


RESEARCH ARTICLE

Open Access



Consensus on the clinical utility of digital mobility outcomes for personalized clinical decision support in parkinson's disease

Alan Castro Mejia^{1†}, Stefano Sapienza^{1*†} , Ivana Paccoud¹, Lisa Alcock^{3,6}, Philip Brown⁴, Heiko Gaßner^{5,12}, Heather Hunter⁴, Walter Maetzler⁸, Anat Mirelman^{9,10,11}, Alice Nieuwboer^{13,14}, Martin Regensburger⁵, Lynn Rochester^{3,4,6}, Sabine Stallforth⁵, Beatrix Vereijken⁷, Alison Yarnall^{3,4} and Jochen Klucken^{1,2}

Abstract

Background Digital mobility outcomes (DMOs) have emerged as novel biomarkers offering objective, quantitative, and examiner-independent outcome measures for clinical studies. Unfortunately, research efforts on DMOs have not yet investigated the domain of clinical utility in Parkinson's disease, i.e. providing evidence of improvements in health outcomes, diagnosis, decision-making, or prevention when compared to e.g. standard-of-care procedures. This manuscript, via a consensus building approach, aims to create a structured conceptual framework to map the knowledge generated by DMOs with clinical domains that could benefit from it.

Methods We conducted a three-round consensus-building study with 12 experts recruited from the Mobilise-D consortium's Parkinson's Disease Working Group. The experts designed and ranked different aspects of the conceptual framework via a 5-level Likert scale for level of agreement. Consensus for the different points evaluated was based on a double threshold: the simultaneous presence of a high level of agreement had to be accompanied by a low level of disagreement. As secondary objectives, the experts were asked to rate the practical application of DMOs by evaluating the timeline to applicability, the foreseen challenges for their implementation in clinical settings, and their main role in the decision-making process.

Results A full consensus on the clinical utility framework was achieved after three rounds. The final framework consisted of three main categories (*Disease Diagnosis*, *Patient Evaluation*, and *Treatment Evaluation*) and six underlying domains (*Enhancing Diagnostic Procedure*, *Predicting Risk*, *Timely Detecting Deterioration*, *Enhancing Clinical Judgment*, *Selecting Treatment*, and *Monitoring Treatment Response*). The experts believed in the next 1–5 years DMOs will play a relevant role in clinical decision making, complementing care knowledge with useful digital biomarkers information. However, the main challenge to address is the definition of clear reference value for DMOs interpretability.

Conclusions This framework provides a structure for subsequent studies to build into by diversifying expert cohorts and expand our findings beyond PD. Additionally, our results support researchers planning future clinical trials

[†]Alan Castro Mejia and Stefano Sapienza are co-primary authors.

*Correspondence:
Stefano Sapienza
stefano.sapienza@uni.lu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

where DMOs can play a valuable role for clinical decision support. Ultimately, this is the first step toward developing guidelines to assess DMOs' clinical utility and support their integration into Real World clinical practice.

Keywords Digital mobility outcomes, Parkinson's disease, Gait, Consensus study, Clinical utility

Background

Mobility, the ability to move freely and easily [1], is a primary constituent of a healthy life, with speed widely considered the sixth vital sign [2]. In patients with Parkinson's disease (PD) who are affected by chronic motor disorders that progress and fluctuate over time, the characterization of mobility bears critical importance for the assessment of overall conditions and subsequent management of treatment. Due to the extreme heterogeneity of clinical presentations in PD, interventions and clinical decisions must be tailored to the patient's characteristics [3]. This practice, defined as personalized care, requires an accurate and comprehensive evaluation of the patient's profile. In this context, mobility-related information assumes a primary role.

Digital mobility outcomes (DMOs) have emerged as novel biomarkers offering objective, quantitative, and examiner-independent outcome measures for clinical studies [1, 4]. Unlike traditional clinical assessments, which are often subjective and vary depending on the examiner's expertise, DMOs provide a standardized approach to capture real-world mobility through digital technologies, such as wearable devices. As such, DMOs are gaining recognition as reliable tools for characterizing patient status, evaluating treatment responses, and supporting regulatory decision-making across various medical conditions [5]. For PD in particular, several studies, including those from the Mobilise-D consortium, have highlighted the ability of DMOs to provide objective, relevant, and validated information that characterizes both symptom severity and progression rates [1, 6–11]. Additionally, promising examples of the clinical applicability of DMOs show that these may differentiate across PD and non-PD groups [12], medication states, targeted gait outcomes (e.g., to detect freezing events) [13], correlate with classical clinical scales such as the MDS-UPDRS [14], and offer clinically meaningful measures of mobility after rehabilitation for long-term assessment [15].

Furthermore, DMOs have gained recognition as endpoints for clinical studies focusing on mobility [16, 17]. The reason is twofold: the necessity of detecting subtle gait changes that cannot be visually captured or quantified and the possibility of extending the monitoring window outside the clinic. This scientific evidence suggests that DMOs should represent an essential tool for clinicians and healthcare professionals.

Unfortunately, research efforts on DMOs have not yet investigated the domain of clinical utility, providing evidence of improvements in health outcomes,

diagnosis, decision-making, or prevention when compared to standard-of-care procedures. A recent literature review focusing on inertial sensors for home monitoring in PD showed that only 15% of studies provided evidence on clinical utility [18], with most presenting findings in qualitative terms. Although DMOs have been touted as playing a complementary role in enhancing current standards of care by supporting clinical decision-making and personalized intervention strategies [6], the lack of clear reference values and difficulty in integrating results into care pathways [6, 19] have hindered their translation from validated technical measures to decision support tools within the clinical management of PD. A myriad of studies demonstrated the technical accuracy of DMOs, and their correlation with standard clinical scales. However, to generate clinical utility, DMO information must be conveyed into actionable items that physicians can apply to individual patient cases. This task involves translating results from highly structured clinical and research studies through the lens of noisy real-world scenarios. Thus, a key first step toward resolving this issue would be to map the knowledge generated by DMOs with clinical domains that could benefit from it.

A structured conceptual framework may bridge the translational gap between research advancements and clinical workflows, encouraging future studies to demonstrate and provide evidence of clinical utility. Therefore, as a primary objective, this expert consensus manuscript aims to build on the clinical and technical knowledge from the Mobilise-D consortium to define the clinical domains where DMOs could be implemented to support personalized care in PD. As a secondary objective, the study seeks to gather expert perspectives on the practical implementation. Ultimately, the goal is to facilitate the translation from *research knowledge* to *clinical utility* of this new information generated by DMOs.

