Supplementary Material

Parkinson's Disease Subtypes: Critical Appraisal and Recommendations

METHODS

Overview

We conducted a systematic review of Parkinson's disease (PD) subtyping studies in order to summarize methods used and reported characteristics of PD subtypes. The methodological quality of each study and the clinical applicability of each identified PD subtype system were also evaluated using standardized approach.

Eligibility criteria

PD subtyping studies were defined as any research study conducted with the purpose of dividing PD patients into subtypes, as stated by its authors. We also included studies in which authors did not state the purpose of PD subtyping as the study's objective but identified distinct groups of PD patients in the research which were thereafter discussed as possible subtypes in the paper. Since the goal of the systematic review was to evaluate the methodologic approach for the identification of a PD subtype classification, we only included the initial report of a given PD subtype classification. We did not include further studies conducting additional evaluations of an already described PD subtype classification. The only exception was the report of follow-up data by the researchers of the initial report, as we considered them part of the same study. We did not exclude studies that used the same database, as long as these studies used distinct defining criteria or approaches to identify a new PD subtype classification system. Studies were conducted with participants diagnosed with PD regardless of the applied diagnostic criteria. We accepted any type

of study design, and both clinical and non-clinical biomarkers as measures. We only included full-text publications written in English. We excluded studies or results within a study which compared PD with healthy controls or other neurological disorders, studies evaluating previously-defined PD subtypes (i.e., tremor dominant vs. PIGD) or using a mixed group of patients as unit of analyses, biomarker studies with no categorical results (no sub-grouping of patients based on the biomarker), studies restricted to a subset of PD patients (e.g., cognitively impaired) and review articles.

Search strategy

We searched for PD subtyping studies in the medical databases PUBMED/MEDLINE and used the following search terms: 'Parkinson Disease'[Mesh] AND ('Subtyp*' OR 'Phenotype'[Mesh] OR 'Phenotyp*' OR 'Biomarkers'[Mesh] OR 'Clinical Feature*' OR 'endophenotyp*'). We reviewed the bibliographic references of included studies and review papers identified through the above search strategy. The initial search was done up to September 2018, and it was complemented by a second search using the same search items up to June 2019. Three expert reviewers (SMF, TAM, CM) were responsible for the initial screening of each abstract identified through the search. Any disagreements regarding abstract inclusion between the two reviewers were discussed with and assessed by the third reviewer for a final decision. Each full-text article of abstracts remaining after the initial screen was reviewed by at least two members of the MDS Task Force, who made final decisions regarding inclusion of the paper and performed data collection.

Data collection

Pairs of reviewers abstracted data from the original studies regarding study design, baseline characteristics, PD subtyping methodology and results were entered into a standardized data sheet. Reviewers could not evaluate their own studies.

Rating schemes

In the absence of available tools to specifically evaluate subtyping studies, we developed our own checklists to assess the methodological quality and clinical applicability of the included studies. The Methodological Quality tool is a 13-item checklist with item scores ranging from 0 to 2. A higher score corresponds to higher methodological quality. The Clinical Applicability tool is an 11-item checklist with items rated as Unknown, Limited/Low, Moderate, and Strong. We drafted an initial version based on available standardized checklists [1]. After reaching a consensus across the Task Force in terms of items and response options, we tested the application of both tools in a sample of included studies (N=5) to refine the tools and finalize the version used in this review. An item on appraisal of statistical methods was included in the Methodological Quality tool. Due to the heterogeneity of the statistical methods used, a pair of statisticians (MM, ML) was given the task of reaching a consensus on rating criteria for this item based on pre-defined statistical evaluative concepts. The same pair of statisticians obtained a final score (0=low quality, 1=intermediate quality, 2=high quality) by consensus, taking into account all aspects of the study and incorporating discussion to arrive at agreement, recording reasons for sub-optimal ratings. For both study data collection and rating schemes, reviewers were requested to extract data and rate the studies independently. Discrepancies were subsequently resolved by consensus between the raters within a pair.

