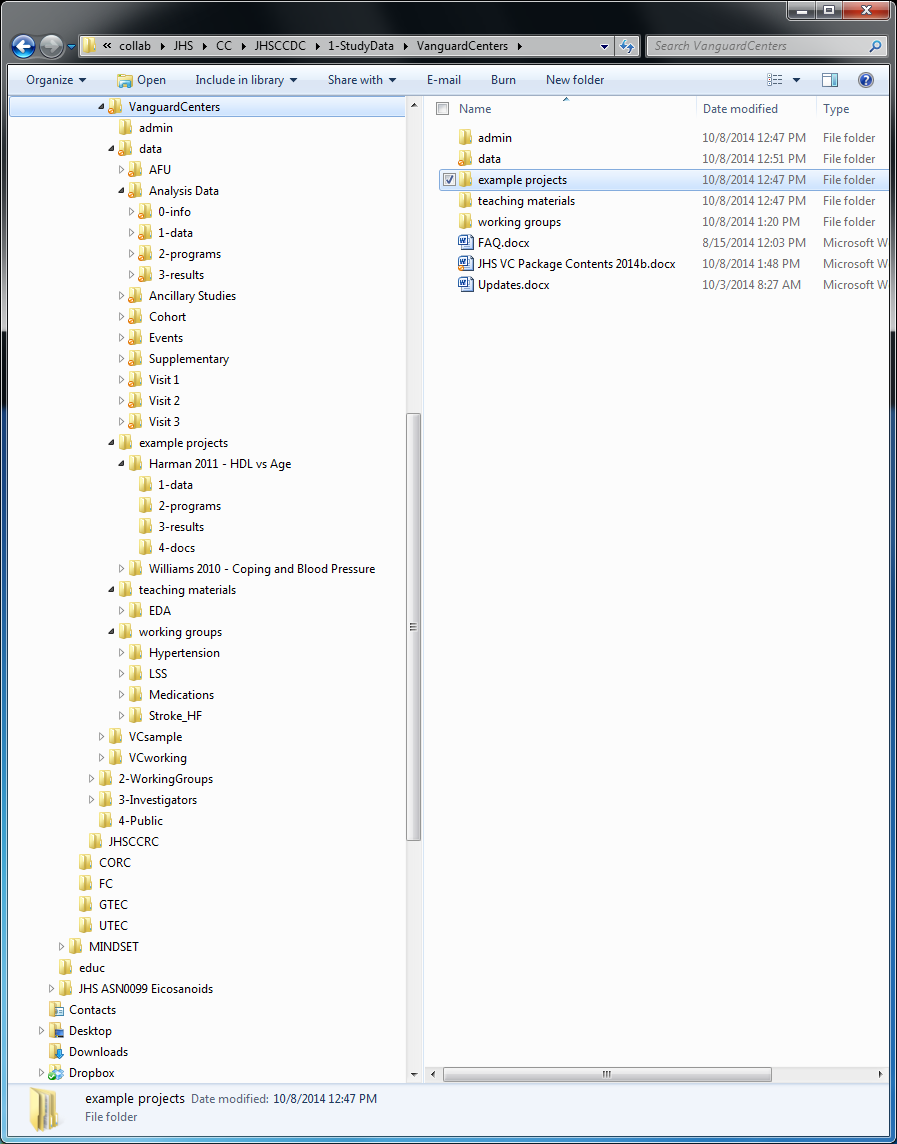
**The Jackson Heart Study Vanguard Center Reproducible Research Package**

The JHS Vanguard Center (VC) research package is updated, compiled and securely distributed by the JHS Coordinating Center Data Core (CCDC). The current VC package contains the following structure and elements:



**INFORMATION FOLDER** (VC description, DDAs, other docs)

**DATA FOLDER**

* + Annual Follow Up data (Shared ARIC & JHS-other Qs)
  + **Analysis-Ready Datasets (Start Here!)**
    - 0-Info (Data Dictionary, Codebooks, Info)
    - 1-Data (Datasets, SAS and Stata)
    - 2-Programs (SAS code to create, examine datasets)
    - 3-Results (SAS ODS results storage folder)
  + Additional Data from Ancillary Studies
  + Cohort data (Deaths, LTFU, annual contact info.) - *through 2013*
  + Adjudicated Events data (CHD, Stroke, HF) - *through 2011*
  + Supplementary and administrative data
  + “Raw” Visit 1 data (exam 1 datasets & codebooks)
  + “Raw” Visit 2 data (exam 2 datasets & codebooks)
  + “Raw” Visit 3 data (exam 3 datasets & codebooks)

