Definitions of Parkinson's Disease and the Major Causes of Parkinsonism, UK Biobank Phase 1 Outcomes Adjudication

Date: March 2018
Version: 1.0

Documentation prepared by:
Kathryn Bush, Kristiina Rannikmae, Tim Wilkinson, Christian Schnier and Cathie Sudlow
On behalf of UK Biobank Outcome Adjudication Group

Definitions of Parkinson's Disease and the Major Causes of Parkinsonism, UK Biobank Phase 1 Outcomes Adjudication

This algorithm includes the subtypes of Parkinson's Disease, Multiple System Atrophy, Progressive Supranuclear Palsy and Other Causes of Parkinsonism.

Algorithms allow a participant to have more than one pathological type.

Data sources on which the algorithm relies are UKB baseline assessment data (verbal interview); linked hospital admissions data (HES APC, SMR01, PEDW); death register data.

Definitions & Abbreviations:

PD	Parkinson's Disease
MSA	Multiple System Atrophy
PSP	Progressive Supranuclear Palsy
CBD	Corticobasal Degeneration
HES APC	Hospital Episode Statistics - Admitted Patient Care (England)
SMR01	Scottish Morbidity Records – General / Acute Inpatient and Day Case Admissions (Scotland)
PEDW	Patient Episode Database for Wales
EHR	Electronic Health Records
Finished Consultant Episode	The basic counting unit for statistics of admitted care Hospital EHR data (= a row of data in the data extracts provided) is a finished consultant episode (FCE).
Code date	The start date of the FCE is taken as the code date.
ICD 9	International Classification of Diseases, Version 9 (SMR only)
ICD 10	International Classification of Diseases, Version 10
Prevalent Case	First known hospitalisation with a relevant diagnostic code prior to recruitment, or self-reported event at recruitment.
Incident Case	First known hospitalisation with a relevant diagnostic code post recruitment, or cause-specific death, in those without indication of prevalent event as defined above.

Background:

Parkinson's Disease (PD) is a progressive degenerative neurological disease, associated with significant morbidity. A diagnosis of PD is made by the presence of the clinical features of Parkinsonism, which include bradykinesia, rigidity, tremor and postural instability.[1&2]

Some of the clinical features of Parkinsonism can also be caused a range of other conditions including Multiple System Atrophy (MSA), Progressive Supranuclear Palsy (PSP), Corticobasal Degeneration (CBD), drug induced Parkinsonism and many other causes. (See Table 1 for full list). Therefore, researchers wishing to study PD may wish to consider these other groups in tandem, as they are groups with significant clinical similarities. It is worth noting that due to the significant clinical overlap between these groups, it is not unusual for people with one of the rarer conditions such as MSA or PSP, to initially receive a diagnosis of PD[3]. In view of this, when using the algorithm to identify participants with PD, researchers may wish exclude participants with a code for PD, who later also receive a code for MSA, PSP or CBD.

There are no disease specific ICD codes for CBD. The codes commonly used are non-specific and were therefore not included in this algorithm: ICD 9 codes 331 (Other cerebral degenerations) and ICD 10 code: G31.8 (Other specified degenerative diseases of the nervous system). The same codes are also used for Lewy Body Dementia and a range of other conditions. There are however specific codes for CBD from GP data (Read codes), which will be included in later versions of this documentation.

Below is a description of the algorithm to identify participants with codes for PD, MSA, PSP and other causes of Parkinsonism, in the UK Biobank population.

A full list of the ICD and Biobank self-report codes used can be found in Table 1 at the end of this document.

The estimated accuracy of the algorithm is included in Appendix 1.

The use of self-report code dates is discussed in Appendix 2.

A. PARKINSON'S DISEASE

(1) Parkinson's Disease prior to baseline assessment ('prevalent Parkinson's Disease')

(a) Parkinson's Disease detected by hospital admission EHR (with or without self-report) : One (or more) of the Parkinson's Disease ICD (9 or 10) codes listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position where either

- The first ICD code date is prior to the date of baseline assessment.

OR

- The participant has self-reported the condition at the baseline assessment, but the first ICD code date is post the date of baseline assessment.

(b) Parkinson's Disease by self-report only: The participant has self-reported Parkinson's Disease at baseline assessment, but without evidence of Parkinson's Disease from linked HES APC, SMR01 or PEDW data (as defined above).