Methods

We conducted a three-round consensus-building study between April 9 and November 27, 2024, to support the integration of DMOs in clinical workflows in PD. As empirical data is unavailable, we chose an electronic-based consensus-building methodology. This method allowed for the development of guidance on the basis of iterative rounds of discussion, voting, and refinement until a predefined expert consensus threshold was reached [20, 21]. Additionally, the electronic format allowed for asynchronous feedback and added flexibility

to ensure the participation of an international expert panel [21].

Expert panel

A single expert panel was recruited from the Mobilise-D consortium's Parkinson's Disease Working Group [10]. This project included 12 international academics with diverse healthcare expertise in PD. Site leaders disseminated the project internally, and interested individuals received a formal enrollment invitation detailing the aim, objectives, and project methodology. Implicit consent was provided by accepting the invitation and voting; anonymity was preserved to ensure equal treatment of expertise across the panel.

Consensus-building procedure

Before each round, the experts received an overview of the goals, including the rating scale and the consensus criteria. All materials provided to the experts were piloted internally with clinicians, clinical researchers, social scientists, and biomedical engineers.

The experts ranked different aspects of a conceptual framework via a 5-level Likert scale for level of agreement (*1 = Strongly Disagree*, *2 = Disagree*, *3 = Neither Agree nor Disagree*, *4 = Agree*, and *5 = Strongly Agree*). A 5-level scale is a robust metric against rounding errors. It measures the intensity of the rater's opinion depending on their level of agreement or disagreement (e.g., “Agree” vs. “Strongly Agree”) [22]. Additionally, an open-ended question was provided at the end of each voting section to gather further expert feedback.

Consensus for the different points evaluated was based on a double threshold: the simultaneous presence of a high level of agreement had to be accompanied by a low level of disagreement. Consensus was thus defined as an average agreement across the experts equal to or above level 4 (i.e., “Agree”), with the simultaneous absence of disagreement (levels 1 and 2). Similar methods have been used in predefined consensus-finding studies to ensure adequate representation of disagreements, even those from minority groups [23–28]. In case of disagreement, experts were asked to describe their concerns in free text and provide insights into the reasons for their disagreement. This feedback was utilized in synergy with the overall agreement to update the framework iteratively.

Initial framework As a starting point for the consensus study, we predefined a framework based on the “treatment cycle” concept described by Klucken et al. [29], where the progression of patient disease is segmented by alternating phases of patient and treatment evaluation. These phases, denominated categories, converge at the clinical consultation, where the physician goes from (a) assessing risks, symptoms, or diseases to (b) selecting and monitoring

the treatment. The primary categories were then subdivided into five secondary domains where DMOs may provide value to the clinical team and thus achieve clinical utility (Fig. 1). For the present study, clinical utility was defined as improving health outcomes, diagnosis, treatment management, or prevention after the use of DMOs [18]. Notably, domains represent distinct clinical contexts and hold different requirements for digital outcomes to provide value. Examples of clinical contexts and potential values were provided for each domain to ensure adequate interpretability.

Rounds 1 and 2 The primary objective of the first round was to find consensus on the established conceptual framework. The experts reviewed a prerecorded workshop describing the “treatment cycles” concept along with the corresponding framework. Then, the experts voted on the overall structure and concept, the contexts, and the clinical values of the DMOs through the 5-level Likert scale described above. The round closed with open-ended suggestions for changes to the framework, including missing domains, updates to the values, or changes to the naming conventions.

Round 3 The primary objective of the third and final round was to review the final framework and present the final structure, taxonomy, and definitions derived from the previous rounds. In contrast to the previous rounds, this final stage was an online interactive session intended to clarify any remaining points from the final framework, allowing for quicker iteration of feedback and sign-off on the study results. Items requiring voting remained anonymous and followed the double-threshold technique as previously employed.

Secondary study objectives We sought expert opinions on three secondary items about the practical application of DMOs by evaluating the timeline to applicability, the challenges foreseen for their implementation in clinical settings, and their main role in the decision-making process. The experts assessed the first two through a single-answer, six-item scale (*currently applicable*, *1 to 5 years*, *5 to 10 years*, *not clinically applicable*, *I do not know*). The third one was evaluated through a multiple choice single-answer question.

Data analysis

Averages were calculated for each voting item. The open-ended questions were grouped by theme and assessed independently of the Likert scale responses. Finally, the participation rate was computed on the basis of the number of enrolled participants and the number of responses received within the allotted time. Descriptive statistics were calculated using a spreadsheet processor. No

