Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 11 (2019) 301-309

Working Group Summaries for European Joint Programming For Neurodegenerative Research (JPND)

# Neuroimaging biomarkers for clinical trials in atypical parkinsonian disorders: Proposal for a Neuroimaging Biomarker Utility System

Thilo van Eimeren[a,b,*], Angelo Antonini[c], Daniela Berg[d], Nico Bohnen[e,f,g], Roberto Ceravolo[h], Alexander Drzezga[a,b,i], Günter U. Höglinger[j], Makoto Higuchi[k], Stephane Lehericy[l], Simon Lewis[m], Oury Monchi[n,o], Peter Nestor[p,q], Matej Ondrus[r], Nicola Pavese[s], María Cecilia Peralta[t], Paola Piccini[u], José Ángel Pineda-Pardo[v], Irena Rektorová[w], María Rodríguez-Oroz[x], Axel Rominger[y], Klaus Seppi[z], A. Jon Stoessl[aa], Alessandro Tessitore[bb], Stephane Thobois[cc,dd,ee], Valtteri Kaasinen[ff,gg], Gregor Wenning[hh], Hartwig R. Siebner[ii], Antonio P. Strafella[jj,kk,ll], James B. Rowe[mm], for the MDS Neuroimaging Study Group and the JPND Working Group ASAP-SynTau

[a]*Multimodal Neuroimaging, Department of Nuclear Medicine, Medical Faculty and University Hospital, University of Cologne, Cologne, Germany*
[b]*German Centre for Neurodegenerative Diseases (DZNE), Bonn-Cologne, Germany*
[c]*Department of Neuroscience, University of Padua, Padua, Italy*
[d]*Department of Neurology, UKSH, Campus Kiel, Christian-Albrechts-University, Kiel, Germany*
[e]*Division of Nuclear Medicine, Department of Radiology, University of Michigan, and VAMC, Ann Arbor, MI, USA*
[f]*Department of Neurology, University of Michigan, Ann Arbor, MI, USA*
[g]*VAMC, Ann Arbor, MI, USA*
[h]*Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy*
[i]*Forschungszentrum Jülich, INM-2, Jülich, Germany*
[j]*German Centre for Neurodegenerative Diseases (DZNE), and Technical University Munich, Department of Neurology, Munich, Germany*
[k]*National Institutes for Quantum and Radiological Science and Technology, Chiba, Japan*
[l]*Institut du Cerveau et de la Moelle épinière – ICM, Centre de NeuroImagerie de Recherche – CENIR, ICM Team "Movement Investigations and Therapeutics", Sorbonne Universités, Inserm U1127, CNRS UMR, Paris, France*
[m]*Brain & Mind Centre, Sydney Medical School, University of Sydney, Sydney, NSW, Australia*
[n]*Department of Clinical Neurosciences, Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada*
[o]*Department of Radiology, Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada*
[p]*Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia*
[q]*Mater Hospital, South Brisbane, QLD, Australia*
[r]*AXON Neuroscience CRM Services SE, Bratislava, Slovak Republic*
[s]*Newcastle Magnetic Resonance Centre & Positron Emission Tomography Centre, Newcastle University, Newcastle upon Tyne, United Kingdom*
[t]*Center for Medical Education and Clinical Research, Section of Neurology, Buenos Aires, Argentina*
[u]*Department of Medicine, Imperial College London, London, United Kingdom*
[v]*hmCINAC, University Hospital HM Puerta del Sur, CEU-San Pablo University, Móstoles, Madrid, Spain*
[w]*First Department of Neurology – Faculty of Medicine and CEITEC MU, Masaryk University, Brno, Czech Republic*
[x]*Department of Neurology, Clinica Universidad de Navarra, Pamplona, Spain*
[y]*Department of Nuclear Medicine, Inselspital, Universitätsspital Bern, Bern, Switzerland*
[z]*Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria*
[aa]*Pacific Parkinson's Research Centre, University of British Columbia, Vancouver, Canada*
[bb]*Department of Medical, Surgery, Neurological, Metabolic and Aging Sciences, University of Campania, "L. Vanvitelli", Caserta CE, Italy*
[cc]*Université de Lyon, Université Claude Bernard Lyon 1, Faculté de Medecine Lyon Sud Charles Merieux, Lyon, France*
[dd]*Hospices Civils de Lyon, Hopital Neurologique Pierre Wertheimer, Neurologie C, Lyon, France*
[ee]*CNRS, Institut des Sciences Cognitives, Bron, France*
[ff]*Department of Neurology, University of Turku, Turku, Finland*
[gg]*Division of Clinical Neurosciences, Turku University Hospital, Turku, Finland*
[hh]*Division of Clinical Neurology, Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria*
[ii]*Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark*

*Corresponding author. Tel.: +492214787570; Fax: +4922147882845.
E-mail address: thilo.van-eimeren@uk-koeln.de

https://doi.org/10.1016/j.dadm.2019.01.011

[jj]E.J. Safra Parkinson Disease Program, Toronto Western Hospital & Krembil Research Institute, UHN, Toronto, Ontario, Canada
[kk]Research Imaging Centre, Campbell Family Mental Health Research Institute, CAMH, Toronto, Ontario, Canada
[ll]University of Toronto, Toronto, Ontario, Canada
[mm]Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

**Abstract**

**Introduction:** Therapeutic strategies targeting protein aggregations are ready for clinical trials in atypical parkinsonian disorders. Therefore, there is an urgent need for neuroimaging biomarkers to help with the early detection of neurodegenerative processes, the early differentiation of the underlying pathology, and the objective assessment of disease progression. However, there currently is not yet a consensus in the field on how to describe utility of biomarkers for clinical trials in atypical parkinsonian disorders.