Setting the date of prevalent Parkinson's Disease Diagnosis:

- If a participant has both an ICD code and a self-report code, the earliest recorded date regardless of source is used.
- If a participant has both an ICD code and a self-report code, but the self-reported date is missing, the ICD code date is used unless it is post the date of baseline assessment, in which case the default missing date is used.
- If the participant has ICD code(s) only, the earliest ICD code date is used.
- If the participant has self-report code(s) only, the earliest self-reported date is used.
- Missing dates are set to 1/1/1900.

(2) Parkinson's Disease following baseline assessment ('incident Parkinson's Disease')

Excluding those with Parkinson's Disease detected prior to baseline assessment:

(a) Parkinson's Disease detected by hospital admission EHR: One (or more) of the Parkinson's Disease ICD (9 or 10) codes in HES APC, SMR01 or PEDW linked records, in the primary or any secondary position, with code date post the date of baseline assessment.

(b) Parkinson's Disease detected by death register only: No ICD codes in HES APC, SMR01 or PEDW linked records, but one (or more) ICD codes in death register records, in the underlying cause or any other position.

Setting the date of incident Parkinson's Disease Diagnosis:

- If a participant has ICD codes in both hospital admission and death register records, the earliest recorded code date regardless of source is used.
- If ICD code(s) recorded in hospital admission only, the earliest ICD code date is used.
- If ICD code(s) recorded in death register only, the date of death is used.

B. PROGRESSIVE SUPRANUCLEAR PALSY

(1) Progressive Supranuclear Palsy prior to baseline assessment ('prevalent Progressive Supranuclear Palsy')

(a) Progressive Supranuclear Palsy detected by hospital admission EHR: One (or more) of the Progressive Supranuclear Palsy ICD (9 or 10) codes listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position where the code date is prior to the date of baseline assessment.

Setting the date of prevalent Progressive Supranuclear Palsy Diagnosis:

- If the participant has more than one ICD code, the earliest ICD code date is used.

(2) Progressive Supranuclear Palsy following baseline assessment ('incident Progressive Supranuclear Palsy')

Excluding those with Progressive Supranuclear Palsy detected prior to baseline assessment:

(a) Progressive Supranuclear Palsy detected by hospital admission EHR: One (or more) of the Progressive Supranuclear Palsy ICD (9 or 10) codes in HES APC, SMR01 or PEDW linked records, in the primary or any secondary position, with code date post the date of baseline assessment.

(b) Progressive Supranuclear Palsy detected by death register only: No ICD codes in HES APC, SMR01 or PEDW linked records, but one (or more) ICD codes in death register records, in the underlying cause or any other position.

Setting the date of incident Progressive Supranuclear Palsy Diagnosis:

- If a participant has ICD codes in both hospital admission and death register records, the earliest recorded code date regardless of source is used.
- If ICD code(s) recorded in hospital admission only, the earliest ICD code date is used.
- If ICD code(s) recorded in death register only, the date of death is used.

C. MULTIPLE SYSTEM ATROPHY

(1) Multiple System Atrophy prior to baseline assessment ('prevalent Multiple System Atrophy')

(a) Multiple System Atrophy detected by hospital admission EHR : One (or more) of the Multiple System Atrophy ICD (9 or 10) codes listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position where the code date is prior to the date of baseline assessment.

Setting the date of prevalent Multiple System Atrophy Diagnosis:

- If the participant has more than one ICD code, the earliest ICD code date is used.

(2) Multiple System Atrophy following baseline assessment ('incident Multiple System Atrophy')

Excluding those with Multi-System Atrophy detected prior to baseline assessment:

(a) Multiple System Atrophy detected by hospital admission EHR: One (or more) of the Multiple System Atrophy ICD (9 or 10) codes in HES APC, SMR01 or PEDW linked records, in the primary or any secondary position, with code date post the date of baseline assessment.

(b) Multiple System Atrophy detected by death register only: No ICD codes in HES APC, SMR01 or PEDW linked records, but one (or more) ICD codes in death register records, in the underlying cause or any other position.

Setting the date of incident Multiple System Atrophy Diagnosis:

- If a participant has ICD codes in both hospital admission and death register records, the earliest recorded code date regardless of source is used.
- If ICD code(s) recorded in hospital admission only, the earliest ICD code date is used.
- If ICD code(s) recorded in death register only, the date of death is used.

D. ALL CAUSE PARKINSONISM (Parkinsonism)

(1) Parkinsonism prior to baseline assessment ('prevalent Parkinsonism')

(a) Parkinsonism detected by hospital admission EHR (with or without self-report) : One (or more) of the All Cause Parkinsonism ICD (9 or 10) codes listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position where either

- The first ICD code date is prior to the date of baseline assessment.