Data analysis

We provide descriptive statistics of the data abstracted and ratings of the Methodological Quality and Clinical Applicability tools. We used chi-square tests (or Fisher's exact test, as appropriate) to compare the relative frequencies of ratings by individual items between two publication periods (1980-2014 vs. 2015-2019) as well as between two methodological approaches (data-driven vs. hypothesis-driven).

REFERENCE

[1] Marras C, Lang A (2013) Parkinson's disease subtypes: lost in translation? *J Neurol Neurosurg Psychiatry* **84**, 409-415.

Supplementary Table 1. 11-item checklist for Clinical Applicability (items rated as Unknown, Limited/Low, Moderate, and Strong) developed for the current systematic review.

Item	Level
Subtyping Algorithm	Strong - Easily applied (availability of measurement and invasiveness) Moderate Limited - Applied with difficulty (measures are invasive or not readily available)
Time Required	Strong - Manageable in one visit (15-30 min) Moderate Limited - Time-consuming
Applicable to Drug-Naïve Stage	Strong - yes Moderate - maybe Limited - no
Cost	Strong - affordable Moderate Limited - Expensive
Subtype Stability	Stable Unstable Unknown
Applicability to general population ***(primarily based on results from external validation but also and recruitment study setting and recruitment basis)***	Strong - High Moderate Limited - Low
Prognostic Value	Strong - High Moderate Limited - Low Unknown
Treatment Implication	Strong - High Moderate Limited - Low Unknown
Clinical importance of differences in the variables used to define subtypes	Strong - High Moderate Limited - Low Unknown
Clinical importance of differences in external clinical/demographic features between subtypes	Strong - High Moderate Limited - Low Unknown
Clinical importance of differences in pathological/biomarker features between subtypes	Strong - High Moderate Limited - Low Unknown

Supplementary Table 2. Individual item scores in the Methodological Quality tool for each study included in the systematic review on PD subtypes.

Study ID (Year, First author)		Disease stages/duration (study population)	Study setting (representativeness)	Recruitment source (generalizability)	Diagnostic methods	Sampling method	Comprehensiveness of data used for subtyping (subtype definition)	Variables compared between subtypes (post hoc)	Statistical methods quality rating	Longitudinal follow-up	Completeness of follow-up	Subtype stability	Algorithm for Classifying individual patients	Validation (internal or external)
1984	Picirilli	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	1	1 = short-term (1 -3 years) OR longer-term but <3 time-points	1 = 50-75% complete	0 = not assessed	1 = provided	0 = not assessed
1986	Santamaria	1 = homogeneous disease stage/duration	0 = single-center or not reported	0 = clinic-based or not reported	2 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	1 = provided	0 = not assessed
1987	Mortimer	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	0 = not described or 1 or 2 not applicable	0 = convenience or not reported	0 = single clinical or biomarker domain	1 = single clinical domain or biomarker	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
1990	Jankovic	1 = homogeneous disease stage/duration	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	1 = provided	0 = not assessed
1994	Friedman	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	0 = not described or 1 or 2 not applicable	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	1 = provided	0 = not assessed
1999	Graham	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
1999	De Ceballos	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	0 = not described or 1 or 2 not applicable	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2005	Lewis	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2009	Reijnders	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	1
2010	White	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2011	Lui	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	1 = use of a test set from the same population
2011	Rooden	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0	0 = not provided	2 = validation in an external population
2012	Flensborg Damholdt	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2013	Dujardin	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2013	Erro	1 = homogeneous disease stage/duration	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	1 = >1 clinical domains or biomarkers	0 = not done	1	1 = short-term (1 -3 years) OR longer-term but <3 time-points	0	0 = not assessed	0 = not provided	0 = not assessed
2014	Kim	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2014	Gong	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	1 = provided	0 = not assessed