**Methods:** To promote standardized use of neuroimaging biomarkers for clinical trials, we aimed to develop a conceptual framework to characterize in more detail the kind of neuroimaging biomarkers needed in atypical parkinsonian disorders, identify the current challenges in ascribing utility of these biomarkers, and propose criteria for a system that may guide future studies.

**Results:** As a consensus outcome, we describe the main challenges in ascribing utility of neuroimaging biomarkers in atypical parkinsonian disorders, and we propose a conceptual framework that includes a graded system for the description of utility of a specific neuroimaging measure. We included separate categories for the ability to accurately identify an intention-to-treat patient population early in the disease (Early), to accurately detect a specific underlying pathology (Specific), and the ability to monitor disease progression (Progression).

**Discussion:** We suggest that the advancement of standardized neuroimaging in the field of atypical parkinsonian disorders will be furthered by a well-defined reference frame for the utility of biomarkers. The proposed utility system allows a detailed and graded description of the respective strengths of neuroimaging biomarkers in the currently most relevant areas of application in clinical trials.



*Keywords:* Biomarker; Trials; PSP; MSA; CBD; CBS; Neurodegeneration; Biomarker; Multicentric; Multisite; Neuroimaging; Harmonization; PET; MRI

## 1. Introduction

Parkinsonian disorders are the most common neurodegenerative diseases after Alzheimer's disease [1]. In about 20% of patients, parkinsonism is not due to Parkinson's disease (PD) pathology, which then is commonly referred to as an atypical parkinsonian disorder (AP) [2,3]. The most frequent forms of underlying neurodegenerative pathologies in AP are progressive supranuclear palsy (PSP), multiple system atrophy (MSA), and corticobasal degeneration (CBD). Neuronal degeneration is generally much more aggressive and symptomatic therapy is much less effective in these disorders than in PD [4]. This does not only lead to a significantly shorter survival but also to a dramatically steeper loss of function, for example, in activities of daily living [5–7] (see Fig. 1). From a neuropathological perspective, parkinsonian disorders are proteinopathies and distinguishable with regard to the form and localization of pathological protein aggregates. In PD, alpha-synuclein accumulation in the form of intraneuronal Lewy bodies occurs progressively and probably largely in an ascending order from the brainstem to thecerebral cortex [8]. MSA is also considered an alpha-synucleinopathy, although protein aggregates mostly appear as cytoplasmic oligodendroglial inclusion bodies [9]. PSP on the other hand is characterized by intracerebral aggregation of tau proteins, predominantly involving isoforms with four microtubule-binding region repeats (4R-Tau), in neurofibrillary tangles, oligodendrocytic coils, and astrocytic tufts [10]. This pathology generally occurs first in the midbrain and the basal nuclei and later also in the cerebral cortex (typically starting in the frontal lobe). In contrast to PSP, 4R-tau pathology in CBD appears more in the form of astrocytic plaques than tufted astrocytes but can also be found as neural inclusions as well as threads in gray and white matter [10]. Importantly, there is considerable clinical and neuropathological overlap between the diseases particularly within subtypes, such as the parkinsonian (MSA-P) and cerebellar variants of MSA (MSA-C) [11] and between PSP-spectrum tauopathies, such as the Richardson's syndrome (PSP-RS), parkinsonism-variant PSP (PSP-P), and pure gait freezing, among others [12].

Current pathophysiological theories propose that a central mechanism of disease progression is the spread of deleterious protein pathologies along functional brain networks [13–15], which opens up the possibility to block this pathogenic cascade by therapeutic intervention. Indeed, new molecular therapy strategies targeting protein aggregations are ready for clinical trials and hold the promise to dramatically improve the prognosis of AP [16,17]. However, the identification of candidates for clinical trials is problematic because accurate early diagnosis of the type of underlying pathology can be extremely difficult. A main reason for this is a strong mismatch between clinical presentation and pathological entity and the existence of a variety of

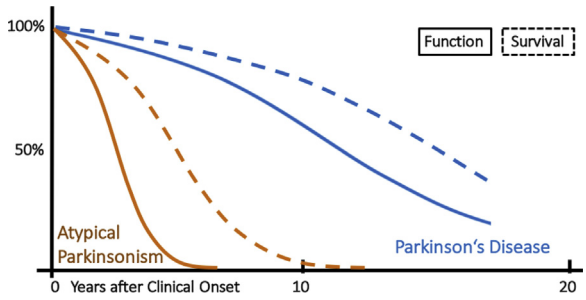


Fig. 1. Illustration of the prototypical clinical trajectories in Parkinson's disease (blue) and atypical Parkinsonism (orange). The trajectories account for the fact that neuronal degeneration is much more aggressive and symptomatic therapy is much less effective in atypical Parkinsonism, which leads to steeper slopes for survival (dashed lines) and even more dramatic drop in functional abilities (solid lines).

overlapping syndromes [12,18]. Although some clinical entities highly correlate with the underlying pathological entity, it is increasingly recognized that in AP, the clinical entity can have limited overlap with a pathological entity and vice versa.

