MarkVCID Detailed Biomarker Kit Protocol:
MRI White Matter Hyperintensity Growth/Regression

Executive Summary

The growth of WMH over time is referred to as the penumbra, derived from subtracting baseline white matter hyperintensity voxels from year one imaging white matter hyperintensity voxels. The penumbra includes both expansion (positive growth change) as well as regression (negative growth change) which can be either summated as an estimate of longitudinal change in response to therapeutic agents or can be analyzed individually in regards to progression, stability, and/or of disease in response to therapeutic intervention. The utility of the penumbra allowing spatial localization of areas of growth and regression that may serve as outcome variables in future clinical trials of VCID is a novel feature of this protocol, that cannot be captured using cross-sectionally derived WMH volumes.

The WMH Growth/Regression Kit focuses solely on post-acquisition processing. The acquisition protocol for this Kit will be the MGH 3-D FLAIR sequence, although it also requires T1-MPRAGE for co-registration of longitudinal images.

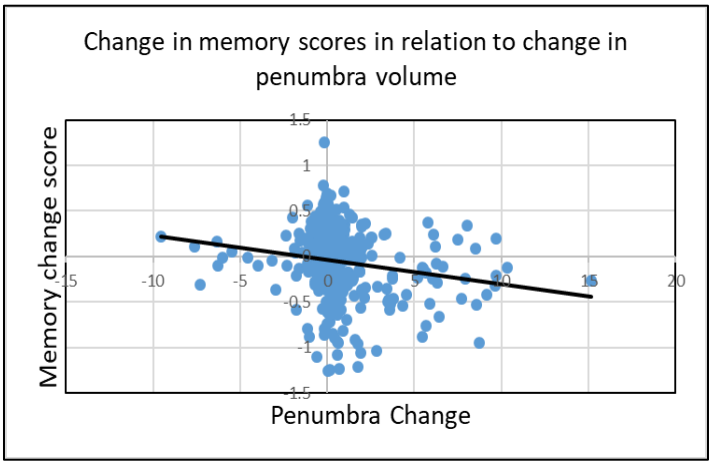

Figure 1. Change in penumbra volume corresponds to meaningful change in memory composite scores calculated using the ADNI protocol (p<0.002).

Our data combined with the longitudinal WMH volume change from ADNI subjects demonstrate that change in WMH volume is significantly associated with changes in memory measured over the same two-year time period (p<0.002, see figure 1). In a regression analysis adjusted for baseline age, sex, and years of education, change in WMH volume was independently and significantly associated with change in memory, such that increases in WMH volume were associated with worsening memory. Given these considerations, it is clear that change in penumbra volume is associated with risk for change in cognitive status over time reflecting clinically relevant outcomes.

Hypothesis #1: The penumbra will reliably, and across raters/technicians, measure longitudinal change in WMH volumes, allowing accurate measurement of penumbra stability, growth or regression.

Individual site validation will include 1) Within center inter-rater reliability assessments of the penumbra protocol using a set of standard longitudinal acquisitions collected by each site as part of the n=6 validation protocol set up by the coordinating center, and 2) Between site reliability assessment using the data acquired from this exercise to analyze cross-site validation of the penumbra protocol.

Expected outcomes: The intraclass correlation coefficient (ICC) among raters for readings of change in WMH volume (a continuous endpoint) will be calculated to test the null hypothesis that the ICC < rho_0 versus the alternative hypothesis that the ICC > rho-1.Here Rho_0 is an unacceptable agreement among raters on the WMH volume while rho_1 is a minimally acceptable agreement among raters. We expect rho_0 = 0.7 and rho_1 = 0.85.

Potential pitfalls: Lack of training and standard application of the protocol is the major challenge. This will be addressed by the UK team sending a trainer to each participating site to implement the protocol and standardize data analysis.

Hypothesis #2: Penumbra growth will be associated with declining cognitive change, whereas regression will be associated with improvements in cognitive outcomes and clinical status.

Expected outcomes: Differences in penumbra change between progressors and regressors will be reflected in composite memory and executive function score change over the one year study period, across disparate consortium populations with SVID.

Potential pitfalls: It is possible that different populations may have unique genetic risks and environmental exposures that may be unique to consortium sites. Such a finding would represent an important discovery that would have implications for the design of future trials of disease modifying therapeutics for VCID and may require modification of the protocol itself or restriction of populations in which the protocol proves valid.

Timeline:

Year 1: Acquisition of longitudinal scans on site cohorts, as well as training on penumbra protocol procedures, and within and between site assessment of penumbra measurement agreement using scans acquired by participating sites as part of the internal validation as mandated by the central (MGH) coordinating center (n=6 per site with repeat scans).

