DA 3

Harmonising large-scale imaging databases to provide integrated assessments of the role of white matter hyperintensities in cognitive aging.

Start date: Jun 2018 **Completion date: 31 Dec 2019**

Overall Discovery Award objectives:

1. Analysis of the distributions of white matter hyperintensities (WMH) in the 3 datasets to identify sources of difference.
2. Development and testing of harmonisation procedures
3. Harmonised WMH measures and harmonisation tools available on the DPUK portal

Deliverables	Milestones	Milestone deadline	Work package dependencies	Person*(s) responsible
Objective 1:				
D1.1 Data available for processing	M1.1.1 Data access Whitehall	M1.1.1 Complete		Griffanti*, Duff, Mackay
	M1.1.2 Data access UKBiobank	M1.1.2 Complete		Griffanti*, Duff, Mackay
	M1.1.3 Non-imaging data matching across datasets for subjects selection	M1.1.3 Complete		Griffanti*, Duff, Mackay
D1.2 Generated imaging derived phenotypes (IDPs)	M1.2.1 Image preprocessing to extract WMH measures (IDPs)	M1.2.1 Complete		Griffanti*, Duff, Mackay
D1.3 Completed analyses reported in a conference abstract (e.g. DPUK, OHBM)	M1.3.1 Identification of sources of difference in WMH in the 3 datasets	M1.3.1 Complete		Griffanti*, Duff, Mackay, Jenkinson,
Objective 2:				
D2.1 Technical report	M2.1.1 Technical development of harmonisation methods	M2.1.1 Complete		Duff*, Griffanti, Jenkinson,
D2.2 Completed harmonisation of DPUK datasets Conference presentation/poster; Manuscript	M2.2.1 Testing / refinement / validation of harmonisation methods on the 3 datasets	M2.2.1 Complete	1.3.1	Duff*, Griffanti, Jenkinson,
Objective 3:				
D3.1 Harmonised WMH measures and harmonisation tools available on the DPUK portal	M3.1.1 Initial code release on github	M3.1.1 Complete	2.1.1	Griffanti*,
	M3.1.2 Final upload of code and WMH measures on DPUK	M3.1.2 Dec 2019 (was Jun 2019) NCE	1.3.1 and 2.2.1	Griffanti*, Duff*

Key updates on delivery against milestones since last report

- M1.1.3. Non-imaging data matching across datasets for subjects selection

The configuration file for the funpack tool (ukbparse in former reports - a Python library for pre-processing of UKB tabular data. https://git.fmrib.ox.ac.uk/fsl/funpack/) is now openly available in the DPUK git repository (https://issues.dpuk.org/eugeneduff/wmh_harmonisation) (see 3.1). The selected UKB non-imaging variables can be

automatically extracted and modified to be comparable to those in WHII (e.g. continuous to categorical variables, units of measure, matching tests for the same function). The configuration file is also customizable to include more/different variables.

- M 1.2.1 Image pre-processing to extract WMH measures (IDPs)

We optimised the pre-processing and WMH volume extraction pipeline minimising/controlling for the sources of variability found in 1.3 on a subset of subjects for which we had manual delineation of WMH (see 2.1). Using this optimised pipeline, we re-processed the data and extracted measures of WMH from Whitehall II and UKB datasets.

- M2.2 Testing / refinement / validation of harmonisation methods on the 3 datasets

In the last report we proposed a processing pipeline that minimises the effect of the non-biological sources of difference. When using this pipeline also in the UKB dataset, we confirmed that BIANCA performance is more similar across datasets than when extracted with the original settings (see project narrative).
A key component of harmonisation is to account for differences in demographics and health/lifestyle status across datasets. The first step towards this is described above – to identify and harmonise demographic and health/lifestyle data across studies. Then, to utilise these data to harmonise datasets, we have explored a number of predictive strategies that model the impact of these data on the imaging measures, which would enables matching of data across studies. Specifically, we identified 9 non-imaging variables that are predictive features of the amount of WMH in individuals in both WHII and UKB. These models account for a significant amount of variance across individuals. We find that a variable accounting for study becomes substantially less necessary when optimal pre-processing steps are used. However, some across-study bias still exist. Thus, these models help to harmonise measurements in the Whitehall study to Biobank. We are now exploring further optimisation of these models.