In sum, there is an urgent need for instruments to help with the early detection of neurodegenerative processes, the early differentiation of the underlying pathology, and the objective assessment of disease progression. In this article, we lay down a conceptual framework aiming to characterize in more detail the kind of neuroimaging biomarkers needed, identify the current challenges in ascribing utility of these biomarkers, and propose criteria for a system that may guide future studies to overcome these challenges.

## 2. Neuroimaging biomarkers for diagnosis

There is plenty of evidence for a relatively large mismatch between clinical and pathological disease entities. While in some cases the clinical syndrome is highly indicative of the underlying pathology (e.g., clinical PSP-Richardson syndrome with 4R-neuroglial pathology), predicting a pathological entity based only on the clinical presentation of the patient is highly problematic in most cases, particularly during the early clinical stage of the disease. In the relatively large cohort of Lee and colleagues, only 35% of patients with corticobasal syndrome actually had CBD as a pathological substrate, while 23% had AD pathology and 13% had PSP pathology [19]. Conversely, the same pathological entity can be associated with a large variety of different syndromes. In their seminal paper, Williams and Lees have described various syndromes and variants of presentations that can be associated with PSP pathology [12]. Considering that the large mismatch between neuropathology and clinical presentation has translated to suboptimal diagnostic accuracy of parkinsonian syndromes [20,21], it currently makes sense to differentiate diagnostic properties of biomarkers according to these categories.

The ultimate goal of a therapeutic intervention is to stop and reverse disease progression, or at least to slow it down. It is increasingly clear that the neurodegenerative cascade can start many years, or even decades, before first clinical symptoms. Therefore, a therapeutic intervention targeting protein aggregation could be more effective, the earlier it is started. However, diagnostic accuracy for clinical entities and even more so for pathological entities is generally low when only few or prodromal symptoms are detectable by the clinician [9,14]. Neuroimaging biomarkers may help increasing diagnostic accuracy in this phase. For an unambiguous definition of what constitutes a diagnostic biomarker, the FDA-NIH Biomarker Working Group has published a compendium of definitions [22]: a diagnostic biomarker is "used to detect or confirm presence of a disease or condition of interest or to identify individuals with a subtype of the disease."

Importantly, the performance of a diagnostic biomarker has to be tested under defined conditions of use. The biomarker can only be put to use for patient selection in clinical trials, if the validity on the individual level in the intention-to-test population (the kind of group to which the diagnostic test will be applied to) is known. For example, a biomarker that distinguishes PSP patients from healthy controls may not be applicable in a scenario, where the intention-to-test population is composed of patients with PSP, but also other forms of parkinsonism. In addition, although it may be helpful to explore the potential of neuroimaging biomarkers in convenience samples (i.e., ad hoc clinic cohorts), one cannot properly infer the performance of the diagnostic biomarker in studies using other sampling settings (e.g., randomized controlled trials, or community cohorts). Moreover, the time point and clinical certainty at which the test is applied should reflect the intended-use scenario. If a diagnostic biomarker performs well in patients who already have a (relatively certain) clinical diagnosis with a fully developed syndrome (e.g., PSP-Richardson Syndrome), this does not mean that it will perform well in patients at the very early stage and with prodromal symptoms or other clinical variants (e.g., PSP-P). It is self-evident that the performance of a diagnostic neuroimaging biomarker should be assessed in a way that is statistically meaningful to be applicable in clinical trials.

The most useful quantifications of diagnostic test performance are the positive predictive value (PPV, i.e., the proportion of positives actually having the condition) and negative predictive value (NPV, i.e., the proportion of those tested negative who actually do not have the condition) in a realistic intention-to-test population. Indeed, predictive values tell us what we want to know: given a positive or negative test, what is the probability, respectively, that the patient does (PPV) or does not have (NPV) a particular disease? Importantly, these values very much depend on the frequency of actual positives or negatives in the study sample (i.e., the prevalence of the target disorder in the study sample) [23]. Therefore, for convenience samples, the assessment of performance in the form of a receiver

operating characteristic curve, that is, ROC curve, or at least sensitivity and specificity of a specific discrimination threshold may be the more sensible choice. Under these preconditions, the usefulness of diagnostic neuroimaging biomarkers is not limited to establishing a clinical diagnosis. A study design, in which the presence of a specific biological condition is used for sample stratification, can also involve target verification (e.g., presence of tau pathology), enrichment strategies, or subtyping of patients (e.g., patients with mild or severe forms of the disease).

## 3. Neuroimaging biomarkers of disease progression

Ultimately, the goal of any therapeutic intervention is to preserve or improve functional abilities by reducing symptoms or the impact of symptoms. However, functional abilities and clinical symptoms have several disadvantages when considered as the primary end point or outcome measure used to assess efficacy in clinical trials. Clinical data can be difficult and costly to generate and are hard to standardize across centers. Moreover, the prospective assessment of clinical events (e.g., falls, hospitalization, nursing home admission) is time consuming and inherently prone to attrition. Moreover, clinical data (e.g., a score in a particular clinical test) may be subject to severe day-to-day or even within-day variability due to unspecific and unrelated factors, such as quality of sleep in the previous night, the composition and amount of breakfast or lunch, and so on. Along similar lines, it is well recognized that life factors, such as education, that are linked to the concept of cognitive reserve have a strong influence on the impact of pathology on clinical performance markers [24,25]. Moreover, clinical scores—such as the PSP-rating scale in its current form—are typically designed for classical presentations of the spectrum of clinical phenotypes associated with a pathological entity (e.g., Richardson syndrome) rather than variant or atypical phenotypes. Even more importantly, it is assumed for AP, just like for practically all neurodegenerative diseases, that the pathogenic cascade starts years to decades before first symptoms. This would mean that in a clinical trial for the preclinical or prodromal phase of AP, waiting for therapeutic effects in clinical symptoms is impracticable. Of note, specific challenges for longitudinal trial design in the preclinical or prodromal phase of neurodegenerative diseases have been addressed by previous working groups (e.g., PreNI, BioLoC-PD) of the EU Joint Program of Neurodegenerative Diseases (JPND; http://www.neurodegenerationresearch.eu).

