
# Visit 5/NCS ANALYSIS MANUAL

September 2015

## Table of Contents

## List of Tables

# 1. INTRODUCTION

## *1.1 Purpose*

The purpose of this working document is to provide guidance to analysts currently working on ARIC cohort manuscripts involving either within visit associations or longitudinal data analysis of trends in participant outcomes across visits. The collaborative document is intended to include information about the visit data and NCS data available to ARIC investigators and provide recommendations for addressing the limitations in ARIC-based analyses. The manual will include drop-in language to be used in methods sections for manuscripts. These recommendations can provide a means to methodological consistency across ARIC manuscripts.

## *1.2. Available data*

### *1.2.1. Visit 5/NCS*

The diagram below shows the ARIC data collection from 1987 to present (Figure 1). In addition to the data collected at the visits, the participants are contacted via telephone semi-annually and asked a short battery of data collection forms (AFU). The cohort is also tracked for cardiac-related events and hospitalizations (ARIC cohort surveillance). The AFU and surveillance data are necessary for Investigators to track the cohort in the long gap between visits 4 and 5. Recommendations will be provided when these datasets may be helpful for imputation.

Comprehensive descriptions of datasets collected for the visits, NCS, AFU, and surveillance are found on the ARIC web site under the Cohort tab on the main menu (http://www2.cscc.unc.edu/aric/).

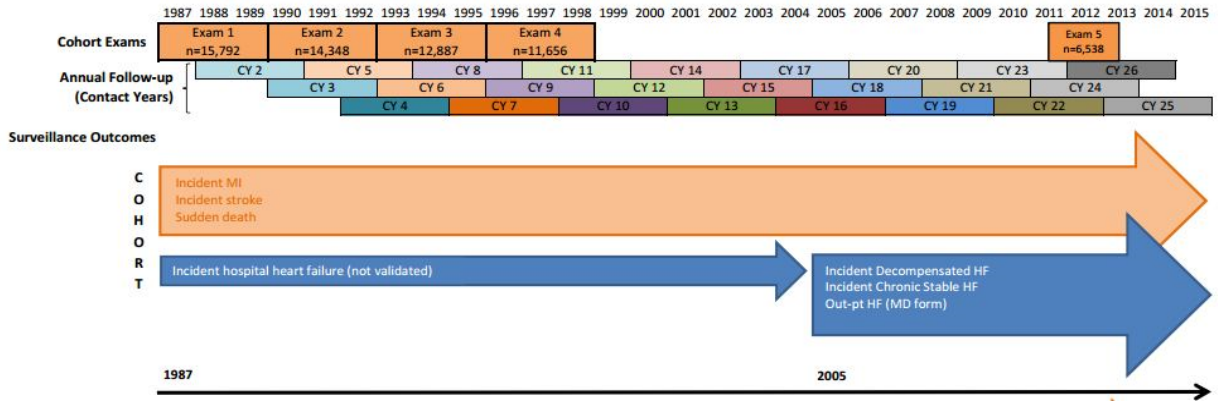



**Figure 1: ARIC data collection from 1987 to present**

### *1.2.2. Datasets*

The ARIC CC has been distributing Visit 5 (NCS stage 1) and NCS stages 2/3 data since January 2014. All data collection is now complete... All supporting dataset documentation, codebooks and derived data dictionaries have been posted to the study website.

There are three types of datasets:

1. Form datasets: Data entered directly into the DMS by the field center or a reading center and retrieved into a corresponding SAS dataset. Two datasets distributed with the form data include the notelog and field status datasets. These datasets include additional information beyond the response, such as comments or the reasons for non-response. The comments are added by the data collector for as needed for each question. The data collector's notes are particularly helpful when trying to determine the reasons for suspicious values contained in the dataset. Some of the visit case report forms were collected during repeat visits. The repeat visit data may be linked to the participant using the RVF dataset.
2. Transfer datasets: Data transferred to the ARIC CC from either the field center (ABI/PWV for example) or a reading center/central laboratory (ECHO for example) and retrieved into a corresponding SAS dataset. Some of the lab datasets have corresponding repeat visit data. The repeat visit data may be linked to the participant using the RVF dataset.

3. Derived datasets: Variables that have undergone transformations, for example calibrated analytes, or new variables created from data found in either the DMS datasets or transfer datasets.

Form Data Sets

Each DMS dataset has a corresponding 'paper' form with instructions [QxQs]. These forms and QxQs are located on the secure study website (http://www2.cscc.unc.edu/aric/). Select Cohort->Visits 1-5 (forms, databooks, dictionaries, manuals). Each form is assigned a 3-character code as its 'name' (for example, the Informed Consent Tracking form is the 'ICT'). To accompany the datasets, each ARIC form has a codebook that contains descriptive information, labels and distributions, of each variable on each Visit 5 data set. The datasets are listed in alphabetical order by form code (3-character code that denotes the form name).

**Table 1: Form Data Sets**

| CODE | FORM |
|---|---|
| V5/NCS Administrative Forms | |
| MAE | Minor Adverse Event Form |
| OIT | Checklist for Observation of Interviewing Technique |
| PSA | Participant Safety Screening |
| RVF | Repeat Visit Form |
| RTS | Recruitment Tracking and Scheduling Form |
| SAE | Serious Adverse Event Form |
| TIC | Telephone Interview for Cognitive Status |
| V5/NCS Stage I Forms | |
| AAT | AAA Technologist Data Collection Form |
| ALC | Smoking and Alcohol Use Form |
| ANT | Anthropometry Form |
| ANT_QC | Anthropometry Form (Quality Control) |
| AQC | Access and Quality of Care Form |
| BIO | Biospecimen Collection Form |
| CES | CES Depression Form |
| DLC | DLCO Management Form |
| MME | Mini-Mental Exam |
| MSR | Medication Survey Form |
| NCS | Stage I Neurocognitive Battery Summary Form |
| NSS | NCS Stage 2/3 Selection |
| PAC | Physical Activity Form |
| PEX | Physical Function Tests |
| PHX | Personal History Form |

| | |
|---|---|
| PWV | Pulse Wave Velocity/ABI Form |
| RSE | Respiratory Questionnaire |
| SBP | Sitting Blood Pressure Form |
| SFE | SF-12 Health Survey |
| SMF | Subjective Memory Form |
| V5/NCS Stage 2 Forms | |
| CDI | Clinical Dementia Rating Informant Interview |
| CDP | Clinical Dementia Rating Subject Interview |
| CDS | Clinical Dementia Rating Summary |
| HIS | Hachinski Ischemic Scale |
| NFH | Neurological Family History |
| NHX | Neurological History |
| NPI | Neuropsychiatric Interview Questionnaire |
| NTR | Neurocognitive Test Repeat |
| PNE | Physical and Neurological Exam-Other |
| UPR | Unified Parkinson's Rating Scale |
| REX | Retinal Exam Form |
| Central Agency Forms | |
| AAA | Over-Reader Data Collection Form |
| DLR | DLCO Reading Center Form |
| ECA | Echocardiograph Alerts Notification Form |
| RPN | Retinal Pathology Notification Form |

Transfer Data Sets

The transfer datasets are data from labs and reading centers. There are no corresponding 'paper' forms. However, the codebook provides distributions of the variables found in those datasets.

**Table 2: Transfer Data Sets**

| Code | File |
|---|---|
| Central Lab Data Files | |
| CBC | Hemogram Data File |
| CHM | Blood Chemistry Date File |
| CYSC | Cystatin C Data File |
| LIP | Lipid Data File |
| THYR | TSH Data File |
| Central Agency Files | |
| ABI | Pulse Wave Velocity and Ankle-Brachial Index Data File |
| AMY | Beta-Amyloid Data File – COMPLETED 12/2014 |
| DCM | DLCO Data File |
| ECG | Electrocardiogram Data File |
| ECH | Echocardiogram Data File |

| | |
|---|---|
| ECH2 | Additional Echocardiogram Variables |
| LSR | Lung Sounds Review |
| PULB | Pulmonary Data File (Post-Bronchodilator) |
| PULP | Pulmonary Data File (Pre-Bronchodilator) |
| RETC0 | Retinal Comments- Right Eye |
| RETC1 | Retinal Comments- Left Eye |
| RETF0 | Retinal Final Data- Right Eye |
| RETF1 | Retinal Final Data- Left Eye |
| RETP0 | Retinal Preliminary Data- Right Eye |
| RETP1 | Retinal Preliminary Data- Left Eye |
| RETV0 | Retinal Vessel Measurement Data- Right Eye |
| RETV1 | Retinal Vessel Measurement Data- Left Eye |
| V1ECG | V5 ECG Methods Applied to V1 ECG |
| V2ECG | V5 ECG Methods Applied to V2 ECG |
| V3ECG | V5 ECG Methods Applied to V3 ECG |
| V4ECG | V5 ECG Methods Applied to V4 ECG |
| Central Agency Files - MRI | |
| MRB | Brain Mask Total Volume |
| MRD | MRI DTI Assessment |
| MRF | MRI Flair - infarct |
| MRM | MRI Microhemorrhage |
| MRS | MRI FreeSurfer |
| MRW | MRI White Matter Hyperintensity |
| BMRB | Stage 3 MRI Reading Methods Applied to Brain MRI AS1999.01 Brain Mask Volume |
| BMRF | Stage 3 MRI Reading Methods Applied to Brain MRI AS1999.01 Infarct |
| BMRS | Stage 3 MRI Reading Methods Applied to Brain MRI AS1999.01 Freesurfer |
| BMRW | Stage 3 MRI Reading Methods Applied to Brain MRI AS1999.01 White Matter Hyperintensity |
| WMRL | Wasserman Substudy Vascular MRI – Qualitative Assessments |
| WMRN | Wasserman Substudy Vascular MRI – Quantitative Assessments |

Derived Data Sets

The derived datasets are a collaborative project for the CC and ARIC working groups (WG). The WG's have supplied specifications for analysis variables and have reviewed and validated the CC's calculations of the newly created variables. The naming convention for all visit derived datasets is to include xxx5v, which xxxx is the name of the dataset, 5 indicates Visit 5 and v indicates the version number. Version 1 of the derived dataset for Visit 5 is DERIVE51.sas7bdat. Only those participants who were defined as completing Stage 1 will have a record in DERIVE51 (N=6538 records). Stage 1 complete status is defined as having either a weight or blood pressure measurement present. Version 1 of the NCS derived dataset is DERIVE_NCS51.sas7bdat. Similarly, only those participants who have completed either Stage 2

or Stage 3 will have a record in DERIVE_NCS51 (N=3123). Stage 2 complete status is defined as being selected to Stage 2 and having the PNE form present in the database. Stage 3 complete status is defined as having any MRI data included in any of the MRI datasets. A third derived dataset has been created by the CC called STATUS51.sas7bdat. STATUS51 (N=15,792 records) includes status indicators for all members of the ARIC cohort at the conclusion of V5/NCS data collection as well as data collected at the time of Visit 5 on participants who did not attend the visit. The dataset is useful for analyses accounting for attrition.

The CC is also distributing a longitudinal analyte dataset (V1_V5_ANALYTES.sas7bdat) containing data from Visits 1-5 and some ancillary studies. A derived dataset containing the neurocognitive battery z-scores collected from V2 through V5/NCS are being compiled in a dataset called V2_V5_CNF.sas7bdat. The documentation for the derived datasets is included on the study website under Cohort->Visits 1-5 (forms, databooks, dictionaries, manuals). These documents will be updated regularly as derived datasets undergo version changes as variables are added and/or reviewed.

**Table 3: Derived Data Sets**

| SAS Dataset Name | Content |
|---|---|
| DERIVE51 (6538 records) | Historically-defined derived variables and newly-defined derived variables using Visit 5 and NCS stage 1 forms, labs, reading centers, some AFU, and some surveillance data. |
| DERIVE_NCS51 (3123 records) | Derived variables from NCS stages 2 and 3. Some variables will be incomplete due to the MCI review process. |
| STATUS51 (15,792 records) | New derived dataset that gives information about the participant's study participation from visit 1 through visit 5 and NCS as well as status at the time of the visit. The dataset is useful for analyses accounting for attrition and includes variables that describe participants who did not attend Visit 5. |
| V1_V5_Analytes (15,792 records) | Combines calibrated labs values for serum creatinine, uric acid, cholesterol, HDL, LDL, triglycerides, glucose, C reactive protein, BNP, troponin, cystatin-C, and eGFR from visits 1 through 5 as well as several ancillary studies. Dataset will be useful for longitudinal analyses. |
| V2_V5_CNF (1 record per patient visit) | Combines DWRT, DSST, WFT raw scores and Z-scores standardized for visit 2 population from visits 2, 3, 4, 5, and ancillary studies, Carotid MRI and Brain MRI. The dataset's purpose is to facilitate NCS longitudinal analyses. |

# 2. GENERAL METHODS SECTIONS FOR PAPERS

## *2.1. Study design*

### *2.1.1. Visit 5*

The ARIC Study is a prospective cohort study investigating the etiology of atherosclerotic disease in a middle-aged, predominantly biracial population. A detailed study design description has been published (1). The cohort was selected by probability sampling in four U.S. communities, Forsyth County, NC; Jackson, MS; northwestern suburbs of Minneapolis, MN; and Washington County, MD. In Jackson only African Americans were recruited whereas in the other centers the racial composition of the cohort reflected that of the community. In 1987-1989, 15,792 men and women aged 45-64 attended the baseline clinic examination (visit 1). There were three subsequent visits at approximately three-year intervals (visit 2 in 1990-1992; visit 3 in 1993-1995; visit 4 in 1996-1998) followed by visit 5 in 2011-2013. Participants have been contacted annually (semi-annually beginning in 2012) since baseline, to obtain information about hospitalizations and for additional data collection.
The ARIC Study protocol was approved by the institutional review board of each participating center and informed consent was obtained from participants at each study visit.

