, that may provide more information than periodic clinic visits. However, DHT measurement are only scientifically and medically useful if they are accurate, reliable and clinically meaningful. Verification and validation, also known as analytical validation and clinical validation, of DHT performance is important to ensure the accuracy and precision of measurements, and the specificity of findings. Given the wide range of clinical manifestations associated with Parkinson's disease and the many tools and metrics to assess them, the challenge is to identify those that may represent a standard for use in clinical trials, and to confirm when digital measurements succeed or fall short of capturing meaningful benefits during drug development.

Keywords: Parkinson's disease, digital health technology, wearable, mobile technology, drug development, FDA, regulatory

Digital health technology (DHT), that includes wearable and environmental sensors, video cameras and other electronic tools to evaluate disease remotely, has provided new opportunities for the measurement of movement and functionality. Accelerometers are present in our actigraphy gadgets, smart watches and smart phones. They can also be customized to provide a continuous 3-dimensional measurement of limb and trunk movement in patients that would not be observed during an examination [1]. Some commercial systems combine accelerometers, gyroscopes and magnetometers into an algorithm, known as inertial measurement units (IMUs) to analyse spatio-temporal parameters [2]. Video cameras, wearable systems, gloves, and other environmental sensors can capture activity and movement.

The impact of Parkinson's disease (PD) on movement is profound and protean. Current standards for evaluating disease severity (e.g., Movement Disorder Society - Unified Parkinson's Disease Rating Scale, MDS-UPDRS) rely on subjective reporting, and some disagreement has been shown between assessments made by investigators and those made by study subjects [3, 4]. Disease rating scales are

*Correspondence to: Leonard Sacks, MD, Associate Director, Clinical Methodology, Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration, 10993 New Hampshire Ave, Bldg, 51, Rm 3342, Silver Spring, MD 20993, USA. Tel.: +1 301 796 8502; E-mail: Leonard.Sacks@fda.hhs.gov and Elizabeth Kunkoski, MS, Health Scientist Policy Analyst, Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration, 10993 New Hampshire Ave, Bldg, 51, Rm 3346, Silver Spring, MD 20993, USA. Tel.: +1 301 796 6439; E-mail: Elizabeth.Kunkoski@fda.hhs.gov.

ISSN 1877-7171/21/$35.00

Table 1
Sensors Being Investigated for Measurement of Parkinson's Disease Symptoms and Assessments (Adapted from Rovini et al. and Rusz et al. [5, 6])

Parkinson's Disease Symptom/Assessment	Sensors and sensor-derived measurements	Type of DHT incorporating sensors
Tremor	accelerometer, electromyograph (EMG), gyroscope, inertial measurement unit (IMU)	smart clothes, smart phone, smart watch, wearable glove systems
Gait and Timed Up and Go (TUG) Test	accelerometer, electrocardiogram (ECG), force sensor, galvanic skin resistance (GSR) sensor, gyroscope, IMU	smart phone, smart watch
Freezing of Gait	ECG, electroencephalogram (EEG), EMG, force sensor, GSR sensor, IMU	earphones, headsets, smart phone, smart watch
Postural Instability	accelerometer, force sensor, gyroscope, IMU	smart phone, smart watch
Upper Limb Motion	accelerometer, EMG, force sensor, gyroscope, IMU	fingers, gloves, pens, smart phone, smart watch, wrists
Other Gait Symptoms (leg agility, rigidity, arm swing)	accelerometer, EMG, inertial sensor	smart phone, smart watch
Motor Fluctuations and On/Off Phases	accelerometer, ECG, EMG, gyroscope, IMU, micro-electromechanical system (MEMS)	smart home, smart phone, smart watch
Functionality Assessments	accelerometer, EMG, environmental sensor, gyroscope, IMU, MEMS	placed in the home environment, smart phone, smart watch, wearable systems
Speech Assessments	acoustic sensor	smart phone, smart watch

also limited by the periodicity of the measurement and recall bias. DHTs now offer the potential for objective measurement of tremor, gait deficits, freezing of gait, postural instability, upper limb motion, leg agility, rigidity, and motor fluctuations. Besides movement, abnormalities in cadence, tonal variation and fluency of speech in PD patients can also be analysed by DHTs. Table 1 describes the spectrum of sensors that are being investigated to capture symptoms and other assessments of PD [5, 6]. In addition to these passive measurements, sensors have been used to challenge patient performance. Tapping tests on a cell phone have distinguished patients with PD from healthy controls while motion detectors can be used to perform timed up-and-go tests [7, 8].