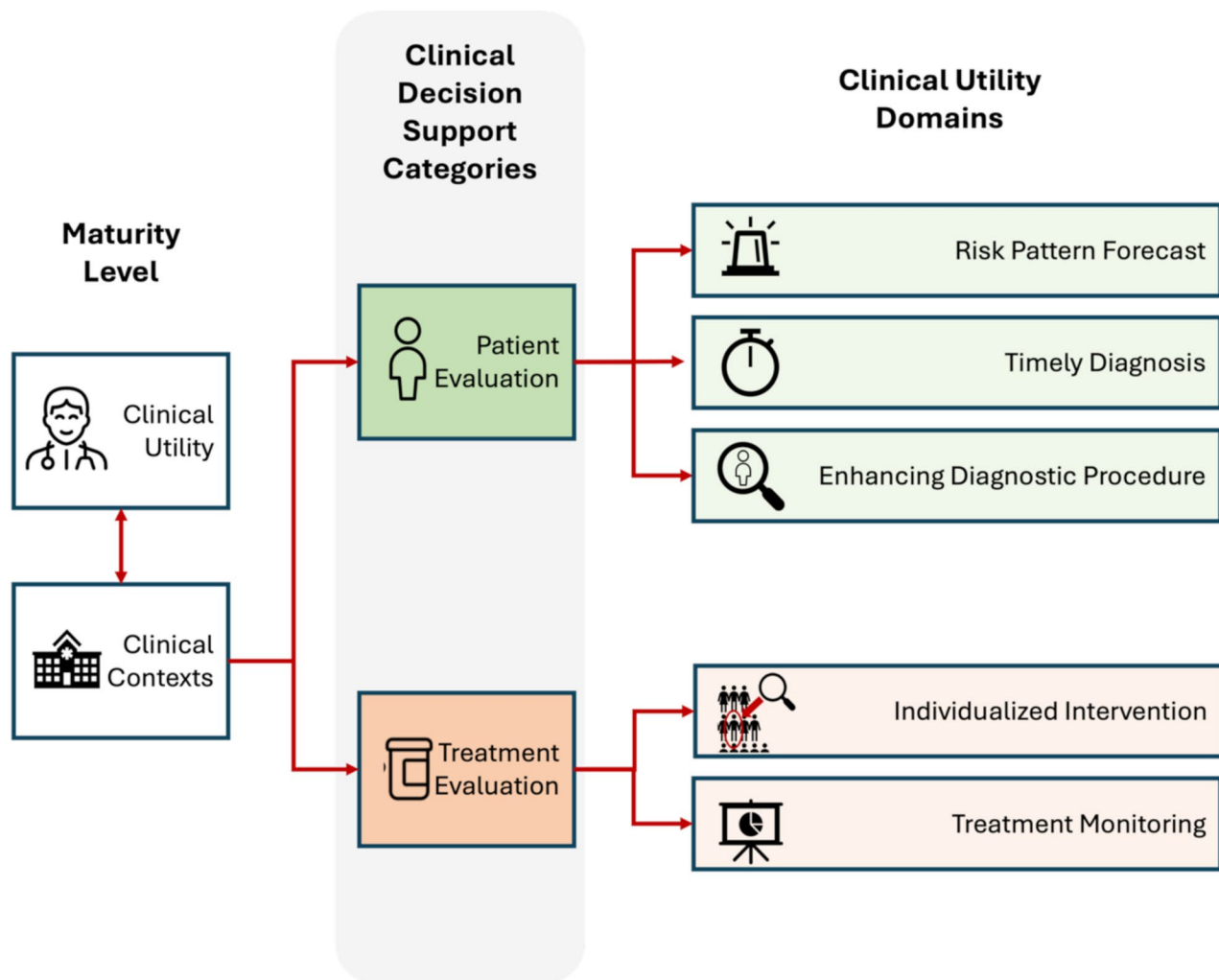


Fig. 1 Predefined Framework based on the treatment cycle concept where the patient management strategy is considered composed by periodic, alternating diagnostic and treatment periods herein represented as *Clinical Decision Support Categories*. The categories are subsequently divided into domains according to the value DMOs can provide. This has been the initial framework presented to the expert panel in round 1

imputation was performed for the replies of the experts who dropped out from the study.

Results

Twelve experts from six academic institutions across Europe received invitations to participate, including 10 (83%) clinical and 2 (17%) technical experts with expertise in neurology, geriatrics, cognition, physiotherapy, biomechanics, and gait analysis (Table 1).

Round 1 results

Round 1 had a 100% response rate (Fig. 2). The experts reached a consensus across the overall framework concept, the two categories, and the *Timely Diagnosis* and *Individualized Intervention* domains (Table S1). From a conceptual applicability standpoint, *Individualized Intervention* was the highest-rated domain, with an overall agreement of 4.64, followed by *Timely Diagnosis* (4.55)

and *Treatment Monitoring* (4.45) (Table S3). Additionally, *Timely Diagnosis* (4.82) and *Enhancing Diagnostic Power* (4.36) received the most approval among the examples evaluating clinical contexts and values (Table. S2-S3).

The most common elements of disagreements in this round were associated with two domains: *Risk Pattern Forecast* and *Treatment Evaluation*. Panel experts suggested changing the naming convention of *Risk Pattern Forecast*, defining the term “individualized,” and narrowing the scope of each domain to an actionable clinical process. Experts expressed concerns regarding the prognostic power of DMOs, which were seen as more capable of detecting subtle deterioration rather than being hallmarks of future impairment. Updates to the *Treatment Evaluation* domains were also requested to account for treatment feedback loops and patient self-management. All measures were retained for voting in Round 2 to ensure alignment with changes derived from

Table 1 Panel expert characteristics. Results are expressed as “mean (Standard Deviation)” or “Unit, Percentage” according to the variable considered

Characteristics	All Panel Experts (n = 12)	Clinical Background (n = 10)	Technical Background (n = 2)
Average Age, years	50.33 (9.54)	50.60 (9.16)	49.00 (7.24)
Sex (n, %)			
• Male	4, 33%	4, 40%	0, 0%
• Female	8, 67%	6, 60%	2, 100%
Education (n, %)			
• Medicine	4	4	0, 0%
• Psychology	1	0	1, 50%
• Physiotherapy	5	5	0, 0%
• Sports Science	2	1	1, 50%
Background (n, %)			
• Clinical	10, 83%	10, 100%	0, 0%
• Technical	2, 17%	0, 0%	2, 100%
Residence (n, %)			
• United Kingdom	5, 42%	4, 40%	1, 50%
• Germany	4, 33%	4, 40%	0, 0%
• Belgium	1, 8%	1, 10%	0, 0%
• Norway	1, 8%	0, 0%	1, 50%
• Israel	1, 8%	1, 10%	0, 0%
Average Experience, years	16.75 (7.24)	17.40 (15.56)	13.50 (2.12)

the feedback provided. Finally, experts requested that disease diagnosis be treated as a separate category and not be embedded within the *Patient Evaluation*.

Concept-related suggestions were split into taxonomy or definition comments and explored separately in the subsequent rounds.

Round 2 results

Round 2 had a lower response rate of 67%, with 8 of the original 12 experts responding. Based on Round 1 feedback, two framework structures were presented to the experts (Fig. S4). 75% of experts voted for a new *Disease Diagnosis* category and corresponding domain to be added to the framework (Fig. S4a) rather than having it nested within the *Patient Evaluation* category. The remaining two experts preferred an updated version of the original structure presented in Round 1 without a new category (Fig. S4b). However, one of the two experts provided open feedback that aligned with the new category structure, increasing the acceptance of the additional category to 87.5%. All other inclusions to the framework were rejected.

Independent of the framework's structure, the experts reached a consensus on the taxonomy, values, and 4 out of 6 domain definitions (Table 2). Additionally, consensus was reached on most research questions, apart from *Enhancing Diagnostic Procedure*, which was not tested as it was the domain corresponding to the new category.