References

(1) The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. American Journal of Epidemiology. 1989; **129**(4): 687-702.

## Table 4: Characteristics of the Cohort at Visit 5

| | | | Age | | BMI(kg/m$^2$) | | SBP(mmHg) | | DBP(mmHg) | | Total Cholesterol (mmol/L) | | LDL (mmol/L) | | HDL (mmol/L) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Description*** | ***Category*** | ***N*** | ***Mean*** | ***Std*** | ***Mean*** | ***Std*** | ***Mean*** | ***Std*** | ***Mean*** | ***Std*** | ***Mean*** | ***Std*** | ***Mean*** | ***Std*** | ***Mean*** | ***Std*** |
| **Overall** | | 6538 | 75.8 | 5.3 | 28.7 | 5.8 | 130.7 | 18.7 | 66.3 | 10.8 | 4.7 | 1.1 | 2.7 | 0.9 | 1.3 | 0.4 |
| **Field Center** | **F** | 1433 | 75.7 | 5.4 | 27.6 | 5.1 | 130.7 | 18.2 | 67.6 | 10.5 | 4.7 | 1.1 | 2.7 | 0.9 | 1.3 | 0.4 |
| | **J** | 1416 | 75.1 | 5.2 | 30.6 | 7.0 | 135.8 | 20.5 | 69.8 | 11.1 | 4.8 | 1.0 | 2.8 | 0.9 | 1.4 | 0.4 |
| | **M** | 1915 | 75.9 | 5.1 | 28.1 | 5.1 | 128.7 | 17.4 | 65.6 | 10.5 | 4.7 | 1.1 | 2.7 | 0.9 | 1.4 | 0.4 |
| | **W** | 1774 | 76.4 | 5.3 | 28.9 | 5.6 | 128.9 | 18.2 | 63.2 | 10.3 | 4.6 | 1.1 | 2.6 | 0.9 | 1.3 | 0.3 |
| **Sex** | **Female** | 3845 | 75.7 | 5.3 | 28.9 | 6.3 | 132.2 | 19.4 | 66.3 | 10.6 | 5.0 | 1.1 | 2.8 | 0.9 | 1.5 | 0.4 |
| | **Male** | 2693 | 76.0 | 5.2 | 28.6 | 4.9 | 128.6 | 17.4 | 66.3 | 11.1 | 4.3 | 1.0 | 2.5 | 0.9 | 1.2 | 0.3 |
| **Race** | **White** | 4977 | 76.0 | 5.3 | 28.2 | 5.2 | 129.1 | 17.8 | 65.2 | 10.5 | 4.7 | 1.1 | 2.7 | 0.9 | 1.3 | 0.4 |
| | **Non-White** | 1561 | 75.1 | 5.2 | 30.5 | 6.9 | 135.7 | 20.5 | 69.7 | 11.1 | 4.8 | 1.0 | 2.8 | 0.9 | 1.4 | 0.4 |

uc642905 by UCCHTC on 26SEP14

## Table 5: Comparison of Risk Factors between Visit 4 and Visit 5

| *Description* | *Category* | *Visit 4* | *V4 with Complete V5* | *Visit 5* |
|---|---|---|---|---|
| **Total Cholesterol (mmol/L)** | **Women** | 5.4 | 5.4 | 5.0 |
| | **Men** | 5.0 | 5.0 | 4.3 |
| **Diabetes (Lower cutpoint 126 mg/dL) (%)** | **Women** | 15.6 | 10.8 | 27.6 |
| | **Men** | 18.0 | 13.1 | 30.5 |
| **Hypertension (Definition 5) (%)** | **Women** | 49.2 | 42.5 | 74.7 |
| | **Men** | 45.8 | 38.5 | 72.5 |
| **Systolic BP (mmHg)** | **Women** | 128.2 | 125.3 | 132.2 |
| | **Men** | 127.2 | 124.4 | 128.6 |
| **Current Smoker (%)** | **Women** | 13.9 | 10.5 | 5.4 |
| | **Men** | 15.8 | 12.2 | 5.8 |

uc642905 by UCCHTC on 26SEP14

*2.1.2. NCS*

The ARIC Neurocognitive Study (ARIC NCS) is integrated operationally with the 5th ARIC examination, in 2011-13, of survivors of the 15,792 middle aged participants first seen in 1987-89, and it evaluates their cognitive performance. Its overall objectives are to determine the prevalence of cognitive impairments and the associations of mid-life vascular risk factors and markers with later-life cognitive impairments and cognitive change. Genetic markers and cerebral imaging features are also studied. Participants are invited for exams in clinic or in their homes or long-term care (LTC) facilities. Those who cannot be examined in person are assessed by telephone. Additional information about participant's cognitive and functional status is sought from informants when necessary. Some participants are invited for further evaluation and brain MR imaging. An expert committee reviews data and classifies dementia, MCI and their subtypes.

ARIC Cohort Visit 5 participants were selected to Stages 2/3 under a stratified random sampling plan designed to oversample for participants with evidence of cognitive impairment ("atypical"). Details of the selection process and the definition of atypical are provided in Manual 17. In brief, 100% of atypical participants (low MMSE score, or low Z-score on any of 5 cognitive domains and definite cognitive decline) as well as 100% of ARIC Brain MRI participants were invited to Stage 2. A random sample of the remaining participants was also invited. Sampling fractions varied by field center and age group (<80, ≥80 years) and were selected to achieve a sample size of 2000 Stage 3 participants. The final sampling fractions are provided below:

**Table 6: Sampling Fractions for Stage 3 Participants**

| Center | Age Group | |
|---|---|---|
| | **< 80** | **≥80** |
| **Forsyth** | 0.18 | 0.36 |
| **Jackson** | 0.65 | 1.0 |
| **Minneapolis** | 0.23 | 0.46 |
| **Washington** | 0.39 | 0.78 |

Since participants in the resulting sample are not equally representative of individuals participating in ARIC V5, weights are recommended to be used to calculate appropriate estimates of population characteristics and their corresponding standard errors. The CC has calculated the weights that take into account the probability of selection. These weights are named S2SAMWT51 and are included in the Stage 2/3 derived dataset, DERIVE_NCS51. The sampling weights are the product of a base weight and an adjustment for refusal. The base weights are the inverse of the empirical sampling fractions and are provided separately with variable name S2BASEWT51. The adjustment for refusal is the inverse of the probability that a sampled participant agrees to participate and completes the exam. This can be estimated by the observed probability of exam completion – these field-center specific probabilities are provided with variable name S2REFADJ51. Note that S2SAMWT51 = S2BASEWT51 x S2REFADJ51. If failure to complete the visit is informative then analysis based on these weights may be biased. An analyst may wish to use more sophisticated methods of calculating the adjustment for refusal to correct for this bias.

Participants were invited to Stage 3 if they were selected to Stage 2, had no contraindications to MRI and (initially) if they attended a clinic visit. This was later revised so that participants completing home visits were invited to Stage 3 as well, but only 10 such participants completed a Stage 3 exam. The base weights for Stage 3 are then the inverse of the proportion of participants completing clinic visits who were selected to Stage 2. The weights were then normalized to the number of participants completing clinic visits (S3BASEWT51). The adjustment for refusal is the inverse of the field center-specific probability of completing the exam (S3REFADJ51), though analysts may choose to re-calculate to account for informative failure to complete the visit. It follows that the Stage 3 sampling weights are S3SAMWT51 = S3BASEWT51 x S3REFADJ51.

### *2.2. Content-specific methods*

#### *2.2.1. Laboratory Analyte Measurements – Advanced Research and Diagnostic Laboratory (University of Minnesota)*

Thryoid Stimulating Hormone, TSH (mIU/L)

TSH was measured in serum using a sandwich immunoassay method on the Roche Elecsys 2010 Analyzer (Roche Diagnostics, Indianapolis, IN 46250) using a sandwich immunoassay method (Roche Diagnostics, Indianapolis, IN 46250). In the first incubation, the patient sample is mixed with a biotinylated monoclonal TSH-specific antibody and a monoclonal TSH-specific antibody labeled with a ruthenium complex to form a sandwich complex. During the second incubation, streptavidin-coated microparticles are added, and the complex becomes bound to the solid phase via interaction of biotin and streptavidin. The microparticles are then captured magnetically, and unbound material is removed. Application of a voltage to the electrode then induces chemiluminescent emission which is measured by a photomultiplier. The amount of light produced is directly proportional to the amount of TSH in the sample. The CV of the method is 3.3%.



Thyroxine (free), fT4 (ng/dL)

Thyroxine (free) was measured in serum on a Roche Elecsys 2010 Analyzer (Roche Diagnostics Corporation) using a competition immunoassay method (Roche Diagnostics, Indianapolis, IN 46250). In the first incubation, patient sample is mixed with T4-specific antibody labeled with a ruthenium complex. Biotinylated T4 and streptavidin-coated microparticles are added during the second incubation. The still-free binding sites of the labeled antibody become occupied, with formation of an antibody-hapten complex. The entire complex becomes bound to the solid phase via interaction of biotin and streptavidin. The microparticles are then captured magnetically and unbound material is removed. Application of a voltage to the electrode then induces chemiluminescent emission which is measured by a photomultiplier. The amount of light produced is inversely proportional to the amount of T4 in the sample. The CV for the method is 2.4%.

Triiodothyronine, T3 (ng/dL)

Triiodothyronine (T3) was measured in serum on a Roche Elecsys 2010 Analyzer (Roche Diagnostics Corporation) using a competition immunoassay method (Roche Diagnostics, Indianapolis, IN 46250). Bound T3 is released from the binding proteins in the sample by 8-anilino-1-naphthalene sulfonic acid (ANS). In the first incubation, T3 in the patient sample reacts with T3-specific antibody labeled with a ruthenium complex. Biotinylated T3 and streptavidin-coated microparticles are added during the second incubation. The still-free binding sites of the labeled antibody become occupied, with formation of an antibody-hapten complex. The entire complex becomes bound to the solid phase via interaction of biotin and streptavidin. The microparticles are then captured magnetically and unbound material is removed. Application of a voltage to the electrode then induces chemiluminescent emission which is measured by a photomultiplier. The amount of light produced is inversely proportional to the amount of T3 in the sample. The CV for the method is 5.2%.



Thyroid peroxidase antibody, anti-TPO (IU/mL)

Thyroid peroxidase antibody (anti-TPO) is measured in serum or plasma on a Roche Elecsys 2010 Analyzer (Roche Diagnostics Corporation) using a competition immunoassay method (Roche Diagnostics, Indianapolis, IN 46250). In the first incubation, the patient sample is mixed with anti-TPO-antibodies labeled with a ruthenium complex. Biotinylated TPO and streptavidin-coated microparticles are added during the second incubation. The anti-TPO antibodies in the sample compete with the ruthenium-labeled anti-TPO antibodies for the biotinylated TPO antigen. The entire complex becomes bound to the solid phase via interaction of biotin and streptavidin. The microparticles are then captured magnetically and unbound substance is removed. Application of a voltage to the electrode then induces chemiluminescent emission which is measured by a photomultiplier. The amount of light produced is inversely proportional to the amount of anti-TPO in the sample. The CV for the method is 10.2% at concentrations below the assay cut-off (34 IU/mL) and 6.0% for concentrations above the assay cut-off.

HbA1c (%)

HbA1c was measured in EDTA whole blood on the Tosoh HPLC Glycohemoglobin Analyzer (Tosoh Medics, Inc., San Francisco CA 94080) using an automated high performance liquid chromatography method. This method is calibrated utilizing standard values derived by the National Glycohemoglobin Standardization Program (NGSP). The laboratory CV is 1.9%.

Creatinine, Serum (mg/dL): Creatinine was measured in serum on a Roche Modular P Chemistry Analyzer (Roche Diagnostics Corporation) using a creatinase enzymatic method (Roche Diagnostics, Indianapolis, IN 46250). In this enzymatic method creatinine is converted to creatine under the activity of creatininase. Creatine is then acted upon by creatinase to form sarcosine and urea. Sarcosine oxidase converts sarcosine to glycine and hydrogen peroxide, and the hydrogen peroxide reacts with a chromophore in the presence of peroxidase to produce a colored product that is measured at 546 nm (secondary wavelength = 700 nm). This is an endpoint reaction that agrees well with recognized HPLC methods, and has the advantage over Jaffe picric acid-based methods that are susceptible to interferences from non-creatinine chromogens. The CV for the method is 2.3%.



Cystatin C (mg/dL)

Cystatin C was measured in serum using Gentian Cystatin C reagent on the Roche Modular P Chemistry analyzer. Serum sample from human is mixed with Gentian Cystatin C immunoparticles. Cystatin C from the sample and anti-Cystatin C from the immunoparticles aggregates. The complex particles created absorb light, and by turbidimetry the absorption is related to Cystatin C concentration via interpolation on an established standard calibration curve. The laboratory inter-assay CV is 3.1% at a value of 0.90 mg/dL and 2.5% at a value of 3.98 mg/dL.

Uric Acid, Serum (mg/dL)

Uric acid was measured in serum using an enzymatic colorimetric assay kit and read on the Roche Modular P Chemistry analyzer (Roche Diagnostics, Indianapolis, IN 46250). In this

method uric acid is oxidized by uricase to produce allantoin, CO2 and peroxide. Then the peroxide produced from this reaction is acted upon by peroxidase in the presence of 4-aminophenazone and TOOS (N-ethyl-N-(2-hydroxy-3-sulfopropyl)-3-methylaniline) to produce a red quinoneimine dye end product. It is a two-point, end-point reaction, with measurement occurring at 546 nm (secondary wavelength 700 nm). The CV for this method is 1.9%

Urine Albumin –UMALI (mg/L)

Albumin was measured in urine using an immunoturbidometric method on the ProSpec nephelometric analyzer (Dade Behring GMBH. Marburg, Germany D-35041). A solution of rabbit-derived anti-human albumin is incubated with the urine specimen. An immunocomplex forms between the antibody and the albumin in the specimen, resulting in an increase in light scatter. The higher the concentration of albumin, the more intense the degree of light scatter. The albumin concentration of the test specimen is determined by comparing its light scatter to that observed using known standards in a calibration curve. The laboratory inter-assay CV is 3.2%.



