The BACH Study Data Archive

The Boston Area Community Health (BACH) Survey study archive contains the Manual of Operations, Forms, Study Data, and Dataset Integrity Check.

BACH is a population-based observational research study that was conducted in 2002-2005 (BACH I). Follow-Up assessments of BACH were conducted between 2008 and 2010 (BACH II), and 2010 and 2012 (BACH III).

Archive files are organized into the following directories:

- Documentation
- Forms
- Data
- Dataset Integrity Check (DSIC)

Documentation

The BACH Documentation directory contains:

- BACH 3 readme.docx – Contains information regarding the BACH 3 dataset, including important variables, and the labeling system and naming conventions for variables
- Bibliography.docx – List of BACH published manuscripts as of September 29, 2015
- Manual of Operations – This folder contains the following documents:
    - 1 – Overview of the Study.doc – Contains information regarding the objectives, research design, and schedule of measurements for the BACH study
    - 1.2 Study Flow Chart .doc – Design of study
    - 2 – Study Portocol_Archive.doc – Protocol for the BACH study
    - 3 – Screening_Archive.doc – Contains information regarding the screening process for the BACH study, including eligibility requirements
    - 4 – Data Collector Manual_4.7-4.13_Archive.doc – Contains instructions for interviewing subjects, both by phone and in person
    - 4.14 – Baseline Interview.doc – Contains instructions for conducting the baseline interview

Forms

The BACH Forms directory contains:

- BACH I – List of forms used in Phase 1 of the BACH study:
    - F0IN_Variable.doc – Baseline Interview
    - F0PM_Variable.doc – Physical Measures form
    - F0SC_Variable.doc – Specimen Collection form
    - F0SM_Vairable.doc – Men's Self-Administered form

- F0SW_Vairable.doc – Women's Self-Administered form
- BACH II – List of forms used in Phase 2 of the BACH study:
  - BACH Freq Codebook fupm_final_version_B.pdf – Physical Measures form (Version B), annotated with means and frequencies for all variables
  - BACH Freq Codebook IAQ_final_version_C.pdf – Follow-Up Interview (Version C), annotated with means and frequencies for all variables
  - BACH II Frequency Codebook Pain_final.pdf – Pain Form (Version C), annotated with means and frequencies for all variables
  - BACH II IAQ w-field codes_Version_A.doc – Follow-Up Interview (Version A), annotated
  - BACH II_IAQ_version_C.pdf – Follow-Up Interview (Version C)
  - BACH II MSAQ w-field codes_version_B.doc – Men's Self-Administered Form (Version B), annotated
  - BACH II PM w-field codes_version_A.doc – Physical Measures Form (Version A), annotated
  - BACH II WSAQ w-field codes_version_B.doc – Women's Self-Administered Form (Version B), annotated
  - CODEBOOK BACH II MSAQ_Version_C.doc – Men's Self-Administered Form (Version C), annotated
  - CODEBOOK BACH II PAIN FORM_Version_C.doc – Pain Form (Version C), annotated
  - CODEBOOK BACH II PM_Version_B.doc – Physical Measures Form (Version B), annotated
  - CODEBOOK BACH II WSAQ_Version_C.doc – Women's Self-Administered Form (Version C), annotated
- BACH III – List of forms used in Phase 3 of the BACH study:
  - IAQ (annotated).doc – Follow-Up Interview
  - Physical Measures Form (Annotated).doc
  - PRE-D Phlebotomy Collection Form_2010_01_06_Annot.doc

Data

The BACH data directory contains the following files, and the following raw datasets in Windows-based *SAS, v.7-8-9* format:

- BACH_1 – Contains the following datasets from the first phase of the BACH study:
  - diuretic.sas7bdat – Dataset from Phase 1 of the BACH study with information about diuretic use for all subjects
  - repositmain2.sas7bdat – Dataset for Phase 1 of the BACH study
  - formats.sas7bcat, formats.sas7bdat – List of formats for Phase 1 of the BACH study
  - readme.txt – Directions for reading in the formats file in SAS
- BACH_II – Contains the following datasets from the second phase of the BACH study:

- bach2publicuse.sas7bdat – Dataset for Phase 2 of the BACH study. This dataset was created as a result of 15 rounds of multiple imputation, and contains 15 records for each ID
- formats.sas7bcat, formats.sas7bdat – List of formats for Phase 2 of the BACH study
- complete_bach2_pre_mi.sas7bdat - Original dataset for Phase 2 of the BACH study. This is the original data, before multiple imputation, and contains one record per subject. Missing data is included.

- BACH_III – Contains the following datasets from the third phase of the BACH study:
  - bach3publicuse.sas7bdat – Dataset for Phase 3 of the BACH study
  - formats.sas7bdat, formats.sas7bdat – List of formats for Phase 3 of the BACH study
  - BACH III Codebook.pdf – Contains formats, frequencies, and means for all variables in the bach3publicuse.sas7bdat dataset
- Contents.pdf – Contents listing of all datasets in the Data directory.

- Readme_Formats.txt - Instructions for reading in the formats files in SAS programs

DSIC

The BACH Data Integrity Check (DSIC) directory contains:

- BACH_I_DSIC.pdf – A report of an examination of the BACH 1 datasets for completeness by statisticians and quality control specialists at the Repository. This dataset was used to replicate Table 1, Table 2, and Table 3 from the paper 'The Association of Nocturia with Cardiac Disease, Diabetes, Body Mass Index, Age and Diuretic Use: Results from the BACH Survey' by Fitzgerald et al. Published results from the BACH 1 data were compared to values recalculated from BACH 1 archived datasets in the NIDDK repository.

- BACH_I_DSIC_Addendum.pdf – Additional analysis for the BACH_I_DSIC.pdf document using the full legacy medications dataset

- BACH_II_DSIC.pdf - A report of an examination of the BACH 2 dataset for completeness by statisticians and quality control specialists at the Repository. This dataset was used to replicate Table 1 from the paper 'Treatment Status and Progression or Regression of Lower Urinary Tract Symptoms among Adults in a General Population Sample' by Maserejian et al. Published results from the BACH 2 data were compared to values recalculated from BACH 2 archived datasets in the NIDDK repository.

- BACH_III_DSIC.pdf – A report of an examination of the BACH 3 dataset for completeness by statisticians and quality control specialists at the Repository. This dataset was used to replicate Table 1 from the paper 'Association of African genetic ancestry with fasting glucose and HbA1c levels in non-diabetic individuals: the Boston Area Community Health (BACH) Prediabetes Study' by Meigs et al. Published results from the BACH 3 data were compared to values recalculated from BACH 3 archived datasets in the NIDDK repository.