CLEANED/DERIVED VARIABLE METADATA TOP SHEET

For Submission to the NSHD Scientific Support Team

Date of submitting documentation	April 2019
Categories of variables*: (may be more than one)	Imaging outcome measure – white matter hyperintensity volumes (global and regional) generated using the cross-sectional Bayesian Model Selection (BaMoS) pipeline
Summary of work undertaken	Cleaning± Description/Rationale: Values were generated using the BaMoS pipeline. The WMH segmentations were then QC'd by CL – further detail is provided below.
Source data file(s)	Insight46_bamos_outcomes.csv Insight46_BaMoSQC_forNSHD.xlxs
Date source file(s) created:	Insight46_bamos_outcomes_cleaned.dta
Names of source variables	Variables are specified below.
Syntax provided	Y
Location of syntax file	bamos_cleaning_forNSHD.do
Date syntax file created:	April 2019
Format of syntax	Stata
Output variables (please list names of new variables created)	Variables are specified below. Bamosfailreason was generated using Insight46_BaMoSQC_forNSHD.xlxs
Output data file provided	Y
Date output file created:	April 2019
Location of output file	Insight46_bamos_outcomes_cleaned.dta
Format of output file	Stata .dta file
Documentation provided	Cleaning: Y Derivation: Y
List any papers in which cleaned/derived variables have been used	Lane et al 2019 Investigating the associations between blood pressure across adulthood and late-life brain structure and pathology in the 1946 British birth cohort: an epidemiological study. Lancet Neurology

BaMoS outcome variables at 69-71y (Phase 1 Insight 46) (2015-7)

This document details all the variables created from the BaMoS pipeline and how they were cleaned (i.e. BaMoS QC).

BaMoS QC

Cross-sectional BaMoS processing[1] was performed using all available T1 and FLAIR images and the output variables extracted. BaMoS segmentations were then visually reviewed in NiftyMidas, overlaid on the resampled FLAIR image. Issues with segmentation that arose e.g mis-segmentation of cortical strokes, excessive temporal lobe artefacts, were flagged and the scans re-run with modified BaMoS scripts, the segmentations for which were then returned and re-reviewed. Choroid plexus which was excessively mis-segmented (as determined on visual review) was manually edited in NiftyMidas and updated outputs generated. Certain pathologies e.g. demyelination, some cortical strokes, vascular abnormalities causing white matter hyperintensities (WMH) not related to cerebral small vessel disease, that were inappropriately segmented were excluded as necessary. Individuals with neurological diagnoses were not otherwise excluded.

Variables:

bamosuseable – whether the WMH segmentation is useable

Coded as:

0	= not useable
1	= useable
-99	= no scan available

bamosfailreason – the reason why a segmentation was not useable (only relevant if bamosuseable=0)

Coded as:

1	= T1 QC fail
2	= FLAIR QC fail
3	= BaMoS segmentation fail
4	= cortical stroke inappropriately segmented as WMH
5	= demyelination inappropriately segmented as WMH
6	= other vascular pathology inappropriately segmented
-99	= not applicable (either bamosuseable=1 or no scan available)

BaMoS outcome variables

Generated outcomes from BaMoS segmentations. For use in analyses, volumetric measures should be adjusted for head size using the total intracranial volume (TIV) measure (**variable x**, **detailed in document x**). It is recommended that this is done by adjusting for TIV within a statistical model, rather than creating a new proportional value "the proportion method".[2,3] Regional lobar values are derived using the GIF v.3 parcellations of the volumetric T1 image[4]. Regional layer values are derived by dividing the subcortical region into 4 equidistant layers according to the normalised distance between the ventricular system and the white matter/cortical grey matter interface adapted from the method described by Yezzi *et al* to compute cortical thickness[5].

lesiontot – global white matter hyperintensity volume (ml). Includes subcortical grey matter but not the infratentorial region.

volwmhfront – white matter hyperintensity volume (ml) in the frontal lobes

volwmhocc - white matter hyperintensity volume (ml) in the occipital lobes

volwmhpar - white matter hyperintensity volume (ml) in the parietal lobes

volwmhtemp - white matter hyperintensity volume (ml) in the temporal lobes

volwmhbg - white matter hyperintensity volume (ml) in the basal ganglia

volwmhlayer1 - white matter hyperintensity volume (ml) in layer 1 (inner)

volwmhlayer2 - white matter hyperintensity volume (ml) in layer 2

volwmhlayer3 - white matter hyperintensity volume (ml) in layer 3

volwmhlayer4 - white matter hyperintensity volume (ml) in layer 4 (outer)

distwmhfront – the proportion of global WMH in the frontal lobes (value 0 – 1)

distwmhocc - the proportion of global WMH in the occipital lobes (value 0 – 1)

distwmhpar - the proportion of global WMH in the parietal lobes (value 0 – 1)