Urine Creatinine (mg/dL)

Creatinine was measured in urine on a Roche Modular P Chemistry Analyzer (Roche Diagnostics Corporation) using a creatinase enzymatic method (Roche Diagnostics, Indianapolis, IN 46250). In this enzymatic method creatinine is converted to creatine under the activity of creatininase. Creatine is then acted upon by creatinase to form sarcosine and urea. Sarcosine oxidase converts sarcosine to glycine and hydrogen peroxide, and the hydrogen peroxide reacts with a chromophore in the presence of peroxidase to produce a colored product that is measured at 546 nm (secondary wavelength = 700 nm). This is an endpoint reaction that agrees well with recognized HPLC methods, and has the advantage over Jaffe picric acid-based methods that are susceptible to interferences from non-creatinine chromogens. The laboratory CV is 4.3% at a concentration of 18.39mg/dL and 1.5% at a concentration of 96.57mg/dL.

Urine Albumin/creatinine Ratio - UMALCR (mg/g Cr)

The urine albumin/creatinine ratio was determined by dividing urinary albumin (mg/L) by creatinine (mg/dL) and multiplying by 0.01 to obtain mg of albumin/g of creatinine.

Vitamin B12 (pg/mL)

Vitamin B12 was measured in serum using a direct chemiluminescent competitive immunoassay method on the Roche Elecsys 2010 Analyzer (Roche Diagnostics, Indianapolis, IN 46250). The sample is first incubated with the vitamin B12 pretreatment 1 and pretreatment 2 during which bound vitamin B12 is released. The pretreated sample is then incubated with the ruthenium labeled intrinsic factor and a vitamin B12-binding protein complex is formed, the amount of which is dependent upon the analyte concentration in the sample. After addition of streptavidin-coated microparticles and vitamin B12 labeled with biotin, the still-vacant sites of the ruthenium labeled intrinsic factor become occupied, with formation of a ruthenium labeled intrinsic factor-vitamin B12 biotin complex. The entire complex becomes bound to the solid phase via interaction of biotin and streptavidin. The reaction mixture is aspirated into the measuring cell where the microparticles are magnetically captured onto the surface of the electrode. Unbound substances are then removed with ProCell. Application of a voltage to the electrode then induces chemiluminescent emission which is measured by a photomultiplier. Results are determined via a calibration curve which is instrument-specifically generated by 2-point calibration and a master curve provided via the reagent barcode. The laboratory CV is 7.39% at a concentration of 469 pg/mL and 8.32% at a concentration of 258 pg/mL.



*2.2.2. Laboratory Analyte Measurements – Atherosclerosis Clinical Research Laboratory (ACRL)*

COMPLETE BLOOD COUNT (CBC)

The ABX Horiba Diagnostics MICROS 60-CS is a fully automated (Microprocessor Controlled) Hematology analyzer. It is used for in-vitro diagnostics testing of whole blood specimens, platelet PRP samples and whole blood component concentrates. The instrument implements both impedance technology and spectrophotometry to determine a Complete Blood Count with 3-part Differential. The 16 parameters are determined with a microsampling of only 10□L. The Micros 60 can analyze approximately 55 samples per hour.

CHOLESTEROL

Assaying total cholesterol in saponified serum extracts using “cholesterol dehydrogenase where the esterase and oxidase are combined into a single enzymatic reagent for the determination of total cholesterol; this is the basis for the Olympus Cholesterol method.

TRIGLYCERIDES

This Olympus Triglyceride procedure is based on a series of coupled enzymatic reactions. The triglycerides in the sample are hydrolyzed by a combination of microbial lipases to give glycerol and fatty acids. The glycerol is phosphorylated by adenosine triphosphate (ATP) in the presence of glycerol kinase (GK) to produce glycerol-3-phosphate. The glycerol-3-phosphate is oxidized by molecular oxygen in the presence of GPO (glycerol phosphate oxidase) to produce hydrogen peroxide ($H_2O_2$) and dihydroxyacetone phosphate. The formed $H_2O_2$ reacts with 4-aminophenazone and N,N-bis(4-sulfobutyl)-3,5-dimethylaniline, disodium salt(MADB) in the presence of peroxidase (POD) to produce a chromophore, which is read at 660/800 nm. The increase in absorbance at 660/800 nm is proportional to the triglyceride content of the sample.



HIGH DENSITY LIPOPROTEIN (HDL) CHOLESTEROL

The Olympus HDL-Cholesterol test (HDL-C) is a two reagent homogenous system for the selective measurement of serum or plasma HDL-Cholesterol in the presence of other lipoprotein particles. The assay is comprised of two distinct phases. In phase one, free cholesterol in non-HDL-lipoproteins is solubilized and consumed by cholesterol oxidase, peroxidase, and DSBmT to generate a colorless end product. In phase two, a unique detergent selectively solubilizes HDL-lipoproteins. The HDL cholesterol is released for reaction with cholesterol esterase, cholesterol oxidase, and a chromogen system to yield a blue color complex, which can be measured bichromatically at 600/700nm. The resulting increase in absorbance is directly proportional to the HDL-C concentration in the sample.

LOW DENSITY LIPOPROTEIN (LDL) CHOLESTEROL, CALCULATED

The Friedwald Formula is use to calculate the LDL cholesterol. The formula is:

[LDL-chol] = [Total chol] - [HDL-chol] - ([TG]/5))

the quotient ([TG]/5) is used as an estimate of VLDL-cholesterol concentration. It assumes, first, that virtually all of the plasma TG is carried on VLDL, and second, that the TG:cholesterol ratio of VLDL is constant at about 5:1 (Friedewald et al. 1972).

GLUCOSE (not measured for ARIC NCS)

In this Olympus procedure, glucose is phosphorylated by hexokinase (HK) in the presence of adenosine triphosphate (ATP) and magnesium ions to produce glucose-6-phosphate (G-6-P) and adenosine diphosphate (ADP). Glucose-6-phosphate dehydrogenase (G6P-DH) specifically oxidizes G-6-P to 6-phosphogluconate with the concurrent reduction of nicotinamide adenine dinucleotide ($NAD^+$) to nicotinamide adenine dinucleotide, reduced (NADH). The change in absorbance at 340/380 nm is proportional to the amount of glucose present in the sample.



HIGH DENSITY C-REACTIVE PROTEIN (CRP)

Latex particles coated with antibody specific to human CRP aggregate in the presence of CRP from the sample forming immune complexes. The immune complexes cause an increase in light scattering which is proportional to the concentration of CRP in the serum. The light scattering is measured by reading turbidity at 572 nm. The sample CRP concentration is determined versus dilutions of a CRP standard of known concentration.

INSULIN (not measured for ARIC NCS)

Immunoassay, sandwich principle, total duration of assay: 18 minutes.

1st incubation: Insulin from 20 µL sample, a biotinylated monoclonal insulin-specific antibody, and a monoclonal insulin-specific antibody labeled with a ruthenium complex[a] form a sandwich complex.

2nd incubation: After addition of streptavidin-coded microparticles, the complex becomes bound to the solid phase via interaction of biotin and streptavidin.

The reaction mixture is aspirated into the measuring cell where the microparticles are magnetically captured onto the surface of the electrode. Unbound substances are then

removed with ProCell. Application of a voltage to the electrode then induces chemiluminescent emission which is measured by a photomultiplier.
Results are determined via a calibration curve which is instrument-specifically generated by 2-point calibration and a master curve provided via the reagent barcode.

NT-proBNP

Immunoassay for the in vitro quantitative determination ol Nterminal pro-Brain natriuretic peptide in human serum and plasma. Sandwich principle. Total duration of assay: 18 minutes. 1st incubation: Antigen in the sample (15;tL), a biotinylated monoclonal NT-proBNP-specific antibody, and a monoclonal NT-proBNP-specific antibody labeled with a ruthenium complexa form a sandwich complex. 2nd incubation: After addition of streptavidin-coated microparticles, the complex becomes bound to the solid phase via interaction of biotin and streptavidin. The reaction mixture is aspirated into the measuring cell where the microparticles are magnetically captured onto the surface of the electrode. Unbound substances are then removed with ProCell, Application of a voltage to the electrode then induces chemiluminescent emission which is measured by a photomultiplier. . Resulls are determined via a calibration curve which is instrument-specifically generated by 2-point calibration and a master curve provided via the reagent barcode. a) Tris(2,2'-bipyridyl)ruthenium(ll)-complex (Ru(bpy)3*)



HIGH SENSITIVE CARDIAC TROPONIN T (HS cTnT)

Immunoassay, sandwich principle. Total duration of assay: 1B minutes. 1st incubation: 50 pL of sample, a biotinylated monoclonal anti-cardiac troponin T-specific antibody, and a monoclonal anti-cardiac troponin T-specific antibody labeled with a ruthenium complex'react to form a sandwich complex. 2nd incubation: After addition of streptavidin-coated microparticles, the complex becomes bound to the solid phase via interaction of biotin and streptavidin. The reaction mixture is aspirated into the measuring cellwhere the microparticles are magnetically captured onto the surface of the electrode. Unbound substances are then removed with ProCell. Application of a voltage to the electrode then induces hemiluminescent emission which is measured by a photomultiplier. Results are determined via a calibration curve which is

instrument-specifically generated by 2-point calibration and a master curve (S-point calibration) provided via the reagent barcode. a) Tris(2,2'bipyridyl)ruthenium(ll)-complex (Ru(bpy) 3. )

*2.2.3. Echo*

Design and methods of echocardiography in ARIC Visit 5 have been previously described in detail.[1] Briefly, studies were acquired in participants attending Visit 5 at all 4 Field Centers by certified study sonographers using uniform imaging equipment and image acquisition protocol. Studies were acquired digitally and sent to the Echocardiography Reading Center at the Brigham and Women's Hospital, where quantitative measures were performed by dedicated Reading Center analysts and independently over-read by staff echocardiographers with both readers blinded to clinical information.

Left ventricular (LV) volumes were calculated by the modified Simpson's method using the apical 4 and 2 chamber views, and LV ejection fraction (LVEF) was derived from volumes in the standard manner. LV dimensions and wall thickness was measured from the parasternal long axis view according to the recommendations of the American Society of Echocardiography (ASE).[2] LV mass was calculated from LV linear dimensions and indexed to body surface area also as recommended by ASE guidelines. LV hypertrophy (LVH) was defined as LV mass indexed to body surface area (LV mass index, LVMi) >115 g/m$^2$ in men or >95 g/m$^2$ in women. Relative wall thickness (RWT) was calculated from LV end-diastolic dimension and posterior wall thickness. Left atrial (LA) volume was measured by the uniplane Simpson's method of discs using apical 4- and 2-chamber views at an end-systolic frame preceding mitral valve opening, and was indexed to body surface area to derive LA volume index (LAVi). E wave, E wave deceleration time (DT), and late transmitral velocity (A wave) were measured by pulsed wave Doppler and the peak lateral and septal mitral annular relaxation velocities (E') were assessed using tissue Doppler imaging, both from the apical 4-chamber view.[3] E/E' ratio, calculated as early transmitral velocity (E wave) divided by E'.



<u>References</u>

(1) [1]Shah AM, Cheng S, Skali H, Wu J, Mangion JR, Kitzman D, Matsushita K, Konety S, Butler KR, Fox ER, Cook N, Ni H, Coresh J, Mosley TH, Heiss G, Folsom AR, Solomon SD.

Rationale and Design of a Multicenter Echocardiographic Study to Assess the Relationship between Cardiac Structure and Function and Heart Failure Risk in a Biracial Cohort of Community Dwelling Elderly Persons: The Atherosclerosis Risk in Communities (ARIC) Study. *Circ Cardiovasc Img* 2014. (epub ahead of print)

(2) [1] Lang RM, Bierig M, Devereux RB et al. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr 2005;18:1440–63.

(3) [1] Nagueh SF, Appleton CP, Gillebert TC, Marino PN, Oh JK, Smiseth OA, Waggoner AD, Flachskampf FA, Pellikka PA, Evangelista A. Recommendations for the evaluation of left ventricular diastolic function by echocardiography. *J Am Soc Echocardiogr* 2009;22:107-33.

*2.2.4. ECG*

Standard 10-second resting 12-lead electrocardiogram at rest was digitally acquired using a GE MAC 1200 electrocardiograph (GE, Milwaukee, Wisconsin) at 10 mm/mV calibration and a speed of 25 mm/s. ECG reading was performed centrally at the Epidemiological Cardiology Research Center, Wake Forest School of Medicine, Winston Salem, North Carolina. All electrocardiograms were initially inspected visually for technical errors and inadequate quality before being automatically processed using GE 12-SL Marquette Version 2001 (GE, Milwaukee, Wisconsin). ECG abnormalities were classified and coded using the Minnesota ECG Classification.

*2.2.5. Pulse Wave Velocity*

Electrocardiogram, bilateral brachial and ankle blood pressures, and carotid and femoral arterial pulse waves were simultaneously measured with a vascular testing device (VP-1000plus, Omron Healthcare) {Cortez-Cooper, 2003 #5308}. This machine was originally developed as a screening device for hypertension (via blood pressure), peripheral artery disease

(via ankle brachial index), and arterial stiffness (via PWV), and this necessitated the use of 4 blood pressure cuffs on each limb. Carotid and femoral arterial pressure waveforms were stored for 30 sec by applanation tonometry sensors attached on the left common carotid artery (via a neck color) and left common femoral artery (via elastic tape around the waist). Bilateral brachial and post-tibial arterial pressure waveforms were stored for 10 sec by extremities cuffs, connected to a plethysmographic sensor and an oscillometric pressure sensor, wrapped on both arms and ankles.