For all these reasons, objective and standardized measures of biomarkers have become an important outcome in many clinical trials. Such biomarker data could support clinical efficacy, provide mechanistical evidence of a disease modifying effect, and may even serve as a primary outcome measure. The term "surrogate end point" or "surrogate marker" is frequently applied to describe this specific use of a biomarker. According to the FDA-NIH definition, "a biomarker measured serially for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) a medical product or an environmental agent" is called a "monitoring biomarker" [22]. If this monitoring biomarker predicts or correlates with a clinical outcome, it may potentially be used as a surrogate end point in a clinical trial. Evidently, the magnitude of change relative to the statistical noise (i.e., effect size) should also be similar or greater for the biomarker measurement than for the clinical measurement. If the effect size of a biomarker is greater than the effect size of clinical outcome measures in a given time period, this may reduce the sample size required to power a therapeutic trial by an order of magnitude [26]. Ideally, short-term changes in the biomarker could anticipate long-term clinical outcome, which would be particularly helpful in presymptomatic or prodromal phase of a disease. Importantly, the use of a biomarker for accelerated approval by the FDA requires that the biomarker is at least a "reasonably likely" surrogate end point. In this case, the end point has to be supported by clear mechanistic and/or epidemiologic rationale. For example, radiologic assessment of tumor size in certain cancer types has been considered reasonably likely to predict an improvement of overall survival [22]. It can therefore be essential that a clear mechanistic rationale is provided, which incorporates the role of the biological process (quantified by the neuroimaging biomarker) within the pathogenic cascade (see Fig. 2).

## 4. Challenges for ascribing utility of a neuroimaging biomarker

There currently is not yet a consensus in the field on which imaging measures have the greatest utility as biomarkers for clinical trials in AP. Recent systematic reviews on available neuroimaging biomarkers for diagnosis and progression of AP disclosed that despite a plethora of studies, these have not yet yielded sufficiently validated biomarkers for diagnosis and disease progression, especially in the early course of disease or newly recognized variants [27–30]. This is problematic not only for finding a consensus but also because standard performance measures for newly developed imaging biomarkers are missing. In many businesses, the process of comparing to an industry standard—called benchmarking—has helped to accelerate development. We call for an increase in academic and industrial engagement in providing validated standards for biomarkers in clinical trials (benchmark). These standards should preferentially be scalable (relatively low cost and available) and robust (resistant to test-retest and multicenter variability). Under these preconditions, inclusion of these standards in national and international guidelines will certainly be facilitated.

The main reasons for the current lack of validation are the limited comparability and generalizability of studies in

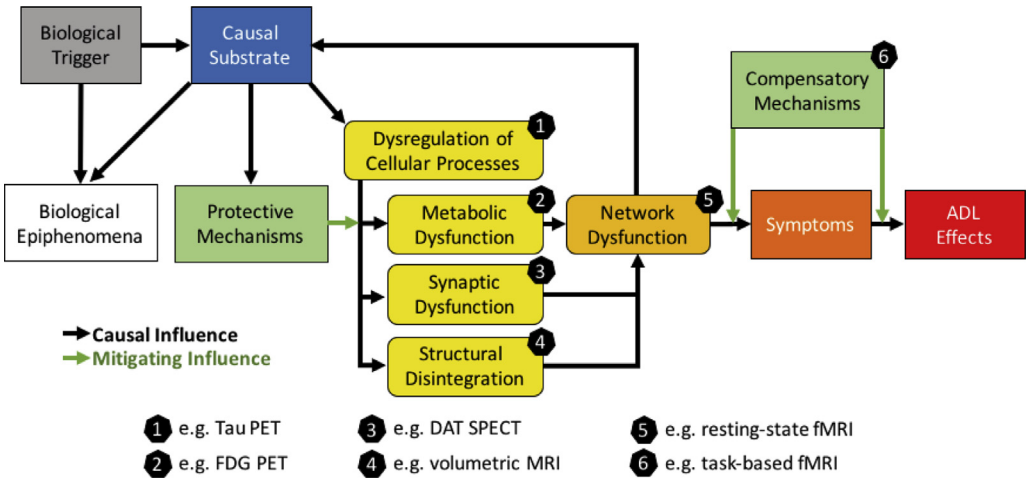


Fig. 2. Illustration of a mechanistic pathophysiological cascade that may occur in neurodegenerative diseases. A certain biomarker may capture more proximal or more distal biological substrates along the cascade. Abbreviations: FDG, fludeoxyglucose; DAT, dopamine transporter.