There is a large spectrum of DHT available for use in a clinical trial. Some DHT meet the definition of a medical device under the Federal Food, Drug and Cosmetic Act while some DHT do not [9, 10]. Generally, clearance or approval of the DHT for use in a clinical trial conducted under an Investigational New Drug (IND) application or an Investigational Device Exemption (IDE) [11, 12] is not a requirement unless the DHT will be marketed independently or as a combination product.

With seemingly limitless combinations of measurements, sensors, tests and algorithms, how are we to choose those that are most useful for drug development?

The first goal is to obtain measurements that have been demonstrated to be accurate and reliable over time and across patients, leading to solid scientific conclusions on drug efficacy. For the approval of drugs, the Food, Drug and Cosmetic Act requires that substantial evidence of effectiveness be provided that would allow experts to conclude that a drug would have the effect described in labelling [13]. To satisfy this regulation, DHTs should allow for a well-defined and reliable assessment of a patient's response to treatment. Verification and validation are important to confirm the accuracy and precision of measurements. Analytical validation and clinical validation ensure the reliability of algorithms that translate accelerometry or other sensor readings into clinical observations (e.g., tremor, falls, steps) [14]. Current standards for measurement of drug effect rely largely on patient reported outcomes, neurological examinations and face-to face consultations and these are useful benchmarks against which to evaluate new measurements.

Not all the manifestations of PD are equally amenable to measurement by DHTs. Some aspects of functionality may be best assessed during in-person visits. On the other hand, for some manifestations, DHT measurements may outperform the assessments of observers. Experiments manipulating the intensity of deep-brain stimulation in PD patients have shown that some sensor measurements are more sensitive and less variable than human scoring on components

of the UPDRS. Some disease features such as freezing of gait are unpredictable and difficult to assess during clinic visits. These have been successfully evaluated using DHT in laboratory conditions and DHT may allow for greater opportunities for detection when used to monitor patients at home [15].

Clinical validation of DHT should ensure that individuals with and without PD can be clearly distinguished. Depending on the characteristic being measured, specificity of measurements may be challenging. Tremor in PD varies in frequency and amplitude. A study by Hossen et al. showed that accelerometers failed to distinguish between the tremor of PD and essential tremor in 10% of subjects [16]. In a review of studies comparing accelerometers to video-recordings to capture freezing of gait, validity values ranged from 73 to 100% for sensitivity, and from 67 to 100% for specificity. The authors concluded that there is a lack of consistency in outcomes measured, methods of assessing validity, and reported results. Given these limitations, the validation of sensor-derived assessments of PD features would benefit from increased collaboration among researchers, aligning data collection protocols, and sharing data sets [17]. Many DHT measurements are in the early stages of research and warrant larger sample sizes with patients with varying stages of disease progression.

Shortfalls in the specificity of sensors suggest that using more than one modality of measurement may be an important strategy, just as current clinical scoring systems measure many facets of the disease. In a study using smartphones to measure five different tasks (voice test, posture test, gait test, finger tapping test and reaction time test) performed by 10 patients with PD and 10 healthy controls, the mean sensitivity of the smartphone measurements for detection of PD was 96.2% (SD 2%) and mean specificity was 96.9% (SD 1.9%) [18]. While multiple measurements add to the richness and specificity of the assessment, the statistical plan used to determine the outcome of a trial using multiple measurements must be prespecified to avoid the pitfalls of multiplicity and the increasing risk of false positive findings when many tests of efficacy are combined [19].

