R E V I E W

New Methods for the Assessment of Parkinson's Disease (2005 to 2015): A Systematic Review

Álvaro Sánchez-Ferro, MD, MSc,[1,2]* Morad Elshehabi, MD, MSc,[3,4] Catarina Godinho, PhD,[5,6,7] Dina Salkovic, MD, MSc,[3,4] Markus A. Hobert, MD,[3,4] Josefa Domingos, MSc,[5,7] Janet MT. van Uem, MSc,[3,4] Joaquim J. Ferreira, MD, PhD,[5,7,8] and Walter Maetzler, MD[3,4]

[1]HM CINAC, Hospital Universitario HM Puerta del Sur, Móstoles, Madrid, Spain
[2]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]Center for Neurology and Hertie Institute for Clinical Brain Research (HIH), Department of Neurodegeneration, University of Tübingen, Tübingen, Germany
[4]DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany
[5]Clinical Pharmacology Unit, Instituto de Medicina Molecular, Lisbon, Portugal
[6]Center of Interdisciplinary Research Egas Moniz (CiiEM), Instituto Superior de Ciências da Saúde Egas Moniz, Monte de Caparica, Portugal
[7]CNS-Campus Neurológico Sénior, Torres Vedras, Portugal
[8]Laboratory of Clinical Pharmacology and Therapeutics, Faculty of Medicine, University of Lisbon, Portugal

ABSTRACT: Background: The past decade has witnessed a highly dynamic and growing expansion of novel methods aimed at improving the assessment of Parkinson's disease with technology (NAM-PD) in laboratory, clinical, and home environments. However, the current state of NAM-PD regarding their maturity, feasibility, and usefulness in assessing the main PD features has not been systematically evaluated.
Methods: A systematic review of articles published in the field from 2005 to 2015 was performed. Of 9,503 publications identified in PubMed and the Web of Science, 848 full papers were evaluated, and 588 original articles were assessed to evaluate the technological, demographic, clinimetric, and technology transfer readiness parameters of NAM-PD.
Results: Of the studies, 65% included fewer than 30 patients, <50% employed a standard methodology to validate diagnostic tests, 8% confirmed their results in a different dataset, and 87% occurred in a clinic or lab. The axial features domain was the most frequently studied, followed by bradykinesia. Rigidity and nonmotor domains were rarely investigated. Only 6% of the systems reached a technology level that justified the hope of being included in clinical assessments in a useful time period.
Conclusions: This systematic evaluation provides an overview of the current options for quantitative assessment of PD and what can be expected in the near future. There is a particular need for standardized and collaborative studies to confirm the results of preliminary initiatives, assess domains that are currently underinvestigated, and better validate the existing and upcoming NAM-PD. © 2016 International Parkinson and Movement Disorder Society

Key Words: Parkinson disease; technology assessment; biomedical; diagnosis; inventions

---

Recently there has been an exponential development of novel methods to improve the assessment of Parkinson's disease (PD) by using different technologies (NAM-PD; Supplementary Figure). Wearable sensors, accelerometers, gyroscopes, magnetometers, smartphones, and a plethora of other electronic systems have been employed to obtain physiological signals related to motor and nonmotor aspects of the disease.[1-3] Although some of the used technologies are not new (the first accelerometer was invented by George Atwood in 1783), they all represent methods that have recently been investigated and used for the

assessment of PD. The reasons for this increased adoption are beyond the scope of this review, but all of these techniques attempt to provide an objective means to measure aspects of the disease.[4,5] Their intended applications range from the improvement and prediction of diagnosis to the assessment of disease progression and the evaluation of care and therapy efficacy.[6]

This systematic review presents a comprehensive evaluation of the NAM-PD investigated in the past 10 years. Our efforts should be considered in light of other recent initiatives.[2,4,7-9] In addition to a narrative[2] and a systematic review[9] by some of the authors listed here, a systematic review on wearable technology for the assessment of gait and balance has recently been published.[7] None of these reviews provide an exhaustive and comprehensive overview of the current state of the NAM-PD field and how far (potentially clinically relevant) quantitative assessment strategies have been developed. This review combines the strengths of all of these previous initiatives. It includes information about both wearable and nonwearable technologies; covers the assessment approaches of the most relevant disease domains and complications; follows a strict, transparent, and exhaustive systematic approach by, for example, including MeSH and non-MeSH terms; presents clinimetric information and—for the first time, to our knowledge—the technology readiness level (TRL[10] of every system reported in the studies evaluated; and provides a spreadsheet of these data extracted from the papers.

After reading this article, clinicians, people with Parkinson's disease (PwP), and other stakeholders will have an overview of how quantitative assessments of PD have developed during the course of the recent decade of innovation, what the options are to date, and what we can expect in the near future from these technologies. To fulfill this latter purpose in particular, a list of systems that are close to being translated into practical use is provided. In addition, the above-mentioned systematic spreadsheet containing the extracted information on the evaluated systems can serve as a living document that will be regularly updated by the authors. Interested readers are encouraged to use this resource to complement the manuscript, which in fact can only provide some initial steps of analysis.