- M3.1.1 Initial code release on GitHub

The configuration file for funpack is now accessible from: https://issues.dpuk.org/eugeneduff/wmh_harmonisation
The repository will be updated with other material (e.g. BIANCA training dataset) by the end of the project.

Summary of plans for the future

- Test the prediction model using WMH volumes extracted with the optimised pipeline and evaluate its accuracy.
- Compare and optimise prediction models and produce Biobank-harmonised WMH estimates for Whitehall dataset.
- Release BIANCA training dataset, code for the prediction model, an implementation guide, and harmonised measures of WMH.
- Write a journal article.

Risks	**Mitigation**
1) The final WMH measures are not completely comparable across datasets	1) Characterise the differences remained in the data after harmonisation to derive a conversion factor or function to link the measures.
2) The model is not robust with respect to changes in the input (i.e. not giving good prediction when we test it using the WMH measures extracted with the optimised pipeline)	2) Explore the conditions under which the model gives good prediction and suggest modifications in other conditions. Define a standard procedure for such cases.

Team members

Ludovica Griffanti, Eugene Duff, Mark Jenkinson, Stephen M Smith, Clare E Mackay

Outcomes

- Configuration file for funpack (https://git.fmrib.ox.ac.uk/fsl/funpack/) to automatically obtain matched variables UKB - WHII available on DPUK git (https://issues.dpuk.org/eugeneduff/wmh_harmonisation)
- Abstract presented to the organisation for human brain mapping (OHBM) conference 2019
- Abstract submitted to European Congress of Radiology (ECR) 2020.
- June 2019, poster presentation at OHBM conference. Title: *"Harmonising measures of white matter hyperintensities across sites and studies: impact of training and manual rating in FSL-BIANCA"*

- 1st July 2019, preliminary results presented by Ludovica Griffanti at the MRI dementia conference 2019, St Anne's College, University of Oxford. Title: *"Harmonising MR analysis across sites "*
- October 2019, abstract submitted to European Congress of Radiology. Title: *"Between- and within-rater agreement in white matter hyperintensitiy segmentation from manual rating and a supervised automated classifier, FSL-BIANCA"*

Project narrative

Since last report our harmonisation work evolved on the two aims we previously identified. The first aim was to identify and correct/control or minimise the sources of non-biological variability in the WMH measures between datasets, mainly due to the image analysis pipeline. The second aspect concerns the modelling of the biological variability between datasets, analysing the relationship between WMH and relevant demographic/clinical/behavioural factors and predicting WMH measures from non-imaging variables.

Towards the first aim, in the last report we proposed a processing pipeline that minimises the effect of the non-biological sources of difference: 1) use masks generated from the same rater whenever possible; 2) perform bias field correction; 3) use a small set of modalities (T1 and FLAIR, not FA), which are more common across studies; 4) use automated thresholding of the lesion probability maps (LOCATE); 5) train BIANCA on data coming from different scanners/studies. Using this pipeline, BIANCA performance is more similar across datasets than when extracted with the original settings, also when comparing WH and UKB data. Moreover the original volume bias that was identified across scanners is significantly reduced after applying the pipeline (Figure 1).

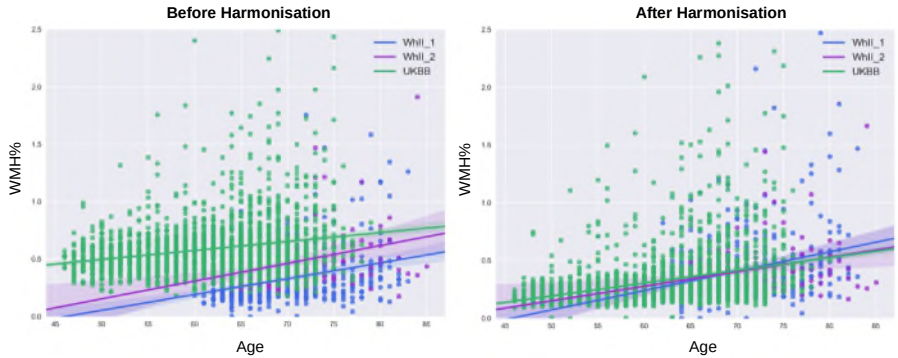

Figure 1. Effect of pre-processing pipeline on WMH volumes for Whitehall scanner 1 data (blue), scanner 2 (purple) and UK Biobank (green). Scatter plot of WMH volumes (% of brain volume) against age shows reduction of biases while maintaining the known biological association with age.

