

ICDB Data Archive

The ICDB Data Archive distributed on this CD-ROM contains: (1) the study protocol, manual of procedures, and other descriptive documentation; (2) study forms; (3) data collected by the study; and (4) data derived by analysis.

The files are organized into the following six directories:

1. Documents
2. Forms
3. SAS individual_data
4. SAS derived_data
5. SAS formats
6. SAS summaries

1. Documents Directory

The Documentation directory contains documents provided by the study, including the protocol and Manual of Operations. Additional descriptive documentation developed by and stored in the NIDDK Data Repository is also included. Documents are stored in PDF format. [Table 1](#) lists the documents and the associated PDF files, and provides a brief description of each document.

2. Forms Directory

The Forms directory contains the forms used for data collection. The forms are provided in Adobe PDF format. [Table 2](#) lists the data collection forms, the PDF file containing the form, and the associated SAS dataset(s) and summary report.

3. SAS Individual Data Directory

The SAS Individual Data directory contains the data collected by the study, comprising 52 datasets in SAS v7 format. Refer to [Table 2](#) for a list of the SAS datasets and the associated data collection form(s) and SAS summary report.

4. SAS Derived Data Directory

The SAS Derived Data directory contains the analysis datasets that have been provided by the ICDB investigators. There are four datasets in SAS v7 format. [Table 3](#) lists the datasets and the SAS summary report for each dataset.

The baseline dataset, **baseline.sas7bdat**, contains baseline demographic information, including selected variables from the baseline; i.e. V_TYPE=0 records from the following CRF forms: SUBJSTAT, BACK, PURG, SYMPTS, SYMHX, INCL, CYST, UROD, VOID, WITHD, REIN, SURVEY, PRIOR, MEDHX, and FHX. The dataset contains 637 records.

The merged analysis dataset for the biopsy data, **biopcyst.sas7bdat**, contains the records from the BIOP, BIOP_REV, CYST, DCYST forms, plus variables in the baseline.sas7bdat dataset. The dataset contains 2385 records.

The longitudinal analyses dataset, **longanal.sas7bdat**, contains the analytic dataset for longitudinal analyses. This data set contains one record per patient per visit date seen. Hence, if a patient's voiding log was 5 days prior to their visit date, the voiding log record and visit information is stored on separate records. The dataset contains 10656 records.

The summary of the voiding log, **sumvlog3.sas7bdat**, contains data for each of the 3 days with derived variables for total number of nocturnal voids, total number of 24 hour voids, total volume, average volume, minimum volume, and maximum volume. The dataset contains 15328 records.

5. SAS Formats Directory

The SAS Formats directory contains the format library for all of the individual and derived datasets and the SAS program used to generate the library.

6. SAS Summaries Directory

The SAS Summaries directory contains a SAS summary report for each of the individual datasets and derived datasets. These reports are text files, and they list the variables in each dataset, along with the variable's data type, length, format, and label. Refer to [Table 2](#) and [Table 3](#) for a list of the SAS datasets and the associated SAS summary reports.