Methods

A systematic search for articles published between January 1, 2005, and January 4, 2016, was conducted in 2 electronic databases (PubMed/Medline and Web of Science). Using a Boolean search structure, 3 blocks of keywords were introduced and connected. The first block was related to the disease ("*Parkinsonian Disorders,*" "*Parkinson(s) disease,*" "*Parkinson Disease, Secondary,*"

TABLE 1. New technologies for the assessment of Parkinson's disease (NAM-PD) published during the past decade (2005-2015) by disease domain

Domain	Number of references describing the NAM-PD (% of all included)
Motor	504 (85.7)
Axial features	212 (36.1)
Bradykinesia	126 (21.4)
Tremor	81 (13.8)
Speech	42 (7.1)
Activity	32 (5.4)
Rigidity	11 (1.9)
Nonmotor[a]	24 (4.1)
Cognition	12 (2.0)
Sleep	9 (1.5)
ANS	2 (0.3)
Neuropsychiatric features	1 (0.2)
Smell	0 (0.0)
Treatment complications	20 (3.4)
Miscellaneous	40 (6.8)
Total, included in the results	588
Not included in the full evaluation process[b]	260
Total number of references evaluated	848

ANS, autonomous nervous system.
[a]Detailed information about studies in different domains is available at http://bit.ly/pd-technologies-database and is presented in a common nonmotor domain in the Results section.
[b]These are review papers, technologies used for other applications (eg, schizophrenia and essential tremors), or references in which a technology used was discarded after the evaluation.

"*Basal Ganglia Diseases,*" "*Parkinsonism*"). The second block included 45 technology terms related to PD assessment (eg, "*Technology /-ies,*" "*Diagnostic Techniques, Neurological,*" "*Assessment,*" "*Patient Outcome Assessment,*" "*Accelerometer,*" "*Smartphone*"). The third block was composed of different keywords describing 12 disease domains. These 12 domains were based on clinical (eg, cardinal PD features) and technical reasoning and on previous reviews.[2,9,11,12] Six domains referred to motor aspects (and are discussed separately in the Results section), 5 to nonmotor aspects, and 1 to treatment-related complications (eg, wearing-off and dyskinesias). Table 1 provides an overview of the number of studies investigating these domains. The vast majority of the identified and evaluated references referred to motor aspects; therefore, the focus of the initial analyses in this manuscript is on these domains. Nonmotor features and treatment-related complications were relatively underrepresented and are therefore not discussed in detail. Papers that were not covered by these 12 domains were collected in a miscellaneous category. This category includes platforms (eg, telemonitoring and telemedicine systems), studies evaluating driving skills, and a breath test to distinguish PwP from controls. No detailed results are reported for this category because of its heterogeneity. A complete description of the search structure and keywords used is provided in the supplementary material section.
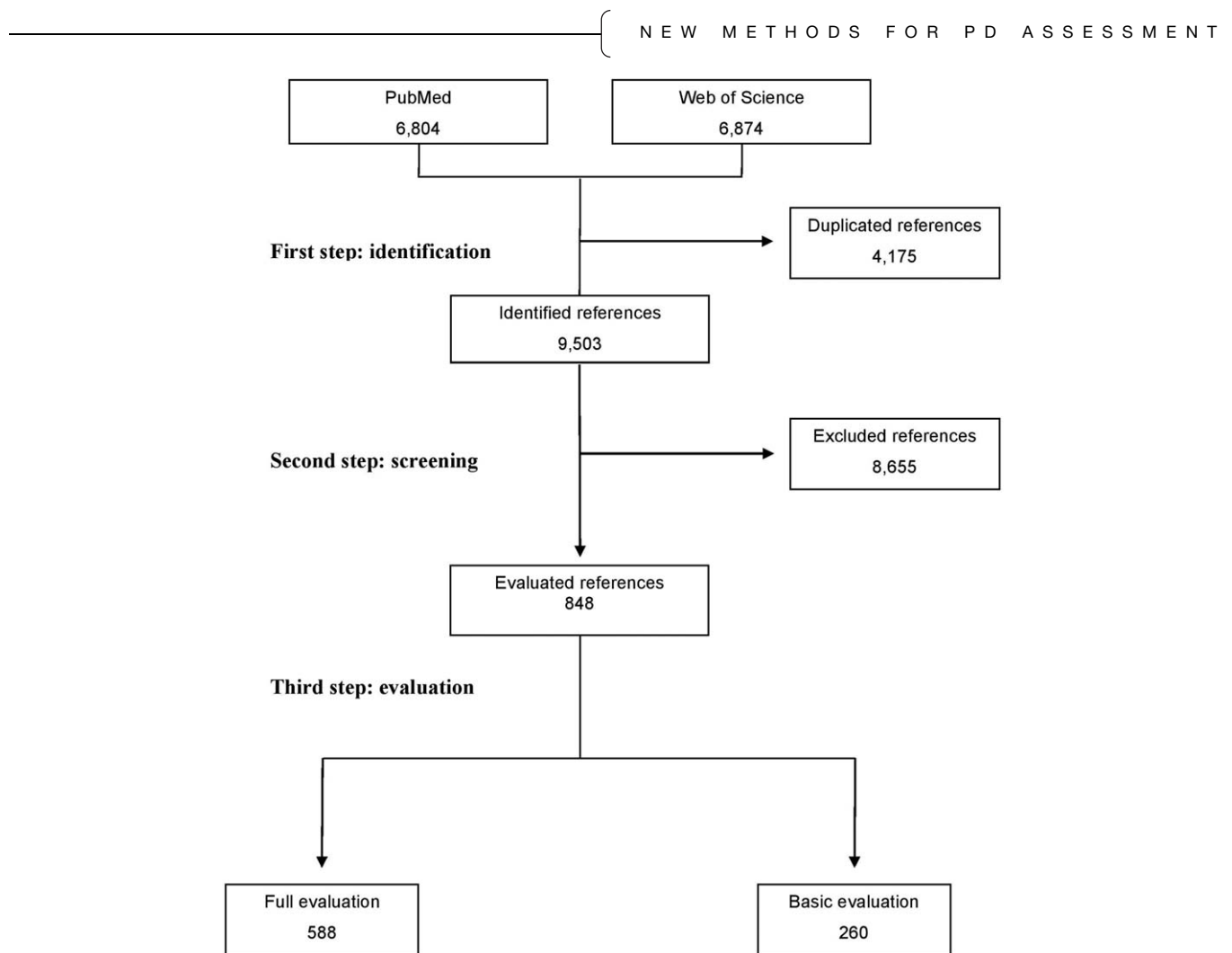
FIG. 1. Flowchart of the systematic search and review process of novel technologies to improve the assessment of Parkinson's disease (NAM-PD) in the past decade (2005-2015).