Pulse wave velocity was calculated from the distance between two arterial recording sites divided by transit time. Transit time was determined from the time delay between the proximal and distal "foot" waveforms. The foot of the wave was identified as the commencement of the sharp systolic upstroke, which was automatically detected by a band-pass filter (5~30 Hz). Time delay between right brachial and tibial arteries (Tba), between carotid and femoral arteries (Tcf), and between femoral and tibial arteries (Tfa) were obtained. The path length from the carotid to the femoral artery (Dcf) was directly assessed in duplicate with a random zero length measurement over the surface of the body with a non-elastic tape measure {Tanaka, 1998 #3182}. The path lengths from the suprasternal notch to brachial artery (Dhb), from suprasternal notch to femur (Dhf), and from femur and ankle (Dfa) were calculated automatically by the machine using the following equations {Yamashina, 2002 #5184}:

Dhb = (0.220 x height {cm} - 2.07)

Dhf = (0.564 x height {cm} - 18.4)

Dfa = (0.249 x height {cm} + 30.7)

PWV were calculated by the following equations:

Carotid-femoral PWV = Dcf / Tcf

Brachial-ankle PWV = (Dhf + Dfa – Dhb) / Tba

The validity and reliability of the automatic device for measuring PWV have been established previously {Cortez-Cooper, 2003 #5308}.



References

(1) Cortez-Cooper MY, Supak JA and Tanaka H. A new device for automatic measurements of arterial stiffness and ankle-brachial index. *Am J Cardiol*. 2003;91:1519-1522.
(2) Tanaka H, DeSouza CA and Seals DR. Absence of age-related increase in central arterial stiffness in physically active women. *Arteriosclerosis, thrombosis, and vascular biology*. 1998;18:127-132.
(3) Sugawara J, Komine H, Hayashi K, Yoshizawa M, Yokoi T, Maeda S and Tanaka H. Agreement between carotid and radial augmentation index: Does medication status affect the relation? *Artery Research*. 2008;2:74-76.
(4) Yamashina A, Tomiyama H, Takeda K, Tsuda H, Arai T, Hirose K, Arai T, Hirose K, Koji Y, Hori S and Yamamoto Y. Validity, reproducibility, and clinical significance of noninvasive brachial-ankle pulse wave velocity measurement. *Hypertension Res*. 2002;25:359-364.

*2.2.6. Spirometry*



Spirometry was conducted in accordance with the American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines (1) using a dry rolling-seal Spirometer (Ohio SensorMed model 827, Ohio Medical Instruction Company, Cincinnati, Ohio). Each spirometer was attached to a computer running dedicated software that provided expiratory curves, calculated lung function parameters and determined the acceptability of the tests (Occupational Marketing, Inc., Houston, TX). The spirometry system has been independently tested and found to exceed ATS spirometry equipment recommendations
All technicians were trained and certified in spirometry procedures. Technicians either participated in a central training, or, in cases of staff turn-over, were trained locally at their clinic location by a centrally-trained supervisor. All technicians were required to take an online course in spirometry and pass a written certification exam with a 70% or higher test. A technician was allowed to retake the written exam one time if their initial score was lower than 70%.

Participants were asked to perform three to eight forced expiratory maneuvers in the seated position in an effort to meet the ATS acceptability and repeatability criteria. The highest value of FVC and $FEV_1$ from the acceptable maneuvers was used. All spirometry exams were

reviewed by one investigator and each test was graded for quality. Only tests with FVC and $FEV_1$ grades of "C" or higher were used in our analysis. Predicted and lower limit of normal values were obtained from the published reference equations derived from NHANES III (2).

References


(1) Miller MR, Hankinson J, Brusasco V, et al.. Standardisation of spirometry. Eur Respir J 2005; 26: 319-338.

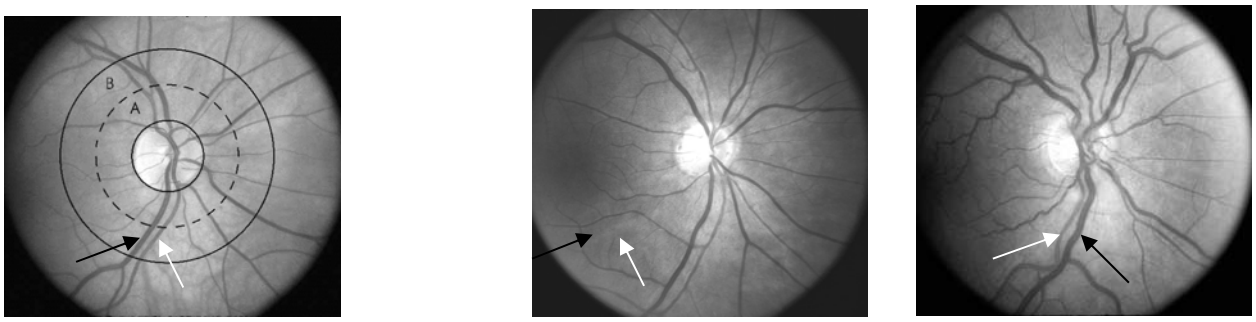
(2) Hankinson JL, Odencrantz JR, Fedan KB. Spirometric values from a sample of the general US population. Am J Respir Crit Care Med 1999; 159: 179-187.


*2.2.7. Retinal - Assessment of Retinal Vessel Diameters, Retinopathy, Focal Retinal Arteriolar Narrowing and Arterio-Venous (A/V) Nicking*

Retinal vessel diameters were measured using a computer-assisted technique based on a standard protocol and formula similar to one used for the ARIC study and other studies.[1] For the assessment of retinal vessel diameters, retinal images of field 1 (centered at the optic nerve head) were used. Trained graders, masked to participant characteristics, measured the diameters of all arterioles and venules coursing through a specified area one-half to one disc diameter surrounding the optic disc using a computer software program which is shown in Figure 1a. On average, between 7 and 14 arterioles and an equal number of venules were measured per eye. Individual arteriolar and venular measurements were combined into summary indices that reflect the average retinal arteriolar and venular diameter of an eye based on the Parr-Hubbard-Knudtson formula.[2] Figure 1b shows an eye with narrow retinal arteriolar diameter and normal retinal venular diameter while Figure 1c shows an eye with normal retinal arteriolar diameter and wide retinal venule diameter in a person with type 1 diabetes.



Figure 1a Figure 1b Figure 1c




1. Three scanned retinal images from eyes of persons with diabetes. a. grid over digitized image centered on right disc showing arterioles (white arrows) and venules (black arrows) coursing through a specified area one-half to one disc diameter (zone B) surrounding optic nerve head; b. right eye with narrow retinal arteriolar diameter and normal retinal venular diameter; c. right eye with normal retinal arteriolar diameters and wide retinal venular diameters.

Graders regularly participated in quality control exercises; the inter- and intra-grader variability was small (interclass and intraclass correlations > 0.90 for central retinal arteriole equivalent [CRAE] and central retinal venule equivalent [CRVE]). Measurements were done independently for each examination and each eye.

The grader assessed the absence, presence and severity of retinopathy lesions by comparing them with standard images. The presence and severity of these lesions were then used to assign an overall disease severity for the eye, based on the ordinal ETDRS diabetic retinopathy severity scale. The component lesions: retinal hemorrhages and microaneurysms (HMA); hard exudate (HE); venous loops (Loops); soft exudates or cottonwool spots (SE); intraretinal microvascular abnormalities (IRMA); venous beading (VB); new vessels on the disc and elsewhere (NVD and NVE); fibrous proliferation on the disc and elsewhere (FPD and FPE); and vitreous and/or preretinal hemorrhage (VH/PRH); are the individual lesions that are used in assigning the diabetic retinopathy severity level. The presence of macular edema and clinically significant macular edema (ME and CSME) were also assessed. These lesions were graded using the Modified Airlie House protocol and definitions adapted for the ETDRS clinical trial.[3]
The presence of other retinal arteriolar characteristics, focal arteriolar narrowing, and arterio-venous (A/V) nicking, was also graded. Focal narrowing was graded by comparing with a standard photograph from the Wisconsin Age-Related Maculopathy Grading protocol in which focal narrowing of small arterioles in the posterior pole (Field 2) involves a total length of 1/3

disc diameter.[4] Focal arteriolar narrowing was graded as absent, questionable, less than the standard, or greater than or equal to the standard for all arterioles more than 900 µm from the disc margin in two standard fields. When there were multiple but separate areas of focal arteriolar narrowing, the composite length of involvement was compared to the standard. For purposes of analyses, two categories were used: absent or questionably present, and present. A/V nicking was graded for all arterio-venous crossings that were more than 900 µm from the disc margin in both fields. A/V nicking was graded as present if there was a decrease in the diameter of the venule on both sides of the arteriole that was crossing it.


References

(1) Hubbard LD, Brothers RJ, King WN, et al. Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the Atherosclerosis Risk in Communities Study. *Ophthalmology* 1999;106:2269-2280.
(2) Knudtson MD, Lee KE, Hubbard LD, et al. Revised formulas for summarizing retinal vessel diameters. *Curr Eye Res* 2003;27:143-149.
(3) Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs-- An Extension of the Modified Airlie House Classification: ETDRS Report #10". Ophthalmology 1991; 98:786-806.
(4) Klein R, Davis MD, Magli YL, et al. The Wisconsin Age-Related Maculopathy Grading System. US Department of Commerce. Springfield,VA: NTIS Accession No. PB91-184267; 1991.




### *2.2.8. MRI*

Purpose

The MRI protocol contains two classes of sequences: (1) those for evaluation of the cerebral vasculature, which are managed by Dr. Wasserman; and (2) those for evaluation of brain anatomy, which are managed by the Mayo Clinic (Dr. Jack's Aging and Dementia Imaging Research (ADIR) Lab). The purpose of the anatomic MRI is to provide quantitative measures or qualitative assessment of brain volume/thickness, cerebrovascular disease, micro hemorrhages, and diffusion which can be correlated with other laboratory, demographic and clinical measures acquired in the study.

Acquisition

All imaging was performed at 3T. The protocol for the anatomic sequences as well as the study data extracted from each sequence was as follows:

*MPRAGE* - used for anatomic measures - i.e., region of interest (ROI)-wise brain volumes, cortical thickness

*Axial T2 Star* - used for quantitative assessment of cerebral micro hemorrhages

*Axial T2 FLAIR* - used for assessment of cerebrovascular disease

*Axial DTI* - used for ROI-wise diffusion measurements (fractional anisotropy and mean diffusivity)

Analysis Procedures

*MPRAGE image pre-processing to correct specific artifacts:* Image pre-processing operations designed to correct intensity inhomogeneity and gradient non-linearity was applied to each set of MPRAGE images prior to image analysis. These methods were developed by the Mayo ADIR Lab for ADNI 1.

*Anatomic ROI-wise measures of brain volume/thickness:* Freesurfer was used to measure ROI-wise measures of brain volume/thickness 2,3. Volume/thickness values were generated for 122 ROIs for each scan and uploaded to the data center.

*Voxel-based morphometry (VBM) [analyses are available to ARIC investigators by request]:* The Mayo ADIR Lab will perform cross-sectional voxel-wise associations using VBM 4. Examples of relevant between-group comparisons with VBM would include voxel-wise comparisons of gray matter (GM) differences between those with significant decline in cognitive function vs. others. Examples of a relevant regression analysis would be the voxel-wise relationship between GM loss and change in score on a specific cognitive test. GM differences between groups or regression analyses are assessed within the general linear model framework of SPM corrected for multiple comparisons.

*Micro hemorrhage assessment:* Micro hemorrhages and areas of superficial siderosis were enumerated and anatomically localized by trained image analysts.

*Qualitative grading for assessment of cerebrovascular disease (CVD):* MR images were viewed on video monitors using locally developed software. Each scan was rated by an experienced

image analyst. Studies were evaluated for presence and location of infarctions in the central GM, hemispheric cortical infarctions and hemispheric white matter lacunar infarctions.
*Measurements of white matter hyperintensity (WMH) volume*: Quantitative measures of WMH volume were derived from the axial FLAIR images. We used a semi-automated algorithm developed in-house 5. Following segmentation, an atlas-based parcellation technique was used to calculate WMH volume in different anatomic compartments: e.g., superficial vs. deep compartments or in named lobes (e.g., frontal, parietal, etc.).
*DTI analysis:* Diffusivity (MD) and fractional anisotropy (FA) maps are created. ROI-wise MD and FA values will be reported using the John Hopkins University atlas.
*Clinical review/alerts:* Review of scans were performed by a neuroradiologist at each field center to screen for clinically significant abnormalities (e.g., acute hemorrhage and/or mass effect). In addition, the Mayo ADIR Lab also QC all incoming scans for medically significant abnormalities, and notified the data center if an abnormality was identified.



Quality Assurance
All scans were evaluated by a trained MR image analyst for protocol compliance and scan quality. This QC data was entered into data forms and transmitted to the data center. Quality problems with any scan resulted in a request for a rescan.

ADIR Lab Contributions
*PI:* Clifford R. Jack., Jr., M.D.

*IT Support (image analysis software):*
Jeff Gunter - prepressing, MCH/SS
Dave Just - MCH/SS
Rob Reid - DTI
Chris Schwarz - WMH, DTI
Matt Senjem - WMH, VBM

*Image Analysts*
Chad Ward - image QA, MCH/SS assessment
Anthony Spychalla - image QA, CVD assessment

Greg Preboske - CVD/WMH assessment
Samantha Zuk - Freesurfer analysis
Kaely Steinert - MRI protocols

*Other*
Kejal Kantarci, M.D. - image interpretation and grading
Denise Reyes – supervision

References


(1) Jack CR, Jr., Bernstein MA, Borowski BJ, et al. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. Alzheimers Dement 2010;6:212-20.