2015	Erro	1 = homogeneous disease stage/duration	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	1	1 = short-term (1 -3 years) OR longer-term but <3 time-points	2 = > 75 % complete	0 = not assessed	1 = provided	0 = not assessed
2015	Fereshtehnejad	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	1	1 = short-term (1 -3 years) OR longer-term but <3 time-points	1 = 50-75% complete	0 = not assessed	0 = not provided	1 = use of a test set from the same population
2015	Lawton	1 = homogeneous disease stage/duration	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2015	Ma	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2015	Seichepine	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	0 = not described or 1 or 2 not applicable	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	1 = provided	0 = not assessed
2015	Tsujikawa	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	1	1 = short-term (1 -3 years) OR longer-term but <3 time-points	2 = > 75 % complete	0 = not assessed	0 = not provided	0 = not assessed
2015	Lee	1 = homogeneous disease stage/duration	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	2	1 = short-term (1 -3 years) OR longer-term but <3 time-points	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	1 = provided	0 = not assessed
2016	Erro	1 = homogeneous disease stage/duration	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2016	van Balkom	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2016	Uribe	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2016	Landau	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	2	2 = longer-term (>3 year) AND >=3 time-points	1 = 50-75% complete	1 = assessed	0 = not provided	0 = not assessed
2016	Berganzo	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	0 = not done	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2017	Belvisi	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	1 = provided	0 = not assessed
2017	Brennan L	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	2	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2017	Eisinger	1 = homogeneous disease stage/duration	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	0 = not done	0	2 = longer-term (>3 year) AND >=3 time-points	0 = cross-sectional or ≤50% complete or not reported	1 = assessed	0 = not provided	0 = not assessed
2017	Fereshtehnejad	1 = homogeneous disease stage/duration	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	1	1 = short-term (1 -3 years) OR longer-term but <3 time-points	2 = > 75 % complete	0 = not assessed	1 = provided	1 = use of a test set from the same population
2018	Lawton	1 = homogeneous disease stage/duration	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	2	2 = longer-term (>3 year) AND >=3 time-points	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	2 = validation in an external population
2018	Vavougios	1 = homogeneous disease stage/duration	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	1	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2018	Alonso-Recio	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	0 = convenience or not reported	0 = single clinical or biomarker domain	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2018	Battista	0 = mixture of stages/disease duration at baseline or not reported	0 = single-center or not reported	0 = clinic-based or not reported	1 = Use of formal diagnostic criteria or diagnosis by an expert neurologist	1 = consecutive or random	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	0	0 = none (cross-sectional) or longitudinal < 1 year	0 = cross-sectional or ≤50% complete or not reported	0 = not assessed	0 = not provided	0 = not assessed
2019	Zhang	0 = mixture of stages/disease duration at baseline or not reported	1 = multi-center	0 = clinic-based or not reported	0 = not described or 1 or 2 not applicable	0 = convenience or not reported	1 = >1 clinical domains or biomarkers	2 = >1 clinical domains or biomarkers;	2	2 = longer-term (>3 year) AND >=3 time-points	2 = > 75 % complete	0 = not assessed	0 = not provided	0 = not assessed

Supplementary Table 3. Individual item scores in the Clinical Applicability tool for each study included in the systematic review on PD subtypes.

