Definitions of Stroke for UK Biobank Phase 1 Outcomes Adjudication

Date: August 2017
Version: 1.1

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

Definitions of Stroke, UK Biobank Phase 1 Outcomes Adjudication

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:

IS	Ischaemic Stroke
IH	Intracerebral Haemorrhage
SH	Subarachnoid Haemorrhage
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:

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. STROKE OF ANY TYPE

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

(a) Stroke detected by hospital admission EHR (with or without self-report) : One (or more) of the Stroke of any type ICD (9 or 10) listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position, with a code date prior to the date of baseline assessment.

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

Setting the date of prevalent Stroke diagnosis:

- If a participant has both an ICD code and a self-report code, the ICD code 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) Stroke following baseline assessment ('incident Stroke')

Excluding those with a stroke prior to baseline assessment:

(a) Stroke detected by hospital admission EHR: One (or more) of the Stroke of any type 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) Stroke 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 Stroke 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. ISCHAEMIC STROKE

(1) Ischaemic Stroke prior to baseline assessment ('prevalent Ischaemic Stroke')

(a) Ischaemic Stroke detected by hospital admission EHR (with or without self-report) : One (or more) of the Ischaemic Stroke ICD (9 or 10) listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position, with a code date prior to the date of baseline assessment.

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

Setting the date of prevalent Ischaemic Stroke diagnosis:

- If a participant has both an ICD code and a self-report code, the ICD code 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) Ischaemic Stroke following baseline assessment ('incident Ischaemic Stroke')

Excluding those with **any** stroke (of any type) prior to baseline assessment:

(a) Ischaemic Stroke detected by hospital admission EHR: One (or more) of the Ischaemic Stroke 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) Ischaemic Stroke 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 Ischaemic Stroke 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. INTRACEREBRAL HAEMORRHAGE

(1) Intracerebral Haemorrhage prior to baseline assessment ('prevalent Intracerebral Haemorrhage')

(a) Intracerebral Haemorrhage detected by hospital admission EHR (with or without self-report) : One (or more) of the Intracerebral Haemorrhage ICD (9 or 10) listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position, with a code date prior to the date of baseline assessment.

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

Setting the date of prevalent Intracerebral Haemorrhage diagnosis:

- If a participant has both an ICD code and a self-report code, the ICD code 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) Intracerebral Haemorrhage following baseline assessment ('incident Intracerebral Haemorrhage')

Excluding those with **any** stroke (of any type) prior to baseline assessment:

(a) Intracerebral Haemorrhage detected by hospital admission EHR: One (or more) of the Intracerebral Haemorrhage 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) Intracerebral Haemorrhage 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 Intracerebral Haemorrhage 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. SUBARACHNOID HAEMORRHAGE

(1) Subarachnoid Haemorrhage prior to baseline assessment ('prevalent Subarachnoid Haemorrhage')

(a) Subarachnoid Haemorrhage detected by hospital admission EHR (with or without self-report) : One (or more) of the Subarachnoid Haemorrhage ICD (9 or 10) listed in Table 1, in HES APC, SMR01 or PEDW linked records in the primary or any secondary position, with a code date prior to the date of baseline assessment.

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

Setting the date of prevalent Subarachnoid Haemorrhage diagnosis:

- If a participant has both an ICD code and a self-report code, the ICD code 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) Subarachnoid Haemorrhage following baseline assessment ('incident Subarachnoid Haemorrhage')

Excluding those with **any** stroke (of any type) prior to baseline assessment:

(a) Subarachnoid Haemorrhage detected by hospital admission EHR: One (or more) of the Subarachnoid Haemorrhage 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) Subarachnoid Haemorrhage 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 Subarachnoid Haemorrhage 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 Stroke