The ability to capture the impact of known effective treatments is another indication that the DHT will be useful in evaluating new treatments. Using wrist and ankle sensors, Pulliam et al. were able to quantify the effect of a dose of levodopa on tremor, bradykinesia and dyskinesia in 13 patients with PD. The measurements made by these sensors correlated with video-recording evaluations made by clinicians [20]. Investigators have reported the ability to distinguish on from off periods, which is another indication the drug effect is captured [21].

A second goal is to ensure that trial endpoints involving DHT measurements represent clinically meaningful responses to a drug; interpreted in FDA regulations as those with an impact on how patients feel, function or survive. The clinical benefit of some sensor readings is self-evident. Weiss et al. found that a 3-day sample of gait recordings using a 3-D accelerometer placed in the middle of the back served as a predictor of falls within the next year [22]. The clinical meaningfulness of other sensor measurements may be less obvious. For example, measurements of tremor may not reflect the functional impairments that patients find most disabling. Early engagement of patients is a cornerstone in determining the relevance of endpoints that involve functional measurements made by DHT [23]. "The Voice of the Patient" is part of FDA's Patient-Focused Drug Development initiative to incorporate perspectives from patients, caretakers and other patient representatives on the most significant effects of PD on their daily lives and experiences with currently available therapies [24]. In addition to patients, engagement of a variety of stakeholders, including caregivers, disease experts and regulatory authorities would be necessary to determine the meaningfulness of certain measurements in a clinical trial.

Challenge tests are helpful to assess activities of daily living in patients with PD. Using a mobile app, Zhan et al. challenged individuals with and without PD to perform various tasks reflecting speech, dexterity, gait, balance, and reaction time and used machine-learning on these tasks to create a PD severity score. The authors aimed to provide a clinically meaningful assessment of patients in their real-world environments [25]. Extensive research has been conducted with machine learning to analyse and predict freezing of gait, tremors and falls. Machine learning algorithms provide new opportunities for long-term monitoring of a drug's effectiveness as well as disease progression [26, 27].

Selection of the metrics best suited to disease evaluation presents another challenge. Just looking at gait characteristics using machine learning, Rehman et al. identified five different clinical characteristics (step velocity, mean step length, step length variability, mean step width, and step width variability) that classified PD [28]. Among the plethora of possible measurements, principle component analyses have

been helpful to whittle down to those that account for most of the variance in the data [29].

Given the complexity of PD and the innumerable possible measurements that can be made, the challenge is to find those that best reflect meaningful responses to treatment and that can be used as a standard in clinical studies. What is an optimal sampling interval to obtain a stable estimate of function? Do we focus on average measurements or outlying measurements? Where do we position sensors, and how many do we need? Not all the pathological features will be captured even when multiple sensors are used, and drugs may also only affect some of these features. Different measurements may be needed for different stages of the disease, and for drugs with different mechanisms of action.

Finally, DHTs should be useable and safe for study participants. In general, large uncomfortable wearables, or DHTs that require fine-motor skills to use them and those that need technological know-how are unlikely to get the necessary cooperation from patients. DHTs need to be physically safe to use, electronically secure, and trustworthy when recording personally identifiable information. FDA regulations are designed to ensure the safety and welfare of subjects enrolled in clinical investigations, detailing requirements for safety reporting and Institutional Review Board supervision and allowing clinical holds when "human subjects are or would be exposed to an unreasonable and significant risk of illness or injury" [30].

From a regulatory perspective, adequate and well-controlled studies are the basis of determining whether there is "substantial evidence" to support the claims of effectiveness for new drugs [31]. The comparative structure of a clinical trial is important to be able to conclude that the measured effect can be attributed to the drug. Randomized and blinded trials showing superiority of the investigational treatment to control, inherently confirm that the sensor is detecting an effect. Such studies may involve parallel arm controls or crossover within individuals. Absent substantial experience with sensors, non-inferiority studies are likely to be difficult to interpret since the effects of comparator drug on the proposed sensor measurements may not be known. Consequently, we may not know whether both arms were effective or ineffective.

Besides their potential for scientific improvements in measurement, DHTs may be able to gather much of the needed study measurements from participants in their home environments, offering a new dimension of convenience for patients. Such decentralized clinical trials may make it much easier for patients with mobility challenges and other personal and practical obstacles in getting to study sites, to participate in clinical research.