The main disagreements were associated with the domains of *Predicting Risk* and *Detecting Timely Deterioration*. In particular, one expert disagreed with the terminology of the first domain, requesting more emphasis on the fact that risks are not exclusively associated with

prodromal phases of diseases but can also be linked to symptoms. Analogously, for *Detecting Timely Deterioration*, concerns were expressed on the definition that originally included an “increased sampling frequency of the clinical assessments.” This wording was considered inadequate and potentially misleading due to its overlap with the electronic interpretation of the concept of “sampling frequency.”

Given the results achieved in Round 2, three items were retained for subsequent voting: the domain definitions for *Predicting Risk* and *Timely Detecting Deterioration* and the research question for *Enhancing Diagnostic Procedure*.

Round 3 results

Round 3 had a consistent response rate of 67% which was comparable to that of the previous round. The framework structure remained stable, with three categories and six underlying domains (Fig. S4a) confirming its robustness. Among the three retained items, experts reached a full consensus on all points. In addition to the consensus items, panel experts provided supplementary suggestions, including aligning the wording of definition and the research question for *Enhancing Diagnostic Procedure*.

Expert-level agreement was achieved on a final clinical utility framework composed of three categories and six underlying domains (Fig. 3). Full consensus was also reached on the overall framework concept, taxonomy, and domain definitions necessary, confirming the role of DMOs in personalized care in PD (Table 3).

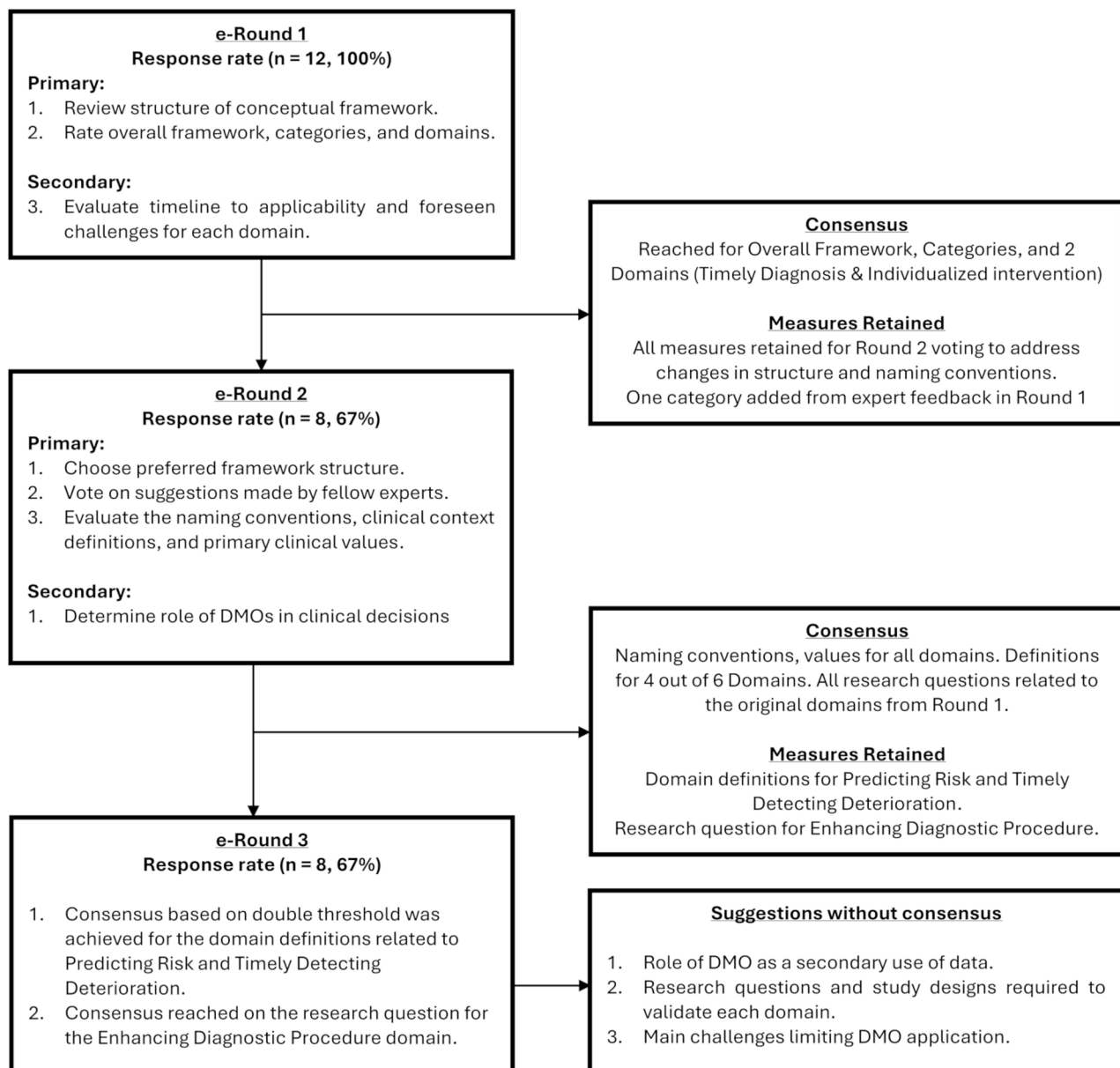


Fig. 2 Flow diagram of the modified consensus methodology. The figure describes the primary and secondary objectives targeted during each round, and the consensus achieved in each iteration

Table 2 Expert consensus results by round. ^aThese symbols denote absence of double threshold consensus. Check (✓) denotes consensus reached. *Denotes averages calculated over 6 experts only due to conceptual differences between V1 and V2 framework

Expert Consensus on Framework Domains	Average Agreement Results							
	Round 1		Round 2			Round 3		
	Concept	Clinical Value	Concept		Clinical Value	Concept		Clinical Value
			Taxonomy	Definition		Taxonomy	Definition	
Enhancing Diagnostic Procedure	-	-	4.50*	4.33*	4.00*	✓	✓	✓
Predicting Risk	4.36 ^a	4.09 ^a	4.50*	3.67* ^a	4.17*	✓	4.38	✓
Detecting Timely Deterioration	4.82 ^a	4.45 ^a	4.38	4.25 ^a	4.50	✓	4.38	✓
Enhancing Symptom Assessment	4.36 ^a	4.09 ^a	4.50	4.50	4.50	✓	✓	✓
Selecting Treatment	4.00 ^a	3.73 ^a	4.50	4.38	4.25	✓	✓	✓
Monitoring Treatment Response	4.27 ^a	4.18 ^a	4.63	4.00	4.25	✓	✓	✓