(2) Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A 2000;97:11050-5.

(3) Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341-55.

(4) Ashburner J, Friston KJ. Voxel-based morphometry--the methods. Neuroimage 2000;11:805-21.

(5) Raz L, Jayachandran M, Tosakulwong N, et al. Thrombogenic microvesicles and white matter hyperintensities in postmenopausal women. Neurology 2013;80:911-8.




*2.2.9. Vascular MRI*

The purpose of the vascular MRI protocol was to estimate the prevalence of intracranial atherosclerotic disease (ICAD), determine whether it is associated with dementia or mild cognitive impairment, and characterize features of ICAD and estimate their associations with risk factors. The exam was carried out on 3T MRI Siemens scanners (Forsyth County: Skyra, 32 channel head coil; Jackson: Skyra, 20 channel head coil; UMN: Trio, 12 channel head coil; and Washington County: Verio, 12 channel head coil). The MRI protocol consisted of a 3-dimensional time-of-flight MR angiogram through the Circle of Willis, centered to include the distal vertebral artery segments inferiorly and the middle cerebral artery branches superiorly (acquired resolution, 0.50 x 0.55 x 0.55 $mm^3$; 152 slices, 8.4cm SI coverage). This was followed

by a 3-dimensional high-isotropic resolution black blood MRI (BBMRI) scan (Qiao et al. J Mag Res Im 2011;34:22-30) oriented in a coronal plane and centered at the Circle of Willis (0.5 x 0.5 x 0.5 $mm^3$; 128 slices, 6.4 cm AP coverage). This vascular protocol was implemented at the end of the ARIC NCS brain MRI protocol, described separately.

MRI images were analyzed by 6 trained analysts certified by successfully completing complex sample cases. Each analyst used picture archiving and communication system (PACS) software (Ultravisual; Emageon, Birmingham, Ala) for the qualitative analysis of the MRA and BBMRI scans. Using the PACS software, the BBMRI and MRA images were co-registered and reconstructed in both short and long axes relative to the flow direction for each vascular territory (RMCA, LMCA, RPCA, LPCA, ACA, Basilar, Vertebral, RICA, LICA). For this analysis, the number of plaques identified for each territory was recorded, with categorical stenosis recorded for the most stenotic plaque per territory.

Quantitative measurements of lumen size and stenosis from the MRA and wall/plaque size from the BBMRI were acquired using LAVA software (LAVA, Leiden University Medical Center, the Netherlands), which uses a deformable tubular model based on Non-Uniform Rational B-Splines (NURBS) surface modeling to contour each vessel segment. This technique provides semi-automated contour detection of the arterial lumen and performs an iterative linear regression fit of the lumen area over the entire segment. Standard vessel segments were measured (e.g., proximal Circle-of-Willis branches such as M1 and basilar artery segments) over a fixed segment length, and the largest plaque identified for each vascular territory in the qualitative assessment was also measured.

Exam reliability was assessed by repeating 102 exams with evidence for plaque. Inter- and intra-observer variability was also assessed by repeat readings. A peer-review process was implemented twice per exam in which an observer re-evaluated each exam read by a different observer and disagreements were arbitrated by the PI.

# 3. STATISTICAL METHODS

## *3.1. Standardizing continuous variables*

### *3.1.1. Calculation of Z scores*

Many ARIC variables are standardized prior to analysis. For consistency of ARIC manuscripts, it is critical that the standardization is done consistently. In this section, we provide the algorithms used for analysis and reference code made available by analysts.

Z-scores for cognitive function tests

Three cognitive function tests have been administered across ARIC visits and ancillary studies – DWRT, DSST and WFT. In NCS manuscripts evaluating cognitive change, Z-scores were calculated to standardize scores to the Visit 2 baseline. Raw scores and Z-scores will be distributed in the dataset V2_V5_CNF.sas7bdat. Analysts may wish to recalculate Z-scores to standardize to a different baseline if Visit 2 is not used or if different eligibility criteria are applied. For reference, the algorithm used for Visit 2 is provided below:
For each test, mean and standard deviation were calculated in the entire Visit 2 population meeting race-center inclusion criteria (white in MN and Washington, black in Jackson, black or white in Forsyth) with a non-missing test score. For subsequent visits, in participants meeting race-center inclusion criteria, Z-scores were calculated by subtracting the mean from the raw score, then dividing by the standard deviation. A global Z-score was calculated for participants with 3 non-missing scores by averaging the individual test Z-scores and re-standardizing.
o



## *3.2. Longitudinal methods*

The purpose of this working document is to provide guidance to analysts currently working on ARIC cohort manuscripts involving longitudinal data analysis of trends or change over time in participant outcomes collected over visits 1 through 5 or a subset of the visits. Key issues include model specification and model assumptions, parameter estimation and

interpretation, and capacity of methods to handle missing visits due to dropout and/or death. These recommendations can provide a means to methodological consistency across ARIC manuscripts. Distinctive features of research questions will require specific solutions or adaptations of the longitudinal methods to be developed within writing or subject-specific working groups. For example, more specific recommendations for analysis of longitudinal cognitive data, when the marginal effect is of interest, are detailed in Appendix A.

An important consideration in this section is the potential for bias resulting from visit non-response due to loss of follow-up due to dropout or death. The potential for bias has substantially increased with the addition of visit 5 that occurs fifteen years after visit 4; the per visit sample sizes for the ARIC cohort are visit 1 (n=15,792), visit 2 (n=14,348), visit 3 (n=12,887), visit 4 (n=11,656) and visit 5 (n=6,538).

Three common classes of longitudinal data models are subject-specific models, population averaged models and transition models. Subject-specific or random effects models are discussed in section 3.2.1, and population-averaged or marginal models are in section 3.2.2. Marginal modeling estimation approaches include generalized estimating equations and inverse-probability-of-attrition weighted estimating equations. Transition models, for example those estimating transition probabilities of elderly populations going from one state of physical functioning or disability to another, are addressed elsewhere (e.,g., Diggle et al. 2002). Sections 3.2.3 and 3.2.4 respectively discuss structural equation models (SEM) and shared parameter models for the joint analysis of longitudinal and survival outcomes.



### *3.2.1 Mixed Effects Models*

Background and Motivation

Mixed effect models – those regression models for longitudinal data including both fixed and random effects – are a class of tremendously popular and versatile approaches to the analysis of longitudinal data. They include linear mixed models (LMM) for continuous outcomes assumed to follow a multivariate normal distribution and generalized linear mixed models (GLMM) for non-normal or discrete outcomes. In the latter case, a popular form is the GLMM for binary outcomes defined by a logit link function and normally distributed random effects, sometimes called the logistic-normal model.

Specification of random effects in mixed effects models serves two main purposes. First, inclusion of random effects alters the interpretations of regression coefficients for the fixed effects in the model, giving them subject-specific interpretations. Thus, mixed effects models are used when subject-specific interpretations are desired. Consider, for example, a dichotomous predictor variable for current smoking status in a logistic-normal model for COPD. The regression coefficient for smoking is a log odds ratio of the effect of smoking versus non-smoking on COPD for a given participant, i.e., if the participant initiates smoking. The distinction between subject-specific interpretations versus population-averaged interpretations is particularly relevant for non-linear models. Thus, results from GLMMS and GEEs are generally not comparable, e.g., the subject-specific log odds ratio parameter for current smoker will tend to be larger in absolute value than the corresponding GEE marginal logs odds ratio parameter (i.e., a marginal model parameter is usually closer toward the null than the corresponding subject-specific parameter.) In contrast, an attractive feature of the LMM (which assumes a normally distributed continuous outcome) is that regression coefficients for fixed effects have both population-averaged and subject-specific interpretations. The implication is that a LMM analysis is sometimes used as a sensitivity analysis for a GEE analysis, or vice versa. while understanding that they have different missingness assumptions as discussed below.



The second important purpose of random effects in mixed effects models is for accounting for the covariance structure across visits of participants. For LMMs, the marginal covariance structure is often easily deduced. For example, the random intercept-only model (with independent errors having constant variance) is equivalent to compound-symmetry, synonymous with equi-correlation or exchangeable correlation.

Specific Recommendations

- **The random intercept-only model is not recommended for ARIC analyses including visit 5 because of the wide visit-spacing**. **Rather, the mixed effects model could include random intercepts and slopes for time, the so-called random coefficients model.** However, if less than four visits are being modelled the approach as described in the next bullet is recommended.

The random coefficients model would have a total of four variance components: intercept variance, slope variance, covariance between intercept and slope (often negative due to floor and/or ceiling effects on the outcome), and error variance. In SAS Proc Mixed, these are specified with the RANDOM statement. Typically, additional random effects, such as for quadratic or cubic terms, are not included given the potential for non-convergence with increasing complexity of the model. However, inclusion of quadratic fixed effects for time may enhance population-average interpretations and may only be restricted by the number of follow-up visits included in the analysis. Similar recommendations for including random intercept and slope terms are applicable to GLMMs.

In a LMM, the covariance structure of the outcomes may alternatively be specified directly with covariance specifications for the error vector corresponding to a subject. In SAS Proc Mixed, these are specified with the REPEATED statement. Such LMMs without random effects have been called general linear models with correlated errors (Diggle et al. 2002).

- **The specification of an unstructured covariance structure (often with a constant variance assumption across visits as deemed applicable to a continuous outcome) is the second recommended approach to handling intra-subject correlation in longitudinal modeling in ARIC.** A disadvantage of the unstructured covariance matrix is the increasing number of parameters with an increasing number of visits that must be estimated. Use of an unstructured covariance could be infeasible (e.g., SAS warnings issued due to convergence issues) for analyses of smaller data sets, say of one or two hundred or less, particularly in the presence of attrition.
    - Alternative more parsimonious covariance structures include those with correlations between visit-pairs that decays over time. However, given the highly unequally spaced visit timings the simple first-order autoregressive (AR-1) correlation structure is not recommended. A suitable structure in this case may be the spatial power structure (i.e., "TYPE=SP(POW)" in the SAS Proc Mixed REPEATED statement) that uses the Euclidiean distance between two time points as the power of the single correlation parameter in combination with a common variance parameter. Generally, model selection should

start with a maximum model by selecting an adequate preliminary mean structure and random-effects structure, then proceed to implement model reduction using a clearly stated model selection criterion to reach a final model; see Cheng et al. (2010) for a tutorial on mixed models including model selection and general good modeling practices such as scaling and centering of variables.

Assumptions and Limitations

Mixed effects model are highly versatile. They can easily accommodate time-varying covariates as well as participants with intermittent missing visits or who dropout. They are applicable to data missing-at-random (MAR), which means that missing outcomes can depend upon the covariates in the model and/or on observed values of outcomes. The MAR assumption underlying this maximum likelihood approach means that missingness is ignorable, in the sense that it is unnecessary to fit a missing data model. This is the case when the missingness depends only upon covariates that are included in the main measurement model (i.e., for Y). Maximum likelihood places the two issues of bias and efficiency on equal footing. If the models are close to correctly specified, then bias will be eliminated, and the solution will be statistically efficient. In sum, mixed effects models are very appealing with respect to their ease of use (particularly LMMs) and they are often fairly robust to violations of MAR. However, if missingness depends on unobserved values of outcomes, then the data are missing-not-at-random (MNAR) violating the assumption of LMMs.

Mixed effects models also have their limitations. In particular, MAR missingness is only truly ignorable when what is missing is part of the multivariate outcome when there are no auxiliary data (i.e., covariates predicting missingness that are not in the main measurement model), and also when likelihood analyses are pursued. If it is a component of the explanatory variables that is missing, then a probability model for the missing covariate is required. In ARIC, data collected at the annual follow-up (AFU) time points as well as prevalent and incident disease and medical conditions could be included in a model for missingness. These covariates are auxiliary data because they are not included in the main measurement model for Y at the study visits. Note that auxilliary data would not have been of interest if the data had been complete (i.e, no missed visits).

•Given the wide spacing between ARIC visits 4 and 5, it is recommended that auxiliary data be used when fitting mixed models, through use of multiple imputation, to account for missing outcome data.

For a more extensive discussion of missing data without unnecessary mathematical distractions, the reader is referred to Rathouz and Preisser (2014). These authors include a basic outline of the process of multiple imputation. In the case of continuous outcomes, LMMs without imputation are generally recommended for mild to moderate levels of missed visits, dropout and loss-to-follow-up due to death.

Generally, inverse-probability-weighting discussed in the context of marginal models in the next section is not compatible with random effects modeling.



References

Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, Muller KE. Real longitudinal data analysis for real people: building a good enough mixed model. Statistics in Medicine, 2010, 29:504-520.

Diggle PJ, Heagerty PJ, Liang K-Y, Zeger SL. Analysis of longitudinal data, second edition. Oxford: Oxford University Press, 2002.

Rathouz PJ, Preisser JS. Missing data: weighting and imputation, in Encyclopedia of Health Economics: Third Edition, Revised and Expanded, ed. A. J. Culyer. Elsevier, Inc, 2014.

### *3.2.2 Marginal Models*

Background

Marginal models are used to obtain population-averaged interpretations for the effect of explanatory variables or risk factors on a longitudinal, categorical, or continuous outcome. Consider, for example, a dichotomous predictor variable for current smoking status in a marginal logistic model for COPD defined as a dichotomous outcome. The regression coefficient for smoking is a log odds ratio of the effect of smoking versus non-smoking on COPD in the population, i.e., for the comparison of a randomly selected smoker to a randomly

selected non-smoker. The standard and popular method for fitting marginal models, particularly for longitudinal categorical outcomes, is generalized estimating equations (GEE). This method of estimation gives valid estimates if the model for the marginal mean is correctly specified. Moreover, valid inference (correct tests and confidence intervals) is obtained with use of the empirical sandwich covariance estimator of the regression coefficients even if the working correlation (or covariance) structure is misspecified. Nonetheless, correct choice of the correlation matrix can increase efficiency of estimation.