The following 3 steps were then performed (Figure 1):

- Identification step: A total of 6,874 papers in the Web of Science and 6,804 in PubMed/Medline were identified. After eliminating duplicates, 9,503 papers entered the next step.
- Screening step: One author (A.S.F.) screened the titles and abstracts. The criteria for the articles included in the next step were (i) use of an NAM-PD, (ii) published in English, and (iii) used in humans.

The exclusion criteria were the following: (i) the use of only questionnaires/scales, (ii) electromyography, (iii) electroencephalography, (iv) imaging techniques, (v) molecular diagnostic/therapeutic methods, (vi) genetic tests, (vii) established speech recognition systems (ie, computerized speech lab), (viii) new developments in deep brain stimulation (DBS), and (ix) magnetic transcranial stimulation. A total of 848 articles entered the next step.

- Evaluation step: This step was performed by 7 authors (A.S.F., M.E., C.G., D.S., M.A.H.,

J.M.T.v.U., and J.D.) using a standardized web-based form designed by A.S.F. (http://bit.ly/pd-technologies). The full text of the 848 articles were evaluated. Articles underwent a basic evaluation (author, journal, volume, pages, year, and type of article) and were not evaluated more extensively (and were not included in the analysis of this review) if they (i) were reviews (n = 39), (ii) reported a therapeutic application of NAM-PD (n = 65), (iii) did not include NAM-PD (n = 129), and (iv) were not used to assess PD (n = 27).

From the remaining 588 original articles, the type of NAM-PD used and the type of researcher using the NAM-PD (developers vs nondevelopers) were defined. The use of an NAM-PD by a nondeveloper reinforced its validation because it confirmed that the NAM-PD can be applied in different settings. In addition, its use by a nondeveloper usually indicates that a system is more ready to be used. Moreover, demographic characteristics (eg, sample size, stage of the disease, inclusion of a control group, and patients with other

diseases) were extracted. Then, the clinimetric characteristics were evaluated. The selection of clinimetric characteristics was based on the literature[13-16] and included accuracy (ie, the degree to which a measurement represents the true value of the attribute being measured, eg, by providing sensitivity and specificity measures[17], agreement (ie, how well a test measures the event of interest when compared with a standard, eg, by using Bland-Altman analyses,[18] intraclass correlation coefficients,[17] and Kappa values[17], responsiveness (sensitivity to change), and repeatability (the stability of the metric over time in a population without an expected change). Despite being used in many studies, pure statistical correlations (ie, Spearman's Rho, Pearson's R, or r^2 from regression models) were not considered adequate to test the clinimetric properties of an NAM-PD and hence are not reported in the results.[17,19]

Moreover, as mentioned previously, we extracted information about the TRL of each original article included in the full evaluation. The TRL scale was developed by the National Aeronautics and Space Administration and has recently been adapted by the European Commission for the Horizon 2020 initiative[10] (Supplementary Table). This scale can be applied by different stakeholders (eg, researchers, clinicians, patients, regulatory and funding agencies, and industrial partners) as a tool to determine how close a technology is to being translated and for what use. Its use in biomedical applications has recently begun.

Development of the Systematic Spreadsheet and Its Proposed Use

Detailed information about all 848 evaluated articles is available at http://bit.ly/pd-technologies-database. The systematic table derived from our evaluation is shared publicly for several purposes. First, interested readers are offered the opportunity to investigate any facet of the current state of NAM-PD because the analyses presented in this manuscript are necessarily limited to certain meta aspects and cannot provide detailed answers on all aspects. Second, the authors of the evaluated technologies can peer review the information systematized by our group. In case of any inconsistency, we are prepared to adapt the spreadsheet accordingly. Third, it can serve as a living document: Because the form used is also shared at http://bit.ly/pd-technologies, any upcoming studies are offered the opportunity to be included in this spreadsheet. To provide an example of the opportunities the spreadsheet provides, we focus in the Results section on the presentation of frequency, type of technology used, demographic aspects, clinimetric properties, and TRL aspects of studies sorted by different motor features. In addition, we include a paragraph each on nonmotor aspects and treatment complications. The data are presented as absolute numbers and, in parentheses, the percentages of the total of all studies that provided the information of interest.

Results

Of the 588 fully evaluated articles, 208 (35%) reported on the validation of NAM-PD by developers, 183 (31%) on NAM-PD performance evaluated by nondevelopers, 157 (27%) on mathematical algorithms, and 40 (7%) on future NAM-PD platforms/ideas.

Studies using wearable NAM-PD covered a relevant proportion of the research performed in this field (211 [37%] included this type of technology), and the number of published studies showed a substantial relative increase during the period evaluated. For example, wearable NAM-PD were used in 3 (18%) of the studies published in 2005 and in 38 (44%) of the studies published in 2015. Of the studies, 45 (8%) confirmed their results in a different dataset.