**EXAMPLE PROJECTS FOLDER**

* + Harman 2011 example project
    - 1-Data (Project datasets)
    - 2-Programs (SAS code to replicate work)
    - 3-Results (SAS ODS results storage folder)
    - 4-Docs (proposal & final publication)
  + Williams 2010 example project (PhD Thesis)

**TEACHING MATERIALS FOLDER**

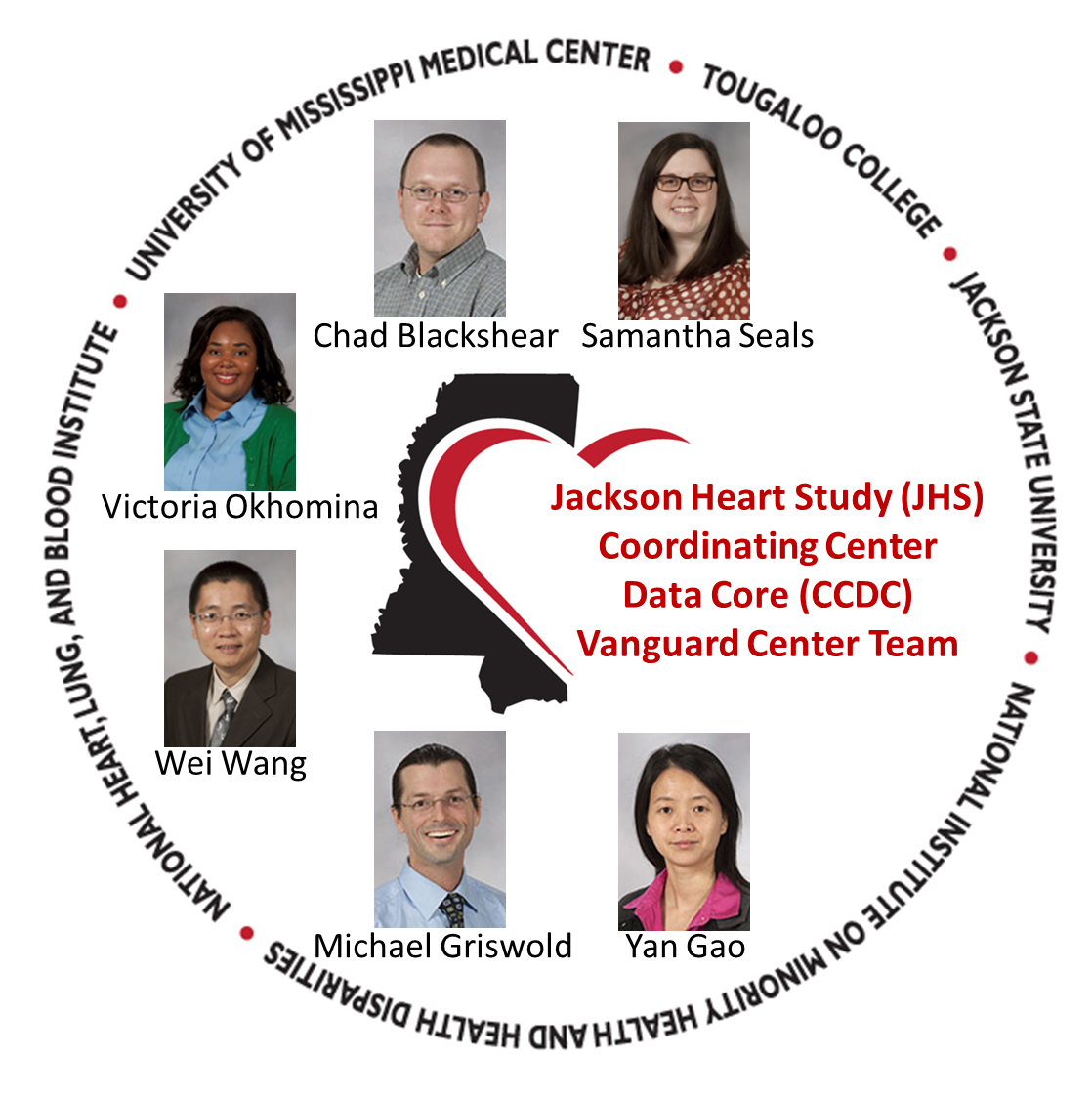
* + Harman 2011 example project
* **WORKING GROUP FOLDERS**
  + HTN WG folder (WG derived variables, programs, info, etc)
  + Life’s Simple 7 WG folder (WG variables, programs, info, etc)
  + Medications WG folder (WG variables, programs, info, etc)
  + Stroke/Heart Failure WG folder (WG vars, pgms, info, etc)

