Digital biomarkers in Parkinson's disease: missing the forest for the trees?

Ashwani Jha MRCP PhD[1], Alberto J. Espay MD MSc[2], Andrew J. Lees FRCP F.Med.Sci PhD[3]

[1]UCL Queen Square Institute of Neurology, London, UK

[2]James J. and Joan A. Gardner Family Center for Parkinson's Disease and Movement Disorders, Department of Neurology, University of Cincinnati, Cincinnati Ohio, USA

[3]Reta Lila Weston Institute of Neurological Studies, Department of Clinical Movement Disorder and Neuroscience, Institute of Neurology, University College London, London, United Kingdom

Correspondence to:

Ashwani Jha

Russell Square House

10-12 Russell Square

1st floor Stroke Research Centre

London, UK WC1B 5EH

ashwani.jha@ucl.ac.uk

Word Count: 1633

Running Title: Digital biomarkers in Parkinson's

Key Words: Digital Health Technology, Outcomes, Personalisation, Parkinson's, Biomarkers

Recently proposed digital measures of Parkinson's are gaining momentum[1,2]. More objective, more precise, and more easily repeated than human expert-based assessments such as the current Gold Standard Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS)[3], these measures – often conceptualized as *digital biomarkers* – promise a new dawn in personalized medicine. Digital tools are already being used as surrogate markers in clinical trials[4] and advocated to improve clinical decision-making[5]. So, is it only a matter of more time, more data, and of course more money before at-home digital assessments become the clinical Gold-standard? In this viewpoint we step-back and re-evaluate what it is we are trying to measure and why we are measuring it. From this perspective we argue that the current pursuit of more precise and objective measures may, if unchecked, lead to ever increasing bias in clinical trials and ever decreasing utility for individual decision-making. We may improve the statistical power of clinical trials but *at the cost of* personalization. We propose a more holistic digital approach as a solution.

The single digital biomarker hypothesis

Many current digital measures aim to represent the severity of Parkinson's in an individual with a *single* number[6–8] although there are notable exceptions[4,9,10]. These proposed digital biomarkers are often more sensitive to clinical change than human expert ratings such as the MDS-UPDRS[11] but the clinical interpretation and utility of these scores remains unclear. Let's define a hypothetical smartphone-derived digital biomarker – a single *overall* severity score that ranges from 0-10 where higher scores represent greater severity. If an intervention reduces an individual's score from 8.5 to 8.1, what does this mean? Often, this is interpreted as an improvement in the clinical features of disease and by proxy, an improvement in the underlying neuronal dysfunction whether modifying individual disease-

related pathophysiological processes or not. We call this interpretation the *single digital biomarker hypothesis* (Figure 1A). Much current regulatory, academic and commercial thought[1] often assumes this hypothesis to be true and is focused on overcoming the technical and practical barriers to validating these measures as digital endpoints. Here we challenge the hypothesis itself and consider the reasons why *improvements in overall digital measures do not necessarily imply overall improvements in clinical or neuronal dysfunction.* We outline some of the conceptual credentials a digital measure must have if it is to be accepted as a digital biomarker. The foundation of our challenge stems from the well-established view of Parkinson's as a multi-dimensional, multi-etiology syndrome.

Parkinson's cannot be measured by a single number

There is no unitary biological marker for Parkinson's disease progression – no equivalent of HbA1c for Diabetes[1]. We can measure *clinical* change but this is difficult. Individuals with Parkinson's suffer in different ways – some may have noticeable hand tremor, others difficulty walking and yet others may be overwhelmed by anxiety. Individual severity, progression and treatment response, therefore, occur unevenly across *multiple* clinical dimensions, likely reflecting multiple underlying biological and environmental factors. The multi-system nature of Parkinson's was established early on and is to some extent reflected in current outcome scales such as the MDS-UPDRS[3], the Parkinson's Disease Quality of Life Questionnaire (PDQ-39)[12] and the Non-Motor Symptoms Scale[13]. But the vast majority of current digital measures, however, focus only on measuring a limited aspect of Parkinson's. Wrist-worn wearables are limited to motor observations at a single joint[5,14,15], whilst smartphone assessments are dominated by motor assessments of one or a few limbs, sometimes with voice[6]. Exceptionally, a wider motor phenotype or even a brief cognitive