OR

- The participant has self-reported the condition at the baseline assessment, but the first ICD code date is post the date of baseline assessment.

(b) Parkinsonism by self-report only: The participant has self-reported Parkinsonism at baseline assessment, but without evidence of Parkinsonism from linked HES APC, SMR01 or PEDW data (as defined above).

Setting the date of prevalent Parkinsonism diagnosis:

- If a participant has both an ICD code and a self-report code, the earliest recorded date regardless of source is used.
- If a participant has both an ICD code and a self-report code, but the self-reported date is missing, the ICD code date is used unless it is post the date of baseline assessment, in which case the default missing date is used.
- If the participant has ICD code(s) only, the earliest ICD code date is used.
- If the participant has self-report code(s) only, the earliest self-reported date is used.
- Missing dates are set to 1/1/1900.

(2) Parkinsonism following baseline assessment ('incident Parkinsonism')

Excluding those with Parkinsonism detected prior to baseline assessment:

(a) Parkinsonism detected by hospital admission EHR: One (or more) of the All Cause Parkinsonism ICD (9 or 10) codes in HES APC, SMR01 or PEDW linked records, in the primary or any secondary position, with code date post the date of baseline assessment.

(b) Parkinsonism detected by death register only: No ICD codes in HES APC, SMR01 or PEDW linked records, but one (or more) ICD codes in death register records, in the underlying cause or any other position.

Setting the date of incident Parkinsonism diagnosis:

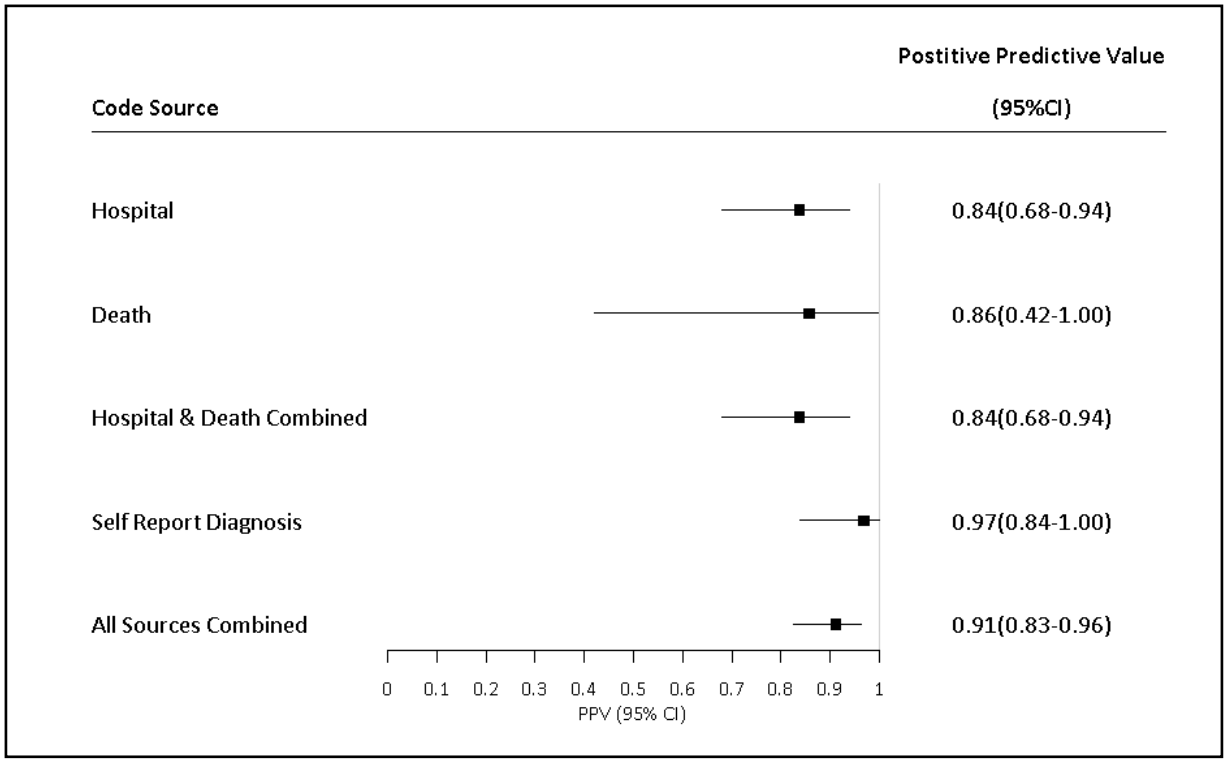
- If a participant has ICD codes in both hospital admission and death register records, the earliest recorded code date regardless of source is used.
- If ICD code(s) recorded in hospital admission only, the earliest ICD code date is used.
- If ICD code(s) recorded in death register only, the date of death is used.