The JHS Vanguard Center project’s mission is to provide collaborators with the data, tools and liaisons to quickly progress from ideas to manuscripts and ancillary studies. We believe it is unique in its approach for encouraging and producing transparent and reproducible research. We do request feedback on the project; please send comments, questions, and suggestions of additional helpful items to include in our next distribution to:

Michael Griswold, Dir. Coordinating Center Data Core (CCDC), [mgriswold@umc.edu](mailto:mgriswold@umc.edu) or [JHSCCDC@umc.edu](mailto:JHSCCDC@umc.edu)

and/or Adolfo Correa (Int. PI), [acorrea@umc.edu](mailto:acorrea@umc.edu) or [JHSCCRC@umc.edu](mailto:JHSCCRC@umc.edu)

This JHS VC package would not be possible without support from the NHLBI & NIMHD, and the long hours and dedicated efforts of the incomparable Jackson Heart Study Coordinating Center Data Core (CCDC) team members including:



Continued Success in Research!

# General

What can Jackson Heart Study data be used for? And how can I get the JHS data?

The JHS data may be used only for the purpose of statistical reporting and analysis. Any effort to determine the identity of any reported case is prohibited. JHS does all it can to assure that the identity of data Participant cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted from the data files. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information.

Therefore, all users will:

1. Sign the approved Data Distribution Agreement (DDA): "JHS VC Investigator DDA.docx"

2. Use the data in these data files for statistical reporting and analysis only.

3. Make no use of any identifiable information discovered inadvertently and advise the JHS of any such discovery.

4. Follow JHS Publications and Presentations Subcommittee (PPS) and the Ancillary Studies Subcommittee (ANSS) Guidelines

For additional information on JHS data and/or JHS Vanguard Centers please email the JHS Coordinating Center (CC):

Data Coordinating Core: JHSCCDC@umc.edu

Research Coordinating Core: JHSCCRC@umc.edu

Where is a good place to begin?

The analysis datasets (analysis1, analysis2, analysis3, analysisLong and analysisWide) contain analysis-ready variables with supporting documentation and are a great place to begin. Further, the validation documentation (Analysis Data > 3-results > validation) contains EDA with outliers highlighted is useful while getting familiar with JHS data. Additionally, we have created an Analysis Data Dictionary (Analysis Data > 4-docs) to assist with understanding the variables available in the analysis datasets.

What data types are not available in the Vanguard Center package?

Due to the highly identifiable nature of the data, the following three data types cannot be made available in the Vanguard Center package:

1. Geocoded data
2. Genetics data – to obtain data, collaborators will have to go through DBgap
3. CMS data

Analyses involving these data types will have to be done at UMMC.

Where are the death data?

The death data are located in two places:

1. CHD-related deaths are in the allevtchd dataset (VanguardCenters > data > Events > 1-data); please see the CFATALDX variable.
2. An exhaustive listing of JHS participant deaths, through 2010, is in the LTFU dataset (VanguardCenters > data > Cohort > 1-data > ltfu2010.sas7bdat).

Are there known differences between the shared-ARIC cohort and the JHS-only cohort?

Since the JHS only subjects and shared-ARIC subjects were recruited at different times, the shared-ARIC cohort is, on average, older. It is suggested to use different baseline hazard function for these two cohorts when performing survival analysis.

What issues need to be considered when performing longitudinal data analyses across all three visits?