UK Biobank Self Report Codes						
Code Type	**Code**	**Biobank Code Text**	**IS**	**IH**	**SH**	**Stroke of any type**
UK Biobank Self Report	Field 20002 Code 1081	Stroke				✓
UK Biobank Self Report	Field 20002 Code 1086	Subarachnoid haemorrhage			✓	✓
UK Biobank Self Report	Field 20002 Code 1491	Brain haemorrhage		✓		✓
UK Biobank Self Report	Field 20002 Code 1583	Ischaemic stroke	✓			✓
ICD 9 Codes						
Code Type	**ICD 9 Code**	**ICD 9 Text**	**IS**	**IH**	**SH**	**Stroke of any type**
ICD 9 Code	430.X	Subarachnoid haemorrhage			✓	✓
ICD 9 Code	431.X	Intracerebral haemorrhage		✓		✓
ICD 9 Code	434.X	Occlusion of cerebral arteries	✓			✓
ICD 9 Code	434.0	Cerebral thrombosis	✓			✓
ICD 9 Code	434.1	Cerebral embolism	✓			✓
ICD 9 Code	434.9	Cerebral artery occlusion, unspecified	✓			✓
ICD 9 Code	436.X[1]	Acute, but ill-defined, cerebrovascular disease	✓			✓
ICD 10 Codes						
Code Type	**ICD 10 Code**	**ICD 10 Text**	**IS**	**IH**	**SH**	**Stroke of any type**
ICD 10 Code	I60	Subarachnoid haemorrhage			✓	✓
ICD 10 Code	I60.0	Subarachnoid haemorrhage from carotid siphon and bifurcation			✓	✓
ICD 10 Code	I60.1	Subarachnoid haemorrhage from middle cerebral artery			✓	✓
ICD 10 Code	I60.2	Subarachnoid haemorrhage from anterior communicating artery			✓	✓
ICD 10 Code	I60.3	Subarachnoid haemorrhage from posterior communicating artery			✓	✓
ICD 10 Code	I60.4	Subarachnoid haemorrhage from basilar artery			✓	✓
ICD 10 Codes Continued Overleaf						

[1] ICD 10: I64 (Stroke, not specified as haemorrhage or infarction) and ICD 9 436 (Acute, but ill-defined, cerebrovascular disease) have been classified as ischaemic stroke because of evidence that the vast majority of these are ischaemic strokes

ICD 10 Codes (Continued)						
Code Type	**ICD 10 Code**	**ICD 10 Text**	**IS**	**IH**	**SH**	**Stroke of any type**
ICD 10 Code	I60.5	Subarachnoid haemorrhage from vertebral artery			✓	✓
ICD 10 Code	I60.6	Subarachnoid haemorrhage from other intracranial arteries			✓	✓
ICD 10 Code	I60.7	Subarachnoid haemorrhage from intracranial artery, unspecified			✓	✓
ICD 10 Code	I60.8	Other subarachnoid haemorrhage			✓	✓
ICD 10 Code	I60.9	Subarachnoid haemorrhage, unspecified			✓	✓
ICD 10 Code	I61	Intracerebral haemorrhage		✓		✓
ICD 10 Code	I61.0	Intracerebral haemorrhage in hemisphere, subcortical		✓		✓
ICD 10 Code	I61.1	Intracerebral haemorrhage in hemisphere, cortical		✓		✓
ICD 10 Code	I61.2	Intracerebral haemorrhage in hemisphere, unspecified		✓		✓
ICD 10 Code	I61.3	Intracerebral haemorrhage in brain stem		✓		✓
ICD 10 Code	I61.4	Intracerebral haemorrhage in cerebellum		✓		✓
ICD 10 Code	I61.5	Intracerebral haemorrhage, intraventricular		✓		✓
ICD 10 Code	I61.6	Intracerebral haemorrhage, multiple localized		✓		✓
ICD 10 Code	I61.8	Other intracerebral haemorrhage		✓		✓
ICD 10 Code	I61.9	Intracerebral haemorrhage, unspecified		✓		✓
ICD 10 Code	I63	Cerebral infarction	✓			✓
ICD 10 Code	I63.0	Cerebral infarction due to thrombosis of precerebral arteries	✓			✓
ICD 10 Code	I63.1	Cerebral infarction due to embolism of precerebral arteries	✓			✓
ICD 10 Code	I63.2	Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries	✓			✓
ICD 10 Code	I63.3	Cerebral infarction due to thrombosis of cerebral arteries	✓			✓
ICD 10 Code	I63.4	Cerebral infarction due to embolism of cerebral arteries	✓			✓
ICD 10 Code	I63.5	Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries	✓			✓
ICD 10 Code	I63.6	Cerebral infarction due to cerebral venous thrombosis, nonpyogenic	✓			✓
ICD 10 Code	I63.8	Other cerebral infarction	✓			✓