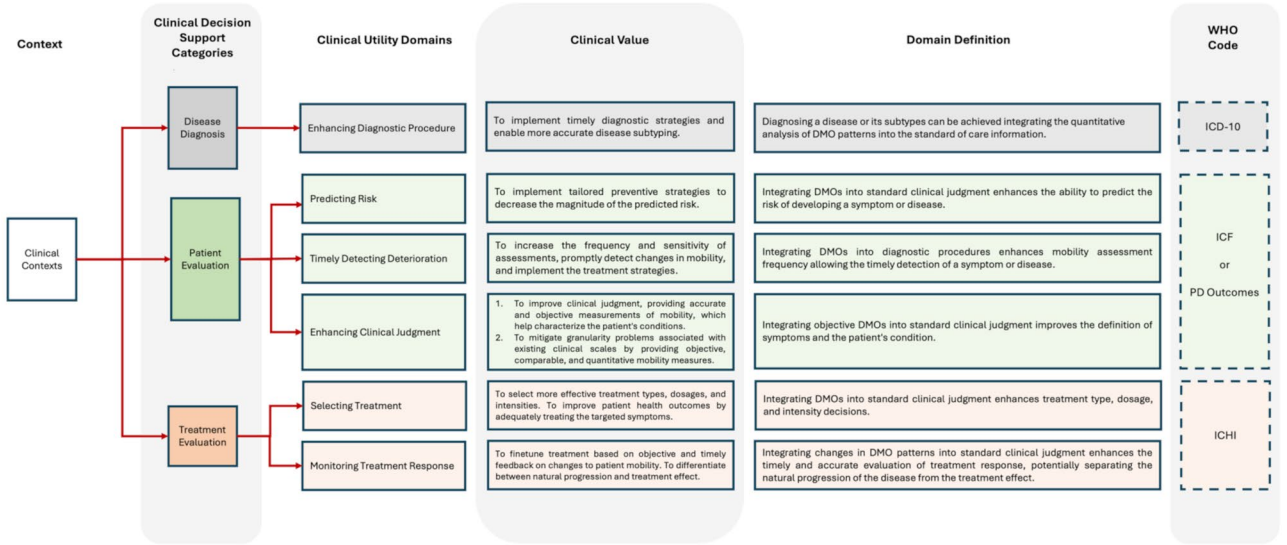


Fig. 3 Final Clinical Utility Framework. The clinical contexts are divided into three main categories which align with internationally recognized existing clinical frameworks used by the world health organization. WHO World Health Organization, ICD-10: International Classification of Disease 10th revision. ICF: International Classification of Functioning, Disability and Health coding system. ICHI: International Classification of Health Interventions

Table 3 Framework taxonomy evolution by round. The table shows the evolution in terms of taxonomy of the different domains and categories according to the feedback from the expert panel

Framework Taxonomy Evolution					
Round 1		Round 2		Round 3	
Categories	Domains	Categories	Domains	Categories	Domains
-	-	Disease	Enhancing Diagnostic Procedure	Disease	Enhancing Diagnostic Procedure
		Diagnosis		Diagnosis	
Patient Evaluation	Risk Patterns Forecast	Patient	Predicting Risk	Patient	Predicting Risk
	Timely Diagnosis	Evaluation	Timely Detecting Deterioration	Evaluation	Timely Detecting Deterioration
	Enhancing Diagnostic Procedure		Enhancing Clinical Judgment		Enhancing Clinical Judgment
Treatment	Individualized Intervention	Treatment	Selecting Treatment	Treatment	Selecting Treatment
Evaluation	Treatment Monitoring	Evaluation	Monitoring Treatment Response	Evaluation	Monitoring Treatment Response

Secondary study results

Throughout the study, the experts did not consider any domains currently applicable in clinical practice, with all the domains voted on average as “1–5 years” away from clinical utility (Table S5). Overwhelmingly, the most significant challenge to clinical applicability was the need for reference values, followed by concerns about the interpretability of DMOs (Table S6). Technology readiness was the main obstacle to implementing DMOs within clinical practice, followed closely by the cost associated with the technologies employed (Table S7). The latter stems from the limited information provided by gait as a single motor manifestation and difficulties in meaningfully visualizing the DMO-derived information. Finally, when asked about the role of DMOs, the experts considered them complementary rather than primary data sources useful for making or verifying a clinical decision (Fig. S8).

Discussion

In this expert consensus manuscript, we developed a conceptual framework to bridge the gap between DMOs for research and DMOs for personalized clinical decision support. Using PD as a use case, we outlined specific clinical utility domains that can benefit from knowledge generated by DMOs.

A full consensus on a final clinical utility framework was achieved based on the feedback of twelve Mobilise-D multidisciplinary experts from six international institutions. This consisted of three main categories (*Disease Diagnosis, Patient Evaluation, and Treatment Evaluation*) and six underlying domains (*Enhancing Diagnostic Procedure, Predicting Risk, Timely Detecting Deterioration, Enhancing Clinical Judgment, Selecting Treatment, and Monitoring Treatment Response*). The framework ultimately defines three areas of application: identification of

the disease, characterization of the patient, and support for treatment decisions.

The initial framework, comprised of only two categories (*Patient Evaluation* and *Treatment Evaluation*) and five domains, achieved consensus within Round 1, validating the appropriateness of the “treatment cycle” concept as a foundation for our work [29]. However, based on expert feedback, a new *Disease Diagnosis* category stemmed from the *Patient Evaluation* branch. The rationale was the conceptual distinction between performing the diagnostic workup (identifying a new disease or disease subtype) and evaluating progression (symptomatic/syndromic workups) within the context of an existing disease [30, 31]. This change closely aligns with clinical practice by emphasizing the clinical utility of DMOs in accurately classifying diseases via the International Classification of Disease (ICD) code [32, 33] or determining a change in the functional and disability health domains structured within the International Classification of Functioning, Disability and Health (ICF) coding system [34]. This alignment extends to the *Treatment Evaluation* branch, where the treatment selection domain allows for reporting health interventions via the International Classification of Health Interventions (ICHI) structure [35].