test is included[7], but these still fall drastically short of a holistic assessment of Parkinson's. Some reports describe their digital biomarker as a *motor* score rather than an overall score but the same issue applies – typically only limited *motor* features are measured in limited parts of the body offering a limited view of the motor state. How does this issue affect the interpretation of digital measures as biomarkers? At the individual level, it affects our ability to take action solely on the basis of the 'hyper-precise' digital measure. It is clear that an intervention that improves our smartphone-derived hypothetical digital score from 8.5 to 8.1 could easily *worsen* an individual's overall condition – measured clinical features such as bradykinesia may have improved slightly, but unmeasured features such as anxiety or dyskinesias may have deteriorated more substantially (Figure 1B). Even with a digital measure, we still need a granular clinical assessment before we can recommend a change to an individual patient's treatment schedule. A selective digital measure will also impact group level inference, for example in interventional clinical trials. An overall digital endpoint that is dominated by upper limb bradykinesia, for example, will be sensitive to interventions that improve hand motor function but relatively insensitive to interventions that improve other aspects of the condition. In the long-term, unless we are careful, digital endpoints may bias translational science in favour of a particular mono- or oligosymptomatic profile.

Digital measures can be confounded

A second problem of the single digital biomarker hypothesis is the vulnerability of motor-centric digital measures to confounding variables (Figure 1C) and other sources of variation[16]. Here, systematic changes in a digital measurement are caused by unmeasured confounds rather than the neural dysfunction of Parkinson's. Unlike expert-rated assessments such as the MDS-UPDRS, digital measures are typically taken at-home without

supervision. Rhythmic hand movements whilst washing dishes or playing a musical instrument may falsely be interpreted as tremor or dyskinesia by an at-home-wearable[14]. Walking carefully on an icy pavement could be misinterpreted as bradykinesia. Many aspects of human life unrelated to the neural dysfunction of Parkinson's will interact in unknown ways with digital measurement[16]. Active task-based measures – where the patient is performing a particular task such as tapping a smartphone screen – provide some level of control against unmeasured environmental confounds but are still susceptible to another kind. Aging and co-morbid diseases such as osteoarthritis will affect movement and confound digital measurements. Going back to our smartphone-derived hypothetical digital measure, a drop from 8.5 to 8.1 could be an improvement in their Parkinson's but may also represent cessation of their trombone-playing practice or improved control of joint-pain in their hands. At the group-level within a randomized controlled trial, one may expect such confounds to 'balance out' in the long run but they would certainly add noise to the digital measure and reduce its sensitivity to change.

What would a digital biomarker look like?

We have described how *single* digital measurements although precise often provide an insufficient or confounded measure of Parkinson's progression. They cannot be used as the sole basis of personalised clinical decision-making. They show promise[4] but risk bias if used without care in clinical trials. How can we address this? A first step would be to develop digital measures that are holistic rather than reductive, measuring *multiple* clinical dimensions with granular detail – similarly to how a clinician would assess a patient. Note that these multiple measures must capture broad aspects of the patient's condition, not just be multiple measures of the same aspect (e.g. hand tremor). With more faithful capture of a

patient's condition, it will be clearer if symptoms have meaningfully worsened in one domain even if they have improved in others. If scores highlight a meaningful worsening of anxiety after an intervention whilst walking has improved, then it may be rational to withdraw the intervention even if our hypothetical single digital biomarker suggests an overall improvement. Unlike a single digital biomarker, therefore, a comprehensive digital assessment has the ability to assist clinical decision-making at the individual level. Likewise, some types of confounded measurement would be more easily brought to light. Consider an intervention for osteoarthritis that improves walking speed by improving joint pain and stability. A single digital biomarker that incorporates a gait assessment may register improvement leading to the false conclusion that the individual's Parkinson's has got better. With multiple measures, however, it may be clearer that walking has improved due to a separate reason (in line with pain reduction and perhaps without associated improvement of hand function or tremor).

Although *necessary*, we do not claim here that increasing the breadth of assessment is *sufficient* for the development of an actionable multi-domain measure – digital or otherwise. The PDQ-39[12] is broad but lacks the granularity required for individual clinical decision-making. The MDS-UPDRS[3] is also broad but perhaps too blunt and unwieldy for routine clinical use (except as part of the assessment for advanced therapies). There are many other questions that novel digital measures will have to also address. We must understand what features are meaningful for patients[17], be concrete in exactly what is being measured[1], and make sure digital tools are safe, robust and reliable[2,16,18]. We must be aware that some aspects such as the number of falls in the last year are countable and easily

measured, whereas other aspects such as the loss of confidence following a fall are not and so harder to quantify. Perhaps many important aspects are immeasurable.