Study ID (Year, First author)		Subtyping Algorithm	Time Required	Applicable to Early, Drug-Naïve Stage	Cost	Subtype Stability	Applicability to general population of PD (primarily based on results from external validation but also and recruitment study	Prognostic Value	Potential Treatment Implication of determining these subtypes	Clinical importance of differences in the variables used to define subtypes	Clinical importance of differences in clinical/demographic features between subtypes compared post hoc	Clinical importance of differences in pathological/biomarker features between subtypes
1984	Piccirilli	Moderate	Limited - Time-consuming (>60 min)	Moderate - maybe	Moderate	Unknown	Moderate	Strong - High	Limited - Low	Strong - High	Moderate	Unknown
1986	Santamaria	Strong - Easily applied (availability of measurement and invasiveness)	Moderate	Strong - yes	Strong - affordable	Unknown	Limited - Low	Unknown	Unknown	Limited - Low	unknown	Unknown
1987	Mortimer	Strong - Easily applied (availability of measurement and invasiveness)	Limited - Time-consuming (>60 min)	Moderate - maybe	Moderate	Unknown	Limited - Low	Unknown	Unknown	Unknown	Limited - Low	Limited - Low
1990	Jankovic	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	Strong - yes	Strong - affordable	Unknown	Limited - Low	Unknown	Unknown	Moderate	Moderate	Unknown
1994	Friedman	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	Strong - yes	Strong - affordable	Unknown	Limited - Low	Limited - Low	Limited - Low	Moderate	Moderate	Unknown
1999	Graham	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Limited - no	Moderate	Unknown	Limited - Low	Unknown	Moderate	Moderate	Moderate	Unknown
1999	De Ceballos	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Limited - no	Limited - Expensive	Unknown	Limited - Low	Limited - Low	Unknown	Unknown	Unknown	Unknown
2005	Lewis	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Limited - no	Moderate	Unknown	Limited - Low	Unknown	Limited - Low	Limited - Low	Unknown	Unknown
2009	Reijnders	Strong - Easily applied (availability of measurement and invasiveness)	Moderate	moderate - maybe	strong - affordable	Unknown	Limited - Low	Unknown	Unknown	Limited - Low	moderate	Unknown
2010	White	Limited -Applied with difficulty (measures are invasive or not readily available)	Moderate	Strong - yes	Strong - affordable	Unknown	Strong - High	Limited - Low	Unknown	Unknown	Limited - Low	Unknown
2011	Rooden	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Limited - no	Moderate	Unknown	Moderate	Unknown	Limited - Low	Moderate	Moderate	Unknown
2011	Lui	Moderate	Limited - Time-consuming (>60 min)	Moderate - maybe	Moderate	Unknown	Moderate	Moderate	Moderate	Moderate	Moderate	Unknown
2012	Flensborg Damholdt	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	Moderate - maybe	Strong - affordable	Unknown	Moderate	Unknown	Unknown	Unknown	Limited - Low	Unknown
2013	Dujardin	Strong - Easily applied (availability of measurement and invasiveness)	Limited - Time-consuming (>60 min)	Strong - yes	Strong - affordable	Unknown	Strong - High	Unknown	Unknown	Strong - High	Strong - High	Unknown
2013	Erro	Strong - Easily applied (availability of measurement and invasiveness)	Moderate	Strong - yes	Strong - yes	Unknown	Moderate	Limited - Low	Unknown	Strong - High	Unknown	Unknown
2014	Gong	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Moderate - maybe	Moderate	Unknown	Limited - Low	Unknown	Unknown	Limited - Low	Moderate	Unknown
2014	Kim	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	moderate - maybe	Strong - affordable	Unknown	Limited - Low	Unknown	Unknown	Unknown	Limited - Low	Unknown