Possible items that need to be addressed before longitudinal analyses are performed are assay types, equipment changes, changes in standard of reporting, changes in units of measurement, etc. These topics will be discussed among working groups and once available, decisions and information will be included in future releases of the VC package.

# AFU

Where did the old AFUlong and AFODcomb datasets go?

The AFUlong dataset now encompasses the AFODcomb dataset. The variable VERS indicates which form data come from: A, B, C, D, E, OA, OB, OC, or OD.

Why is there no AFOE data?

Though AFUE has been implemented, JHS is still using AFOD.

How current is the AFU data?

The AFU data is now pulled close to the vanguard center package release date. The current AFU data is current through July 2014.

Where can I find information about the variables in AFUlong?

There is now a data dictionary for AFUlong; it provides the variable name, question wording on AFU questionnaire, and the variable number for each of the AFU forms. This data dictionary shows where data is missing (denoted by a black box) and moved to the AFO form (denoted by O and its question number).

# Analysis Data

How are the analysis datasets derived?

Variables that have been derived by the JHS CCDC are added to the analysis dataset. Documentation of these datasets is available in the analysis data dictionary, in the data > Analysis Data > 4-docs folder. This analysis data dictionary gives the description and definition of the derived variables as well as the “raw” variables the derivation utilizes. References for definitions, when available, are listed in the “reference” column of the data dictionary.

# Ancillary Studies

What is an ancillary study?

An ancillary study is an investigation which is not part of the Jackson Heart Study (JHS) protocol but uses all or a subgroup of the JHS cohort, samples, or data collected by JHS. In most cases, an ancillary study will involve the acquisition of additional interviews or examinations of study participants as well as analysis of blood, urine, tissue or other samples, or images previously collected, which are not compiled as part of the standard JHS data set.

Ancillary studies are subject to the same policies, reviews and approvals as the core JHS protocol. The highest priority will be given to studies which:

1. Have the highest scientific merit,

2. Do not interfere with the main study objectives

3. Produce the least burden on participants

4. Have objectives directly related to the study

5. Require the unique characteristics of the cohort

# Cohort

What does “LTFU” mean?

LTFU means “Lost to Follow-Up” and can happen in one of three ways:

1. Adjudicated death of participant, or
2. Participant tenders a hard refusal, indicating that he/she no longer consents to participation of any kind in the JHS,
3. Participant is censored; no contact has been made with participant since the date of previous contact.

The LTFU dataset contains a variable for the date of last contact and the type of last contact.

Are there any active potential validation items for the cohort dataset?

There are some instances of duplication in the cohort dataset based on subject ID and date.

* Exam date and AFU can occur on same date. Some participants have their clinic visit then go by office to do AFU in same trip
* We are verifying the instances of multiple AFU forms filled out on same date

There are some instances that AFU predates the first clinic exam visit

* AFU can predate clinic exam visits due to ARIC/JHS shared cohort

There are some instances where there appear to be skipped contact years

* It is possible that the participants were contacted the maximum number of times and never get in contact with them

# Events

Why are some participants missing follow-up information in the incident datasets?

Participants meeting the following conditions will have missing follow-up data in the incident datasets:

* prevalent disease (CHD, stroke, HF) at baseline;
* no consent for surveillance.

Please see the following link for more information: <https://www.jacksonheartstudy.org/Portals/0/Documents/pdf/EventsSummary2012.pdf?ver=2016-01-05-121301-883>

How are events monitored in JHS?

JHS obtains hospitalization data from two sources:  the annual follow-up (AFU) and the hospital discharge index from all catchment area hospitals.  Both of these sources can trigger an event investigation.  For example, a participant hospitalized outside of the JHS catchment area (like Memphis, TN) may report a hospitalization for heart failure.  This would not show up on our hospital discharge index, so this would be investigated separately, requiring an authorization from the participant to send to the out of state hospital when requesting a copy of records related to that hospitalization.  The self-reported data from AFU is often reconciled with the hospital discharge index and often match up (reported cases by participants match closely with hospitalizations on the hospital discharge index).  There are instances when the hospital discharge index is the more reliable indicator of a hospitalized event.  Older participants may not remember details of hospitalizations like dates.

