

**Collaborative Pediatric Critical Care Research Network**  
**Critical Pertussis Study**  
**Overview of Public Use Datasets**

This document provides an overview of the general principles used in creating the public use datasets for the *Eunice Kennedy Shriver* National Institute for Child Health and Human Development Collaborative Pediatric Critical Care Research Network (CPCCRN) Critical Pertussis Study. There are a total of 20 datasets available as both CSV and SAS® (.SAS7BDAT) files (with formats and labels). Accompanying documentation and resources provided include the study protocol, the annotated Case Report Form (CRF), and, for each dataset, 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 annotated CRF should be referenced frequently during analysis as this is the most complete description of all variables included in each 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 all consented patients, excluding those who withdrew permission prior to baseline data collection. See the protocol for a detailed list of inclusion/exclusion criteria and for additional details of study conduct.
2. The datasets are primarily based on raw data as collected by the clinical sites. In addition, we have included calculated survey scores for the Infant & Toddler Quality of Life Questionnaire (ITQOL), Child Health Questionnaire (CHQ), and Mullen Scales of Early Learning. All variables are described in the annotated CRF.
3. Variables related to the one year Mullen assessment are provided in two datasets. These datasets are identical except that one uses the child's age for the T-scores and the other uses the child's age after adjustment for prematurity if the child was < 37 weeks gestational age at birth. The primary analyses of one year outcomes for this study were based on the scores adjusted for prematurity.
4. Open text fields and other variables have been reviewed for sensitive or identifying information and modified as needed. Dates were recoded to be the number of days since PICU admission.
5. Within each dataset, the SubjectID is a masked identifier. The annotated CRF provides information as to the structure of each dataset and the unique identifier for each record.
6. Daily data collection during the PICU course utilized a "study day" (Day 0, 1, etc.) that is defined differently than simple calendar day based on the time of admission to the PICU. See the annotated CRF and study protocol for details.

## LIST OF PUBLIC USE DATASETS AVAILABLE

- Demographics (DEMOGRAPHICS)
- Eligibility and Consent (ELIGIBILITY)
- Subject Status (SUBJECTSTATUS)
- Infant & Toddler Quality of Life (ITQOL)
- Child Health Questionnaire (CHQ)
- Subject and Birth Information (SUBJECTINFO)
- Subject and Birth Information – Baby Problems (BABYPROBLEMS)
- Pertussis Illness Information (PERTILLNESS)
- Initial PICU Chart Abstraction (INPICUCA)
- Initial PICU Chart Abstraction - Immunization Status (IMMUNESTATUS)
- Pertussis Laboratory Data (PERTLAB)
- Documented Past Medical History (PASTMEDHX)
- Daily PICU Data Collection (DAILYDATA)
- Daily PICU Data Collection – Bacterial and Fungal Microbiology Reports (BACTFUNGMICRO)
- Daily PICU Data Collection – Hematology (HEMATOLOGY)
- PICU Discharge Information (PICUDC)
- One Year Follow Up (ONEYEAR)
- Mullen Assessment (without adjustment for prematurity) (MULLEN)
- Mullen Assessment (with adjustment for prematurity) (MULLENADJ)
- Death Information (DEATH)