distwmhtemp - the proportion of global WMH in the temporal lobes (value 0 – 1)

distwmhbg – the proportion of global WMH in the basal ganglia (value 0 – 1)

distwmhlayer1 - the proportion of global WMH in layer 1 (value 0 – 1)

distwmhlayer2 - the proportion of global WMH in layer 2 (value 0 – 1)

distwmhlayer3 - the proportion of global WMH in layer 3 (value 0 – 1)

distwmhlayer4 - the proportion of global WMH in layer 4 (value 0 – 1)

percentwmhfront – the percentage of total white matter in the frontal lobes considered to be white matter hyperintensity (range 0 – 100)

percentwmhocc - the percentage of total white matter in the occipital lobes considered to be white matter hyperintensity (range 0 – 100)

percentwmhpar - the percentage of total white matter in the parietal lobes considered white matter hyperintensity (range 0 – 100)

percentwmhtemp - the percentage of total white matter in the temporal lobes considered white matter hyperintensity (range 0 – 100)

percentwmhbg - the percentage of total volume in basal ganglia considered white matter hyperintensity (range 0 – 100)

For all of the variables above, the outcome measure is either a volume in mls (vol), proportion (dist) or percentage (percent), and is otherwise coded as:

-99 = not available (either bamosuseable=0 or no scan available)

Summary statistics

bamosuseable	Freq.	Percent
No scan available (-99)	31	6.18
Not useable (0)	16	3.19
Useable (1)	455	90.64

bamosfailreason	Frequency	Percent
Not applicable (-99)	486	96.81
T1 fail (1)	3	0.60
FLAIR fail (2)	3	0.60
Bamos segmentation fail (3)	1	0.20
Cortical stroke mis-segmented (4)	5	1.00
Demyelination mis-segmented (5)	3	0.60
Other vascular anomaly mis-segmented (6)	1	0.20

Variable	Obs	Mean	Std. Dev.	Min	Max
lesiontot	455	5.109443	5.441696	.2661919	33.66793
volwmhocc	455	.654722	.6169186	0	5.820504
volwmhfront	455	2.55895	2.91357	.0847699	18.22171
volwmhpar	455	1.069738	1.578511	.0013348	10.0226
volwmhtemp	455	.53425	.6169533	.0163922	4.18243
volwmhbg	455	.291783	.3828594	.0013348	2.595435
volwmhlayer1	455	1.680457	1.406311	.0702311	8.167364
volwmhlayer2	455	1.317616	1.739918	.0323781	14.00983
volwmhlayer3	455	1.025574	1.537759	.0136052	10.76657
volwmhlayer4	455	1.085797	1.329666	.0319939	9.112107
distwmhocc	455	.1749019	.1068451	0	.5984535
distwmhfront	455	.4905785	.1389233	.1478342	.8764374
distwmhpar	455	.1587049	.0957124	.0014029	.5342677
distwmhbg	455	.0647237	.0435142	.0021898	.3443137
distwmhtemp	455	.1110911	.0573211	.0132275	.3828413
distwmhlayer1	455	.394021	.1336008	.0945299	.7496052
distwmhlayer2	455	.2259022	.0772718	.0515649	.4997641
distwmhlayer3	455	.1653936	.0707513	.0219116	.4312162
distwmhlayer4	455	.2146832	.1054891	.0446742	.647063
percentwmhocc	455	1.135649	.9654228	0	7.283258
percentwmhfront	455	1.249054	1.406807	.0381781	9.244668
percentwmhpar	455	1.113462	1.607811	.0016583	9.820433
percentwmhtemp	455	.5646139	.6478538	.0179838	4.883367
percentwmhbg	455	.7077708	.8949738	.0032964	6.353084

References

1. Sudre CH, Cardoso MJ, Bouvy WH, Biessels GJ, Barnes J, Ourselin S. Bayesian model selection for pathological neuroimaging data applied to white matter lesion

segmentation. IEEE Trans Med Imaging [online serial]. 2015;34:2079–2102. Accessed at: http://ieeexplore.ieee.org/document/7078891/. Accessed June 25, 2017.

2. Malone IB, Leung KK, Clegg S, et al. Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance. Neuroimage. Elsevier; 2015;104:366–372.

3. Barnes J, Ridgway GR, Bartlett J, et al. Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage [online serial]. 2010;53:1244–1255. Accessed at: http://www.ncbi.nlm.nih.gov/pubmed/20600995. Accessed September 1, 2017.

4. Cardoso MJ, Modat M, Wolz R, et al. Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. IEEE Trans Med Imaging. 2015;34:1976–1988.

5. Yezzi AJ, Prince JL. An eulerian pde approach for computing tissue thickness. IEEE Trans Med Imaging [online serial]. 2003;22:1332–1339. Accessed at: http://www.ncbi.nlm.nih.gov/pubmed/14552586. Accessed August 13, 2018.