Although experts reached a consensus according to the double threshold rule, some concerns remained. A critical area of disagreement was associated with the feasibility of using DMOs to provide individualized interventions and treatment monitoring. Specifically, experts questioned the level of granularity provided by DMOs, with one expert stating that it appears impossible to detect monthly impairment changes at the single-patient level due to the extreme heterogeneity of PD clinical presentations. A second point of divergence was the prognostic power of DMOs. Some experts consider these digital biomarkers unable to provide any insight into future progression, as their information is limited to a snapshot of the present time.

These concerns were reflected in the answers related to the “Time to Applicability” rather than in the elements of the framework (Table S5). Nevertheless, it is essential to note that only a few experts considered DMOs to be immediately applicable in generating clinical utility, with most agreeing technological implementation is between 1 and 5 years from applicability. Interestingly, even if DMOs can generate clinical utility, the panel unanimously concluded that DMOs do not represent the primary source for making a clinical decision. Thus, experts agreed that DMOs play a key secondary, complementary role in the decision process or in verifying its outcomes.

A core objective of this framework is to facilitate the translation of the research knowledge of DMOs (i.e., symptomatic characterization and outcomes generated in research clinical trials) into clinical utility applicable

to usual care. The latter can support clinical decisions through digital health information, which can be integrated as objective measures into diagnostic digital medical devices to tailor individualized interventions based on a patient’s profile. Additionally, this new information can complement the patient’s perspective and self-awareness of their symptoms and clinical experiences, thus directly supporting personalized care not only from the healthcare professional perspective but also from the patient’s side [36].

However, to achieve these goals, crucial challenges must be overcome to integrate DMOs into digital medical devices (e.g., FDA/EMA-certified tools) that can provide clinical decision support. Expert-derived feedback revealed that DMO interpretability, appropriate reference data, and meaningful visualizations for decision-making are the primary challenges in achieving clinical utility. These results align with what was previously reported by Sharma et al. [37] and complement the cost and data accuracy elements presented by Ambrens et al. [6]. Additionally, differences across medical procedures, heterogeneous reimbursement strategies, diverse levels of physician digital health literacy, are all aspects that increase the complexity of making DMOs applicable across different clinical settings and healthcare ecosystems and align with similar challenges reported throughout interdisciplinary care in PD [38]. For this reason, a significant amount of effort is still required in clinical research to translate DMOs into real-world clinical tools. This effort should come in parallel with continuous harmonization and update of the regulatory apparatus by national authorities [39].

It is essential to note that our framework is not yet complete. Its six clinical utility domains can guide future clinical studies seeking to implement DMOs within patient treatment cycles for personalized care. The individual research questions and results of such clinical studies will play a vital role in generating an additional layer of “subdomains” in our framework, where more precise, practical medical applications, contexts, and interpretations of DMOs’ values will be described. An example of a potential subdomain associated with our *Treatment Evaluation* category, and more specifically the *Selecting Treatment* branch, is the study of Somerset et al., where DMOs have successfully been utilized to tailor the deep brain stimulation parameters of patients with PD [40].

This study has some potential methodological limitations. First, the large dropout between the first and the second round, from 12 to 8 experts, could have partially biased the consensus results by reducing the amount of feedback received. Second, the electronic nature reduced the possibility of providing real-time feedback and requesting immediate clarification if the feedback was unclear. Additionally, it prevented experts from building

upon ideas from fellow panel members. Third, the association between the panel experts and the longitudinal Mobilise-D clinical validation study [10] in parallel with the unbalance between clinical and technical backgrounds (10 vs. 2, see Table 1 for additional details) may bias their perspective toward the applicability and clinical utility of the DMOs, as described in the paradox of choice. It is important to underline, however, that clinical experts involved in this study still present robust technical skills, crafted through participation in numerous top quality DMOs analysis. Furthermore, the iterative workflow increased confidence in the objectivity of the expert feedback nevertheless, future studies are required to validate our findings with larger and more heterogeneous expert cohorts. Fourth, the small, discipline-unbalanced expert panel may limit the generalizability of the consensus findings, suggesting that future studies should involve broader expert groups across healthcare professions to ensure optimal identification of translational challenges and clinical applications that have not been considered by this work. However, our study achieved an average participation of 75%, with the number of rounds and participation rate consistent with those of other consensus-building studies [6, 20, 25]. Finally, our framework focuses exclusively on the clinical aspects of clinical utility. Further studies should investigate the legal, ethical, and regulatory dimensions influencing and governing the translation from research utility to clinical utility.

Conclusions

Through a structured consensus-building process, we developed the first conceptual framework defining clinical utility domains where objective DMOs could support personalized care in PD. Experts reached a consensus on clinical utility domains aligned with disease diagnosis, patient evaluation, and treatment management, underscoring areas where DMOs have the potential to generate clinical value for the patient and the clinician.

From a consensus perspective, despite the study's limitations, this framework provides a preliminary structure for subsequent studies to confirm our results with diverse expert cohorts and expand our findings beyond PD. From a DMOs clinical utility point of view, our results support researchers planning future clinical trials with contexts where DMOs can play a valuable role for clinical decision support in the management of PD, and insights on challenges associated with them. Ultimately, this is the first step toward developing guidelines to assess DMOs' clinical utility and support their integration into Real World clinical practice.

Abbreviations

PD	Parkinson's disease
DMOs	Digital Mobility Outcomes
ICD	International Classification of Disease

ICF	International Classification of Functioning, Disability and Health
ICHI	International Classification of Health Interventions

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42466-025-00426-8>.

Supplementary Material 1

Acknowledgements

We would like to thank all the members of the Digital Medicine Group for supporting the process of creation of the different rounds with valuable feedback.

Author contributions

JK developed the initial framework. ACM, SS, and JK developed the consensus study. ACM and SS interpreted the data. IP provided expert feedback on the consensus methodology and presentation. BV, RM, LR, PB, AN, LA, AN, AY, SS, HG, HH, WM, and AM were panel experts. All the authors contributed to the manuscript, read, and approved the final manuscript.

Funding

This research was funded in whole, or in part, by the Luxembourg National Research Fund (FNR), grant reference 14146272 & 17981757. For the purpose of open access, and in fulfilment of the obligations arising from the grant agreement, the author has applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission.