There is also the PHF form that we send to physicians to fill out regarding physician office diagnosis of heart failure (AFU – D Questions 31, 32, 33\*). This helps us to establish incident HF outside the hospital.   If a participant indicates that a physician said they had HF or a weak heart, we ask them for permission to obtain information from their physician who first diagnosed it.  This is a short form.

Does a positive response to one or more questions on the annual follow-up survey (e.g., were you hospitalized for heart failure?) trigger an investigation of inpatient records?

If a participant reports being hospitalized for heart failure, this would be passed on to hospital surveillance for further investigation by matching it up to the discharge list or requesting records from an out of catchment hospital.

In JHS, are any HF hospitalizations identified that were not triggered from a self-reported annual follow-up question?

Yes, there are some hospitalized heart failure events that are identified solely by hospital discharge index.  We tend to try to rely less heavily on the self-reported data.  Another instance of less than optimal reliability on only self-reported data would be the cases where we were unable to contact a participant for an interview over the course of the year.  With the hospital discharge index, JHS would still be able to capture any hospitalizations related to CHD and heart failure.

What are the questions that trigger hospitalization surveillance?

The following questions would trigger a HF incident investigation: AFU – D 34\* and 35\*.  The remaining questions refer to specifics about a hospitalization for HF over the past year (since the last interview).

Note that for previous versions of the AFU, the AFU question numbers may have changed.

How is adjudication handled in JHS?

Adjudication for CHD, HF and stroke is completed using the same process as defined by the ARIC study, through the data coordinating center at the University of North Carolina. Hospital records are pulled for JHS participants and ICD codes are examined for specific events.

* HF diagnosis form: <http://www2.cscc.unc.edu/aric/sites/default/files/public/forms/HDXB_Form_012313.pdf>
* Stroke diagnosis form: <http://www2.cscc.unc.edu/aric/sites/default/files/public/forms/SDXB_Form.pdf>
* MI diagnosis form:

<http://www2.cscc.unc.edu/aric/sites/default/files/public/forms/CDXE%202013JUL10.pdf>

* HFA form (HF): <http://www2.cscc.unc.edu/aric/sites/default/files/public/forms/HFAC_for%20web_062512.pdf>
* HRA form (MI):

<http://www2.cscc.unc.edu/aric/sites/default/files/public/forms/HRA_Form.pdf>

* CHI form (MI and HF): <http://www2.cscc.unc.edu/aric/sites/default/files/public/forms/CHI_Form.pdf>
* STR form (stroke): <http://www2.cscc.unc.edu/aric/sites/default/files/public/forms/STR_form.pdf>

When did adjudication begin for events?

Stroke and CHD events have been adjudicated since the beginning of JHS. Heart failure, however, did not begin the adjudication process until 2005.

How does the adjudication date for heart failure affect analyses?

Because of the left censoring of heart failure events, possible definitions and suggested analytic approaches for incident heart failure are currently under discussion with the heart failure working group. Once defined, this information will be released in a future VC package.

What happens if one of the shared ARIC participants has an event before enrolling in JHS?

Participants were excluded if they have a stroke/CHD event before enrolling JHS.

# Supplementary

Why was the day of every participant’s birth set to 15 in the elga dataset? And how can I calculate the participant’s actual birthday?

Due to HIPAA regulations, the participant's date of birth in these datasets is based on a "mock" date of birth; the day of every participant's birth was set to 15. For this reason, some degree of error is expected in the age calculation based on the datasets provided in the Vanguard Center package. See the "frozen data" folder for an age variable that is calculated on the participant's actual birthday.

# Visit 1 “Raw Data”

What are the dates for Visit 1?

Visit 1 clinic exams began 09/26/2000 and concluded 03/31/2004.

What data was collected at Visit 1?

In the Vanguard Center package, Data > Visit 1 contains all of the data collected at Visit 1. For corresponding forms, please see the JHS Visit 1 form website, located at <http://jacksonheartstudy.org/jhsinfo/ForResearchers/FormsManuals/Exam1Forms/tabid/109/Default.aspx>

Where is the heart failure diagnosis variable?