Basic disease severity information using Hoehn and Yahr stages was reported in 247 (42%) of the studies. PwP in stage 2 were most often included (in 216 of those studies), followed by those in stage 3 (193), 2.5 (154), 1 (120), 4 (77), and 5 (11).

The ages of the participants were included in 394 (76%) of the studies that tested individuals, and 110 studies (20%) included information on the statistical assumptions (eg, whether the data followed a Gaussian distribution).

The technologies used were accelerometers and gyroscopes alone or combined, which we categorized as inertial measurement units (IMUs), digitizing tablets, optoelectronic systems, electromagnetic trackers, force plates, virtual reality systems, software-based technologies, smartphones, audio recorders, insoles, computer vision systems, smart gloves, videogame consoles and ad hoc solutions.

NAM-PD for the Assessment of Motor Aspects

Axial Features

Number of Studies. A total of 70 articles (33%) reported on NAM-PD used by nondevelopers, 68 (32%) on validation aspects, 65 (31%) on a mathematical algorithm, and 9 (4%) on future NAM-PD platforms/ideas.

Type of Technologies. The most frequently used systems (106 articles, 50%) were IMUs, followed by optoelectronic systems (20 articles, 9%).

Demographic Aspects. Of the evaluated studies, 147 (71%) included 30 or fewer PwP. The most frequently included mean/median age group ranged from 61 to 70 years (101 studies, 69%), followed by 71 to 80

years in 26 (18%) of the studies, and 51 to 60 years in 20 (14%) of the studies.

Clinimetric Properties. A total of 49 studies (23%) performed data diagnostics (eg, normality assumption tests) prior to conducting statistical analyses. The most frequently reported reference standard was the motor section of the Unified Parkinson's Disease Rating Scale (UPDRS-III), which was employed in 33 (16%) of the studies. Accuracy was reported in 98 studies (48%). The sensitivity and specificity of the NAM-PD were higher than 80% in 27 (28%) and 35 (36%) of them, respectively. The areas under the receiver operating characteristics curve (AUCs) and the predictive values/likelihood ratios were provided in 8 (4%) and 5 (2%) of the studies, respectively. Agreement with a reference standard was described in 26 (13%) of the articles. Of these articles, 19 (73%) provided evidence of adequate agreement. Responsiveness was investigated in 31 (15%) of the articles. The most frequent method used to induce a change was pharmacological intervention, which was present in 13 (42%) of these studies. Repeatability was assessed in 6 articles (3%).

Technology Readiness. Of the studies, 85 (45%) included NAM-PD with a TRL ≤ 3, and NAM-PD with TRL9 were used in 18 (8%) of the studies.

Bradykinesia

Number of Studies. Of the studies, 72 (57%) reported on the validation of NAM-PD in characterizing bradykinesia. Thirty-two (25%) involved NAM-PD used by nondevelopers, 17 (13%) tested a mathematical algorithm for NAM-PD, and 5 (4%) described an NAM-PD or idea.

Type of Technologies. The most frequently used systems were IMUs (39 studies, 31%), followed by a combination of various methods (14 studies, 11%).

Demographic Aspects. Of the studies, 89 (71%) included 30 or fewer PwP. The most frequently included mean/median age group range was 61 to 70 years (73 studies, 85%). With one exception, the mean/median age of the groups was between 50 and 80 years in all studies.

Clinimetric Properties. Of the studies, 18 (15%) performed data diagnostics prior to conducting any statistical analyses. The most frequently reported reference standard was the UPDRS-III, which was employed in 41 (34%) of the studies, followed by expert diagnosis (eg, classification of PD according to established criteria) in 36 (30%) studies.

The accuracy of the NAM-PD was reported in 45 (37%) of the studies. The sensitivity and specificity of

the results generated by NAM-PD were higher than 80% in 10 (22%) and 7 (15%) of the studies, respectively. Of the studies, 11 (9%) and 5 (4%) provided AUCs and predictive values/likelihood ratios, respectively. Agreement with a reference standard was described in 10 (8%) of the studies. A total of 8 (60%) of them provided evidence of an adequate agreement. Responsiveness and repeatability were investigated in 32 (26%) and 6 (5%) studies, respectively. For the former, the most frequent method applied to induce a change was a pharmacological intervention (21 studies, 66%).

Technology Readiness. A total of 94 studies (74%) included NAM-PD with a TRL ≤ 3, and 9 (7%) of the studies used NAM-PD that reached TRL9.

Tremors

Number of Studies. A total of 33 studies (41%) reported the validation of NAM-PD in characterizing tremors, 23 (28%) tested a mathematical algorithm for NAM-PD, 20 (25%) involved NAM-PD used by nondevelopers, and 5 (6%) described an NAM-PD or an idea.

Type of Technologies. The most frequently used systems were IMUs (38 studies, 47%), followed by smartphones, which were used in 11 (14%) of the studies.

Demographic Aspects. A total of 52 (67%) studies included 30 or fewer PwP. The most frequent mean/median age group in the studies ranged from 61 to 70 years (29 studies, 10%), and in all studies, the mean/median age of the participants was between 50 and 80 years.

Clinimetric Properties. A total of 15 studies (20%) performed data diagnostics prior to conducting statistical analyses. The most frequently reported reference standard was PD classification according to established criteria (14 studies, 18%), followed by the UPDRS-III (12 studies, 16%).