We also further explored the impact of manual segmentation on the final measures of WMH volume, thanks to the availability of repeated manual masks on data from WH dataset, Verio scanner. On 24 subjects we now have manual WMH segmentation performed by two raters (R1, R2) and repeated by the second rater a year later (R2a, R2b). The manual masks were used to train BIANCA and the automated WMH masks were generated using a leave-one-out approach. We then calculated between- and within-rater agreement on manually and automatically segmented masks using Dice index and compared the results with paired t-tests.

We found that the agreement between BIANCA outputs generated with masks from different raters is similar to the manual between-rater agreement. BIANCA outputs trained with masks from R2(a, b) were more consistent than within-rater agreement of manual masks (Figure 2). This suggests that if the examples provided to BIANCA are sufficiently in agreement, the automated tool improves the consistency of the output. Moreover it highlights the need to standardise the definition of WMH, especially if automated tools are planned to be used in multi-centre studies.

We included this analysis in an abstract submitted to the European Congress of Radiology (ECR) 2020.

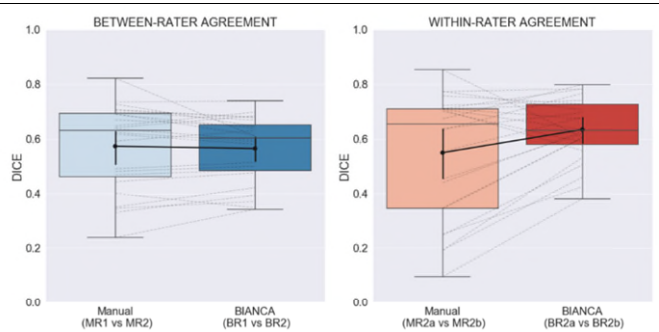

Figure 2. Results of (left) between- and (right) within-rater agreement from manual rating and BIANCA output (M=manual masks; B=BIANCA output; R=rater).

Regarding the second aim, we have now released the configuration file for the funpack (formerly ukbparse) tool (https://git.fmrib.ox.ac.uk/fsl/funpack/), which extracts the UKB non-imaging variables we selected to be relevant for our study on WMH and modifies them to be comparable to those in WHII (e.g. continuous to categorical variables, units of measure, matching tests for the same function)(https://issues.dpuk.org/eugeneduff/wmh_harmonisation). The configuration file is customizable to include more/different variables.

Using this tool, for the purpose of this study we selected from UKB only the subjects that had no missing data in any of the variables of interest, identifying ~2000 subjects entering into further analyses. We explored the relationship among the comparable non-imaging variables and worked on a model to predict the expected pattern of WMH for individuals from specific demographic, clinical, and behavioural bands.

For our initial model (based on Elastic Net) we identified 9 non-imaging variables that are predictive features of the amount of WMH in both WHII and UKB. They include demographic factors (age), clinical measures (blood pressure, BMI, medications), and behavioural (reaction time and digit-coding test). Using these 9 variables, we then used a Gaussian process regression model to obtain a prediction of WMH in terms of distribution rather than single points. This model gave good prediction both within dataset (when trained on 75% of the data and tested on a 25% holdout sample) and when then trained on one dataset and tested on the other.

We are now refining the model and we will integrate the WMH measures obtained with the optimised pipeline with the prediction model from non-imaging data. We aim to expand the number of variables utilised, and to assess for subtle features of the distribution of WMHs for specific categories of individuals. For example, WMHs may be particularly clinically valuable in younger individuals, where at-risk outliers can be identified.