Definitions of Dementia and the Major Diagnostic Pathologies, UK Biobank Phase 1 Outcomes Adjudication

Date: March 2018
Version: 1.0

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

Definitions of Dementia and the Major Diagnostic Pathologies, UK Biobank Phase 1 Outcomes Adjudication

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:

AD	Alzheimer's Disease
VD	Vascular Dementia
FTD	Frontotemporal Dementia
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 (SMR01 only)
ICD 10	International Classification of Diseases, Version 10
Prevalent Case	First known hospitalisation with relevant diagnostic code prior to recruitment or, self-reported diagnosis at recruitment
Incident Case	First known hospitalisation with relevant diagnostic code post recruitment or cause-specific death, in those without a prevalent code, as defined above

Background:

The World Health Organization (WHO) defines dementia as:

'...a syndrome – usually of a chronic or progressive nature – in which there is deterioration in cognitive function (i.e. the ability to process thought) beyond what might be expected from normal ageing...' [1]

The WHO estimates that in 2017 around 47 million people worldwide have dementia, and there are nearly 10 million new cases every year[1]. There are multiple subtypes of dementia, which are determined according to clinical criteria, imaging and biomarkers during life and pathologically, usually at post-mortem.

This algorithm has been developed to identify participants with codes for dementia in the UK Biobank population. It has been developed to identify participants with codes for any cause of dementia and the specific pathological subtypes of Alzheimer's Disease, Vascular Dementia and Frontotemporal Dementia.

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

When participants enrolled in the UK Biobank study, they underwent a verbal interview with research nurse, in which they could 'self-report' medical conditions. The participants were asked if they had a history of 'Dementia or Alzheimers or Cognitive Impairment' – to which they answered "Yes" or "No". This code is therefore not specific for dementia, or Alzheimer's disease, but has been included in the list of 'All Cause Dementia' for completeness.

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. ALZHEIMER'S DISEASE

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

(a) Alzheimer's Disease detected by hospital admission EHR: One (or more) of the Alzheimer'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 the code date is prior to the date of baseline assessment.

Setting the date of Prevalent Alzheimer's Disease diagnosis:

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

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

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

(a) Alzheimer's Disease detected by hospital admission EHR: One (or more) of the Alzheimer'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) Alzheimer'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 Alzheimer'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. VASCULAR DEMENTIA

(1) Vascular Dementia prior to baseline assessment ('prevalent Vascular Dementia')

(a) Vascular Dementia detected by hospital admission EHR: One (or more) of the Vascular Dementia 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 Vascular Dementia diagnosis:

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

(2) Vascular Dementia following baseline assessment ('incident Vascular Dementia')

Excluding those with Vascular Dementia detected prior to baseline assessment:

(a) Vascular Dementia detected by hospital admission EHR: One (or more) of the Vascular Dementia 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) Vascular Dementia 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 Vascular Dementia 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. FRONTOTEMPORAL DEMENTIA

(1) Frontotemporal Dementia prior to baseline assessment ('prevalent Frontotemporal Dementia')

(a) Frontotemporal Dementia detected by hospital admission EHR : One (or more) of the Frontotemporal Dementia 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 Frontotemporal Dementia diagnosis:

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

(2) Frontotemporal Dementia following baseline assessment ('incident Frontotemporal Dementia')

Excluding those with Frontotemporal Dementia detected prior to baseline assessment:

(a) Frontotemporal Dementia detected by hospital admission EHR: One (or more) of the Frontotemporal Dementia 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) Frontotemporal Dementia 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 Frontotemporal Dementia 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 DEMENTIA

(1) Dementia prior to baseline assessment ('prevalent dementia')

(a) Dementia detected by hospital admission EHR (with or without self-report): One (or more) of the All Cause Dementia 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) Dementia by self-report only: The participant has self-reported dementia, but without evidence of dementia from linked HES APC, SMR01 or PEDW data (as defined above).

Setting the date of prevalent dementia 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) Dementia following baseline assessment ('incident dementia')

Excluding those with dementia detected prior to baseline assessment.