Year 2: Continue to acquire longitudinal scans on site cohorts. Completion of reliability, repeatability, and reproducibility analysis on scan data (n=6) mandated by the central coordinating hub (MGH).

Year 3: Final collection of longitudinal scans on site cohorts, individual site validation analyses, and further central analysis of pooled cohort penumbra values in relation to cognitive change over the study MRI acquisition period.

Brief Description of Biomarker Kit

UKY Kit C- Penumbra Protocol is a post-acquisition, semi-automated image co-registration and processing protocol developed to identify annual longitudinal change in T2 hyperintensity signals in the subcortical white matter.

This biomarker is designed to examine longitudinal change in WMH progression, stability or regression of WMH lesions that may provide insights into underlying mechanisms of disease and potential effects of future therapeutic agents to treat VCID.

Usefulness of this biomarker requires longitudinal imaging at annual time points, although adaptability for both shorter and more extensive between-imaging time scales should be amenable to analysis.

Software/program requirements: Post-acquisition processing systems requires MIPAV, FSL, AFNI, and FreeSurfer (or at least the MINC toolbox from MNI) installed. The FSL, AFNI and FreeSurfer software packages only exist for Mac and linux operating systems (see below under hardware requirements). A separate toolbox license is required for MATLAB in order to fit the image intensity histogram.

Hardware requirements: The protocol requires Mac or Linux platforms.

Site personnel needs: In order to perform inter-rater reliability analyses, at least two independent analysts at each site are necessary for validation studies in UH3 year 3, and at least six subjects will have repeated baseline scans within 1 week of their initial baseline scans at each site. A single qualified rater is needed for the construct validity analysis after each site acquires longitudinal MRI on MarkVCID prospective subjects.

Between rater and site validation of methodology: To perform an initial site-dependent interrater analyses and cross site interrater analyses each site will acquire 6 baseline scans repeated one week apart.

Longitudinal image acquisition: Additional validation examining the relationship between clinical outcomes and penumbra change will require each site to independently acquire images separated by a one-year timeframe. The window for this timeframe should include +/- 2 months which will provide an appropriate interval to allow uniform analysis of the data.

Additional clinical data required: No additional clinical data required is expected within the core data set of the MarkVCID consortium collection. Other data elements required that have been included in the recommendations for consortium participation as core data elements are included in appendix A.

Participating sites

All of the consortium sites have agreed to cross-site validate the University of Kentucky Kit C as part of the UH3 transition report: JHU, CHARGE, Rush, UCSF, UNM, USC, and the protocol development site UKY.

Protocol for MRI acquisition

No special acquisition sequences are need for this kit as it relies on post-acquisition processing only. Standard T1-MPRAGE and the MGH 3-D FLAIR sequences will be used. These have been specified as core imaging sequences for the consortium by the Imaging subcommittee and the Executive Committee.

Additional data collection required for analysis

Clinical variables: Clinical outcomes that may show one-year longitudinal change include the MOCA, CDR sum of boxes, and diagnosis. Specific cognitive tests we would like to validate against the total, growth and regression penumbras include the supraspan word list recall (CVLT(-SF), HVLT, SEVLT) . Demographic variables required for adjusting analyses include age, education, gender, race/ethnicity. These variables are all included in the proposed minimum data set for discovery as part of the MarkVCID consortium.

Protocol for MRI analysis

Included as appendix B.

Step-by-step analytic plan

Procedures:

1) Begin longitudinal enrollment and annual acquisition of scans and clinical data at each consortium site.
2) Acquire baseline and one week repeat 3D-FLAIR scans on (n=6) subjects.
3) Finalize image and clinical data acquisition in each consortium site to allow baseline and one-year follow-up scans/data.
4) Send UK trainer to each engaged site to review and implement the penumbra protocol procedures.
5) Process site-specific acquired scans using the penumbra protocol to complete the reliability, repeatability and reproducibility analysis based on (n=6) subjects with baseline and one-week scans and one-year follow-up scans.
6) Perform clinical outcome association studies on site specific data sets.
7) Analyze aggregate site-specific data to allow further refinement of the protocol and potential subject I/E criteria for validity (only if any site fails to validate).
8) Exploratory analyses to be performed to examine a broader array of cognitive and clinical risk factors for progression of the penumbra and clinical consequences of penumbra change over time.

Inclusion/exclusion criteria:

As this protocol is designed to evaluate longitudinal change across the spectrum of progression of SVID, there are no absolute restrictions on inclusion of subjects for individual site recruitment into this protocol.

We do however require a full spectrum of baseline white matter hyperintensity abnormalities in order to validate the protocol across all stages of SVID. Subjects may be included with normal cognition, MCI, or dementia provided that the distribution of baseline images provided represents the full spectrum of SVID pathology from low levels (Fazeka grade 1), to moderate levels (Fazeka grade 2), to severe levels (Fazeka grade 3). These should be approximately equal in distribution.