As noted in the analysis data dictionary, direct questions on heart failure history were not collected at Visit 1. However, there is a heart failure question asked in the annual follow up questionnaire but responses were not adjudicated until January 1, 2005. We are currently working with the heart failure working group to derive a heart failure variable and once derived, will be included in a future release of the Vanguard Center package.

Why did only a small sample (276 participants) have both spot urine sodium and 24-hour urine sodium measured? Why are there so many participants (750) who have 24-hour urine sodium only? In looking at the lab dates for both measures, we see that the range of dates for spot urine sodium was 6/2002 to 3/2004, and for 24-hour urine sodium was 9/2000 to 4/2004. Why was spot urine sodium started 2 years later than when 24-hour urine sodium started?

Due to the nature of the 24-hour urine collection, there was a low participation rate. Thus, JHS switched to spot urine testing later in V1.

*NOTE: There are 2275 with only spot urine sodium, 750 with only 24-hour urinary sodium and 276 participants have data for both spot urine sodium and 24-hour urine sodium.*

# Visit 2 “Raw Data”

What are the dates for Visit 2?

Visit 2 clinic exams began 10/10/2005 and concluded 12/30/2008.

What data was collected at Visit 2?

In the Vanguard Center package, Data > Visit 2 contains all of the data collected at Visit 2. For corresponding forms, please see the JHS Visit 2 form website, located at <http://jacksonheartstudy.org/jhsinfo/ForResearchers/FormsManuals/Exam2Forms/tabid/110/Default.aspx>

# Visit 3 “Raw Data”

What are the dates for Visit 3?

Visit 3 clinic exams began 02/26/2009 and concluded 01/31/2013.

What data was collected at Visit 3?

In the Vanguard Center package, Data > Visit 3 contains all of the data collected at Visit 3. For corresponding forms, please see the JHS Visit 3 form website, located at <http://jacksonheartstudy.org/jhsinfo/ForResearchers/FormsManuals/Exam3Forms/tabid/124/Default.aspx>

How was the MRI data collected?

The collection of the Visit 3 MRI data actually began in Visit 2. For this reason, there are two copies of the “MRIB” dataset, one in the Visit 2 “raw data” and one in the Visit 3 “raw data”. The copy in the Visit 2 “raw data” houses only the MRI results for participants that had an MRI during the exam 2 time window (2005-2008). Also worth noting, some of the participants that had an MRI during the exam 2 time window were given another MRI during the exam 3 time window.

How were participants selected to have MRI data collected?

When they came to clinic for the exam they were asked if they would like to participate in the MRI, contingent upon whether the following conditions were met:

1. They were less than 350 lbs.
2. If Not pregnant. (Pregnancy test was done for female < 45.)
3. No Pace maker
4. No Gun Shot- metal in the body

Further, the participant must meet the criteria to be eligible for MRI with contrast as determined when Clinic staff and the MRI Technologist complete the JHS MRI Safety Screening Form; this includes all the standard exclusion criteria related to implanted devices and foreign bodies.

Additional Exclusion Criteria:

1. Kidney failure
2. Kidney transplant
3. Liver failure
4. Having a known allergy to contrast agents-commonly referred to as “Gadolinium”

How is creatinine (serum and urine) measured in visit 3?

Dates of use: 2006 - current

Creatinine is measured in serum, EDTA plasma or urine by the Roche enzymatic method (Roche Diagnostics, Indianapolis, IN 46250) on a Roche Modular P Chemistry Analyzer. (Roche Diagnostics Corporation). In this enzymatic method, creatinine is converted to creatine by creatinase. Creatine is then acted upon by creatinase to form sarcosine and urea. The sarcosine is converted to hydrogen peroxide which reacts with a chromophore to produce a colored product that is measured colorimetrically. This method has the advantage over the Jaffe method in that it is not susceptible to interferences from no-creatinine chromogens. The method is calibrated using a National Institute of Standards and Technology (NIST) standard traceable to reference material SRM 909b (Isotope Dilution Mass Spectroscopy (IDMS)). The laboratory CV is 2.3% for serum and EDTA plasma; for urine, 4.3% at a concentration of 18.39mg/dL and 1.5% at a concentration of 96.57mg/dL.

Other important information from an analytical perspective is how creatinine was calibrated in this exam and in exam 1. We are still trying to get information on how creatinine was measured in exam 1.