Specific Recommendations

•The specification of an unstructured correlation structure is recommended for marginal longitudinal modeling in ARIC in consideration of the unequally spaced study visits and the particularly wide spacing of visits 4 and 5.

A major limitation of GEE is the assumption that missing longitudinal outcomes are missing completely-at-random (MCAR); while MCAR means that missingness is completely random the terminology has been used in longitudinal data analysis to include the situation of covariate-dependent missingness whereby missingness due to attrition may depend upon the covariates in the model but not on the outcomes themselves whether observed or unobserved. This is a strong assumption in any setting but should be seriously questioned in ARIC with its substantial attrition at visit 5. In general, a model for the missing data (via multiple imputation), or a model for the missingness process (via inverse probability weighting), should be investigated with the use of auxiliary data as covariates, that is, a rich set of covariates that include the variables in the measurement model of interest but are not limited to those variables. This section focuses on the use of inverse weights with the goal of unbiased estimation of longitudinal trajectories or change over time in health outcomes in ARIC.

Choice of exposure and covariates in the main measurement model for Y

When assessing change over time it is natural to think in terms of trajectories of the mean outcome and how these trajectories vary according to baseline covariates including age, sex, race (or race-center), education, smoking, etc. Depending upon the number of visits available, linear, quadratic or linear splines terms could be used to define trajectories. The model would include interactions of time with baseline characteristics to determine how change over time in

the mean outcome varies according to subgroups or baseline factors. For example, the main goal of Weuve et al. (2012) was to assess the effect of baseline smoking status on change in cognitive status. For assessing causal effects, time-varying covariates may be of interest; however, time-updating of covariates must be theoretically motivated and use of analytical methods accounting for the possibility of time-varying confounding must be considered. For example, Weuve et al. (2012) considered the effect to smoking at baseline on subsequent cognitive decline; therefore, only baseline variables could act as confounders and time-updated covariates were not used in the main measurement model. In Power et al. (2013), time-updated confounding was plausible and supported by the data; therefore IP weighting was used to mitigate bias due to time-varying confounding. In the case of acute events, time varying covariates are generally appropriate; however, use of time-varying covariates for non-acute events, especially those with no definitive onset date (e.g. dementia, COPD) is appropriate only in very specific situations.



Motivation for weighting

The general practice of weighting observed data to account for missing data comes from weighting for non-response in sample surveys to correct for bias. To quote Rathouz and Preisser (2014):

"…when non-responders differ from responders according to their distribution of measured characteristics, the responders' data is weighted so that analysis restricted to the complete data sample would resemble analysis of the combined data of responders and non-responders had it been observed. Use of weights assumes there exists knowledge of some variables, such as demographic characteristics, on the non-responders so that they can be placed in groups or bins with responders. The bin-specific inverse probability of being a responder is then calculated and used as the weight for observations in that bin. The same principle can be applied in a prospective study where the outcome Y at the end of the follow-up period is missing for some subjects. The general idea is to weight records inversely to their probability of being observed such that observed data with a low likelihood of being observed receive relatively high weight. Incorporating continuous as well as categorical baseline variables instead of bins, a logistic regression model can be used to estimate the probability of being observed at

follow-up for each study participant. The resulting inverse probability is then the weight used in the complete-data analysis for the outcome."

The same general method of constructing weights applies for missing data patterns for repeated measures outcomes.

Inverse-probability-weighting estimating equations for dropouts in longitudinal studies

This section describes longitudinal data analysis with marginal models using inverse-probability-weighting to correct for bias resulting from dropouts (Robins, Rotnitzky and Zhao, 1995). Recommendations are given for ARIC analyses, with special relevance to those involving visit 5. The literature on these methods distinguishes between data that have only monotone patterns of missingness in the responses versus data that include intermittent missingness patterns. The latter involve more complex methods and are infrequently used so they will not be covered here. In a monotone pattern of missingness, once a participant misses a visit, he/she is assumed to be a dropout and has no other subsequent visits. Typically, there will be some subjects with intermittent missingness patterns that arise from missing a visit but then return for a later visit. The most popular IPW approaches involve only monotone data patterns and so their use requires a strategy to deal with participants who have intermittent missingness patterns:



• For IPW longitudinal data analysis of ARIC cohort (or NCS), we recommend omitting any observed visit that occurs after the first missed visit. Even after such subject-visit deletions, the analysis will involve at least one observation for every ARIC participant who has a baseline observation. The baseline visit will often be visit 1, but it could be a later visit depending upon the outcome being analyzed.

The advantage of this common strategy is that assumptions underlying a model for dropout are weaker than assumptions required for a model that allows intermittent data patterns. If it is only the outcome that is missing at a visit, then imputation of intermittent missing outcomes might be considered as a part of an IPW analysis, particularly, if this type of missing is so infrequent that simple imputation procedures could be defended. Otherwise, multiple imputation of missing outcomes (Paik, 1997) for visits that were attended could be incorporated as a part of the IPW analysis. An example is where lung function is the outcome

and the subject attends a visit (has not dropped out of the study) yet either does not show up for spirometry or does not produce a valid pulmonary function test. Generally, IPW analysis should do a good job of recovering the deleted information so that the extra effort of incorporating multiple imputation into an IPW analysis may not be worth the effort.

After a monotone data set has been produced as part of preliminary data processing, the IPW analysis can be implemented (see Preisser et al. 2000, Weuve et al. 2012, Hernan et al. 2000, Power et al. 2013 (note appendix) for examples and further discussion). The methodology we describe assumes the dataset cannot have missing data for any of the covariates. In the case of minimal covariate missingness simple imputation procedures could be used; otherwise a missingness indicator could be considered in a model for missingness even though use of such indicators in the main measurement model for Y may be questionnable. There are three general steps:

Step 1: Determine weights through a model (usually logistic regression) for dropouts.

Step 2: Apply the weights in a set of weighted estimating equations and solve them to estimate the regression parameters in the model for the marginal means.

Step 3: Calculate the empirical sandwich estimator of the variance-covariance matrix of the estimated regression parameters preferably taking into account that the weights were estimated.



The choice of variables to be included in the missing data model is an important one that is discussed below. First, we discuss four options for implementing IPW, starting with the most common approach:

1. Independence IPW estimating equations with observation-specific weights and conservative estimation of standard errors. This method uses routine GEE software such as STATA or SAS. See Fitzmaurice, Laird and Ware (2004) for details of applying this method in SAS.
2. Independence IPW estimating equations with observation-specific weights and consistent (asymptotically unbiased) estimation of standard errors. This method uses routine software for model fitting along with bootstrapped estimates of standard errors (e.g., Weuve et al. 2012) or it uses a SAS macro WGEE (Preisser et al. 2014).
3. Non-independence IPW estimating equations with observation-specific weights and

consistent estimation of standard errors. This method uses the SAS macro WGEE.

4. Independence or Non-independence IPW estimating equations with cluster-specific weights and conservative or consistent estimation of standard errors.

The fourth method, which is based on cluster-specific weights, most closely resembles the bin-method of adjustment for non-response in sample surveys. It produces a single weight for each participant and the weight is assigned to each of his/her observed visits. This method is not recommended since it can produce much less efficient estimates in finite samples than the IPW estimating equations based on observation-specific weights (Preisser, Lohman, and Rathouz, 2002). This leads to the first recommendation in the choice of IPW analysis of the monotone data set.

• For IPW longitudinal data analysis of ARIC cohort (or NCS), we recommend commonly used observation-specific weights (e.g., Weuve et al. 2012), which are equal to the inverse of the cumulative probability of being observed at the visit.

Details of the computation of these weights are provided elsewhere (e.g., Preisser et al. 2014, Power et al. 2013 (appendix), Hernan et al. 2000). It is basically a three-step process:

Step 1: From the fit of the logistic regression, determine visit-specific probabilities of being observed at the visit conditional upon not dropping out at the previous visit.

Step 2: Calculate the cumulative (unconditional) probabilities of not dropping out by the visit as a product of the conditional probabilities up to the visit.

Step 3: The visit-specific weight is the inverse of the cumulative probabilities up through the visit. The weight is taken as 1 for the first visit since all subjects (kept in the data set) are observed at the first visit.

The second decision involves choice of either an independence or a non-independence (non-diagonal) working correlation structure. For large cohort studies such as ARIC, there is little efficiency loss with use of an independence working correlation matrix. Use of weighting based on the missingness model actually recovers much of the potential efficiency loss of independence estimating equations (Tchetgen et al. 2012b). The second recommendation for IPW analysis follows.

• For IPW longitudinal data analysis of ARIC cohort (or NCS), we strongly recommend use of an

independence working correlation matrix.

Caution: use of a non-independence working correlation structure (such as any of the usual ones -exchangeable, AR-1, unstructured) with routine GEE or generalized linear model software such as SAS PROC GENMOD will give biased regression parameter estimates. This error is very easy to make; for example, Preisser et al. (2002) acknowledged use of biased IPW estimating equations in a previous published article when the exchangeable correlation matrix was inappropriately used. Tchetgen et al. (2012c) made a similar acknowledgment reporting a 28% change in the estimate for the association of never smokers and cognitive change with the exchangeable correlation matrix (used by Weuve et al., 2012) relat ive to the valid estimates based on the independence working correlation matrix. These references provide a clear warning against inappropriate use of a non-diagonal working correlation structures in IPW analysis. Having issued this warning, we note that it is possible to validly use a non-diagonal working correlation structure for IPW analysis, which can be done with the SAS macro WGEE. This approach requires all covariates to be completely observed (even at unobserved visits), which means they are baseline (cluster-level) covariates or time-varying covariates that are available external to the data collection process, such as time or functions of time. Details of implementation of IPW using macro WGEE are provided elsewhere (Preisser et al. 2014). Given the minimal efficiency loss with use of independence working correlation, the following recommendations are made depending upon the software used:



• IPW analysis with routine software: For analysis of the ARIC cohort (or NCS) using independence estimating equations with observation-specific weights, we recommend handling covariance estimation in one of two ways: (i) use the conservative standard errors given by the software while acknowledging as a limitation of the method that they are likely too large (by 15% in the simulation study of Preisser et al. 2002) (ii) otherwise, use bootstrapping for more accurate variance estimation.

• IPW analysis with SAS macro WGEE: For analysis of the ARIC cohort (or NCS) using independence (or non-independence) estimating equations implemented using observation-specific weights, use the correct (consistent) standard errors provided in the output.

Selection of covariates and strategies for building a missingness model

IPW analyses can accommodate a wide range of variables in the missingness model, including outcome measures at prior visits and covariates from the main measurement model as well as auxiliary covariates. The modeling process for missingness could include a separate model for each study visit (baseline visit excluded), which may have similarity if not equivalency to a single overall missingness model that includes interactions of visits with all other explanatory variables. Some guidelines for variable selection are:

• In IPW ARIC analysis, include covariates and prior outcomes from the measurement model into the missingness model, but don't restrict selection to these. Because you have to model the conditional probability of dropout at each visit, given survival through the previous visit, use information available between visits to predict dropout, such as AFU data and incident health events (e.g., stroke, CHD, etc).

• Generous inclusion of interactions in dropout models is recommended, particularly interactions involving race/center and age.

• When death is modeled as a process separate from dropout (see next section), we recommend use of the same AFU data as for the dropout model.

• Given the spacing between visits 4 and 5, consider developing a pseudo-visit between visits 4 and 5 in the model for dropout to improve the estimation of the dropout probabilities (SAS code is available in one ARIC application). For a measurement model including only some of the visits (e.g., spirometry is only collected at visits 1, 2 and 5), consider including visits 3 and 4 in the dropout model.

• For assessing the strength of covariates in the dropout model, use standard model selection criteria (e.g., deviances, p-values) as a guideline for including variables, but also include non-significant variables because these can also have an effect on the weights. Furthermore, there is no penalty to pay in terms of efficiency loss by overfitting the missingness model (Robins et al. 1995). Preisser et al. (2000) show this in an analysis of smoking trends with repeated binary responses subject to dropout.

• Balance consideration for dropout model complexity with increasing instability of weights. If the ratio of the largest weight to the smallest weight exceeds 20, then stabilized weights are recommended.

The development of the dropout model will involve some creativity. Sharing of ideas across ARIC studies is encouraged to improve analyses and provide some consistency across studies. However, differences in approaches are expected, particularly when the outcomes of interest are collected at different visits.

Death versus Dropout

Another issue for ARIC cohort and NCS studies, and the longitudinal study of aging populations generally, is whether attrition due to death should be handled differently than attrition due to dropout. There are two dimensions to this issue. The first dimension is a recognition that the reasons for death are likely different than the reasons for dropout.

• For IPW analysis of the ARIC cohort (or NCS), it is recommended to fit separate models for dropout and death, and to develop weights based on the consideration of distinct reasons for attrition.

The second dimension relates to the use of these estimated visit-specific probabilities, one for death and the other for dropout given being alive, in the construction of weights in the IPW analysis. The choice of whether to apply weights for total attrition (death + drop out) or only for non-death drop-out is not obvious. Regardless, the choice has some implication in the interpretation of the findings. Findings weighted for non-death drop out only hope to recover the association or to estimate trends in the case of perfect follow-up conditional upon being alive (Kurland and Heagerty 2005). Issues of generalizability must be considered - who dies in your population is often specific to location, population, and year of birth - and it is possible to find no association or an inverse association between your exposure and outcome in a population even if a causal effect exists at the level of the individual - i.e. if the exposure truly does cause the outcome - if most of those who would get the outcome due to exposure die of something else first. Findings weighted for both death and drop out hope to recover the association in the case where the only possible cause of death is your exposure (or alternately, people can die, but except for death due to the exposure, who dies is a random process). While this is helpful for understanding etiology, it must be admitted that the association corresponds to a hypothetical population that does not and cannot exist.