Inclusion of subjects with no white matter hyperintensities at baseline are inappropriate for the current study.

Sites should ensure that a wide range of representation for cerebrovascular risk factors as well as cognitive performance and diagnoses are represented in the sample use for this analysis although there are no absolute requirements in this regard for any participating site.

Need for control populations: as this protocol involves evaluation of within-subject longitudinal change, there is no need for inclusion of SVID-negative subjects at baseline in the protocol.

Statistical analysis plans:

1) Inter-rater and inter-site validation studies will be performed to determine biomarker agreement using intra-class correlation coefficient techniques.
    a. Penumbra volume calculations: Penumbra volumes are calculated from the methods described above and in the accompanying appendices. Within subject penumbra volumes include both positive (growth or progression) and negative (shrinkage or regression) voxels. For the primary analysis, these measures should be summated to provide an overall change volume, although we note that the ability to examine progression and regression individually is an inherent strength of the protocol that may provide unique data on interventions designed to either slow or stop growth (i.e. modulation of CVD risks…) vs. repair or augment regression of WMH (i.e. anti-inflammatory or pro-angiogenic agents…).
    b. Inter-rater intraclass correlations between two raters at each site will allow within site validation of the protocol. The statistical analysis required should be well known to individual site statisticians, but for sites requiring assistance, the UK statisticians, Dr. Kryscio and Abner will be available to assist. The measure of agreement is based on the intraclass correlation coefficient (ICC) among raters for readings of change in WMH volume (a continuous endpoint). We will test the null hypothesis that the ICC < rho_0 versus the alternative hypothesis that the ICC > rho-1. We have chosen Rho_0 = 0.7 as an unacceptable agreement among raters on the WMH volume while rho_1= 0.85 will represent a minimally acceptable agreement among raters. All sites should adhere to these agreement rules, and should consider $p \leq 0.05$ as sufficient for validation of the mechanics of the protocol.
    c. Inter-site intraclass correlations using the same images (n=40, provided by UK and distributed through the central coordinating site at MGH) and dataset derived from the intra-site correlation experiment will allow this analysis at no additional effort from individual participating coordinating sites. Engaged sites should be prepared to share their intra-site intraclass validation raw data with the UK biostatisticians (as well as the statisticians at all other cooperating sites and the central coordinating site) for this inter-site validation study. UK intends to perform the analysis and share authorship with other involved sites, but we also feel strongly that the raw data should be shared across all participating sites.
2) Construct validity analysis will be performed at the individual site level, but again raw data will be shared between all sites to allow more refined data analysis of construct validity for subpopulations of subjects that may require the assembly of a larger dataset than that collected at any individual site.
    a. Penumbra volume composite scores should be calculated as described above, including a summated score of all voxels. Subjects with summated penumbra scores < 0 will be classified as *regressors*. Subjects with summated penumbra scores > 0 will be classified as *progressors*. Any subject with complete stability (penumbra summated score =0 will be excluded from the analysis at each site. We expect the number of subjects with stable penumbra summated volumes to be < 0.5%, so no adjustment in the subject number described in the power analysis is necessary for the intra-site sample size.
    b. Memory (supra-span word list delayed free recall) scores: Unfortunately, no existing validated memory composite score for the MarkVCID cognitive measures has been developed. While the ADNI data composite memory scores utilize both supraspan word list and paragraph recall scores, the memory composite for which our power analysis is based also includes memory measures from the ADAS-Cog which will not be

collected in MarkVCID. As such, CVLT/ delayed recall z-scores will be used as distinct outcomes rather than as a composite measure in the initial intra-site validation analyses.

c. Linear regression will be used to analyze the data, with change in memory (supra-span wordlist delayed free recall) analyzed as the primary validation outcome variable using longitudinal one-year penumbra change as the study variable. All models will be adjusted for covariates age, sex, and education; power calculations assume up to 3 additional covariates may be included in the models (additional covariates for analyses should be utilized uniformly across intra-site analyses. Site specific cohort attributes may need to be taken into account if they represent significant confounds and so the statistical analysis plan remains open to adjust for such variables as needed). This sample size will allow individual sites to validate the penumbra protocol without pre-selecting subjects in order to relate penumbra change with meaningful clinical outcomes on memory, processing speed, and executive function measures. Intra-site construct validity will be considered valid for p-values ≤ 0.05 for the association penumbra change with the primary memory outcome measure in this exploratory analysis. The UK statistical team is readily available to help site statisticians conduct the analysis in a uniform manner as needed.