We suggest beginning with a multi-domain digital diary, perhaps focusing on interpretable countable observations (number of falls, episodes of urinary incontinence, hours slept in the daytime, etc) where possible. With appropriate care, this coarse but curated and actionable resource could provide the data required to develop more complex and interactive digital tools that incorporate objective measures. A similar development pipeline has already been proposed for the monitoring of clinical fluctuations in Parkinson's[19]. This may seem a backwards step to developers of digital health technologies who seek to "upgrade the practice of medicine to one that is high-definition[20]". They seek to reduce human inter-rater error and increase objective measurement precision in order to provide better clinical endpoints for trials. But in the pursuit of a more precise aggregate progression score, the individual has slipped out of focus. *Comprehensiveness* more so than precise but unidimensional measurement is required for actionable, personalised care decisions. This will be familiar, even mundane, to expert clinicians who perform individualised multi-dimensional assessments routinely as part of their clinical practice. But somewhere along the digital development pathway this wisdom has been lost in translation.

Acknowledgements:

The authors would like to thank Professors Kailash Bhatia and Marcelo Merello for their input and for inviting this viewpoint following a presentation and panel discussion at the Movement Disorders Clinical Practice Conference 2022.

Author Roles:

1. Research project: A. Conception, B. Organization, C. Execution;

2. Statistical Analysis: A. Design, B. Execution, C. Review and Critique;

3. Manuscript Preparation: A. Writing of the first draft, B. Review and Critique;

AJ: 1A, 1C, 3A, 3B

AJE: 1C, 3B

AJL: 1C, 3B

Disclosures:

Funding Sources and Conflict of Interest

AJ has been involved in the development and clinical assessment of a smartphone-based tool for Parkinson's disease (cloudUPDRS). AJ is funded by the Wellcome Trust (213038) and supported by the National Institute of Health Research University College London Hospitals Biomedical Research Centre

Financial Disclosures for the Previous 12 months

AJ has received travel expenses from Merz to attend a scientific course. AJE has received grant support from the NIH and the Michael J Fox Foundation; personal compensation as a consultant/scientific advisory board member for Neuroderm, Amneal, Acadia, Acorda, Bexion, Kyowa Kirin, Sunovion, Supernus (formerly, USWorldMeds), Avion Pharmaceuticals, and Herantis Pharma; personal compensation as honoraria for speakership for Avion and Amneal; and publishing royalties from Lippincott Williams & Wilkins, Cambridge University Press, and Springer. He cofounded REGAIN Therapeutics (a biotech start-up developing

nonaggregating peptide analogues as replacement therapies for neurodegenerative diseases) and is co-owner of a patent that covers synthetic soluble nonaggregating peptide analogues as replacement treatments in proteinopathies.

Ethical Compliance Statement:

The authors confirm that the approval of an institutional review board was not required for this work and informed consent was not necessary. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

References:

1. Stephenson D, Badawy R, Mathur S, Tome M, Rochester L. Digital Progression Biomarkers as Novel Endpoints in Clinical Trials: A Multistakeholder Perspective. *J Parkinsons Dis*. 2021;11(s1):S103-S109. doi:10.3233/JPD-202428
2. Espay AJ, Hausdorff JM, Sánchez-Ferro Á, et al. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. *Mov Disord*. 2019;34(5):657-663. doi:10.1002/mds.27671
3. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord*. 2008;23(15):2129-2170. doi:10.1002/mds.22340
4. Lipsmeier F, Taylor KI, Postuma RB, et al. Reliability and validity of the Roche PD Mobile Application for remote monitoring of early Parkinson's disease. *Sci Reports*

2022 121. 2022;12(1):1-15. doi:10.1038/s41598-022-15874-4

5. Woodrow H, Horne MK, Fernando C V., et al. A blinded, controlled trial of objective measurement in Parkinson's disease. *npj Park Dis 2020 61*. 2020;6(1):1-10. doi:10.1038/s41531-020-00136-9