Accuracy was reported in 35 studies (46%). The sensitivity and specificity were higher than 80% in 17 (49%) and 14 (40%) studies, respectively. AUCs and predictive values/likelihood ratios were provided in 8 (10%) and 2 (2%) of the studies, respectively. Agreement with a reference standard was described in 5 studies (7%), with all of them providing evidence of an adequate agreement. Of the studies, 6 (8%) investigated responsiveness (3 by pharmacological intervention), and 1 (1%) assessed repeatability.

Technology Readiness. Of the studies, 50 (68%) included NAM-PD considered to represent TRL1 to TRL3. Two studies (2%) included NAM-PD of TRL9.

Speech

Number of Studies. A total of 30 studies (71%) tested a mathematical algorithm for NAM-PD to characterize speech, 9 (21%) reported NAM-PD used by nondevelopers, and 3 (7%) studies validated an NAM-PD.

Type of Technologies. The most frequently used systems were audio recorders (8 studies, 19%), followed by software-based solutions (6 studies, 14%). A total of 24 studies (57%) investigated mathematical algorithms without an additional hardware or software component.

Demographic Aspects. Of the studies, 24 (62%) that provided information on cohort size included 30 or fewer PwP. The most frequent mean/median age group ranged from 61 to 70 years (24 studies, 92%), and the mean/median age of the participants in all studies was between 50 and 80 years.

Clinimetric Properties. A total of 8 studies (19%) performed data diagnostics. The most frequently reported reference standard was PD classification according to established criteria (16 studies, 38%), followed by UPDRS-III (7 studies, 17%). Accuracy was reported in 24 studies (57%). The sensitivity and specificity of NAM-PD were shown to be higher than 80% in 9 (38%) and 6 (25%) studies, respectively. AUCs and predictive values/likelihood ratios were provided in 8 (19%) and 2 (5%) of the studies, respectively. Agreement with a reference standard was described in 6 (14%) of the articles, and 4 provided evidence of an adequate agreement. Responsiveness was investigated in 5 studies (12%; 3 studies used disease progression as the reference measure), and repeatability in 4 studies (10%).

Technology Readiness. Of the studies, 26 (65%) used NAM-PD evaluated as TRL $\leq$ 3. NAM-PD that were proven in an operational environment (TRL9) were not found.

Physical Activity

Number of Studies. A total of 18 (56%) studies involved NAM-PD used by nondevelopers to characterize physical activity, 5 (16%) reported on the validation of NAM-PD, 5 (16%) tested a mathematical algorithm for NAM-PD, and 4 (12%) described an NAM-PD or an idea.

Type of Technologies. The most frequently used systems were IMUs (22 studies, 69%).

Demographic Aspects. Of the studies, 19 (59%) included 30 or fewer PwP. The most frequent mean/median age group ranged from 61 to 70 years (15

studies, 71%), and in all studies, the mean/median age of the participants was between 50 and 80 years.

Clinimetric Properties. Of the studies, 6 (21%) performed data diagnostics prior to the statistical analyses. The most frequently reported reference standard was the UPDRS-III, which was used in 2 (7%) of the studies.

Accuracy was reported in 13 studies (46%). Of them, 4 (31%) reported sensitivity levels higher than 80%, and 2 (15%) identified specificity levels higher than 80%. AUCs and predictive values/likelihood ratios were provided in 3 (9%) and 1 (3%) of the studies, respectively. Agreement with a reference standard was described in 4 (13%), with 2 of these studies providing evidence of adequate agreement. Responsiveness was investigated in 5 (16%; 2 studies used disease progression and 2 DBS treatment as the reference measure) and repeatability in 2 (6%) of the studies.

Technology Readiness. Of the studies, 14 (50%) included NAM-PD considered to be TRL $\leq$ 3, and 6 (21%) used NAM-PD that were TRL9.

Rigidity

Number of Studies. Of the studies, 6 (55%) reported the validation of an NAM-PD that characterized rigidity, 4 (36%) involved a NAM-PD used by nondevelopers, and 1 (9%) tested a mathematical algorithm for NAM-PD.

Type of Technologies. The most frequently used systems were IMUs (4 studies, 36%).

Demographic Aspects. Of the studies, 7 (64%) included 30 or fewer PwP. The most frequently included age group ranged from 61 to 70 years (4 studies, 50%), and the mean/median age of the participants in all studies was between 50 and 80 years.

Clinimetric Properties. Of the studies, 3 (27%) performed data diagnostics. The most frequently reported reference standard was the UPDRS-III (3 studies, 27%).

Accuracy was reported in 2 (18%) of the studies, and none showed a sensitivity or specificity higher than 80%. AUCs and predictive values/likelihood ratios were not provided in any of the studies. Responsiveness and repeatability were investigated in 3 (27%; 2 used pharmacological interventions as the reference measure) and 2 (18%) of the studies, respectively.

Technology Readiness. Of the studies, 6 (55%) used NAM-PD considered to be of a TRL $\leq$ 3, and none of TRL9.

NAM-PD for the Assessment of Nonmotor Aspects

Number of Studies

Of the studies, 15 (63%) involved NAM-PD used by nondevelopers, 7 (29%) reported the validation of a NAM-PD, and 2 (8%) tested a mathematical algorithm for NAM-PD.

Type of Technologies

The most frequently used systems were IMUs (7 studies, 29%), followed by virtual reality-based systems (6 studies, 25%) and software-based solutions (5 studies, 21%).