combination with a general sparsity of confirmatory reports and a practical nonexistence of confirmation of type and distribution of pathology (postmortem) [27]. Comparability is certainly hindered by a great variability in technical procedures for acquisition and processing, but also in the way that the potential utility of the biomarker is described. Although many studies refer to a potential use as a biomarker in the conclusions, only a few of them provide an operational definition as to how exactly the biomarker should be used. Moreover, the statistical means by which the utility is described are variable and sometimes inadequate. For example, although the potential of the biomarker as a diagnostic tool is discussed, measures of test accuracy (e.g., area under the ROC curve) are not reported. Generalizability, on the other hand, is limited by the fact that studies typically focus on specific phenotypes (e.g., PSP with Richardson Syndrome, PSP-RS), making its application in other variants questionable (e.g., PSP with progressive gait freezing). Another important limitation of generalizability is relatively small sample sizes, typically from a single clinical center. In addition, an exploratory approach in convenience samples—while certainly sometimes sensible in novel approaches—limits applicability in a clinical trial due to potential selection biases. An instrument to address these issues would be to manage cohorts for imaging studies as a "mock" clinical trial, including stringent and reproducible study designs detailed in trial protocols.

We suggest that the advancement of standardized neuroimaging in the field will be furthered by a well-defined reference frame for the utility of biomarkers. We argue that the currently most relevant areas of application include diagnostic biomarkers for early clinical diagnosis, diagnostic biomarkers for a specific pathology, and monitoring biomarkers for disease progression that qualify as surrogate end points.

## 5. Criteria for a biomarker utility system: the E-S-P methodology

To resolve these issues and to reach an international consensus on propositions to overcome the main challenges, we held two JPND-sponsored workshops (Cologne, Germany, and Vancouver, Canada) in 2017. As a consensus outcome, we recommend a formal utility description system in response to the challenges outlined previously. As a conceptual framework, we propose a graded system for the description of utility of a specific neuroimaging measure.

Our selection process for the categories and grades of the system was mainly guided by what are currently the most useful parameters for a clinical trial in AP in the near future. This approach was chosen to provide the necessary balance between complexity and applicability of the system. We therefore included separate categories for the ability to accurately identify an intention-to-treat patient population early in the course of disease (Early), to accurately detect a specific underlying pathology (Specific), and the ability to monitor disease progression (Progression). We aimed to provide a relatively intuitive system with relatively easy operational definitions (see Fig. 3).

### 5.1. Early diagnosis

We started from the position that a diagnostic biomarker has to be more accurate than a quasi-simultaneous clinical examination for it to be of particular use. Because current diagnostic criteria are still tailored to clinical entities, we focused on biomarker properties regarding "the earlier, the better" diagnosis of clinical entities. We propose three grades:

1. the biomarker accurately identifies a clinical entity in absence of clinical symptoms

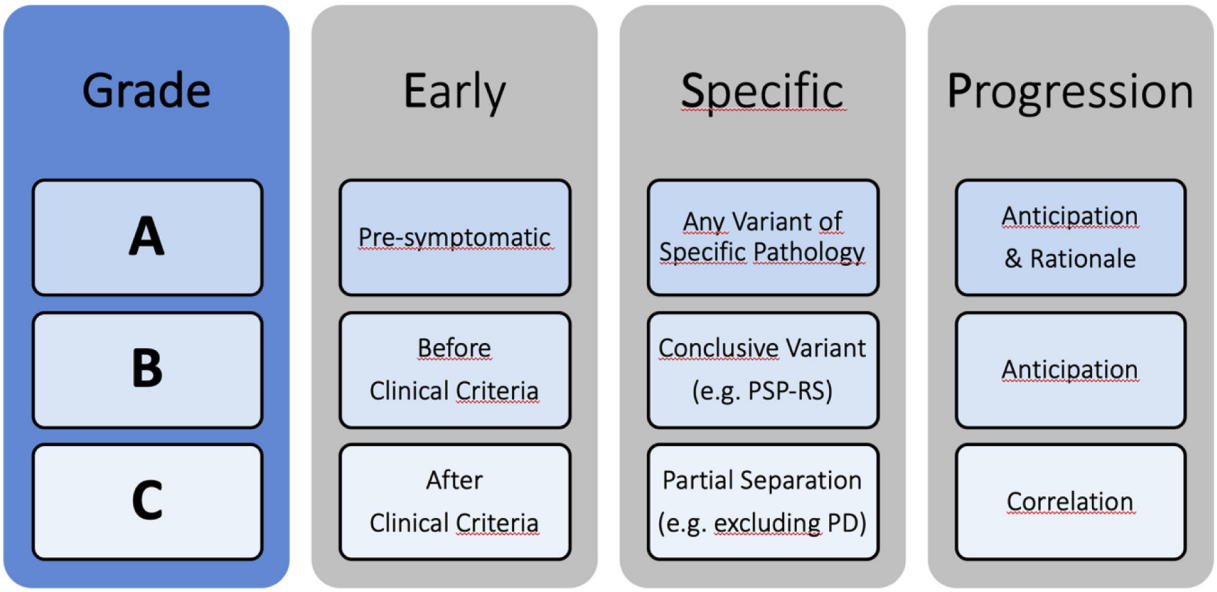


Fig. 3. Proposition of a graded Neuroimaging Biomarker Utility System for clinical trials. Early: the capability of a biomarker to detect a clinical entity early in the course of disease. Specific: the capability to increase the likelihood of a specific underlying pathology. Progression: the capability to serve as a surrogate end point in a clinical trial. Abbreviations: PSP-RS, progressive supranuclear palsy–Richardson Syndrome; PD, Parkinson's disease.