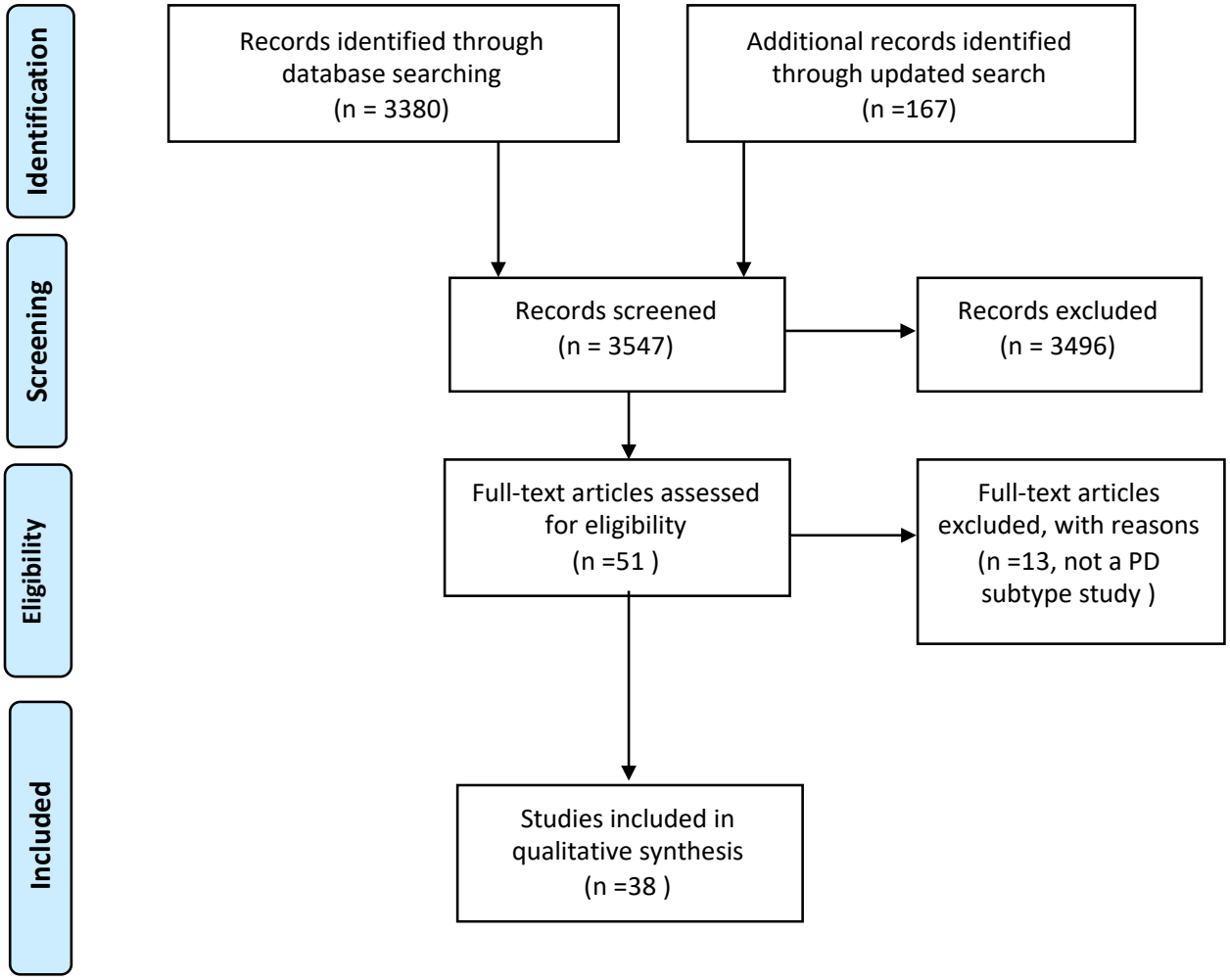
2015	Erro	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	Strong - yes	Strong - affordable	Unknown	Moderate	Limited - Low	Limited - Low	Moderate	Limited - Low	Moderate
2015	Fereshtehnejad	Moderate	Limited - Time-consuming (>60 min)	Strong - yes	Moderate	Unknown	Limited - Low	Strong - High	Unknown	Unknown	Unknown	Unknown
2015	Lawton	Strong - Easily applied (availability of measurement and invasiveness)	Limited - Time-consuming (>60 min)	Strong - yes	Strong - yes	Unknown	Moderate	Unknown	Unknown	Strong - High	Limited - Low	Unknown
2015	Lee	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	Strong - yes	Strong - affordable	Unknown	Limited - Low	Unknown	Unknown	Moderate	Moderate	Unknown
2015	Seichepine	Moderate	Strong - Manageable in one visit (15-30 min)	Strong - yes	Strong - affordable	Unknown	Limited - Low	Unknown	Limited - Low	Limited - Low	Limited - Low	Unknown
2015	Tsujikawa	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Moderate - maybe	Limited - Expensive	Unknown	Strong - High	Moderate	Limited - Low	Strong - High	Strong - High	Strong - High
2015	Ma	Strong - Easily applied (availability of measurement and invasiveness)	limited - Time-consuming (>60 min)	moderate - maybe	strong - affordable	Unknown	moderate	Unknown	Unknown	Limited - Low	moderate	Limited - Low
2016	Erro	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Strong - yes	Moderate	Unknown	Moderate	Unknown	Limited - Low	Moderate	Moderate	Unknown
2016	van Balkom	Moderate	Limited - Time-consuming (>60 min)	Moderate - maybe	Moderate	Unknown	Moderate	Unknown	Unknown	Moderate	Moderate	Unknown
2016	Berganzo	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Limited - no	Limited - Expensive	Unknown	Limited - Low	Unknown	Unknown	Unknown	Unknown	Unknown
2016	Uribe	Moderate	Limited - Time-consuming (>60 min)	Strong - yes	Moderate	Unknown	Strong - High	Moderate	Moderate	Limited - Low	Limited - Low	Unknown