Demographic Aspects

Of the studies, 19 (79%) included 30 or fewer PwP. The participants had a mean/median age ranging from 61 to 70 years in 17 (85%) of the studies that reported this information.

Clinimetric Properties

Of the studies, 3 (13%) performed data diagnostics prior to conducting any statistical analyses. The most frequently reported reference standards were validated tests/scales in 7 (29%) and expert diagnosis in 5 (21%) of the studies.

Accuracy was reported in 5 studies (21%). The sensitivity and specificity were shown to be higher than 80% in 2 (40%) of them. AUCs were reported in 2 (8%), and predictive values/likelihood ratios in 1 (4%) of the studies. Agreement with a reference standard was described in 4 studies (17%). Two of them provided evidence of an adequate agreement. Responsiveness and repeatability were not investigated in any of the studies.

Technology Readiness

A total of 11 (46%) studies assessed NAM-PD considered to be of a TRL ≤ 3, and 3 (27%) NAM-PD of TRL9.

NAM-PD for the Assessment of Treatment Complications

Number of Studies

Ten studies (50%) tested a mathematical algorithm for NAM-PD, 6 (30%) involved NAM-PD used by non-developers, and 2 (10%) reported the validation of an NAM-PD.

Type of Technologies

The most frequently used systems were IMUs (13 studies, 65%).

Demographic Aspects

Of the evaluated studies, 11 (55%) included 30 or fewer PwP. The included participants had a mean/median age ranging from 61 to 70 years in the 5 (63%) studies reporting this information.

Clinimetric Properties

Of the studies, 1 (5%) performed data diagnostics prior to the statistical analyses. The most frequently reported reference standards were expert diagnosis and total score on the UPDRS (both with 3 studies, 16% each).

Accuracy was reported in 13 (68%) of the studies. The sensitivity and specificity were higher than 80% in 3 (23%) and 4 (31%) of the studies, respectively. AUCs and predictive values/likelihood ratios were provided in 4 (20%) and 1 (5%) of the studies, respectively. Agreement with a reference standard was described in 1 study, which provided evidence of adequate agreement. Responsiveness and repeatability were investigated in 3 (15%) and 2 (10%) studies, respectively. For the former, the methods used to induce a change included DBS treatment (n = 1) and pharmacological intervention (n = 1).

Technology Readiness

Of the studies, 7 (41%) assessed NAM-PD considered to be of a TRL ≤ 3, and none included an NAM-PD of TRL9.

NAM-PD That Completed the Proof-of-Concept Stage (TRL ≥ 5)

A complete list of the systems that completed the proof of concept stage and had some initial validation in the laboratory can be found in the supplementary material section.

Discussion

This work provides a panoramic overview of a decade of innovation in the evaluation of Parkinson's disease (2005-2015). The breadth of the study is reflected in the substantial number of articles considered in this systematic approach (Supplementary Figure).

The evaluation of NAM-PD by disease domain, the systematic approach, the inclusion of nonwearable technologies, and the introduction of the TRL in this field of scientific activity complement the existing initiatives on this topic.[2,7-9] As previously mentioned, this comprehensive initiative aims to provide the interested reader with the opportunity to extract relevant data from the spreadsheet for personal purposes and to set the stage for a common collection of qualitative data in this highly dynamic research field. The analysis provided in this manuscript presents only an example

of how the data can be used and lays the groundwork for some conclusions that can be considered in the further development of the field.

Number and Type of Studies Per Disease Domain

Within the different disease domains, studies that primarily presented validation efforts were the most frequent type. The exceptions were axial features, physical activity, and nonmotor domains, for which studies focusing on nondevelopers' evaluations of NAM-PD performance represented the largest portion. For speech, several of the papers reported mathematical algorithms. All of these exceptions can be explained by different reasons. For axial features, the specific development of NAM-PD started early, which enabled already relatively mature NAM-PD to be available to nondevelopers. For physical activity, the existence of wearable systems intended for other applications (eg, physical activity monitoring systems for the general public) were available to nondevelopers for use as NAM-PD. For nonmotor assessments, our observations suggest that validation efforts have been neglected, and a particular emphasis should be placed on developing validated systems for the assessment of nonmotor features of PD. Interestingly, for speech, the existence of open databases may have favored the development of new mathematical algorithms.

An interrelated fact pertains to the number of references identified per domain. Axial features were the most represented domain, with 213 articles included in the detailed evaluation process. This highlights the importance of this domain in the assessment of PD from a clinical and PwP point of view; however, the heterogeneity of this category (comprising different aspects of the disease, eg, posture, gait, freezing of gait, and balance) may also have contributed to this result. The second most frequently studied domain was bradykinesia (127 articles). This emphasis may be related to the importance of bradykinesia in the assessment of PD[20] and its ability to be measured with existing hardware (eg, IMUs)[1]. In contrast, NAM-PD for rigidity and nonmotor features were rarely investigated and thus represent a huge opportunity for technology developers given the relevance of these domains in the management of PD and the quality of life of PwP.[21,22]

Although not specifically presented in this systematic review, the positive evaluations of individual NAM-PD platforms that measure several aspects of the disease (see below) and the underinvestigation of certain domains support the findings of previous literature (eg, Ref. 23) in that such platforms seem more meaningful. Of note, 110 (19%) of the 588 studies evaluated more than 1 disease domain.