2. the biomarker accurately identifies a clinical entity in presence of symptoms but before clinical criteria are met
3. the biomarker accurately identifies a clinical entity only after clinical diagnostic criteria are met. It should be noted that diagnostic biomarkers of this category may still have utility, for example, when the accuracy with respect to long-term follow-up diagnosis is increased by including the biomarker in the assessment of the patient.

### 5.2. *Assessment of accuracy*

For a clinical diagnosis in AP, it seems reasonable that the gold standard should be the diagnosis of a movement disorders specialist after a long-term follow-up. For optimal assessment of utility of a biomarker, the specialist should be blinded to the biomarker outcome, to avoid biasing the clinical decision. The exact statistical value that appropriately describes accuracy of the test may vary, but there are certain rules-of-thumb: (1) specificity may be more important than sensitivity when the biomarker should be used for sample stratification in a clinical trial; (2) for some scenarios, the predictive value of the test is more informative than its sensitivity and specificity. The predictive value strongly depends on the a priori probability of the disease in the patients undergoing the test (which can be influenced by referral bias and regional prevalence). On the other hand, the same will apply for the predictive value of the clinical examination. Therefore, we propose that the predictive value of the biomarker should always be seen in relation to the predictive value of the clinical examination.

### 5.3. *Specific pathological substrate*

Because therapeutic strategies are increasingly targeting an underlying pathology (e.g., tau), it is critically important to address specificity regarding pathology as a separate category of utility. Strategies for patient stratification may vary and using clinical variants with very high specificity for an underlying pathology (e.g., PSP-RS) may be chosen by some trial sponsors. Most recent anti-amyloid trials in AD, however, used negative amyloid-PET scans as an exclusion criteria to only include patients with significant amyloid plaque pathology. One may argue that patient selection criteria in these clinical trials are likely to influence approval to the point that in a clinical setting, biomarker positivity is a prerequisite of "on-label" drug application and possibly remuneration. Along these lines, if efficacy for antitau therapies has been demonstrated only in PSP-RS patients, treating patients with other variants, however plausible, may be considered "off label." Ideally, a neuroimaging biomarker would be 100% accurate for a specific pathology. Realistically, a neuroimaging biomarker that increases the likelihood of a certain kind of pathology (enrichment) to an extent that efficacy of the drug may be demonstrated would still be of great use. Therefore, we propose a grading system that is oriented at the likelihood of a certain kind of pathology:

1. the biomarker identifies a specific pathology regardless of the clinical variant
2. the biomarker identifies a specific pathology only in the context of a specific clinical variant (e.g., PSP-RS)
3. the biomarker identifies partially differentiating features of pathology (e.g., synuclein pathology in

PD and MSA vs. tau pathology in PSP and CBD) which, although not specific, may still be useful for some trial contexts.

Our recommendations for measuring accuracy are similar to those for the previous category. Notable exceptions are the gold standard, which should be the pathological diagnosis based on (postmortem) histopathological examination.

### 5.4. Progression tracking

For this category, we were guided by the potential of a biomarker to qualify as surrogate end point. As outlined previously, a biomarker may qualify as a surrogate end point, if it (1) correlates with or even anticipates clinical progression and (2) shows superiority over clinical measurements in terms of practicability, precision, effect size, or any combination of these. Importantly, clinical progression may refer to range or severity of symptoms as measured by clinical examination.

Within this framework, we define "correlation" as follows: in a given time period, the effect size of the biomarker change over time significantly correlates with a clinical measure of progression. We define "anticipation" as follows: the effect size of the biomarker in a given time period significantly correlates with the effect size of a relevant clinical progression measure in a later time period. In other words, short-term biomarker changes anticipate long-term changes in clinical outcomes. The latter property is of critical importance for progression markers in absence of significant changes of clinical measures (e.g., presymptomatic phase). Moreover, a clear mechanistic rationale will dramatically increase the likelihood of acceptance of a biomarker as a surrogate end point in a clinical trial [22]. The main reason for demanding this property of a biomarker is that a correlation between the biomarker and clinical measures is not in itself evidence for a causal relationship but could be a biological epiphenomenon. The relationship could be due to a shared, more proximal causal agent along the pathological cascade or even due to the biomarker representing protective or compensational mechanisms (see Fig. 2). Combining these quality categories, we propose the following grades:

1. the biomarker anticipates clinical progression and a clear mechanistic rationale for a causal relationship exists
2. the biomarker anticipates clinical progression, in the absence of evidence of a causal mechanism
3. the biomarker correlates with clinical progression and is superior for use in a trial (e.g., larger effect size).

Note that a biomarker conveying information about future clinical severity in longitudinal data, but taken only at a single time point, is not included here. This sort of biomarker belongs in a different category (i.e., prognostic biomarker, not a progression biomarker) that is not discussed in this article as it is not based on longitudinal progression.

## 6. Strengths and limitations

Our aim was to promote the development of neuroimaging biomarkers in the field of AP by proposing a conceptual framework guiding the assessment and description of the utility of a biomarker in clinical trials. We primarily tailored our recommendations to current demands, which may very well be debatable or change over time. We are very aware that not all aspects of interest for neuroimaging biomarkers are covered here. We did not cover biomarkers specifically conveying information about the risk for developing a medical condition (susceptibility or risk biomarker), the likelihood of an individual to have faster or slower disease progression (prognostic biomarker), and many more potentially useful types of biomarkers. In addition, we did not specifically discuss the potential of a combined set of biomarkers (e.g., fluid/"wet" biomarkers plus neuroimaging). However, our goal was to find an appropriate balance between the necessary complexity and an intuitive applicability.

