ARTS Detailed Biomarker Kit Protocol

1. Executive Summary

Small vessel disease (SVD) encompasses a range of common pathologies (arteriolosclerosis, cerebral amyloid angiopathy (CAA), atherosclerosis, small and microscopic infarcts and bleeds, and white matter (WM) disease) that we and others have shown are associated with cognitive impairment, decline, and dementia. The Rush-IIT ARTS biomarker kit contains a classifier that has been trained to predict the presence of (moderate to severe) arteriolosclerosis based on MRI measures of white matter hyperintensities (WMH) and diffusion anisotropy, and demographic features. The higher the score, the higher the likelihood of arteriolosclerosis and therefore higher risk of cognitive decline. Sites are given the biomarker (computer software) and are asked to collect MRI data, and input MRI DICOM data into the biomarker. The biomarker will generate a score that is high when the likelihood that the person under study suffers from arteriolosclerosis is high, and the biomarker will generate a score that is low when the likelihood that the person under study suffers from arteriolosclerosis is low. The site will have to repeat the process on multiple participants and test the primary hypothesis that the score generated by the biomarker is associated with faster cognitive decline. The biomarker is in the category "Risk", as it aims to identify those at highest risk for vascular cognitive decline.

The required procedure and timeline for validation of the biomarker are as follows:

- Collect data for each participant (involves one MRI scan and two cognitive assessments) (Sections 4 & 5 below) (Years 3-5)

- Input MRI DICOM data into biomarker and obtain a score for each participant (involves running the biomarker software) (Section 7 below) (Years 3-5)

- Perform statistical analysis testing the association of biomarker score with cognitive decline (involves conventional statistical modeling) (Sections 7 & 8 below) (Year 5)

2. Brief Description of Biomarker Kit

The ARTS biomarker kit is a software package that takes as input MRI DICOM data and outputs a score representing the likelihood a human subject suffers from arteriolosclerosis. It requires the collection of MPRAGE, FLAIR and DTI data. It also requires the collection of simple demographic information, i.e. age and sex. Once this information has been entered into the software of the biomarker, the software outputs a score.

3. Participating Sites

UCSF/UCD/UCLA
UKY
UNM
Rush-IIT
CHARGE

4. Protocol for MRI acquisition

3D MPRAGE, 3D FLAIR and DTI sequences per MarkVCID MRI protocol.

5. Additional data collection required for analysis

The following demographic information needs to be collected for each participant: age at the time of the scan (years), and sex (binary; female=0, male=1). This information is used as input to the software of the biomarker.

Cognitive assessments at two timepoints are required for testing the performance of the biomarker, one at the time of the MRI scan and a second two years later. The cognitive assessments should follow the MarkVCID protocol. The cognitive assessments are not used as input to the biomarker, but only for the evaluation of the performance of the biomarker.

6. MRI data preparation

For each participant, MRI data must be organized in separate DICOM folders for each sequence. Therefore, each participant's folder must have the following subfolders:

- Folder containing only MPRAGE DICOM files
- Folder containing only FLAIR DICOM files
- Folder containing only DTI DICOM files
- Folder containing only DTI reverse polarity DICOM files

(*Since ARTS utilizes orientation information from the imaging data, it is recommended to use DICOM folders instead of NIFTI files as input. If, however, DICOM folders are not available, NIFTI files can be used as input, but the user must ensure that the orientation information and labels are correct*).

7. Step-by-step analytic plan

The following analytic plan should be tested in persons with vascular risk factors and no Alzheimer's (if relevant biomarkers are available). The analytic plan involves four steps: 1) calculate the change in cognition between two time points, 2) prepare the data that will be input to the biomarker, 3) run the biomarker, and 4) perform statistical analysis testing the association of biomarker score with change in cognition. These steps are described in detail below.