Weuve et. al (2012) describe an application of IPW for estimating the baseline effects of

smoking on cognitive decline whereby two models for attrition are fit, the first is a model for the probability of death at a visit, and the other is for dropout conditioned on being alive at the visit. They take the product of these two probabilities at each visit and then they take the inverse product of the cumulative probabilities as the weights.

•If ARIC investigators wish to weight for both death and drop out, then the approach of Weuve et al (2012), implemented with and independence working correlation matrix will give valid inference under correctly specified models for drop out and death.

•If ARIC investigators want to make inference to the ARIC population for individuals conditional upon being alive, then the weights should based upon dropouts only. Use of weighted GEE with an independence working correlation matrix will give valid inference under correctly specified models for dropout.

The question of weighing up the dead as well as the missing seems to be an open question for ARIC investigators for which no single recommendation can be given at this time (Chaix et al. 2012; Tchetgen et al. 2012a). A final general caution is raised in regards to seemingly innocuous refinements of established approaches, which as we have seen in the case of using non-independence estimating equations, can introduce bias into the estimation method based on use of routine software.



References

Chaix B, Evans D, Merlo J, Suzuki E. Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. Epidemiology 2012; 23:129-131.

Fitzmaurice GM, Laird NM and Ware JH. Applied Longitudinal Analysis, second edition. New York: John Wiley & Sons, 2004.

Hernan, M A, Brumback B, Robins JM (2000). "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men." Epidemiology 11(5): 561-570.

Kurland BF, Heagerty PJ. Directly parameterize regression conditioning on being alive: analysis of longitudinal data truncated by deaths. Biostatistics 2005; 6:241-258.

Paik MC. The generalized estimating equation approach when data are not missing completely

at random, 1997; 92:1320-1329.

Power, M C, Tchetgen Tchetgen EJ, Sparrow D, Schwartz J, Weisskopf MG (2013). "Blood pressure and cognition: factors that may account for their inconsistent association." Epidemiology 24(6): 886-893.

Preisser JS, Galecki A, Lohman K, Wagenknecht L. Analysis of smoking trends with incomplete longitudinal binary responses. Journal of the American Statistical Association, 2000; 95:1021-31.

Preisser JS et al. A SAS/IML program for marginal mean modeling of longitudinal data with dropouts missing at random using inverse probability-weighted estimating equations. 2014 (in preparation).

Preisser JS, Lohman K, Rathouz PJ. Performance of weighted estimating equations in longitudinal studies with dropouts missing at random. Statistics in Medicine 2002; 21:3035-3054.

Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association 1995; 90:106-121.

Tchetgen Tchetgen EJ, Glymour MM, Shpitser I, Weuve J. To Weight or Not to Weight? On the relation between inverse-probability weighting and principal stratification for truncation by dealth. Epidemiology 2012a; 23:132-137.

Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins J. A cautionary note on specification of the correlation structure in inverse-probability-weighted estimation for repeated measures. (February 2012b). Harvard University Biostatistics Working Paper Series. Working Paper 140. http://biostats.bepress.com/harvardbiostat/paper140.

Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins J. Specifying the correlation structure in inverse-probability-weighting estimation for repeated measures. Epidemiology 2012c;644-646 [Letter].

Weuve J, Tchetgen Tchetgen EJ, Glymour M, Beck TL, Aggarwal NT, Wilson RS, Evans DA, Mendes de Leon CF. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. Epidemiology 2012; 23:119-128.

# Appendix A: ARIC ANALYSIS RECOMMENDATIONS: ANALYSES OF BASELINE EXPOSURE AND COGNITIVE CHANGE (updated 7/27/2015)

## *PART I. GENERAL RECOMMENDATIONS*

The purpose of this working document is to provide guidance to analysts currently working on ARIC NCS manuscripts of baseline risk factors and subsequent cognitive change over visits 2, 4, and 5. The recommendations in this document apply specifically to analyses modeling the relation of an exposure at visit 2 (baseline) with change in cognition, measured by cognitive test scores on the digit-symbol substitution, word fluency, and word recall from visits 2, 4, and 5. Lessons learned thus far are included, and may serve as guidance for future analyses of this type in ARIC NCS. Usefully, these recommendations can provide a means to methodological consistency across our manuscripts, which will help us more directly compare results across exposures. These guidelines should not be construed as mandatory for all ARIC manuscripts, which is unrealistic. However unconstrained flexibility in analysis models, especially when used to obtain statistical significance, has a higher risk of leading to false conclusions (Ioannidis, 2005 PLOS Medicine), hence analytic plans should be specified in advance.

These recommendations are not intended to apply to analyses considering other questions, such as the association between time-varying exposure and cognitive change or the relation of exposure to heterogeneity in individual patterns of cognitive change. While some recommendations, primarily the exclusions and choice to use z-scores computed in the entire population, may be applicable when considering other questions, very different methods, especially in terms of how one models the correlation across repeated cognitive measures, are likely to be needed for these alternate analyses. We expect that analysts considering such questions will work with their writing group and the NCS Analysis committee to develop recommendations for how to analyze the data appropriately for their particular question.

**Recommendations and Rationale**

- Exclusions: We recommend excluding individuals with the following characteristics: Not white in Washington County or Minnesota; not African-American in Jackson; not white or African American in North Carolina; missing education.
    - *Additional exclusions that may be commonly applied*: having no exposure measure or cognitive tests (please see section on Dealing with Informative Attrition towards the end of this document), excluding data from individual study visits (not persons) when the individual reports taking CNS altering medications (variable to-be-available in the derived dataset).
- Exposure: Analyst choice
- Outcome (Cognitive test performance at visits 2, 4, and 5):

- We currently recommend use of test-specific and global z-scores for digit-symbol substitution, word fluency, and word recall based on the mean and standard deviation of the scores in the entire population at baseline (visit 2)
  - z-scores are available from the Data Coordinating Center as a derived variable. Use of these z-scores will allow comparison of the estimated effects of different risk factors across studies; however, if exposure data is available on only a subset of the Visit 2 participants, it may be worth deriving z-scores specific to the population used in the analysis, to retain the interpretation of the z-scores as a 1 standard deviation change in the baseline score in that population.
  - Test-specific z-score $= \frac{\text{Test score} - \text{Mean score of that test at Visit 2}}{\text{SD of that test at Visit 2}}$
  - The global z-score is the sum of all 3 test-specific z-scores, standardized to Visit 2:

$$\frac{(\text{Sum of 3 test specific zscores}) - (\text{Mean of the sum of the 3 test specific z scores at Visit 2})}{\text{SD of the sum of the 3 test specific z scores at Visit 2}}$$

*Comment on use of z-scores in race-stratified and race-combined models*

- *We would like to be able to estimate the effect in the entire population, independent of race (i.e. a single model estimating the effect in both white and black participants, referred to as a "race-combined model" in this text.) However, there is concern that cognitive test scores may mean different things by race, i.e. that a white person and a black person with the same cognitive status would not receive the same cognitive test score due to cultural factors or educational background. Using z-scores computed separately by race in a single race-combined model is not acceptable. It makes the strong assumptions that race-specific z-scores have equivalent meaning <u>and</u> that there is no real difference in mean cognition by race (e.g. despite higher prevalence of diabetes in one race, the mean true cognitive performance is expected to be the same). As this second assumption does not hold in ARIC, we do not recommend using race-specific z-scores in a race-combined model. Use of z-scores computed in the entire population does not solve the problem that the meaning of cognitive scores may differ by race – especially when the standard deviation of scores differs by race – but does not make the additional assumption of similar true mean across groups. However, it should be noted that the three cognitive tests administered repeatedly at visits 2, 4 and 5 represent a set of tests that are much more likely to have the equivalent meaning across races, as compared to some of the other tests administered in the expanded battery at visit 5 (e.g. logical memory). As such, this seems to be a reasonable assumption and supports the use of z-scores computed based off visit 2 scores in the entire population.*
- *Race-stratified z-scores may be acceptable in the case of race-stratified models. However, we currently recommend using z-scores calculated in the entire population in race-stratified models for the sake of consistency of interpretation across models and manuscripts.*

- Exploratory Data Analysis:
  - We recommend use of spaghetti plots and visit-specific means and other univariate statistics to understand the trajectories of cognitive data.
    - Such exploratory data analyses should be repeated by exposure status and race
- Timescale:
  - We recommend using time on study as the primary time scale, and encourage analysts to consider adjustment for age at baseline, interaction between time on study with age at baseline, the potential for non-linearity in both time on study and age at baseline variables as described below.
  - *Based on completed work, the following model specification for time/age is recommended:*
    - Model time since baseline (visit 2) using a linear spline with a knot at 6 years (the approximate time of visit 4)
    - Center baseline age at the mean
    - Begin with both a linear and a quadratic term for age at baseline and interactions between both age at baseline and the time on study variables. Consider dropping these interactions or the quadratic age at baseline term if non-significant. Keep the quadratic term for age at baseline regardless of its significance if interactions with quadratic age at baseline are retained.
    - If you follow these recommendations, the parameters of interest to report are as follows:
      - Modification of slope 1 by exposure (coefficient of the interaction between exposure and first spline term)
      - Adjusted difference, by exposure status, in cognitive decline over the first 6 years of the study (per year or per 6 years)
      - Modification of slope 2 by exposure (coefficient of the interaction between exposure and second spline term)
      - Adjusted difference in cognitive decline from the $6^{th}$ year onward (per year or per 6 or 14 years)
      - Difference in model estimated 20 year change by differences in exposure status
      - P-value for impact of exposure on entire cognitive trajectory (Likelihood ratio, Chi-Squared / Wald, or F test)

- Covariates:
  - We currently recommend the following:

- Include demographics, dummy variables for study center (or race-center in race combined models), a priori theoretical confounders, and needed interactions of these covariates with time on study
    - *Note: To date, papers with major vascular risk factors as exposures of interest have considered the interaction between other major vascular risk factors and time as covariates. This is important as it allows for control of confounding of associations with cognitive change as well as confounding of associations with cognitive level.*
    - *e.g., If diabetes is a covariate and time is modeled using a linear spline with knot at year 6, include diabetes, diabetes*timespline1, and diabetes*timespline2 as covariates in the model*
- Center all continuous variables

*Comments on parameterizations of time on study, age, and covariates:*

- *Thus far, model-estimated cognitive change over time on study does not appear to be linear. Use of a linear spline with at knot at visit 4 (year 6) captures this non-linearity well given our relatively sparse data (three measures over follow-up).*
  - *Although this document largely discusses analyses using a single knot at 6 years, whether the use of more than one knot more appropriately describes the data should be carefully evaluated (see Diagnostics)*
  - *Note: Linear spline terms can be parameterized in several ways. In this document "slope 1" refers to the mean slope from 0 to 6 years and "slope 2" refers to the mean slope from year 6 to the end of the study. Alternate parameterizations can be used. For example, you can parameterize this spline as the linear change across the period of the study and the change in slope after the knot point. Analysts may choose to scale both so that beta estimates with respect to calendar time correspond to a per-year or a per 6-year change.*
- *Inclusion of quadratic terms may not be the most appropriate way to model non-linearity of age and inclusion of squared continuous variables can cause problems with collinearity or undue influence of very large quadratic terms in modeling.*
- *For longitudinal data analysis models, centering continuous predictors at a given value (e.g., mean) in general has been recommended in the research literature and can be a useful technique (Enders 2007 Psychological Methods; Kraemer 2004 International Journal for Psychiatric Research).*
  - *Centering baseline continuous variables at a given value (e.g., mean) provides for a meaningful interpretation for coefficients of terms whose order is less than the maximum (as a quantity at the centered value of the covariate). The interpretation of the highest order term is the same as for an uncentered covariate. For example, for a model with age (in years) and age-squared terms and no interactions with age, the interpretation of the age coefficient is the mean difference in outcome per one year of*

- Accounting for repeated measures design
  - Analysts may use either generalized estimating equations (GEE) models or mixed effects models. We recommend either:
    - A GEE model with unstructured correlation matrix for the within-person observations and robust variance estimates.
      - *An unstructured correlation matrix is appropriate here because our sample is large enough and has a flexible structure that includes other structures as special cases, including compound symmetry (which is equivalent to a random intercept only model). Given this flexibility and lack of a reason why particular correlation structure should be the truth, it is likely to fit the data better than use of less flexible options.*
      - *Robust variance estimates are appropriate given the large sample size and our uncertainty in how best to account for the correlation across repeated measures. In smaller subsample analyses where robust variance options may cause estimation difficulties or otherwise be unwarranted due to their reliance on asymptotic convergence (large sample theory), reliance on the assumed, model-based variance estimates may be necessary*
    - Mixed effects models with random intercept and random slopes. If the recommendation for linear splines to parameterize time on study is used, the corresponding mixed effects model will include a random intercept, a random slope for spline 1 and a random slope for spline 2. Under this parameterization, we recommend use of an independence covariance matrix for the random effects.
      - *Specifying an independence covariance matrix for the random effects here is appropriate and preferable to the alternatives. Given our sparse data (effectively 3 time points for cognitive data, although the mixed effects model does take advantage of the misaligned measurements) estimating an unstructured covariance matrix for the above random effects parameterization is not recommended – as you would be trying to estimate 6 parameters (3 variances and 3 covariances) this model likely will not converge properly. It is also preferable to other, simpler structures. Unlike a compound symmetry covariance structure, the independence structure estimates unique variances for each random effect; this allows for variation in the variance of the outcome data (cognitive data) that is dependent on time.*
      - *The random intercept, random slope for spline 1 and random slope for time 2 specification has also been shown in example analyses to fit the data better than an alternate random effects structure including a random*

*intercept and a single random slope for time on study with an unstructured random effects covariance matrix (which also requires estimating 3 variance-covariance parameters).*