(a) Dementia detected by hospital admission EHR: One (or more) of the All Cause Dementia 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) Dementia 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 dementia 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 Dementia and Specified Subtypes

UK Biobank Self Report Codes						
Code Type	**Code**	**UK Biobank Code Text**	**AD**	**VD**	**FTD**	**All Cause Dementia**
UK Biobank Self Report	Field 20002 Code 1263	Dementia/Alzheimers/Cognitive Impairment				ü
ICD 9 Codes						
Code Type	**ICD 9 Code**	**ICD 9 Code Text**	**AD**	**VD**	**FTD**	**All Cause Dementia**
ICD 9 Code	290.2	Senile dementia, depressed or paranoid type				ü
ICD 9 Code	290.3	Senile dementia with acute confusional state				ü
ICD 9 Code	290.4	Arteriosclerotic dementia		ü		ü
ICD 9 Code	291.2	Other alcoholic dementia				ü
ICD 9 Code	294.1	Dementia in other conditions classified elsewhere				ü
ICD 9 Code	331.0	Alzheimer's disease	ü			ü
ICD 9 Code	331.1	Pick's disease			ü	ü
ICD 9 Code	331.2	Senile degeneration of brain				ü
ICD 9 Code	331.5	Creutzfeldt-Jakob disease				ü
ICD 10 Codes						
Code Type	**ICD 10 Code**	**ICD 10 Code Text**	**AD**	**VD**	**FTD**	**All Cause Dementia**
ICD 10 Code	A81.0	Sporadic Creutzfeldt-Jakob disease				ü
ICD 10 Code	F00	Dementia in Alzheimer's disease	ü			ü
ICD 10 Code	F00.0	Dementia in Alzheimer's disease with early onset	ü			ü
ICD 10 Code	F00.1	Dementia in Alzheimer's disease with late onset	ü			ü
ICD 10 Code	F00.2	Dementia in Alzheimer's disease, atypical or mixed type	ü			ü
ICD 10 Code	F00.9	Dementia in Alzheimer's disease, unspecified	ü			ü
ICD 10 Code	F01	Vascular dementia		ü		ü
ICD 10 Code	F01.0	Vascular dementia of acute onset		ü		ü
ICD 10 Code	F01.1	Multi-infarct dementia		ü		ü
ICD 10 Code	F01.2	Subcortical vascular dementia		ü		ü
ICD 10 Code	F01.3	Mixed cortical and sub-cortical vascular dementia		ü		ü
ICD 10 Code	F01.8	Other vascular dementia		ü		ü
ICD 10 Code	F01.9	Vascular dementia, unspecified		ü		ü
ICD 10 Code	F02	Dementia in other diseases classified elsewhere				ü
ICD 10 Code	F02.0	Dementia in Picks disease			ü	ü
ICD 10 Code	F02.1	Dementia in Creutzfeldt-Jacob disease				ü
ICD 10 Code	F02.2	Dementia in Huntington’s disease				ü

ICD 10 Codes Continued Overleaf

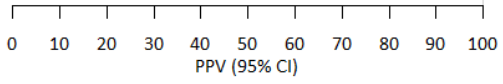
ICD 10 Codes (Continued)						
Code Type	**ICD 10 Code**	**ICD 10 Code Text**	**AD**	**VD**	**FTD**	**All Cause Dementia**
ICD 10 Code	F02.3	Dementia in Parkinson's disease				ü
ICD 10 Code	F02.4	Dementia in HIV disease				ü
ICD 10 Code	F02.8	Dementia in other specified diseases classified elsewhere				ü
ICD 10 Code	F03	Unspecified dementia				ü
ICD 10 Code	F05.1	Delirium superimposed on dementia				ü
ICD 10 Code	F10.6	Mental and behavioural disorders due to use of alcohol - amnesic syndrome				ü
ICD 10 Code	G30	Alzheimer's disease	ü			ü
ICD 10 Code	G30.0	Alzheimer's disease with early onset	ü			ü
ICD 10 Code	G30.1	Alzheimer's disease with late onset	ü			ü
ICD 10 Code	G30.8	Other Alzheimer's disease	ü			ü
ICD 10 Code	G30.9	Alzheimer's disease unspecified	ü			ü
ICD 10 Code	G31.0	Circumscribed brain atrophy			ü	ü
ICD 10 Code	G31.1	Senile degeneration of brain				ü
ICD 10 Code	G31.8	Other specified degenerative diseases of nervous system				ü
ICD 10 Code	I67.3	Binswanger's disease		ü		ü

Appendix 1

A 2018 systematic review from the UK Biobank outcomes adjudication team looked at the accuracy of identifying dementia cases with routinely-collected healthcare data. It found that for all-cause dementia, positive predictive values (PPVs) ranged from 33-100%. Sensitivities (relative to all true dementia cases in the population) ranged from 21-86%. PPVs for Alzheimer's disease (range 57-100%) were generally higher than for vascular dementia (range 19-91%).[2]

A subset of 17,000 Biobank participants has been studied for an (as yet unpublished) validation study, looking at the accuracy of codes in UK Biobank for identifying participants with dementia. The positive predictive values (PPVs) for different code groups are detailed in a forest plot below, with 95% confidence intervals (Clopper Pearson Exact method).