For each participant, the following information should already be available from Sections 5 & 6 above:
- Demographics
    - ✓ Age at time of MRI (years)
    - ✓ Sex (binary; female=0, male=1)
- Imaging
    - ✓ MPRAGE DICOM folder
    - ✓ FLAIR DICOM folder
    - ✓ DTI DICOM folder
    - ✓ DTI reverse polarity DICOM folder

In order to run ARTS, the Singularity software must first be installed. With Singularity software, imaging data are processed inside a virtual environment that has all of the necessary programs and files. This means that the user does not need to install any additional neuroimaging software. Instructions for installing Singularity can be found at:
- Linux: https://sylabs.io/guides/3.5/user-guide/quick_start.html#quick-installation-steps
- MacOS: Appendix A.1 of this document

Step 1 – Calculate the change in cognition between two time points

Two cognitive assessments are required per participant to test the primary hypothesis, one at the time of the MRI scan and a second 2 years later. The cognitive test of interest is the Trail Making Test Part B (raw score; total number of seconds to complete, 0-150; #8b on page 31 of MarkVCID Case Report Form for Clinical Data Elements, version 1.0, May 2018):

Based on the two cognitive assessments (from the two timepoints), calculate the change in Trails B score as follows: *measurement at second timepoint - measurement at first timepoint*, (a reduction in cognition over time should be a positive number, and an increase in cognition over time should be a negative number). Exclude participants that received the maximum score at the first cognitive assessment because it is not possible to measure a decline (i.e. it is not possible to get a higher score at the second timepoint).

Step 2 – Prepare the data that will be input to the biomarker

Information about the demographic and imaging data must be entered into a comma separated value (.csv) or excel spreadsheet (.xls or .xlsx) file which will be input to the biomarker script; this file is referred to as the input

file. In this input file, each row corresponds to a single participant, and multiple participants can be included in the same csv file (for batch processing). The information in each row must be as follows:

Column 1: Participant ID (can have letters, numbers, underscores, but no symbols; must be unique)
Column 2: Age at the time of MRI (in years)
Column 3: Sex (binary; 1 for male, 0 for female)
Column 4: Absolute path to MPRAGE DICOM folder
Column 5: Absolute path to FLAIR DICOM folder
Column 6: Absolute path to DTI DICOM folder
Column 7: Absolute path to DTI reverse polarity DICOM folder*
   * Although not advisable, ARTS can also run without DTI reverse polarity data. If such data is not available for a participant, leave this column blank for that participant, and DTI processing for that participant will be performed without EPI distortion correction.

(*Since ARTS utilizes orientation information from the imaging data, it is recommended to use DICOM folders instead of NIFTI files as input. If, however, DICOM folders are not available, NIFTI files can be used as input, but the user must ensure that the orientation information and labels are correct.*
If using NIFTI (.nii or .nii.gz) files as input, the information in each row of the input file must be as follows:
   Column 1: Participant ID (can have letters, numbers, underscores, no symbols; must be unique)
   Column 2: Age at the time of MRI (in years)
   Column 3: Sex (binary; 1 for male, 0 for female)
   Column 4: Absolute path to MPRAGE NIFTI file
   Column 5: Absolute path to FLAIR NIFTI file
   Column 6: Absolute path to DTI NIFTI file (this file must be in the same folder as bval and bvec, i.e. columns 7 & 8)
   Column 7: Absolute path to DTI bval file
   Column 8: Absolute path to DTI bvec file
   Column 9: Phase encoding direction for DTI. Options are: vertical or horizontal (case sensitive)
   Column 10: Absolute path to DTI reverse polarity NIFTI file* (this file must be in the same folder as bval and bvec,
      i.e. columns 11 & 12)
   Column 11: Absolute path to DTI reverse polarity bval file*
   Column 12: Absolute path to DTI reverse polarity bvec file*
      * If reverse polarity data is not available for a participant, leave these columns blank for that participant, and DTI processing for that participant will be performed without EPI distortion correction.

Step 3 – Run the biomarker

The script that runs ARTS can be found in the downloaded ARTS folder and can be executed as follows:

run_ARTS.sh <Path to input file> <Path to output directory>

Note that all arguments with a path are absolute paths. Running the script without any input arguments also displays information about the expected input arguments.

Once ARTS has finished running, the output directory will contain A) one subdirectory per participant (containing intermediate pre-processing files, final results, and files necessary for quality checks) and B) a "score_batch" csv file. A) The subdirectory of each participant includes an "analysis" folder which contains a file named "score.csv" that includes the biomarker score of that participant. B) The "score_batch" csv file in the output folder contains the Participant IDs and biomarker scores of all those participants that successfully completed ARTS. The "score_batch" csv file in the output folder has a suffix that indicates the date and time when the ARTS script was executed: [yyyymmdd]T[hhmmss]. For example, the filename "score_batch_20200216T144658.csv" indicates that ARTS was started on February 16th 2020, at 2:46:58 pm.