6. Zhan A, Mohan S, Tarolli C, et al. Using smartphones and machine learning to quantify Parkinson disease severity the mobile Parkinson disease score. *JAMA Neurol*. 2018;75(7):876-880. doi:10.1001/jamaneurol.2018.0809

7. Omberg L, Chaibub Neto E, Perumal TM, et al. Remote smartphone monitoring of Parkinson's disease and individual response to therapy. *Nat Biotechnol*. 2021. doi:10.1038/S41587-021-00974-9

8. Lo C, Arora S, Lawton M, et al. A composite clinical motor score as a comprehensive and sensitive outcome measure for Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2022;93(6):617-624. doi:10.1136/JNNP-2021-327880

9. Jha A, Menozzi E, Oyekan R, et al. The CloudUPDRS smartphone software in Parkinson's study: cross-validation against blinded human raters. *npj Park Dis*. 2020;6(1):1-8. doi:10.1038/s41531-020-00135-w

10. Rochester L, Mazzà C, Mueller A, et al. A Roadmap to Inform Development, Validation and Approval of Digital Mobility Outcomes: The Mobilise-D Approach. *Digit biomarkers*. 2020;4(Suppl 1):13-27. doi:10.1159/000512513

11. 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. doi:10.1016/J.PARKRELDIS.2014.02.022

12. Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease.

Qual Life Res. 1995;4(3):241-248. doi:10.1007/BF02260863/METRICS

13. Chaudhuri KR, Martinez-Martin P, Brown RG, et al. The metric properties of a novel non-motor symptoms scale for Parkinson's disease: Results from an international pilot study. *Mov Disord*. 2007;22(13):1901-1911. doi:10.1002/mds.21596

14. Powers R, Etezadi-Amoli M, Arnold EM, et al. Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson's disease. *Sci Transl Med*. 2021;13(579). doi:10.1126/scitranslmed.abd7865

15. Elm JJ, Daeschler M, Bataille L, et al. Feasibility and utility of a clinician dashboard from wearable and mobile application Parkinson's disease data. *NPJ Digit Med*. 2019;2(1). doi:10.1038/S41746-019-0169-Y

16. Roussos G, Herrero TR, Hill DL, et al. Identifying and characterising sources of variability in digital outcome measures in Parkinson's disease. *NPJ Digit Med*. 2022;5(1). doi:10.1038/S41746-022-00643-4

17. Manta C, Patrick-Lake B, Goldsack JC. Digital Measures That Matter to Patients: A Framework to Guide the Selection and Development of Digital Measures of Health. *Digit Biomarkers*. 2020;4(3):69-77. doi:10.1159/000509725

18. Badawy R, Hameed F, Bataille L, et al. Metadata Concepts for Advancing the Use of Digital Health Technologies in Clinical Research. *Digit Biomarkers*. 2019;3(3):116-132. doi:10.1159/000502951

19. Vizcarra JA, Sánchez-Ferro Á, Maetzler W, et al. The Parkinson's disease e-diary: Developing a clinical and research tool for the digital age. *Mov Disord*. 2019;34(5):676-681. doi:10.1002/MDS.27673

20. Steinhubl SR, Topol EJ. Digital medicine, on its way to being just plain medicine. *npj Digit Med 2018 11*. 2018;1(1):1-1. doi:10.1038/s41746-017-0005-1