2016	Landau	Strong - Easily applied (availability of measurement and invasiveness)	Limited - Time-consuming (>60 min)	Strong - yes	Strong - affordable	Strong - Stable over long-term period	Strong - High	Moderate	Moderate	Moderate	Limited - Low	Unknown
2017	Belvisi	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	Strong - yes	Strong - affordable	Unknown	Moderate	Unknown	Unknown	Unknown	Limited - Low	Unknown
2017	Brennan L	Moderate	Limited - Time-consuming (>60 min)	Strong - yes	Strong - affordable	Unknown	Strong - High	Unknown	Unknown	Strong - High	Strong - High	Unknown
2017	Eisinger	Strong - Easily applied (availability of measurement and invasiveness)	Strong - Manageable in one visit (15-30 min)	Strong - yes	Strong - yes	Limited - Unstable with substantial shift over short-term period	Strong - High	Unknown	Unknown	Strong - High	Unknown	Unknown
2017	Fereshtehnejad	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Strong - yes	Moderate	Unknown	Limited - Low	Limited - Low	Unknown	Strong - High	strong - High	Limited - Low
2018	Lawton	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Moderate - maybe	Limited - Expensive	Unknown	Limited - Low	Moderate	Unknown	Moderate	Strong - High	Unknown
2018	Vavougios	Moderate	Moderate	Moderate - maybe	Unknown - an algorithm to classify individuals was not reported	Unknown	Moderate	Unknown	Moderate	Unknown	Unknown	Unknown
2018	Alonso-Recio	Strong - Easily applied (availability of measurement and invasiveness)	Limited - Time-consuming (>60 min) (>60 min)	Limited - no	Moderate	Unknown	Moderate	Strong - High	Moderate	Moderate	Unknown	Unknown
2018	Battista	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Limited - no	Unknown - an algorithm to classify individuals was not reported	Unknown	Limited - Low	Limited - Low	Limited - Low	Limited - Low	Limited - Low	Limited - Low
2019	Zhang	Limited -Applied with difficulty (measures are invasive or not readily available)	Limited - Time-consuming (>60 min)	Strong - yes	Strong - affordable	Unknown	Limited - Low	Moderate	Limited - Low	Moderate	Moderate	Strong - High