Important information:

• ARTS can be used on High Performance Clusters that have Singularity installed, to quickly process large amounts of data. To install Singularity on a Linux cluster, see this guide:

https://sylabs.io/guides/3.5/admin-guide/installation.html

- Inside the output directory of ARTS, there is a folder for each participant. These folders contain the intermediate files for ARTS. They also contain a QC folder that includes useful images for assessing quality of the preprocessing steps, e.g. the quality of FA maps, quality of registration of FA to a template, quality of brain segmentation, quality of WMH segmentation etc. <u>It is advisable to review these files for each participant</u>.

- Participants that have an error during ARTS processing will not have a biomarker score. In this case, the score_batch csv file will contain fewer participants than the input file.

- The runtime of ARTS depends on the number of participants in the input file, the resolution of the MRI data, and the number of processing cores used for processing. By default, ARTS will use all available cores on the local workstation. You can limit the number of cores used by modifying the N_cpu variable inside the run_ARTS.sh script. N_cpu must be an integer.

- Ensure that a minimum of 6 GB of RAM is available during ARTS processing to avoid out of memory errors.

FAQ for this step (Step 3):

Q: ARTS was interrupted in the middle of processing an input file. I know which participant was the last one to be completed. How should I restart ARTS?

A:

1. Create a new input file containing only the participants that did not complete processing.

2. Run ARTS with the new input file. For example,

    run_ARTS.sh <Path to <u>new</u> input file> <Path to output directory>

3. There will be two score_batch files at the end of this process. One containing the results generated before ARTS was interrupted, and one containing the results of the second run. The time-stamps in the suffix of the two score_batch filenames reveal which file corresponds to each run.

Q: The terminal printed that a participant was completed, but with errors. How do I find out what the error is?

A: The participant's log file contains the console output of ARTS. This log file is located in the participant's folder inside the output folder and is named "processing.log". If the text near the bottom of the log file has notifications that a process was killed, it is likely that this is an out of memory error. Run ARTS again with at least 6 GB of RAM available.

For other errors or if the error persists, please send the log file to: Konstantinos_Arfanakis@rush.edu

Q: For one of the participants, the WMH mask generated by ARTS is inaccurate. I have my own WMH mask that is more accurate. How can I use my WMH mask instead?

A: Firstly, small errors in the WMH mask are not a concern since they have insignificant effects on the biomarker score. In case the WMH mask generated automatically by ARTS is largely inaccurate, you can use your own (more accurate) WMH mask. Just check that your WMH mask has the same orientation information as the FLAIR image inside the participant's QC folder (located at <Output directory>/<Participant ID>/QC/WMH_processing/flair.nii.gz). This check is important, as different orientation information may cause your WMH mask to not overlap with voxels affected by WMH in the FLAIR image. For reference, orientation information in the FLAIR image is generated from dcm2niix.

To use your WMH mask with ARTS:

1. Copy your WMH mask named "WMH_mask_custom.nii.gz" into the participant's WMH folder located at "<Output directory>/<Participant ID>/user_input". The "WMH_mask_custom.nii.gz" must be a binary image and must be in the participant's FLAIR space.

2. Run the run_ARTS.sh command with the usual input arguments, but this time add a third input argument, "custom". For example,

   run_ARTS.sh <Path to input file> <Path to output directory> custom

3. All participants in the input file will be quickly reprocessed, since all the processing is already completed. Those with a custom WMH mask in their WMH folder will have an updated biomarker score in the new score_batch file.

Step 4 – Test the association of biomarker score with change in cognition, or level of cognition at baseline

After all previous steps have been completed for all participants, statistical analysis will be used to assess the association of biomarker score with faster cognitive decline (primary hypothesis) or level of cognition at baseline (secondary hypothesis).

More specifically, to test the primary hypothesis, we will use linear regression with two-year change in Trails B score as the dependent variable and biomarker score at baseline as the independent variable, controlling for years of education (we will not control for age and sex as they have already been used as features in the biomarker).