Research was supported by the Mobilise-D project, which received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 820820. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

This paper presents independent research supported by the NIHR Newcastle Biomedical Research Centre (BRC). The NIHR Newcastle Biomedical Research Centre (BRC) is a partnership between Newcastle Hospitals NHS Foundation Trust, Newcastle University, and Cumbria, Northumberland and Tyne and Wear NHS Foundation Trust and is funded by the National Institute for Health and Care Research (NIHR).

Lynn Rochester, Lisa Alcock, Alison Yarnall are supported by funding from the National Institute for Health and Care Research (NIHR) Senior Investigator Awards (2020–2024; 2024–2028); and the NIHR Newcastle Biomedical Research Centre (BRC), the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at Newcastle upon Tyne Hospitals NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Heiko Gassner is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1483 – Project-ID 442419336, EmpkinS, and Project-ID 438496663, Mobility_APP. He is further supported by the Fraunhofer Internal Programs under Grant No. Attract 044-602140 and 044-602150 as well as Grant No. SME 40-09311.

Data availability

The dataset supporting the conclusions of this article is available in the repository: <https://doi.org/10.17881/g865-hf21>.

Declarations

Ethics approval and consent to participate

Ethical approval was not required for this study as it did not involve research on human participants. The consensus offered contributions to the research topic where expert opinions were voluntary.

Consent for publication

Not applicable.

Competing Interests

ACM, SS, IP, BV, MR, LR, PB, AN, AY, SST, HG, HH, WM, and AM declare no competing interests. LA reports consultancy activity with Hoffmann-La Roche Ltd. outside of this study. JK acknowledges shares in Portables HCT GmbH, manufacturer of the gait sensor system providing digital mobility outcomes. The companies did not influence the study, nor and the outcomes of the study.

Author details

¹Luxembourg Centre for Systems Biomedicine, Digital Medicine Group, University of Luxembourg, Esch-sur-Alzette, Luxembourg

²Centre Hospitalier de Luxembourg, Luxembourg City, Luxembourg

³NIHR Newcastle Biomedical Research Centre, Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne, UK

⁴The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

⁵Department of Molecular Neurology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

⁶Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

⁷Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

⁸Department of Neurology, University Hospital Schleswig-Holstein and Kiel University, Kiel, Germany

⁹Laboratory for Early Markers of Neurodegeneration (LEMON), Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

¹⁰School of Medicine & Health Sciences, Tel Aviv University, Tel Aviv, Israel

¹¹Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

¹²Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

¹³Neurorehabilitation Research Group (eNRGy), Department of Rehabilitation Sciences, KU Leuven, Leuven, Vlaams-Brabant, Belgium