Table 1. Code Lists for PD, MSA, PSP and All Cause Parkinsonism

UK Biobank Self Report Codes						
Code Type	**Code**	**Biobank Code Text**	**PD**	**MSA**	**PSP**	**All Cause Parkinsonism**
UK Biobank Self Report	Field 20002 Code1262	Parkinson's disease	ü			ü
ICD 9 Codes						
Code Type	**ICD 9 Code**	**ICD 9 Text**	**PD**	**MSA**	**PSP**	**All Cause Parkinsonism**
ICD 9 Code	332.0	Paralysis agitans	ü			ü
ICD 9 Code	332.1	Secondary parkinsonism				ü
ICD 9 Code	333.0	Other degenerative diseases of basal ganglia				ü
ICD 10 Codes						
Code Type	**ICD 10 Code**	**ICD 10 Text**	**PD**	**MSA**	**PSP**	**All Cause Parkinsonism**
ICD 10 Code	G20	Parkinson's disease	ü			ü
ICD 10 Code	G21	Secondary parkinsonism				ü
ICD 10 Code	G21.0	Malignant neuroleptic syndrome				ü
ICD 10 Code	G21.1	Other drug induced secondary parkinsonism				ü
ICD 10 Code	G21.2	Secondary parkinsonism due to other external agents				ü
ICD 10 Code	G21.3	Post encephalitic parkinsonism				ü
ICD 10 Code	G21.4	Vascular parkinsonism				ü
ICD 10 Code	G21.8	Other secondary parkinsonism				ü
ICD 10 Code	G21.9	Secondary parkinsonism unspecified				ü
ICD 10 Code	G22	Parkinsonism in diseases specified elsewhere				ü
ICD 10 Code	G23.0	Hallervorden-Spatz disease				ü
ICD 10 Code	G23.1	Progressive Supranuclear Palsy			ü	ü
ICD 10 Code	G23.2	Multiple system atrophy, parkinsonian type [MSA-P]		ü		ü
ICD 10 Code	G23.3	Multiple system atrophy, cerebellar type [MSA-C]		ü		ü
ICD 10 Code	G23.8	Other specified degenerative diseases of basal ganglia (Calcification of basal ganglia Neurogenic orthostatic hypotension [Shy-Drager]				ü
ICD 10 Code	G23.9	Degenerative diseases of basal ganglia, unspecified				ü
ICD 10 Code	G25.9	Extrapyramidal and movement disorder, unspecified				ü
ICD 10 Code	G26	Extrapyramidal and movement disorders in diseases classified elsewhere				ü
ICD 10 Code	G90.3	Multi-system degeneration		ü		ü

Appendix 1

A subset of 20,000 Biobank participants has been studied for a (pending publication) validation study, looking at the accuracy of codes in UK Biobank for identifying participants with Parkinson's Disease. In total 78 participants were identified with specific codes for Parkinson's Disease. It was not possible to comment upon the accuracy of identifying participants with MSA/PSP/CBD, due to the small number identified (<5). The plot below illustrates the positive predictive values of codes according to source and their confidence intervals:

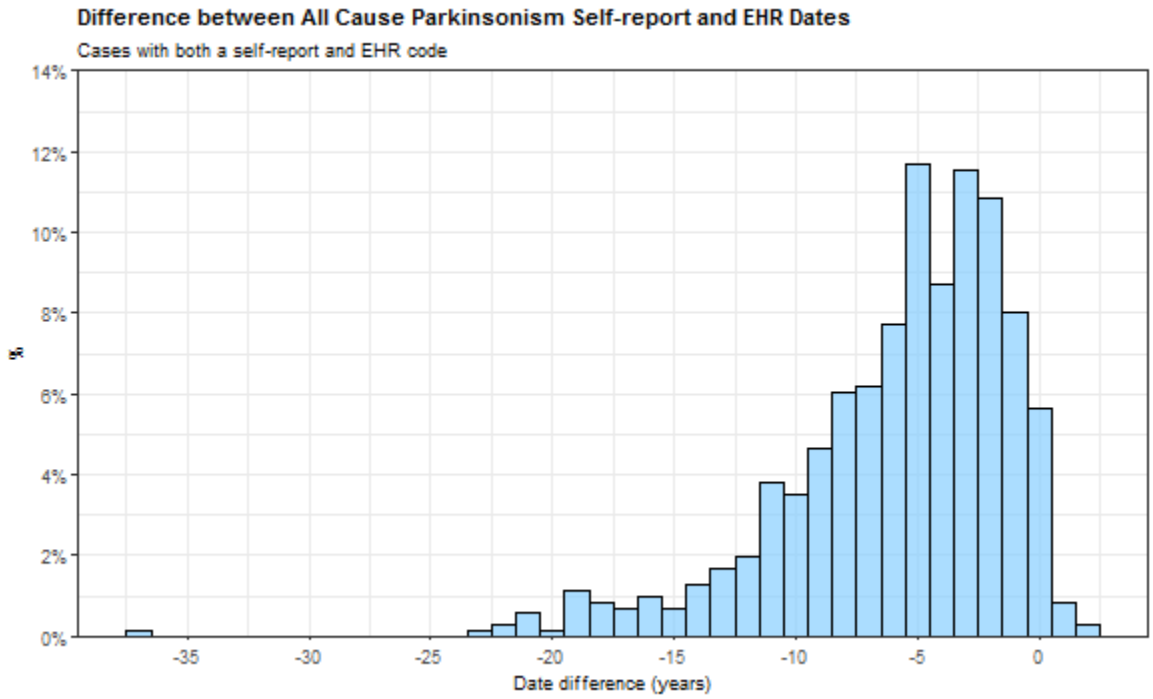


This data also reflects a recent (as yet unpublished) systematic review by the UK Biobank outcome adjudication group, which looked at the positive predictive value (PPV) and sensitivity of using routinely collected healthcare data to identify PD and Parkinsonism. For PD it showed PPVs ranging from 56-90% in hospital and death data sets. 53-87% in prescription data and 81-90% in primary care data. Combining diagnostic and medication codes increased PPV. Sensitivities ranged from 15-73% in single data sets.

Appendix 2

The self-report date is taken from the UK Biobank field 20008 ("Interpolated Year when non-cancer illness first diagnosed"). At the nurse led interviews, nurses were instructed to record either a year or an age at which the diagnosis occurred. Where an age was provided, a best-fit fractional year was then calculated.

For cases that have both a self-report and EHR code, this algorithm assigns the earliest of the two code dates as the event date for the case. The histogram below shows the difference (in years) between self-report and EHR dates for the subset of All Cause Parkinsonism cases that have both. Negative values indicate that the self-report date is earlier than the EHR. In the vast majority of cases (96%), the earliest date is the self-reported date.



Acknowledgements:

We would like to acknowledge the contributions of the 'UK Biobank Follow-up and Outcomes Working Group', whose work provided the foundations of this document:

Chair: John Danesh, Cambridge University,
Naomi Allen, UK Biobank, Oxford University,
Mark Atkinson, Swansea University,
Ekaterini Blaveri, Cancer Research UK,
Rachael Brannan, National Cancer Intelligence Network,
Carol Brayne, Cambridge University,
Sinead Brophy, Swansea University,
Nish Chaturvedi, University College London,
Rory Collins, UK Biobank, Oxford University,
Simon deLusignan, Surrey University,
Spiros Denaxas, University College London,
Parul Desai, Moorfields Eye Hospital,
Sophie Eastwood, University College London,
John Gallacher, Cardiff University,
Harry Hemingway, University College London,
Matthew Hotopf, Kings College London,
Martin Landray, Oxford University,
Ronan Lyons, Swansea University,
Mark McGilchrist, Dundee University,
Henrik Moller, Kings College London,
Terence O'Neil, Manchester University,
Mike Pringle, Nottingham University,
Tim Sprosen, Oxford University,
David Strachan, St George's University, London,
Cathie Sudlow, UK Biobank, Edinburgh University,
Frank Sullivan, Dundee University,
Rebecca Woodfield, Edinburgh University,
Qiuli Zhang, UK Biobank, Edinburgh University,
Secretariat: Robin Flaig, UK Biobank Edinburgh University.

References:

1. Jankovic J. Parkinson's disease: clinical features and diagnosis: J Neurol Neurosurg Psychiatry. 2008;79:368-276).

2. Postuma, RB, Berg, D, Stern, M et al. MDS Clinical Diagnostic Criteria for Parkinson's Disease. Mov Disord. 2015; 30: 1591–1601

3. Hughes AJ, Daniel SE, Ben-Shlomo Y, et al. The accuracy of diagnosis of parkinsonian syndromes in a specialized movement disorder service. Brain2002;125:861–70