Types of Technologies Used

The types of systems used to assess specific disease domains were highly heterogeneous. However, IMUs were the primary choice in most of the domains, and publications using this type of NAM-PD are increasing at a disproportional rate. The reason for this development is obvious: The use of wearable technology in the general population[2] and the miniaturization of sensors have created a new era in PD assessment.[24] At their current stage of development, IMUs are primarily used to assess bradykinesia,[25] tremors,[26] and axial features.[27] Virtually all cardinal motor features (perhaps with the exception of rigidity, for which only a small number of studies are currently available) can be quantified with these systems, at least in a research setting. Nonetheless, their use in routine clinical care needs to be defined (see also Ref. 28).

Clinimetric Testing and Lessons Learned From the Field

A particularly important aspect for the successful future development of the field toward clinical applicability is to define useful standards and systems to prevent the overinvestigation of specific aspects in studies performing very similar experiments with slightly different NAM-PD as well as to foster topics that are currently underinvestigated. Based on the results of this systematic evaluation of the recent decade of NAM-PD research, consideration of the following points may contribute to this process.

First, many of the evaluated validation articles did not report information on aspects that are essential to understand the context in which the validation occurred. Variables such as age, disease stage, and validity parameters were inconsistently reported. These parameters are crucial to ensure that the systems' performances are replicable and can be adequately interpreted when other PD populations are investigated. The Standards for Reporting of Diagnostic Accuracy guideline (STARD) could serve as an adequate starting point to define the minimum set of variables that should be included in future NAM-PD validation studies.[16]

Second, the studies that are currently available in the field do not sufficiently reflect the full PD spectrum. For example, most of the studies included a relatively low number of participants. Moreover, most studies focused on PwP aged in their 60s, with relatively low Hoehn and Yahr stages. Larger studies that are representative of the full spectrum of the disease are needed to ensure that the performance of the NAM-PD will be generalizable. This need could be facilitated by collaborative initiatives. These initiatives could also be essential in independently replicating pilot results.[29] In the studies evaluated in this review,

only 8% confirmed their results in an independent dataset.

Third, an inherent limitation of the NAM-PD field is the absence of quantitative standards. Most of the studied domains used the UPDRS-III as a reference. Although acceptable, the limitations of this approach have already been discussed.[15,30-32] Our recommendation is for key players in the field to standardize the validation of upcoming NAM-PD, ideally with quantitative and objective parameters, and to foster open datasets of already published studies.

Fourth, most of the studies used statistical correlations instead of recommended methodologies to prove diagnostic validity. Simple correlation approaches have relevant drawbacks.[17-19] From our perspective, diagnostic validation requires the assessment of the performance of a new system at the individual level. A detailed description of the methods that can be used is beyond the scope of this review; however, forming multidisciplinary teams exposed to this methodology is highly recommended. This multidisciplinary approach could also increase the applicability of NAM-PD in realistic clinical scenarios (which is a key aspect for the future translation of these technologies), and there is clear evidence that this development process will benefit from the participation of PwP.[23,33]

Fifth, 87% of the included studies were implemented in clinical or laboratory settings and not in simulated or real-life settings. As part of the validation process, developers should be encouraged to test their systems in the actual environment of their intended use.

Technology Readiness

A final exploratory aspect of our review was the assessment of the TRL in evaluating the readiness of an NAM-PD[10] (Supplementary Table). Based on our evaluation, a surprisingly low percentage of 6% of the studies included NAM-PD that were TRL9. Only an NAM-PD at this level justifies the hope of including this system in clinical assessments in due course. Most of these systems are commercialized for research and not for clinical application. They include mainly accelerometers but also force plates and motion analysis systems. Few software-based technologies and mathematical algorithms reached a $TRL \geq 5$. When we examined the disease domains assessed by the most mature systems, the most represented domains were axial features, followed by bradykinesia, physical activity, and tremors. A final aspect that needs to be stressed is that no systems for home-based assessments reached TRL9.

Study Limitations

Our evaluation has limitations. First, we tried to cover all NAM-PD studies investigating the 12 PD domains of interest that had been published in the past 10 years, and we might have missed relevant articles by using this approach. Specifically, the defined period for which the search was conducted may have led to the exclusion of relevant (earlier) studies and (different) NAM-PD. Nonetheless, considering the short cycle of technological innovation[34] and the thorough selection process of the domains of interest, it seems unlikely that this limitation had a relevant impact on the general results and our conclusions. Second, the inclusion of more than 1 rater may have introduced inconsistencies into the data. However, 3 of the authors (A.S.F., M.E., and D.S.) counterchecked for these inconsistencies and addressed them. Special emphasis was placed on the TRL rating (all articles were counterchecked by 1 author) and the implications of labeling a system as mature. Moreover, as the database is shared publicly, any feedback on individual studies/inconsistencies could be integrated in updated versions.

Conclusion

This systematic review provides an overview of the current options in the field of quantitative assessment of the most important PD features and what these technologies can provide in the near future. It may serve as a foundation for the design and performance of future studies in the field and may help increase their quality and impact. In these studies, particular emphasis should be placed on the combined efforts of scientists to increase cohort sizes, provide measurements in naturalistic environments, prevent the overinvestigation of singular topics, enhance the maturation processes of assessment systems, and consider validation issues. Such efforts have the most probable potential of revolutionizing the diagnosis and treatment of PD for both PwP and those who treat them. ■

References

1. Ciuti G, Ricotti L, Menciassi A, Dario P. MEMS Sensor technologies for human centred applications in healthcare, physical activities, safety and environmental sensing: a review on research activities in Italy. Sensors 2015;15(3):6441-6468.

2. Maetzler W, Domingos J, Srulijes K, Ferreira JJ, Bloem BR. Quantitative wearable sensors for objective assessment of Parkinson's disease. Mov Disord 2013;28(12):1628-1637.