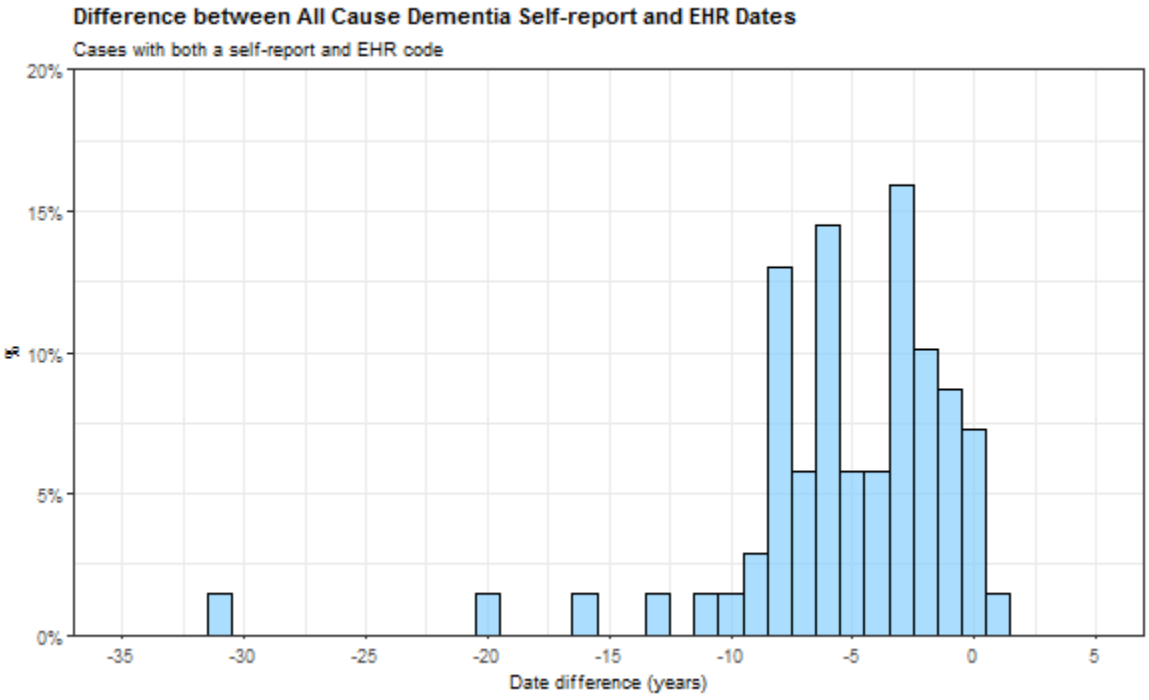
Code Source & Combination	True Positives With Dementia	All Partcipants With Codes	Postitive Predictive Value (95%CI)
All Cause Dementia			
Primary care	92	106	86.8%(78.8-92.6)
Hospital	48	55	87.3%(75.5-94.7)
Mortality	8	10	80.0%(44.4-97.5)
Hospital & Mortality	49	58	84.5%(72.6-92.7)
All Sources Combined	99	120	82.5%(74.5-88.8)
Alzheimer's Disease			
Primary care	40	54	74.1%(60.3-85.0)
Hospital	15	22	68.2%(45.1-86.1)
Mortality	2	4	50.0%(6.80-93.2)
Hospital & Mortality	17	24	70.8%(48.9-87.4)
All Sources Combined	45	63	71.4%(58.7-82.1)
Vascular Dementia			
Primary care	6	13	46.2%(19.2-74.9)
Hospital & Mortality	2	6	33.3%(4.3-77.7)
All Sources Combined	7	16	43.8%(19.8-70.1)
FTD/PDD/DLB Combined			
All Sources Combined	8	12	66.7%(34.9-90.1)



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 Dementia 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. World Health Organization 2017. Dementia Fact Sheet. Accessed 29 November 2017, http://www.who.int/mediacentre/factsheets/fs362/en/

2. Wilkinson T, Ly A, Schnier C, Rannikmäe K, Bush K, Brayne C, Quinn TJ, Sudlow CLM. Identifying dementia cases with routinely-collected health data: a systematic review. Alzheimer's & Dementia: The Journal of the Alzheimer's Association.