Supplementary Table 4. Item score distribution in the Methodological Quality tool administered to PD subtype studies included in the systematic review and dichotomized according to year of publication. A higher item score corresponds to a better methodological quality for a given item.

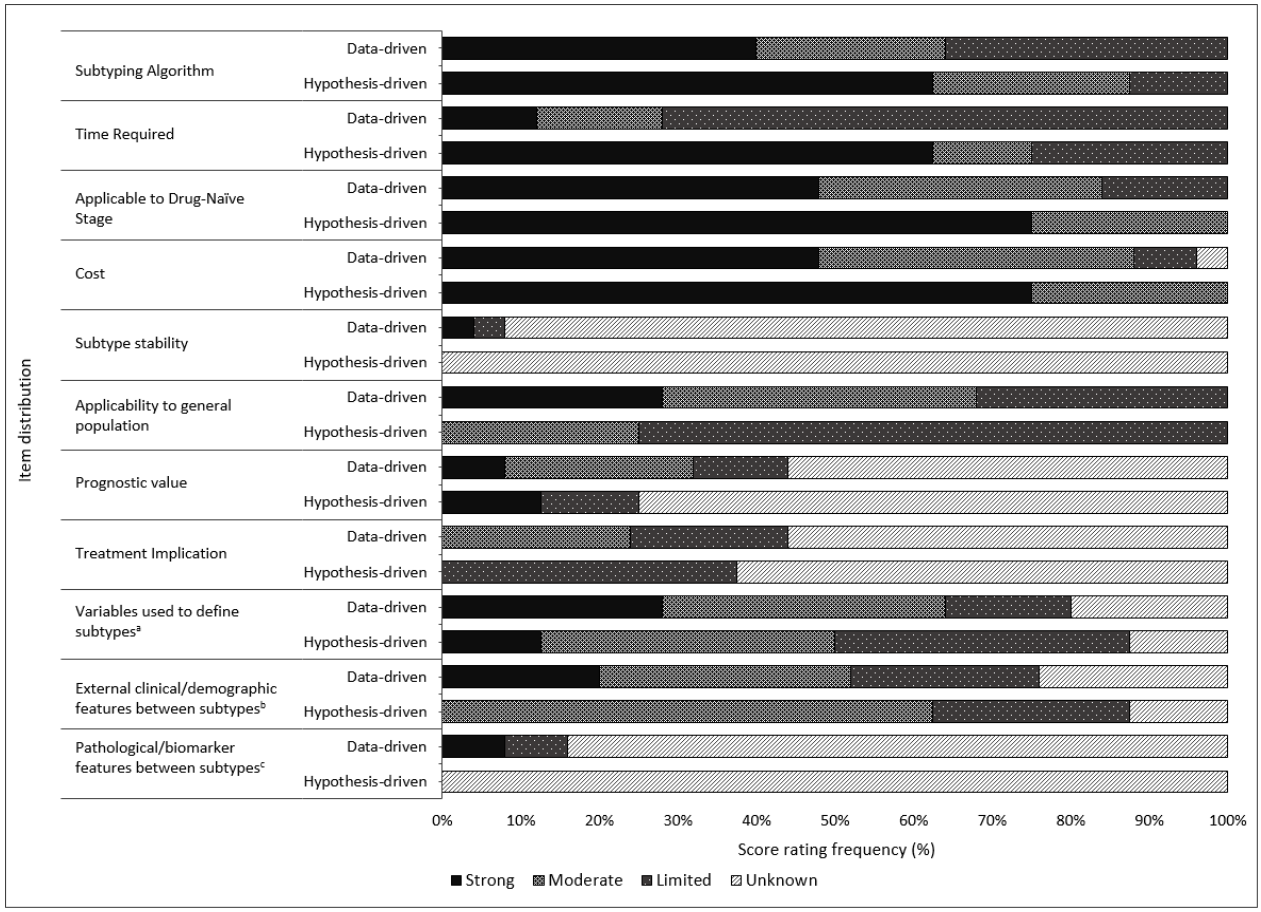
	Item													
Date of publication	**Score**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1980-2014 (n=17)	0	14 (82.4)	12 (70.6)	17 (100)	3 (17.6)	12 (70.6)	9 (52.9)	1 (5.9)	7 (41.2)	15 (88.2)	16 (94.1)	17 (100)	12 (70.6)	14 (82.4)
	1	3 (17.6)	5 (29.4)	0 (0)	13 (76.5)	5 (29.4)	8 (47.1)	1 (5.9)	7 (41.2)	2 (11.8)	1 (5.9)	0 (0)	5 (29.4)	2 (11.8)
	2	-	-	-	1 (5.9)	-	-	15 (88.2)	3 (17.6)	0 (0)	0 (0)	-	-	1 (5.9)
2015-2019 (n=21)	0	13 (61.9)	10 (47.6)	21 (100)	2 (9.5)	18 (85.7)	10 (47.6)	2 (9.5)	5 (23.8)	12 (57.1)	15 (71.4)	19 (90.5)	16 (76.2)	18 (85.7)
	1	8 (38.1)	11 (52.4)	0 (0)	19 (90.5)	3 (14.3)	11 (52.4)	0 (0)	7 (33.3)	5 (23.8)	2 (9.5)	2 (9.5)	5 (23.8)	2 (9.5)
	2	-	-	-	0 (0)	-	-	19 (90.5)	9 (42.9)	4 (19)	4 (19)	-	-	1 (4.8)

1. Disease stages/duration (study population), 2.Study setting (representativeness), 3.Recruitment source (generalizability)., 4. Diagnostic methods, 5. Sampling method, 6. Comprehensiveness of data used for subtyping (subtype definition), 7. Variables compared between subtypes (post hoc), 8. Statistical methods quality rating, 9. Longitudinal follow-up, 10. Completeness of follow-up, 11.Subtype stability, 12.Algorithm for classifying individual patients. 13.Validation (internal or external)