Sensors cannot capture certain aspects of a face-to-face interview, a physical examination of patient's balance or muscle tone. With increasing use of digital measurements in clinical trials, it will be important to ensure that we do not ignore these aspects of the disease. There are situations where sensors are more accurate and sensitive than human raters, and situations where human raters are more discerning and specific than sensors. Careful studies will be needed to demonstrate when digital measurements succeed or fall short of capturing meaningful benefits during drug development.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

DISCLAIMER

The opinions expressed in this article are those of the authors and are not intended to reflect the position of the Food and Drug Administration.

REFERENCES

[1] Brognara L, Palumbo P, Grimm B, Palmerini L (2019) Assessing gait in Parkinson's disease using wearable motion sensors: A systematic review. *Diseases* **7**, 18-32.

[2] Esser P, Dawes H, Collett J, Feltham MG, Howells K (2012) Validity and inter-rater reliability of inertial gait measurements in Parkinson's disease: A pilot study. *J Neurosci Methods* **205**, 177-181.

[3] Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stern MB, Dodel R, Dubois B, Holloway R, Jankovic J, Kulisevsky J, Lang AE, Lees A, Leurgans S, LeWitt PA, Nyenhuis D, Olanow CW, Rascol O, Schrag A, Teresi JA, van Hilten JJ, LaPelle N, Movement Disorder Society UPDRS Revision Task Force (2008) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov Disord* **23**, 2129-2170.

[4] Seidel SE, Tilley BC, Huang P, Palesch YY, Bergmann KJ, Goetz CG, Swearingen CJ (2012) Subject-investigator reproducibility of the Unified Parkinson's Disease Rating Scale. *Parkinsonism Relat Disord* **18**, 230-233.

[5] Rovini E, Maremmani C, Cavallo F (2017) How wearable sensors can support Parkinson's disease diagnosis and treatment: A systematic review. *Front Neurosci* **11**, 555-596.

[6] Rusz J, Hlavnička J, Tykalová T, Novotný M, Dušek P, Šonka K, Růžička E (2018) Smartphone allows capture of speech abnormalities associated with high risk of

developing Parkinson's disease. *IEEE Trans Neural Sys Rehabil Eng* **26**, 1495-1507.

[7] Lee CY, Kang SJ, Hong S-K, Ma H-I, Lee U, Kim YJ (2016) A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson's disease. *PLoS One* **11**, e0158852.

[8] Palmerini L, Mellone S, Rocchi L, Chiari L (2011) Dimensionality reduction for the quantitative evaluation of a smartphone-based timed up and go test. *Annu Int Conf IEEE Eng Med Biol Soc* **2011**, 7179-7182.

[9] The Federal Food, Drug and Cosmetic Act section 201(h) Definitions; generally (21 USC 321(h)), https://uscode.house.gov/view.xhtml?req=(title:21section:321edition:prelim). Accessed on October 15, 2020.

[10] Food and Drug Administration. Ask a question about digital health regulatory policy, https://www.fda.gov/medical-devices/digital-health/ask-question-about-digital-health-regulatory-policies. Dated November 2019, Accessed on December 1, 2020.

[11] Code of Federal Regulations Title 21 – Food and Drugs, Part 312 Investigational New Drug Application (21 CFR 312), https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=312. Accessed on October 15, 2020.

[12] Code of Federal Regulations Title 21 – Food and Drugs, Part 812 Investigational Device Exemptions (21 CFR 812), https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=812. Accessed on October 15, 2020.

[13] The Federal Food, Drug and Cosmetic Act section 505(d) Grounds for refusing Application; Approval of Application; "Substantial Evidence" defined (21 U.S.C. §355(d)). https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title21-section355&num=0&edition=prelim. Accessed on October 15, 2020.

[14] Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, Godfrey A, Godino JG, Gujar N, Izmailova E, Manta C, Peterson B, Vandendriessche B, Wood WA, Wang KW, Dunn J (2020) Verification, analytical validation, and clinical validation (V3): The foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). *NPJ Digit Med* **3**, 55.