To test the secondary hypothesis, we will use linear regression with baseline Trails B score as the dependent variable and baseline biomarker score as the independent variable, controlling for years of education (we will not control for age and sex as they have already been used as features in the biomarker).

8. Power analysis

Here we provide information about the power to detect important associations of our biomarker kit with cognitive decline. We will calculate Spearman's correlation coefficient between the change in cognition and biomarker score. Power (2-sided) depends on the sample size, alpha, and the population correlation of the change in cognition with the biomarker score. The sample size at Rush will be 400 (ignoring available data); among sites that have agreed to test our biomarker kits, the smallest sample size is 120 (UKY), and the largest is >1000 (CHARGE). At 80% power, the Rush sample size can detect a population correlation of 0.14, the UKY sample size can detect a population correlation of 0.25, and the CHARGE consortium can detect a population correlation of 0.09. In summary, we anticipate strong power for important effect sizes.

9. Reproducibility

In the following, we provide information about testing the reproducibility of our biomarker kit a) across imaging sessions on the same MRI scanner, and b) across users of our biomarker kit working on the same data. We consider the intra-class correlation coefficient (ICC) as a suitable measure.

a) Test/retest reliability (same participant and same scanner, two different scans). We calculate the number of subjects necessary to reject the null hypothesis ($\rho_0 \leq 0.7$) for the ICC with effect sizes at $\rho_1=0.9$ (relatively big effect size), 0.8 (small effect size). In the figure, the

N vs Power by ρ1
ρ0=0.70000 K=2 Alpha=0.05 F Test

ρ1
● 0.80000
● 0.90000
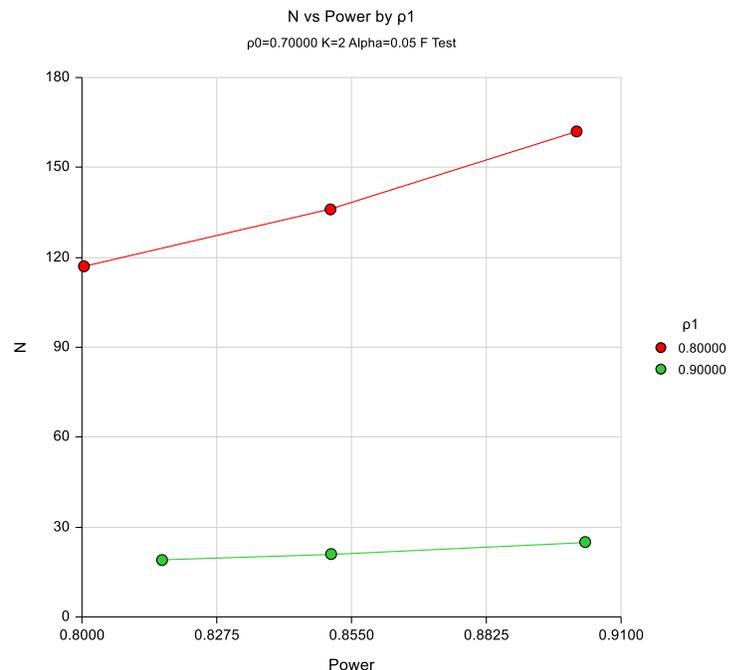
sample size was calculated at varying powers from 80% to 90% for each effect size. The green line is for $\rho_1=0.9$ and the red for $\rho_1=0.8$. For instance, N is 19 with effect size $\rho_1=0.9$ and 80% power, N=21 with $\rho_1=0.9$ and 85% power, and N=25 with $\rho_1=0.9$ and 90% power. To estimate overall test/retest reliability across centers, we evenly divide the total number of participants required by the number of centers. The number of subjects required per center (total number of centers is 4 to 5; see Section 3 above) with $\rho_1=0.9$ and 80% power is thus about 6. The number of subjects required for $\rho_1=0.8$ is about 120 with 80% power and is higher if higher power is desired. We however expect a high test/retest reliability ($\rho_1 \geq 0.9$), and thus propose the effect size at $\rho_1=0.9$ and the sample size accordingly. Therefore, 6 older adults will be asked to be scanned twice on the same scanner, with the two scans occurring approximately one week apart.

b) Inter-rater reliability (same participant, same scan, two different users of the biomarker kit). When a different user uses the biomarker kit on the same data from the same participant, assuming that the protocol of the biomarker kit is followed without errors, the only variation in output would originate from variations in WMH segmentation, registration, or diffusion tensor fitting. However, we do not expect these variations to be substantial when using the exact same raw images. We therefore anticipate the ICC to be higher than 0.93 for inter-rater reliability.