¹⁴Leuven Brain Institute (LBI), Leuven, Belgium

Received: 14 May 2025 / Accepted: 2 September 2025

Published online: 18 September 2025

References

- Kirk, C., Küderle, A., Micó-Amigo, M. E., Bonci, T., Paraschiv-Ionescu, A., Ullrich, M., et al. (2024). Mobilise-D insights to estimate real-world walking speed in multiple conditions with a wearable device. *Scientific Reports*, 14(1), 1754.
- Middleton, A., Fritz, S. L., & Lusardi, M. (2015). Walking speed: The functional vital sign. *Journal of Aging and Physical Activity*, 23(2), 314–322.
- Höglinger, G., German, P. G., Committee, Bähr, M., Becktepe, J., Berg, D., Brockmann, K., et al. (2024 June). Diagnosis and treatment of parkinson's disease (guideline of the German society for Neurology). *Neurol Res Pract*, 6(1), 30.
- Mancini, M., Afshari, M., Almeida, Q., Amundsen-Huffmaster, S., Balfany, K., Camicioli, R., et al. (2025). Digital gait biomarkers in parkinson's disease: Susceptibility/risk, progression, response to exercise, and prognosis. *Npj Park Dis*, 11(1), 51.
- Jaeger, S. U., Wohlrab, M., Schoene, D., Tremmel, R., Chambers, M., Leocani, L., et al. (2022). Mobility endpoints in marketing authorisation of drugs: What gets the European medicines agency moving? *Age and Ageing*, 51(1), afab242.
- Ambrens, M., Delbaere, K., Butcher, K., Close, J., Gonski, P., Kohler, F. (2025 Jan). Wearable Technology in Mobility and Falls Health Care: Finding Consensus on Their Clinical Utility and Identifying a Roadmap to Actual Use. *J Geriatr Phys Ther*. 8 [cited 2025 Mar 24]; Available from: <https://journals.lww.com/https://doi.org/10.1519/JPT.00000000000000434>
- Farzanehfard, P., Woodrow, H., & Horne, M. (2021). Assessment of wearing off in parkinson's disease using objective measurement. *Journal of Neurology*, 268(3), 914–922.
- Isaacson, S. H., Boroojerdi, B., Waln, O., McGraw, M., Kreitzman, D. L., Klos, K., et al. (2019 July). Effect of using a wearable device on clinical decision-making and motor symptoms in patients with parkinson's disease starting transdermal rotigotine patch: A pilot study. *Parkinsonism & Related Disorders*, 64, 132–137.
- Hadley, A. J., Riley, D. E., & Heldman, D. A. Real-World evidence for a Smart-watch-Based parkinson's motor assessment app for patients undergoing therapy changes. *Digit Biomark* 2021 Sept 8;5(3):206–215.
- Mikolaizak, A. S., Rochester, L., Maetzler, W., Sharrack, B., Demeyer, H., Mazzà, C. (2022). Connecting real-world digital mobility assessment to clinical outcomes for regulatory and clinical endorsement—the Mobilise-D study protocol. Phillips T, editor. *PLOS ONE*. 17(10):e0269615.
- Micó-Amigo, M. E., Bonci, T., Paraschiv-Ionescu, A., Ullrich, M., Kirk, C., Soltani, A., et al. (2023 June). Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. *J NeuroEngineering Rehabil*, 14(1), 78.
- Shah, V. V., McNames, J., Mancini, M., Carlson-Kuhta, P., Nutt, J. G. (2020 July) El-Gohary, M. Digital Biomarkers of Mobility in Parkinson's Disease During Daily Living. *J Park Dis*. 28;10(3):1099–111.
- Corrà, M. F., Atsraei, A., Sardoreira, A., Hansen, C., Aminian, K., Correia, M. (2021 June). Comparison of Laboratory and Daily-Life Gait Speed Assessment during ON and OFF States in Parkinson's Disease. *Sensors*. 9;21(12):3974.
- Schlachetzki, J. C. M., Barth, J., Marxreiter, F., Gossler, J., Kohl, Z., Reinfelder, S., et al. (2017). Wearable sensors objectively measure gait parameters in parkinson's disease. *PLoS One*, 12(10), e0183989.
- Hug, A., Spingler, T., Pleines, V., Heutheaus, L., Schoenfeld, M. A., Hauptmann, B., et al. (2025). Exploring the relationship of clinical walking tests with 8-months inertial measurement unit (IMU)-based real world mobility tracking in stroke and spinal cord injury survivors. *Neurol Res Pract*, 7(1), 30.
- Raccagni, C., Sidoroff, V., Paraschiv-Ionescu, A., Roth, N., Schönherr, G., Eskofier, B., et al. (2024). Effects of physiotherapy and home-based training in parkinsonian syndromes: Protocol for a randomised controlled trial (MobilityAPP). *British Medical Journal Open*, 14(5), e081317.
- Caballol, N., Bayés, À., Prats, A., Martín-Baranera, M., & Quispe, P. (2023). Feasibility of a wearable inertial sensor to assess motor complications and treatment in Parkinson's disease. Suzuki K, editor. *PLOS ONE*. 18(2):e0279910.
- Sapienza, S. (2024). Assessing the clinical utility of inertial sensors for home monitoring in parkinson's disease: A comprehensive review.
- Xu, S., Kim, J., Walter, J. R., Ghaffari, R., & Rogers, J. A. (2022). Translational gaps and opportunities for medical wearables in digital health. *Science Translational Medicine*, 14(666), eabn6036.
- Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017 Sept). Guidance on conducting and reporting Delphi studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine*, 31(8), 684–706.
- Nasa, P., Jain, R., & Juneja, D. (2021 July). Delphi methodology in health-care research: How to decide its appropriateness. *World J Methodol*, 20(4), 116–129.
- Lehmann, D. R., & Hulbert, J. (1972). Are Three-Point scales always good enough? *Journal of Marketing Research*, 9(4), 444–446.
- WHO handbook for guideline development 2nd Edition. [cited 2024 Oct 21]. Available from: <https://www.who.int/publications/i/item/9789241548960>
- Barrios, M., Guilera, G., Nuño, L., & Gómez-Benito, J. (2021). Consensus in the Delphi method: What makes a decision change? *Technol Forecast Soc Change*, 163, 120484.
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., et al. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401–409.
- Foth, T., Efstathiou, N., Vanderspank-Wright, B., Ufholz, L. A., Dütthorn, N., Zimansky, M., et al. (2016). The use of Delphi and nominal group technique in nursing education: A review. *International Journal of Nursing Studies*, 60, 112–120.
- Keeney, E., Thom, H., Turner, E., Martin, R. M., & Sanghera, S. (2021). Using a modified Delphi approach to gain consensus on relevant comparators in a Cost-Effectiveness model: Application to prostate cancer screening. *Pharmacoeconomics*, 39(5), 589–600.
- Nakanishi, N., Liu, K., Kawauchi, A., Okamura, M., Tanaka, K., Katayama, S. Instruments to assess post-intensive care syndrome assessment: a scoping review and modified Delphi method study. *Crit Care*. 2023 Nov 7 [cited 2025 July 21];27(1). Available from: <https://ccforum.biomedcentral.com/articles/https://doi.org/10.1186/s13054-023-04681-6>
- Glucken, J., Krüger, R., Schmidt, P., & Bloem, B. R. (2018). Management of Parkinson's Disease 20 Years from Now: Towards Digital Health Pathways. Brundin P, Langston JW, Bloem BR, editors. *J Park Dis*. 8(s1):S85–94.
- Hausmann, D., Zulian, C., Battegay, E., & Zimmerli, L. (2016). Tracing the decision-making process of physicians with a decision process matrix. *BMC Med Inform Decis Mak*, 16(1), 133.

31. Whelehan, D. F., Conlon, K. C., & Ridgway, P. F. (2020). Medicine and heuristics: Cognitive biases and medical decision-making. *Ir J Med Sci* 1971 -, 189(4), 1477–1484.
32. Classifications [cited 2024 Nov 11]. Available from: <https://www.who.int/standards/classifications>
33. Importance of ICD [cited 2024 Nov 11]. Available from: <https://www.who.int/standards/classifications/frequently-asked-questions/importance-of-icd>
34. International Classification of Functioning Disability and Health (ICF). [cited 2024 Nov 11]. Available from: <https://icd.who.int/dev11/l-icf/en#/http://id.who.int/icd/entity/1688850216>
35. International Classification of Health Interventions [cited 2024 Nov 11]. Available from: <https://www.who.int/standards/classifications/international-classification-of-health-interventions>
36. Fomo, M., Borga, L., Abel, T., Santangelo, P., Riggare, S., Klucken Empowering capabilities of people with chronic condition supported by digital health technologies: A scoping review. *Journal of Medical Internet Research*.
37. Sharma, Y., Cheung, L., Patterson, K. K., & Iaboni, A. (2024). Factors influencing the clinical adoption of quantitative gait analysis technology with a focus on clinical efficacy and clinician perspectives: A scoping review. *Gait & Posture*, 108, 228–242.
38. Lummer, C., Eggers, C., Becker, A., Demandt, F., & Warnecke, T. (2024). Parkinson Netzwerke Deutschland e.v. Interdisciplinary network care collaboration in parkinson's disease: A baseline evaluation in Germany. *Neurol Res Pract*, 6(1), 5.
39. Boers, M., Rochereau, A., Stuwe, L., Miguel, L. S., Klucken, J., Mezei, F. (2025). Classification grid and evidence matrix for evaluating digital medical devices under the European union landscape. *Npj Digit Med*. May 24 [cited 2025 July 11];8(1). Available from: <https://www.nature.com/articles/s41746-025-01697-w>
40. Somerset, J., Hammersley, B., & Bonello, M. (2019). Can the quantified timed up and go (QTUG) device support decision making for patients undergoing deep brain stimulation? *WILEY 111 RIVER ST. HOBOKEN 07030 – 5774, NJ USA*.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.