[15] Heldman DA, Espay AJ, LeWitt PA, Giuffrida JP (2014) Clinician versus machine: Reliability and responsiveness of motor endpoints in Parkinson's disease. *Parkinsonism Relat Disord* **20**, 590-595.

[16] Hossen A, Muthuraman M, Al-Hakim Z, Raethjen J, Deuschl G, Heute, U (2013) Discrimination of Parkinsonian tremor from essential tremor using statistical signal characterization of the spectrum of accelerometer signal. *Biomed Mater Eng* **23**, 513-531.

[17] Silva de Lima AL, Evers LJW, Hahn T, Bataille L, Hamilton JL, Little MA, Okuma Y, Bloem BR, Faber MJ (2017) Freezing of gait and fall detection in Parkinson's disease using wearable sensors: A systematic review. *J Neurol* **264**, 1642-1654.

[18] Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, Little MA (2015) Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. *Parkinsonism Relat Disord* **21**, 650-653.

[19] Dmitrienko A, D'Agostino RB Sr, (2018) Multiplicity considerations in clinical trials. *N Engl J Med* **378**, 2115-2122.

[20] Pulliam CL, Heldman DA, Brokaw EB, Mera TO, Mari ZK, Burack MA (2018) Continuous assessment of Levodopa response in Parkinson's disease using wearable motion sensors. *IEEE Trans Biomed Eng* **65**, 159-164.

[21] Heijmans M, Habets JGV, Herff C, Aarts J, Stevens A, Kuijf ML, Kubben PL (2019) Monitoring Parkinson's disease symptoms during daily life: A feasibility study. *NPJ Parkinsons Dis* **5**.

[22] Weiss A, Herman T, Giladi N, Hausdorff JM (2014) Objective assessment of fall risk in Parkinson's disease using a body-fixed sensor worn for 3 days. *PLoS One* **9**, e96675.

[23] Clinical Trials Transformation Initiative. CTTI Recommendations: Effective engagement with patient groups around clinical trials, https://www.ctti-clinicaltrials.org/files/pgctrecs.pdf. Dated October 2015, Accessed on December 1, 2020.

[24] Food and Drug Administration, The voice of the patient – Parkinson's Disease, https://www.fda.gov/media/124392/download, Dated April 2016, Accessed on December 1, 2020.

[25] Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, Elson MJ, Spear KL, Glidden AM, Little MA, Terzis A, Dorsey ER, Saria S (2018) Using smartphones and machine learning to quantify Parkinson disease severity: The mobile Parkinson disease score. *JAMA Neurol* **75**, 876-880.

[26] Pardoel S, Kofman J, Nantel J, Lemaire ED (2019) Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: A review. *Sensors (Basel)* **19**, 5141-5177.

[27] Pedrosa TI, Vasconcelos FF, Medeiros L, Silva LD (2018) Machine learning application to quantify the tremor level for Parkinson's disease patients. *Procedia Computer Science* **138**, 215-220.

[28] Rehman RZU, Del Din S, Guan Y, Yarnall AJ, Shi JQ, Rochester L (2019) Selecting clinically relevant gait characteristics for classification of early Parkinson's disease: A comprehensive machine learning approach. *Sci Rep* **9**, 17269.

[29] Palmerini L, Mellone S, Rocchi L, Chiari L (2011) Dimensionality reduction for the quantitative evaluation of a smartphone-based timed up and go test. *Annu Int Conf IEEE Eng Med Biol Soc* **2011**, 7179-7182.

[30] Code of Federal Regulations Title 21 – Food and Drugs, Part 312 Investigational New Drug Application, Clinical holds and requests for modification, Grounds for imposition of clinical hold (21 CFR 312.42(b)(i)), https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=312.42. Accessed on October 15, 2020.

[31] Code of Federal Regulations Title 21 – Food and Drugs, Part 314 Applications for FDA Approval to Market a New Drug, Adequate and well-controlled studies (21 CFR 314.126), https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.126. Accessed on October 15, 2020.