Importantly, the proposed utility system allows for the fact that a single biomarker may have strengths in one domain (e.g., as a biomarker for progression) while it may not excel in others (e.g., specificity). This system theoretically also allows us to describe the utility of a combination of different biomarkers. However, the categories "Specific" and "Progression" may depend on the time of application with respect to the phase of a disease.

We would have liked to compare our conceptual framework in the field of atypical parkinsonian disorders with systems developed for the description of biomarker utility in other neurodegenerative diseases. However, we are not aware of similar conceptual frameworks.

We would like to emphasize that our propositions are a conceptual framework that aims to be a stimulus for continued discussion in the field. Our hope is that this framework can form a basis for a future consensus among a wider range of scientists and other stakeholders.

## Acknowledgments

This work was supported by the EU Joint Program of Neurodegenerative Diseases (JPND, http://www.neurodegenerationresearch.eu).

Thilo van Eimeren received honoraria from Lilly, Shire and is supported by grants from DFG, Leibniz Foundation, JPND. Angelo Antonini received consultancies from BIAL, Lundbeck, UCB Pharma, Zambon, AbbVie and honoraria for talks from AbbVie, BIAL, Lundbeck, UCB Pharma, Mundipharma, Zambon and is supported by grants from Horizon2020, JPND, Chiesi Pharmaceuticals,

AbbVie. Daniela Berg received honoraria from Biogen, BIAL, Lundbeck, UCB Pharma GmbH, AbbVie, Zambon and is supported by grants from Michael J. Fox Foundation, Janssen Pharmaceutica N.V., German Parkinson's Disease Association (dPV), BMWi, BMBF, Parkinson Fonds Deutschland gGmbH, UCB Pharma GmbH, TEVA Pharma GmbH, EU, Novartis Pharma GmbH, Lundbeck, Damp foundation. Nico Bohnen received grants from the NIH (RO1 NS070856 and P50 NS091856), US Department of Veterans Affairs (I01 RX000317), the Michael J. Fox Foundation, Axovant Sciences and Chase Pharmaceuticals. Roberto Ceravolo received fees for talks and consultancy from General Electric, Zambon, UCB Pharma. Günter U Höglinger received research support from Neuropore, GE Health; Research collaborations with Orion, Prothena; consultant for Abbvie, Alzprotect, Asceneuron, Biogen, Bristol-Myers Squibb, Novartis, Roche, UCB; honoraria for scientific presentations from Abbvie, Roche, Teva, UCB, Zambon. Stephane Lehericy received grants from "Investissements d'avenir" (grant number ANR-10-IAIHU-06 and ANR-11-INBS-0006). Research grant from Biogen Inc. Simon Lewis is supported by an NHMRC-ARC Dementia Fellowship #1110414. Matej Ondrus is an employee of AXON Neuroscience SE. Irena Rektorova has received honoraria as associate editor (Parkinsonism and related disorders) from Elsevier Publishers, Amsterdam, The Netherlands, and is supported by grants H2020-MSCA RISE, IMI H2020, and grants from the Agency of Health Research of the Ministry of Health, Czech Republic. Klaus Seppi received personal fees from Teva, UCB, Lundbeck, AOP Orphan Pharmaceuticals AG, Roche, Grünenthal, and Abbvie, honoraria from the International Parkinson and Movement Disorders Society, research grants from FWF Austrian Science Fund, Michael J. Fox Foundation, and International Parkinson and Movement Disorder Society. Stephane Thobois reports grants from Fondation pour la Recherche Médicale, from France Parkinson, Fondation Neurodis, ANR, PHRC and personal fees from UCB, Medtronic, Teva, St Jude, Novartis, Aguettant, Zambon, Abbvie. Valtteri Kaasinen has received honoraria from Abbvie Finland, Orion Pharma, GE Healthcare, Teva, Abbvie, NordicInfu Care Ab; consultant for Zambon. Gregor Wenning has received honoraria from Abbvie Biogen Lundbeck Novartis Ono Prana Theravance. Hartwig R Siebner has received honoraria as speaker and consultant from Sanofi Genzyme, Denmark and as senior editor (NeuroImage) from Elsevier Publishers, Amsterdam, The Netherlands, and royalties as book editor from Springer Publishers, Stuttgart, he has received a research fund from Biogen-Idec. Antonio P. Strafella is supported by Canadian Institutes of Health Research (MOP-136778) and Canada Research Chair Program. James B Rowe has received honoraria/consultancy fees from Biogen and Asceneuron, and research grants from AZ-Medimmune, Lilly, Janssen and the Dementias Platform UK. His principal funding is from Wellcome Trust, Medical Research Council and the National Institute for Health Research.

## RESEARCH IN CONTEXT

1. Future directions: The article proposes a novel conceptual framework for neuroimaging biomarkers for clinical trials in atypical parkinsonian disorders. As the outcome of international consensus meetings, the proposal includes a graded system of the ability of a biomarker to accurately identify an intention-to-treat patient population early in the course of disease (Early), to accurately detect a specific underlying pathology (Specific), and the ability to monitor disease progression (Progression). This well-defined reference frame for the utility of biomarkers may stimulate standardized use of neuroimaging biomarkers in clinical trials.
2. Systematic review: We examined literature on the description of biomarker utility in general and in neurodegenerative disorders, with a specific emphasis on atypical parkinsonian disorders.
3. Interpretation: We did not find applicable definitions or descriptions in the literature and provide an expert consensus.