Figures

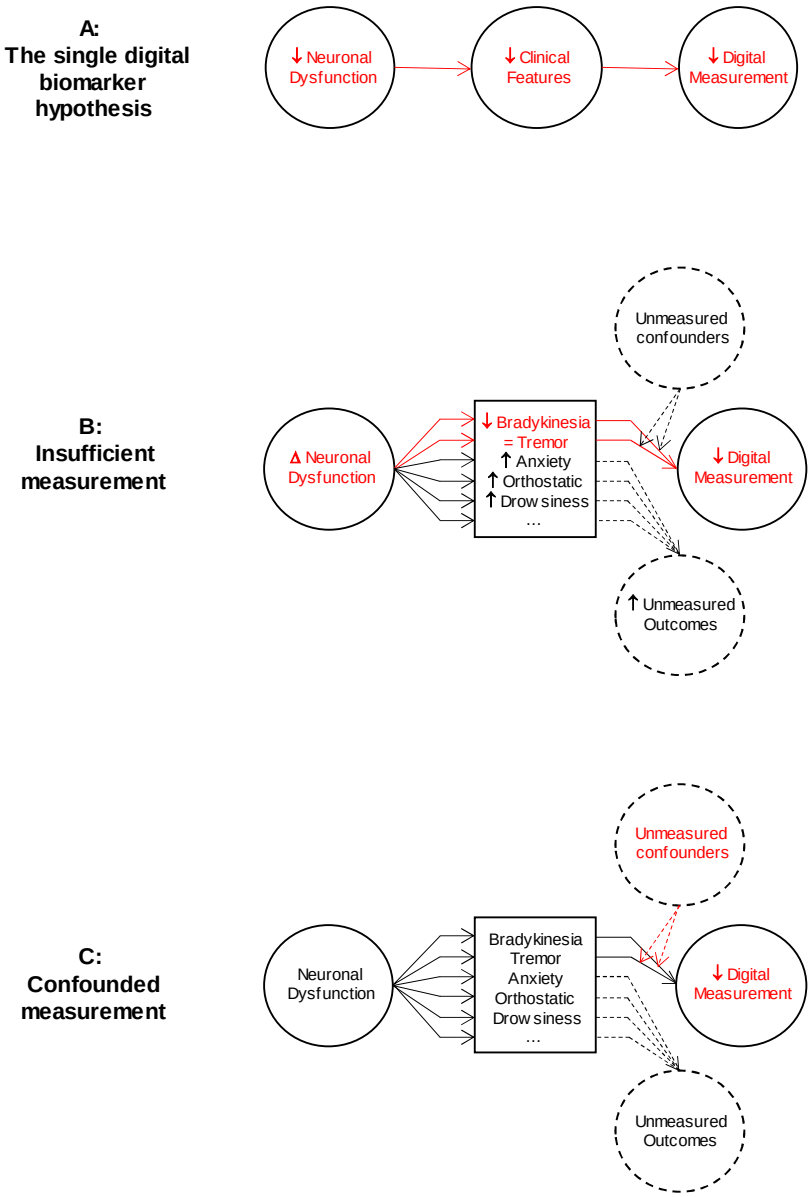
A:
The single digital biomarker hypothesis
↓ Neuronal Dysfunction
↓ Clinical Features
↓ Digital Measurement
B:
Insufficient measurement
Unmeasured confounders
Δ Neuronal Dysfunction
↓ Bradykinesia
= Tremor
↑ Anxiety
↑ Orthostatic
↑ Drowsiness
...
↓ Digital Measurement
↑ Unmeasured Outcomes
C:
Confounded measurement
Unmeasured confounders
Neuronal Dysfunction
Bradykinesia
Tremor
Anxiety
Orthostatic
Drowsiness
...
↓ Digital Measurement
Unmeasured Outcomes

Figure 1: A: The single digital biomarker hypothesis assumes that overall digital measures reflect clinical and neural dysfunction. Given a hypothetical digital measure of disease (higher is worse), a reduction in a digital measure (↓) can be interpreted as clinical improvement with associated reduction of neural dysfunction (↓). **B:** A broader conceptualisation of Parkinson's as a multi-domain syndrome, however, provides two alternative reasons for this reduction in digital measure. Current measures incorporate only a few of the many clinical features associated with Parkinson's. In this example, the digital measure incorporates bradykinesia and tremor but doesn't incorporate other clinical features such as anxiety, orthostatic hypotension and drowsiness (only a few out of many domains are shown to aid legibility, others are represented with an ellipsis). Because interventions unevenly affect different clinical domains, a reduced digital measure (↓) due to improved measured bradykinesia may hide coincident meaningful worsening (↑) of unmeasured anxiety, orthostatic hypotension and drowsiness. The digital measure reports an overall improvement but the overall condition is in reality worse and the underlying neuronal dysfunction worse or the same (Δ). The digital measure is insufficient. **C:** Similarly, unmeasured confounders can affect motor-centric digital measures. Treated osteoarthritis of the knees may improve gait speed and subsequently reduce a digital measure incorporating it without any associated changes to the underlying Parkinson's. *The arrows between the nodes of the graphs above represent directed causal links. Red labels and arrows represent the flow of information entering the digital measurement, whilst black labels and arrows represent information that the measure fails to incorporate. 'Neuronal dysfunction' here is a broad term that doesn't necessarily refer to the individualized molecular/biological aetiology.*

291