3. Bonato P. Clinical applications of wearable technology. IEEE Eng Med Biol Soc Annu Conf 2009;2009:6580-6583.

4. Hobert MA, Maetzler W, Aminian K, Chiari L. Technical and clinical view on ambulatory assessment in Parkinson's disease. Acta Neurol Scand. 2014;130(3):139-147.

5. Maetzler W, Rochester L. Body-worn sensors-the brave new world of clinical measurement? Mov Disord 2015;30(9):1203-1205.

6. Oung QW, Muthusamy H, Lee HL, et al. Technologies for assessment of motor disorders in Parkinson's disease: a review. Sensors 2015;15(9):21710-21745.

7. Hubble RP, Naughton GA, Silburn PA, Cole MH. Wearable sensor use for assessing standing balance and walking stability in people with Parkinson's disease: a systematic review. PLoS ONE 2015;10(4):e0123705.

8. Pasluosta CF, Gassner H, Winkler J, Klucken J, Eskofier BM. An emerging era in the management of parkinson's disease: Wearable Technologies and the Internet of things. IEEE J Biomed Heal Informatics 2015;19(6):1873-1881.

9. Godinho C, Domingos J, Cunha G, et al. A systematic review of the characteristics and validity of monitoring technologies to assess Parkinson's disease. J Neuroeng Rehabil 2016;13(1):24.

10. European Comission. Technology readiness levels (TRL). http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf. Accessed on January 7, 2016.

11. Ferreira JJ, Godinho C, Santos AT, et al. Quantitative home-based assessment of Parkinson's symptoms: the SENSE-PARK feasibility and usability study. BMC Neurol 2015;15:89.

12. Serrano JA, Larsen F, Isaacs T, et al. Participatory design in Parkinson's research with focus on the symptomatic domains to be measured. J Parkinsons Dis 2015;5(1):187-196.

13. Gunzler SA, Pavel M, Koudelka C, Carlson NE, Nutt JG. Foot-tapping rate as an objective outcome measure for Parkinson disease clinical trials. Clin Neuropharmacol 2009;32(2):97-102.

14. Galna B, Barry G, Jackson D, Mhiripiri D, Olivier P, Rochester L. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. Gait Posture 2014;39(4):1062-1068.

15. Heldman DA, Espay AJ, LeWitt PA, Giuffrida JP. Clinician versus machine: Reliability and responsiveness of motor endpoints in Parkinson's disease. Parkinsonism Relat Disord 2014;20(6):590-595.

16. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem 2003;49(1):1-6.

17. Delgado Rodríguez M, Llorca Díaz J, Doménech Massons JM. Validation studies for diagnostic tests and prognostic factors. 5a ed. Barcelona: Signo; 2012;1–213.

18. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol 2003;22(1):85-93.

19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307-310.

20. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 1992;55(3): 181-184.

21. Schrag A. What contributes to quality of life in patients with Parkinson's disease? J Neurol Neurosurg Psychiatry 2000;69(3):308-312.

22. Barone P, Antonini A, Colosimo C, et al. The PRIAMO study: a multicenter assessment of nonmotor symptoms and their impact on quality of life in Parkinson's disease. Mov Disord 2009;24(11): 1641-1649.

23. Stamford JA, Schmidt PN, Friedl KE. What engineering technology could do for quality of life in Parkinson's disease: a review of current needs and opportunities. IEEE J Biomed Heal Informatics 2015;19(6):1862-1872.

24. LeMoyne R, Coroian C, Mastroianni T. Quantification of Parkinson's disease characteristics using wireless accelerometers. ICME International Conference on Complex Medical Engineering. New York: IEEE; 2009:480-484.

25. Cancela J, Pansera M, Arredondo MT, et al. A comprehensive motor symptom monitoring and management system: the bradykinesia case. Conf Proc IEEE Eng Med Biol Soc 2010;2010:1008-1011.

26. Giuffrida JP, Riley DE, Maddux BN, Heldman DA. Clinically deployable Kinesia technology for automated tremor assessment. Mov Disord 2009;24(5):723–730.

27. Ellis RJ, Ng YS, Zhu S, et al. A validated smartphone-based assessment of gait and gait variability in Parkinson's disease. PLoS ONE 2015;10(10):e0141694.

28. Maetzler W, Klucken J, Horne MK. A clinical view on the development of technology-based tools in managing Parkinson's disease [published online ahead of print June 7, 2016]. Mov Disord.

29. Jasny BR, Chin G, Chong L, Vignieri S. Data replication & reproducibility. Again, and again, and again. Introduction. Science 2011;334(6060):1225-1225.

30. Goetz CG, Stebbins GT. Assuring interrater reliability for the UPDRS motor section: utility of the UPDRS teaching tape. Mov Disord 2004;19(12):1453-1456.

31. Post B, Merkus MP, de Bie RM A, de Haan RJ, Speelman JD. Unified Parkinson's Disease Rating Scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? Mov Disord 2005;20(12):1577-1584.

32. Goetz CG, Stebbins GT, Wolff D, et al. Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device. Mov Disord 2009;24(4):551-556.

33. van Uem J, Isaacs T, Lewin A, et al. A viewpoint on wearable technology-enabled measurement of wellbeing and health-related quality of life in Parkinson's disease. J Park Dis 2016;6(2):279-287.

34. Klepper S. Entry, exit, growth, and innovation over the product life cycle. Am Econ Rev 1996;86:562-583.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.