

The following are data handling instructions for the Public Use Dataset of the NIDDK-funded Evaluating Predictors & Interventions in Sphincter of Oddi Dysfunction (SOD) (EPISOD - U01 DK074739). The dataset is comprised of patients enrolled into the EPISOD and EPISOD2 studies.

### **General Notes**

- Files
  - ReadMe PDF file (pdf)
  - FTP Checklist (pdf)
  - 43 dataset SAS files: including 40 CRF datasets, 1 derived dataset, 1 data dictionary and 1 code list SAS file
  - Contents file – pdf of proc contents by form
  - Data Collection Schedule (pdf)
  - 40 Annotated Case Report Forms pdf file
  - 1 pdf file of EPISOD publications
- Datasets
  - A complete listing of all variables, their description and their codes may be found in the data dictionary file and code list file.
  - There is a separate dataset for each included form.
  - Forms that were collected only once during the study will have one row per subject.
  - Forms that were collected multiple times during the study will have more than one row per subject.
  - Forms that contain logs may have more than one row per subject.
  - A complete listing of which forms were collected at each study visit can be found in “EPISOD Data Collection Schedule”.
  - Seven CRF datasets were excluded from this public use dataset:
    - Forms 34 and 38 captured patient contact information
    - Form 24 was a UB04 billing statement
    - Forms 22 and 23 captured information on the concomitant medication and treatments during follow up. Due to the complexity of this form, the information that was used for the EPISOD analysis was derived and included in the sas file titled, ‘Derived’
    - Form 04 captured information on prior medication use but due to the data quality the study team refrained from analyzing these data and including these data in the EPISOD public use dataset
    - Form 05 captured information on physical exam at baseline as was never used for data analysis or reporting
- Subject and Site IDs
  - Subject and Site IDs are unique identifiers for each subject and are not the original ID used by the study team.
- Variable names
  - Raw data are named in the format F##Q##, where F## is the form number and Q## is the question number. For multiple choice questions (ex. Question05 on Form33), the data are named in the format F##Q##M#, where M# is the option number. The data dictionary provides the variable names. For logs, the data are

named in the format F##CQ#, where C means the question is in the log. For the date variables, the date in CRFs are named in the format F##Q##\_d and the date of form collected are named in the format Formdate\_F##.

- Derived variables are included in some datasets, indicated below in Dataset Specific Notes.
- Dates and times
  - All dates are replaced with calculated days from the start date (the randomization date for EPISOD patients or the initial treatment date for EPISOD2 patients). These variables are expressed as the number of days from the start date.
  - Negative times indicate the event occurred before the start date.
  - If the day of a date is missing, the date is treated as missing since the scenario is rare.
- Text fields, including clinical narratives and general comments, are not included in the public use dataset in order to avoid releasing potential identifying patient information.

### **Dataset Specific Notes**

- Derived: dataset including the study number(EPISOD OR EPOSOD2), the treatment arm randomized ('ztreatmentnm'), days from study start to re-intervention, the primary outcome ('success'), the three primary analysis criteria indicators, per protocol population indicator, the site ID and the real treatment received ('real\_trt').
- Form 00: Subject Enrollment Form
- Form 01: Inclusion Exclusion Criteria
- Form 02: Functional Biliary Disorders Module
- Form 03: Medical History
- Form 06: Vitals
- Form 07: Complete Blood Count (Baseline)
- Form 08: Liver/Pancreas Test (Baseline)
- Form 09: HCG Blood Serum-H.Pylori Test
- Form 10: Abdominal Imaging
- Form 11: MINI International Neuropsychiatric Interview
- Form 12: RAPID
- Form 13: Beck's Depression Inventory II
- Form 14: HADS
- Form 15: SF-36 (v2)
  - Questions 23-31 used the official version1 SF-36 form.
- Form 16: Rome III Diagnostic Module
- Form 17: Sphincter of Oddi Manometry– Unblinded
- Form 18: ERCP– Unblinded
- Form 19: Trauma Scale
- Form 20: Coping
- Form 21: Baseline Resource Utilization Questionnaire
- Form 25: Adverse Events
  - Including MedDRA Preferred term - variable 'name4'
- Form 26: Follow Up Resource Utilization Questionnaire
- Form 27: Review for Re-Intervention
- Form 28: Liver/Pancreas Test (Follow Up)
- Form 29: Complete Blood Count (Follow Up)
- Form 30: Best Guess (Study Coordinator)

- Form 31: Best Guess (Subject)
- Form 32: Best Guess (Evaluating Physician)
- Form 33: Unscheduled Clinic Assessment
- Form 35: End of Study
- Form 36: RAPID– Central Caller
- Form 37: Subject Randomization Form– Unblinded
- Form 39: RAPID START
- Form 40: Post Procedure Labs
- Form 41: Subject Treatment Form- Unblinded
- Form 42: EPISOD2 Subject Treatment Form
- Form 43: EPISOD2 Sphincter of Oddi Manometry
- Form 44: EPISOD2 ERCP
- Form 45: Adverse Events- Unblinded
  - Including MedDRA Preferred term - variable 'name4'
- Form 46: Monthly Abdominal Pain Assessment

NOTE: These SAS data files are based on EPISOD study database frozen on OCT 18<sup>th</sup>, 2013. All potential identifying fields have been removed.