ICD 10 Codes Continued Overleaf

ICD 10 Codes (Continued)						
Code Type	**ICD 10 Code**	**ICD 10 Text**	**IS**	**IH**	**SH**	**Stroke of any type**
ICD 10 Code	I63.9	Cerebral infarction, unspecified	✓			✓
ICD 10 Code	I64.X[1]	Stroke, not specified as haemorrhage or infarction	✓			✓

Appendix 1

The estimated accuracy of algorithmically defined stroke events is based on two different systematic reviews of published studies on coded and self-reported data conducted on behalf of the UK Biobank Stroke Outcomes Group. The selected ICD codes from hospital and death data are estimated to produce positive predictive values (PPVs):

- for any stroke of around 85-90%;
- for ischaemic stroke of around 80-90%
- for intracerebral haemorrhage of around 75-95%
- for subarachnoid haemorrhage of around 85-95%

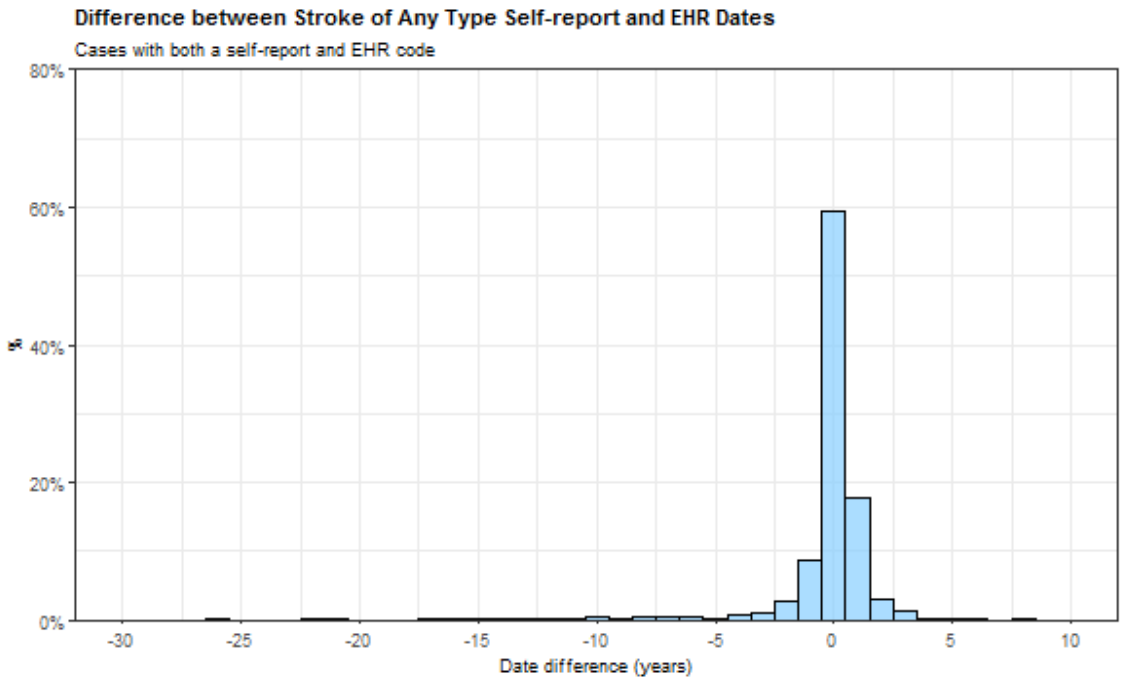
The PPV of stroke events prior to recruitment identified by self-report alone is likely to be lower (<70%), increasing to 75% or more if transient ischaemic attacks are considered true positives. The PPV of self-report for specific pathological types of stroke is uncertain.

Further direct validation studies in UK Biobank participants are ongoing and additional information on accuracy of event identification will be added to this documentation as it becomes available.

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 EHR code date 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 Stroke (any type) cases that have both. Negative values indicate that the self-report date is earlier than the EHR. There is a good level of agreement between the dates from the two sources, with 78% of cases having a self-report date within 1 year of the EHR 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.