## References

[1] Elbaz A, Carcaillon L, Kab S, Moisan F. Epidemiology of Parkinson's disease. Rev Neurol (paris) 2016;172:14–26.
[2] Horvath J, Burkhard PR, Bouras C, Kövari E. Etiologies of Parkinsonism in a Century Long Autopsy Based Cohort. Brain Pathol 2013;23:28–33.
[3] Jellinger K. The pathology of parkinsonism. In: Marsden CD, Fahn S, eds. Mov Disord. London: Butterworths & Co.; 1987. p. 124–65.
[4] Litvan I. What is an Atypical Parkinsonian Disorder? In: Litvan I, ed. Atypical parkinsonian disorders: clinical and research aspects. Totowa, NJ: Humana Press; 2005. p. 1–10.
[5] Wenning GK, Geser F, Krismer F, Seppi K, Duerr S, Boesch S, et al. The natural history of multiple system atrophy: a prospective European cohort study. Lancet Neurol 2013;12:264–74.
[6] Constantinescu R, Richard I, Kurlan R. Levodopa responsiveness in disorders with parkinsonism: a review of the literature. Mov Disord 2007;22:2141–8.
[7] O'Sullivan SS, Massey LA, Williams DR, Silveira-Moriyama L, Kempster PA, Holton JL, et al. Clinical outcomes of progressive supranuclear palsy and multiple system atrophy. Brain 2008; 131:1362–72.
[8] Braak H, Del Tredici K, Rüb U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003;24:197–211.

[9] Cykowski MD, Coon EA, Powell SZ, Jenkins SM, Benarroch EE, Low PA, et al. Expanding the spectrum of neuronal pathology in multiple system atrophy. Brain 2015;138:2293–309.
[10] Kovacs GG. Invited review: neuropathology of tauopathies: principles and practice. Neuropathol Appl Neurobiol 2015;41:3–23.
[11] Gilman S, Wenning G, Low PA, Brooks D, Mathias C, Trojanowski J, et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology 2008;71:670–6.
[12] Williams DR, Lees AJ. Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges. Lancet Neurol 2009;8:270–9.
[13] Palop JJ, Chin J, Mucke L. A network dysfunction perspective on neurodegenerative diseases. Nature 2006;443:768.
[14] Cope TE, Rittman T, Borchert RJ, Jones PS, Vatansever D, Allinson K, et al. Tau burden and the functional connectome in Alzheimer's disease and progressive supranuclear palsy. Brain 2018;141:550–67.
[15] Hoenig MC, Bischof GN, Seemiller J, Hammes J, Kukolja J, Onur ÖA, et al. Networks of tau distribution in Alzheimer's disease. Brain 2018;141:568–81.
[16] Boxer AL, Yu J-T, Golbe LI, Litvan I, Lang AE, Höglinger GU. Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. Lancet Neurol 2017;16:552–63.
[17] Castro Caldas A, Levin J, Djaldetti R, Rascol O, Wenning G, Ferreira JJ, et al. Critical appraisal of clinical trials in multiple system atrophy: Toward better quality. Mov Disord 2017;32:1356–64.
[18] Höglinger GU, Respondek G, Stamelou M, Kurz C, Josephs KA, Lang AE, et al. Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. Mov Disord 2017;32:853–64.
[19] Lee SE, Rabinovici GD, Mayo MC, Wilson SM, Seeley WW, DeArmond SJ, et al. Clinicopathological correlations in corticobasal degeneration. Ann Neurol 2011;70:327–40.
[20] Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G. Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. Neurology 2016;86:566–76.
[21] Joutsa J, Gardberg M, Röyttä M, Kaasinen V. Diagnostic accuracy of parkinsonism syndromes by general neurologists. Parkinsonism Relat Disord 2014;20:840–4.
[22] FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016. Available from http://www.ncbi.nlm.nih.gov/books/NBK326791/.
[23] Gallagher EJ. The problem with sensitivity and specificity…. Ann Emerg Med 2003;42:298–303.
[24] Stern Y. What is cognitive reserve? Theory and research application of the reserve concept. J Int Neuropsychol Soc 2002;8:448–60.
[25] Hoenig MC, Bischof GN, Hammes J, Faber J, Fliessbach K, van Eimeren T, et al. Tau pathology and cognitive reserve in Alzheimer's disease. Neurobiol Aging 2017;57:1–7.
[26] Jack C, Shiung M, Gunter J, O'brien P, Weigand S, Knopman D, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurology 2004;62:591–600.
[27] Whitwell JL, Höglinger GU, Antonini A, Bordelon Y, Boxer AL, Colosimo C, et al. Radiological biomarkers for diagnosis in PSP: where are we and where do we need to be? Mov Disord 2017;32:955–71.
[28] Strafella AP, Bohnen NI, Perlmutter JS, Eidelberg D, Pavese N, Van Eimeren T, et al. Molecular imaging to track Parkinson's disease and atypical parkinsonisms: new imaging frontiers. Mov Disord 2017;32:181–92.
[29] van Eimeren T, Bischof GN, Drzezga A. Is Tau Imaging More Than Just Upside-Down 18F-FDG Imaging? J Nucl Med 2017;58:1357–9.
[30] Strafella AP, Bohnen NI, Pavese N, Vaillancourt DE, van Eimeren T, Politis M, et al. Imaging markers of progression in Parkinson's disease. Mov Disord Clin Pract 2018;5:586–96.