10. <u>Plan for longitudinal data collection analysis</u>

Although the proposed biomarker is based on a single MRI scan and generates a single score for the likelihood that arteriolosclerosis is present, it could also be used with longitudinal MRI, to generate multiple scores over time (one score per MRI scan). A participant with scores that increase more over time than the scores of another participant, would have a higher chance of cognitive decline and progression to mild cognitive impairment (MCI) or dementia.

11. <u>Plan for reporting outcomes</u>

Our biomarker software will be made available initially to the consortium for testing, and eventually to the community.

12. <u>Plan for sharing data, samples/images, protocols</u>

We are ready to support cross site testing of other sites' biomarkers via a large cohort of older persons that are followed annually with cognitive assessment, examinations, and interviews, biennial in-vivo MRI, baseline and annual blood draws, ex-vivo MRI, and pathology. These persons are participants of the Rush Memory and Aging Project (MAP) (R01AG17917) and Religious Orders Study (ROS) (P30AG010161), two longitudinal, epidemiologic clinical-pathologic cohort studies of aging that recruit non-demented individuals and have high follow-up rates and high autopsy rates.

APPENDIX

A.1) Installing Singularity on MacOS

These instructions demonstrate how to install Singularity on a Mac via a virtual machine (VM) that is running Ubuntu. Although installing "Singularity Desktop" is an easy option for macOS (through a conventional dmg package), "Singularity Desktop" cannot run ARTS due to its limited functionality. While most of the steps provided here are in the official Singularity guide (https://sylabs.io/guides/3.5/admin-guide/installation.html#mac), additional steps are provided for setting up the VM for ARTS.

1. Install Homebrew using the following command:

```
/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"
```

2. With Homebrew, install Virtualbox and Vagrant:

```
brew cask install virtualbox

brew cask install vagrant

brew cask install vagrant-manager
```

3. Set the current directory to a directory where the VM settings can be saved. Once in this directory, execute the following commands:

```
mkdir singularity-vm

cd singularity-vm

vagrant init sylabs/singularity-3.5-ubuntu-bionic64 --box-version 20191206.0.0

vagrant up
```

4. With a text editor, change the properties of the virtual machine by modifying the configuration file called "Vagrantfile" in the same directory (singularity-vm).

   First navigate to line 58 of the Vagrantfile and modify that section as follows:

```
config.vm.provider "virtualbox" do |vb|
   vb.customize ["modifyvm", :id, "--ioapic", "on"]
   vb.memory = "10000"
   vb.cpus = "2"
 end
```

```
# At least 6 GB of memory is required to run ARTS. Also set number of CPUs to your preference.
```
Then navigate to line 48 and modify that section of the Vagrantfile as follows:

```
config.vm.synced_folder "/Users/Downloads/ARTS", "/ARTS"
config.vm.synced_folder "/MarkVCID/data", "/input_data"

# This assumes that the downloaded and unzipped ARTS folder is located in /Users/Downloads, and
# the imaging data you will be working on are located in /MarkVCID/data.
```

5. Restart the current VM to apply the changes to the Vagrantfile.

```
vagrant reload
```

A.2) Running ARTS on MacOS

The following step assumes that the folder /MarkVCID/data is synced to the /input_data folder inside the VM based on settings in the Vagrantfile. Similarly, the folder /Users/Downloads/ARTS is synced to the /ARTS folder inside the VM.

1. The paths to the MRI DICOM data defined inside the input csv file should refer to the absolute paths inside the VM. For example, the absolute path /MarkVCID/data/participant1/T1W should be changed to /input_data/participant1/T1W.

2. Save the modified input csv file to a folder that is synced to the VM, like /MarkVCID/data.

3. Change the current directory to the vagrant folder (singularity-vm), and enter into the VM.

```
vagrant ssh
```

4. Run the ARTS command with the appropriate input csv file and output directory relative to the VM. The output directory should be in a folder that is synced.

```
/ARTS/run_ARTS.sh <Path to input csv file in VM> <Path to output directory in the VM>
```