"Application of Transcriptional Signatures for Diagnosis of Febrile Infants within the Pediatric Emergency Care Applied Research Network"
Targeted Issues Grant Public Use Dataset (PUD)
Overview of Public Use Datasets

This document provides an overview of the general principles used in creating the public use dataset for the PECARN 2009-2010 Biosignatures analysis. A single dataset is available as both a CSV and a SAS® (.SAS7BDAT) file (with formats and labels). Accompanying documentation and resources provided include the study protocol, the data dictionary, culture and viral review and follow-up process documentation and a PDF document summarizing variables in the associated dataset (i.e., frequency distributions or descriptive statistics). For SAS software users, an example script to set up the SAS library and apply SAS formats is also provided. The data dictionary should be referenced frequently during analysis as this is the most complete reference of all variables included in the dataset. Please see the attached Research Data Use Agreement (RDUA) for a description of intended use and disclaimer.

GENERAL PRINCIPLES FOR CREATION OF DATASET

1. The population for the public use dataset is a convenience sample of infants enrolled in the PECARN Biosignatures study from 2009 to 2010. See the protocol for a detailed list of screening criteria for the Biosignatures study and the manuscript for a description of how infants were selected into the convenience sample.

2. The dataset is primarily based on raw data as collected by the clinical sites, culture and viral test results as reviewed by study investigators (see Essential study documentation), and the results of microarray analyses. All variables are described in the data dictionary.

3. Open text fields and other variables have been reviewed for sensitive or identifying information and modified as needed.

4. Within the dataset, the TIGSubjectID is a masked identifier for infants and geo_accession is an identifier for the corresponding microarrays in the GEO repository (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64456). The data dictionary describes the structure of the dataset and the unique identifier for each record.

LIST OF PUBLIC USE DATASETS AVAILABLE

- Clinical characteristics, infection status and Biosignature classification (TIG).