- The choice between GEE and mixed effects models depends on several factors and should be driven by the research question of interest. The following table should help guide the choice:

| GEE Models | Mixed Effects Models |
|---|---|
| Appropriate for estimating marginal effects; <u>cannot</u> provide individual-level prediction | Appropriate for estimating marginal effects <u>and</u> providing individual-level prediction |
| May be <u>less</u> sensitive to high leverage of design whereby Visit 5 follows Visit 4 by ~14 years | May be <u>more</u> sensitive to high leverage of design whereby Visit 5 follows Visit 4 by ~14 years |
| Must be used for IPAW analyses | More compatible with MICE (and is required for dead MICE – see Part II) |
| Assumes MCAR | Assumes MAR |

- <u>*Note*</u>: *These recommendations do not apply to questions considering time-varying exposures. Such analyses have issues of endogeneity. Use of a working correlation matrix other than independence is subject to biases, and naïve longitudinal data analyses using an independence correlation matrix are likely to be biased as well. Other methods (e.g. marginal structural models based on a GEE model with independence correlation matirx) are required*
- <u>*Note*</u>*: If using inverse probability of attrition weighting* (*IPAW) as a sensitivity analysis to account for informative dropout, because the IPAW is conducted using SAS, the final GEE without IPAW should also be programmed in SAS, especially if reporting 3 decimal places. Differences between estimates comparing SAS to STATA have been observed at the 3rd decimal place, and we believe that SAS and STATA are using slightly different GEE estimators.*

- <u>Model building recommendations</u>:
  - Look at your data

- Given that a goal of ARIC is to identify novel risk factors for CVD in both blacks and whites, run both race-combined and race-stratified models
- Consider interactions between all covariates and time on study
    - *Because we chose all our confounders a priori, we are making the assumption that they are related both to the exposure and to the outcome (change in cognition over time on study). Thus, it is important to allow the covariates to be related to change in cognition over time. However, these relations may not be evident in our data. It may be appropriate to drop interactions if the covariate by time interaction is not significant*

- Diagnostics:
  - The primary diagnostic plot is a lowess plot (including the scatter) of the residuals from the model vs. time. Time should be treated in 3 different ways: (1) total time (V2-V5); (2) Visit 4 only ; and (3) Visit 5 only
    - Plots should be created for the entire population and for whites and blacks separately
    - These plots should be repeated, stratifying by the categorical exposure of interest (within each race group) (i.e., a separate lowess for each exposure category)
      - Examples: diabetes and no diabetes; smokers, formers smokers and nonsmokers; hypertension, pre-hypertension, no hypertension (all for the total population and also stratified by race)
    - The lowess should be horizontal at zero. If it is not, a model including additional knots for the linear spline for time should be fit (see point after next for recommended placement), findings evaluated for stability, and diagnostics repeated, to ensure a spurious association is not being induced by lack of model fit
    - The simpler model (one knot at year 6) will be acceptable if the model with additional knots does not substantially alter the exposure – cognitive change associations of interest (provided that the diagnostics for this second model suggest adequate fit)
    - Recommended knots to allow differential slopes between Visits 4 and 5: $5^{th}$ percentile of time at Visit 4,the $95^{th}$ percentile of time at Visit 4, and the $5^{th}$ percentile of time at Visit 5
      - This recommendation may not be ideal for all analyses; analysts are encouraged to use residual plots from his/her analysis to determine the number and location of possible additional knots
  - Other diagnostics:
    - Summarize mean residuals
      - Overall, by visit (V2, V4, V5) and by covariates in the model (e.g., age groups, race, sex, etc.)
      - Mean should be 0
      - If the model is under- or over-fitting for a particular group, additional modeling flexibility for that group (e.g. interactions) may need to be incorporated in the model
    - Scatterplots

        - If interactions between continuous variables are used in the model, look at scatterplots between these two variables – there should be data in each quadrant, want a spherical spread - e.g. age v time, bmi v time
    - Residual plots versus interaction terms should be examined
    - Additional residual plots:
        - histogram/Q-Q plot for normality of errors
        - residual v predicted for model misspecification (look for no systematic pattern, mean 0, constant variance)
        - residual v confounding covariates (to assess whether form of covariate is appropriate), adding a lowess line – want it to be horizontal at 0
        - partial residual plots versus time
    - Influence plots
    - *Note: Residuals for weighted models (e.g., IPAW) would require additional methods. The recommendation of this working group is that evaluation of the residuals from the unweighted models is adequate.*
    - *Note: Raw residuals may be used to evaluate inadequacy of the mean model; for residual plots; however, to accurately evaluate influence, a Cholesky-type decomposition is needed*

- <u>Dealing with informative attrition</u>: For risk factor-cognitive decline analyses where dropout over time differs by exposure status, we recommend analysts use available methods to mitigate bias due to selective attrition.
    - Using MICE estimates of missing cognitive scores (for both Visit 4 and Visit 5) for persons who were alive at the time of the visit easily accommodates the use of data from dementia ascertainment and appears to work well in practice. Specific recommendations on how to implement analyses using MICE are detailed in Part II. Discussion of alternatives to MICE can be found in the box below.

*Alternatives to MICE*

- *Inverse probability of attrition weighting on corresponding GEE model with independent covariance matrix and robust standard errors (please see note in section about repeated measures model choice about using SAS to estimate the GEE both with and without IPAW)*
  - *This may be preferable if you have more confidence in our ability to model missingness than in our ability to model cognitive test scores.*
  - *May be adapted to allow computation of estimands that directly address issues of competing risks (e.g. survivor average causal effect, or SACE estimate – Tchetchen Tchetchen 2014 Stat Med 33 (21): 3601-28)*
  - *IPAW may also be preferable for alternate study questions that do not estimate the impact of a baseline value of a risk factor on cognitive decline (e.g. studies of prevalence at Visit 5, studies considering time-varying exposures during follow-up that utilize inverse probability of exposure weights)*
- *Shared parameter models*
  - *Such models allow linkage with time to dementia analyses and may better address situations where we have reason to believe that the data are MNAR.*
- *Alternate sensitivity analyses:*
  - *Vary the extent to which missing values for those lost to follow up differ on average from those not lost to follow up—also interaction between exposure and this—and evaluate at what points of these hypothesized values the inference changes. Our current recommendation for implementing this strategy is to use multiple imputation (simulation/sensitivity analysis) adding in biases associated with loss to follow-up, both overall and differentially according to exposure.*
  - *It may also be useful to assign scores for drop-outs known to have dementia at a time proximate to the diagnosis, using empirical data for such assignment.*

## *PART II. IMPLEMENTING MICE (MULTIPLE IMPUTATION BY CHAINED EQUATIONS) FOR ADDRESSING INFORMATIVE ATTRITION*

The following recommendations from the ARIC Analysis Working Group are intended for risk factor-cognitive decline studies in which the outcome is the trajectory of global cognitive decline measured using a combination of DSST, DWRT, and WFT scores at Visits 2, 4, and 5. This method would be broadly applicable for studies of cognitive change with exposures measured at Visit 4 and in which global cognition is measured only at Visits 4 and 5.

These recommendations may also possibly be applicable to studies of test-specific cognitive decline measured at Visits 2, 4 and 5, but work up to this point has only focused on methods to predict the global z-score.

### A. Missing Outcome Data

1. **Recommendation**: For risk factor-cognitive decline analyses in which the research question of interest is to quantify the difference in rates of cognitive change by level of the risk factor, and in which dropout over time differs by exposure status, we recommend a primary analysis that incorporates MICE estimates of missing cognitive scores (for both Visit 4 and Visit 5) for persons who were alive at the time of the visit. As a sensitivity analysis, we recommend the additional inclusion of cognitive scores of persons who died during follow-up, estimated by MICE at a time 6 months prior to the date of death. Note, a simulation study is underway to evaluate whether this sensitivity analysis is likely to be conservative or anti-conservative; therefore, this recommendation may change based on the results of this simulation.

   **1.a. Recommendation**:
   For participants with missing cognitive outcome data, but who are alive at the start of a given visit, we recommend imputing a cognitive score at the time of the median observed date for that visit (July 1997 for Visit 4 and May 2012 for Visit 5). We recommend imputing missing cognitive scores for both Visit 4 and Visit 5.

   **1.b. Recommendation:** We recommend imputing scores for participants who died at a time 6 months prior to death, regardless of the date of death. Note: Dementia ascertainment is limited to death occurring in 2004 (when participants were at least 60 years of age) and later. Again, please note that this recommendation may change subject to the results of a simulation study that is currently in development.

### B. Missing Exposure and Covariate Data

2. **Recommendation:** For a given risk factor-cognitive decline analysis, we recommend investigators think carefully about possible mechanisms giving rise to missing exposure and missing covariate data. We recommend imputing missing exposure and covariate data if missing data is likely to be largely non-informative as a function of the covariates included in the imputation model. However, if the mechanism for missingness for a given variable is likely to be quite informative (i.e., dependent on the unobserved values of the variable), simpler imputation methods or sensitivity analyses may be warranted. For example, if laboratory values are missing because they are below the lower limit of detection, these values should not be imputed with MICE, but may be set to a value less that the lower limit of detection in a sensitivity analysis.

### C. Analytic Sample

3. **Recommendation:** Previous recommendations from the Analysis Workgroup stated that only participants with complete outcome data at the baseline visit (i.e., all three test-specific cognitive scores: DSST, DWRT, WFT) were to be included in the analysis. With MICE, we recommend additionally including those participants who attended the baseline visit regardless of whether the participant completed cognitive testing at that visit. Inclusion of these participants is important, as the reason for not completing testing is often informative. However participants with cognitive testing missing at all visits might be excluded.

### D. Building the Imputation Model

4. **Recommendation:** When building the imputation model, all variables that will be included in the analysis should be included in the model. Investigators should use the provided annotated STATA code.

5. **Recommendation:** For practical reasons, we recommend conducting exploratory data analysis with the complete (i.e., non-missing) dataset or with an imputed dataset consisting of 5-10 imputations. Results examining the number of imputations, showed very little change in estimated 20-year change after 5-10 imputations. However, for the final analysis, we recommend using a data set with a minimum of 25 imputations. This recommendation is based on the percentage of missing data in the analysis to develop the imputation model. In general, more missingness requires a larger number of imputations.

6. **Recommendation:** Regarding choice of model to estimate the association of the risk factor of interest with cognitive change over time, we recommend following Analysis Workgroup guidelines previously described in this document.

### E. Troubleshooting

We have found error messages in the MICE imputation (e.g., vce is not positive definite, non-concave likelihoods) are often due to difficulty in imputing afu variables. Use of the omit statement to drop a collinear variable [e.g., (logit, omit(i.collinear_varaible)) afu_variable] or complete omission of an afu variable may help the imputation model to run. However, the Workgroup recommends retaining afu variables that are known to be associated with cognitive decline, including but not limited to afu_diab1, afu_htn1, and afu_proxy1. This error message can also occur when imputing variables that are simple linear combinations of other variables (e.g. imputing HDL cholesterol, which is calculated from measured total and LDL coholeterol). Such variables should be omitted from the MICE and should be calculated after imputation of the contributing variables (i.e. passively imputed).

We have found that sometimes a given mixed model will run for days (we stopped after 7), compared to an average of less than an hour. This may be a case of the model failing to converge (but no error messages are displayed). If a mixed model runs for more than a couple hours, we recommend rerunning the MICE with a different seed value or a longer burn-in period.

## F. Analysis with Imputed Data

**7. Recommendation:** We recommend a nested model building process that allows investigators to compare how the inclusion of imputed data affects model estimates. Estimates included in the final manuscript are at the discretion of the writing team.

**MODEL 1**: Complete case. Only participants with complete exposure, outcome and covariate data are included in the model.

**MODEL 2**: Imputed covariate data. Participants with complete exposure and outcome data are included in the model. Includes imputed values for missing covariate data.

**MODEL 3**: Imputed exposure and covariate data. Participants with complete outcome data are included in the analysis. Includes imputed values for missing exposure and covariate data. May be omitted if there are few missing exposure values.

**MODEL 4**: Full MICE in living participants. Includes imputed covariate and exposure data, as well as imputed cognitive outcome data for participants who were alive at the time of a visit, but did not complete cognitive testing at that visit.

**MODEL 5**: Sensitivity analysis: 'Dead mice'. Includes imputed covariate and exposure data, as well as imputed cognitive outcomes for all participants.

In ARIC, participants with lower cognitive scores at baseline have higher mortality rates and, if living, are less likely to return for exams than participants with higher cognitive scores. Because dropout is informative, we would expect estimated rates of cognitive change to be faster in Models 4 and 5 than in Models 1-3.

If dropout also differs by exposure status, we would also expect that estimated differences in the rates of cognitive change comparing exposed to unexposed may be greater in Models 4 and 5 than in Models 1-3.

Although estimates are expected to change, drastic differences may indicate an incorrectly specified imputation model.

As a matter of good practice, we recommend simple exploratory analyses be undertaken prior to the imputation to determine if the exposure of interest is related to mortality or to loss to follow-up in living participants.

**8. Recommendation:** Models 1 and 2-4 (incorporating MICE) may be used with either model choice described previously in this document (GEE with unstructured correlation matrix and robust variance estimates or linear mixed models with random intercept and random slope). However, because of the unequal follow-up times introduced as a consequence of including imputed scores at a time 6 months prior to death, random effects models are necessary with 'Dead MICE' (MODEL 5), as GEE models require balanced data.

### G. Model Checking

7. **Recommendation:** In addition to standard model diagnostics, including plots of residuals vs. time, we recommend the following diagnostics for the MICE imputation:
   a. Comparison of imputed values to observed values (distribution of scores for continuous variables, frequency table for binary and categorical variables)
   b. Trace Plots
   c. 20% validation sample (see provided Stata code)

## *PART III. EXAMPLE CODE*

Example STATA and/or SAS code following these recommendations will be available in a later draft. Reference to this code will help promote consistency across analyses in terms of both variable coding (e.g. having the same reference category and same parameterization of the spline terms) and analysis (e.g. running the model).