Sample size calculation (individual site level)

1) Intra- & inter-site reliability study: The primary goal of the study is to assure that a given image is read in the same way by different raters within sites or at different sites in the VCID project. The measure of agreement is based on the intraclass correlation coefficient (ICC) among raters for readings of change in WMH volume (a continuous endpoint). We will test the null hypothesis that the ICC < rho_0 versus the alternative hypothesis that the ICC > rho-1. We have chosen Rho_0 = 0.7 as an unacceptable agreement among raters on the WMH volume while rho_1= 0.85 will represent a minimally acceptable agreement among raters. Setting the level of significance at p ≤ 0.05, and desired power at 90%, the validation study will require approximately n=37 pairs of longitudinal scans which is achievable with the 7 consortium sites engaged (anticipated n=6 per site for a total of 42 scans collected for reliability, repeatability and reproducibility analyses mandated by the coordinating center.
2) Construct validity study: While single center validation of any measurable biomarker for use in clinical trials can be limited by the number of subjects available, we have been tasked by this through our funding mechanism. Fortunately, penumbra change over time provides a natural experiment with which to conduct this exercise. As almost all subjects experience some change in penumbra over time, there is likely to be a natural distribution of 35% of subjects experiencing regression and 65% of subjects experiencing progression of the penumbra. Our power analysis and sample size calculation is based on this distribution which has been replicated in both our own internal data set as well as in the ADNI cohort. In order to detect a clinically relevant change in memory function (based on ADNI memory composite scores; Crane et al., 2012), approximately 126 subjects at each site will require longitudinal imaging measures. This will provide 82% power to detect the association between penumbra change and memory change, assuming a partial correlation between change in supraspan word list delayed free recall z-scores (the use of z-scores will allow standardization of memory measures across sites in further exploratory inter-site analyses) and change in penumbra of -0.25. Linear regression will be used to analyze the data, with change in memory z-scores as the outcome variable and penumbra change as the study variable. All models will be adjusted for covariates age, sex, and education; power calculations assume up to 3 additional covariates may be included in the models. This sample size will allow individual sites to validate the penumbra protocol without pre-selecting subjects in order to relate penumbra change with meaningful clinical outcomes on a memory composite.
3) While we note that our proposed cohort size of 120 individuals falls below the number needed to validate our own protocol, we accept the fact that we will have to modify our budget to allow for the additional subjects required for protocol validation as part of our obligation to the consortium. We will accomplish this through leveraging internal funds and collection of images through ancillary studies as needed in order to accomplish this goal. We hope that are collaborating sites will choose to do the same if their proposed cohort size is less than that required to complete the construct validation analysis.

Plan for longitudinal data collection analysis

Longitudinal data collection will include core acquisition sequences including T1-MPRAGE and 3-D FLAIR sequences at one-year intervals as well as standard clinical data set collection. Supraspan word list delayed recall measures will also be required in order to complete intra-site validation of this biomarker.

Plan for reporting outcomes

Over the 3rd year of the UH3 funding mechanism, we should have accumulated longitudinal scans from all consortium sites that have been analyzed with this methodology. This should allow the development of international abstracts and presentations as well as completion of manuscripts representing the study in regards to both the generalizability of the protocol in disparate populations as well as the validity of the protocol in relation to the primary memory outcome measure over time in the study population. Findings from this study are under no restriction in regards to data sharing and so may be posted online by the central coordinating center at any time that they become available.

Plan for sharing data, samples/images, protocols

1) This proposal requires initial sharing of n=6 reliability, repeatability and reproducibility scans mandated by the coordinating at baseline as well as our penumbra protocol with all consortium sites. There are no restrictions on this sharing within the consortium.
2) The multi-site validation proposed will include sharing of data and images from all consortium sites engaged which is part of the funding agreement coordinated by the central coordinating consortium site.

APPENDIX A. Clinical variables required for protocol validation

1. Age
2. Education
3. Diagnosis
4. Gender
5. Race/ethnicity
6. ApoE status (as available)
7. MOCA
8. CDR global score
9. CDR sum of boxes
10. Trailmaking A
11. Trailmaking B
12. Craft Story delayed recall
13. Supraspan word list delayed recall (CVLT, CVLT-SF, HVLT, SEVLT word list…etc) converted to z-score for population
14. History of hypertension, hyperlipidemia, diabetes, tobacco use, stroke/TIA, MI, CHF
15. Blood pressure at baseline and one-year followup
16. Lipid panel measurements at baseline and one-year followup
17. Fasting glucose and HgbA1c at baseline and one-year followup
18. Active vs. remote tobacco and exposure in pack/years
19. Remote vs. stroke/TIA in last year
20. Remote vs. MI in last year
21. Remote vs. CHF in last year