Supplementary Figure 1. Study flowchart as per PRISMA guidelines

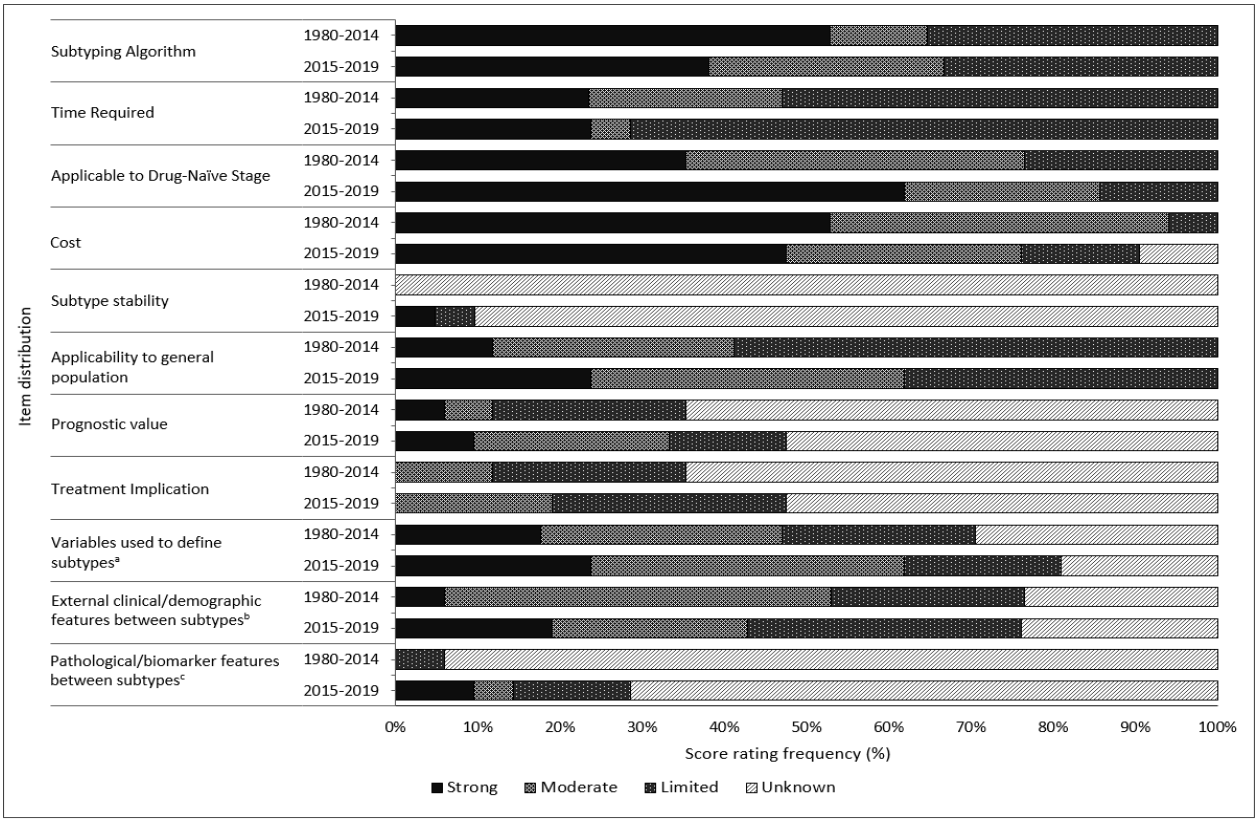


Supplementary Figure 2. Item score distribution in the Clinical Applicability tool of subtyping studies using a data-driven or hypothesis-driven approach.



a: Clinical importance of differences in the variables used to define subtypes. b: Clinical importance of differences in external clinical/demographic features between subtypes. c: Clinical importance of differences in pathological/biomarker features between subtypes.

Supplementary Figure 3. Item score distribution in the Clinical Applicability tool and dichotomized according to year of publication.



a: Clinical importance of differences in the variables used to define subtypes. b: Clinical importance of differences in external clinical/demographic features between subtypes. c: Clinical importance of differences in pathological/biomarker features between subtypes.