This readme file documents the programs that are contained in this repository in support of the PCORnet Antibiotics and Childhood Growth Study publications.

The logical order of the programs and packages is as follow:

Distributed Packages:

1. Antibiotics SSDC Wave1.zip
2. Antibiotics SSDC Wave2.zip
3. Antibiotics All-Aims Individual-level Query.zip
4. Antibiotics Aim1 Distributed Regression Analytic Package.zip
5. Antibiotics Maternal-Child Linkage Query.zip

Analytic codes:

1. Antibiotics Aim 1 Analytic Package Part 1.sas
2. Antibiotics Aim1 Analytic Package Part 2.sas
3. Antibiotics Maternal-Child Linkage analytic package part 1.sas
4. Antibiotics Maternal-Child Linkage analytic package part 2.sas

The *distributed packages* are zip files that are sent to data-contributing sites for execution. The sites need to unzip the package, update the paths to match their local file/folder structure, and execute the package. No other modification is required. The packages include all the necessary folder structures, input files, and SAS programs (version 9.3 or newer). They are assigned a unique request tracking identifier and typically contain the following standardized folder structure:

* **sasprograms**: folder containing the master SAS program that must be edited and then executed locally. The master SAS program is the only program to edit.
* **dmlocal**: folder containing output generated by the request that should be saved locally but not returned to the analysis center. Outputs may be used locally or to facilitate follow-up queries.
* **drnoc**: folder containing output datasets and logs generated by the request that should be returned to the analysis center. The tables consist of aggregate output and transfer the minimum information required to answer the analytic question. None of these files contain patient-level data.
* **infolder**: folder containing all input and lookup files needed to execute the package. Input files are created for each request by the analysis center; the contents of this folder should not be edited. This folder contains all the SAS codes in the macros subfolder.

Some other useful notes about the distributed packages in the repository:

* **Sasprograms/master.sas**: This is where paths and common data model (CDM) file names are specified. The participating data-contributing site will update the path and CDM dataset file names and execute the package. It is the only file to be edited by the data-contributing sites. This program calls the run\_programs.sas code.
* **Infolder/run\_programs.sas**: In this code, all request- and study-specific parameters are defined (e.g., age groups, weight/height minimums and maximums values) by the study investigators. The program executes the runXYZ.sas code which launches the request/analysis (e.g., runabxwave1.sas).
* **Infolder/macros/runXYZ.sas**: this is the core code that runs the sequence of macros. It is very useful to the user who wishes to follow the analysis step by step.

The *analytic codes* (also called local codes) are used to process the returned datasets from the data-contributing sites to produce the network-wide results.

Below is a high-level description of the files contained in the repository.

**Antibiotics SSDC Wave1.zip**

The Antibiotic SSDC Wave 1 analysis aims to characterize a specific cohort of patients to inform the primary analyses of the PCORnet Antibiotics and Childhood Growth Study. Following cohort identification, it undergoes both demographic/vital and drug exposure ad-hoc code checks. Specifically, using PCORnet’s Table Diagnostic tool, counts, frequencies, and distribution of key uncharacterized table and variables of interest are reported (including missingness and outliers). Table Diagnostic is a reusable tool that gathers basic record counts and assesses variable formats, lengths, and missingness allowing for rapid study-specific data conformance of any PCORnet CDM Table. Both Antibiotic Wave 1 and Table Diagnostic execute against local SAS data stored in PCORnet CDM V3 format.

**Antibiotics SSDC Wave2.zip**

Similar to Wave1, the Antibiotics SSDC Wave 2 analysis aims to characterize CDM tables and variables needed for the PCORnet Antibiotics and Childhood Growth Study that were not part of the prior data characterizations done in wave 1. In wave 2, a new cohort is formed from the PCORnet patient population using modified specific inclusion/exclusion criteria and it is further characterized using specific ad-hoc analyses**.** Results gave the study team a full description of the study cohort used in the primary analyses. This code executes against local SAS data stored in PCORnet CDM V3 format.

**Antibiotics All-Aims Individual-level Query.zip**

This code is the individual-level data pull for further analyses in the main study and for conducting the benchmark pooled patient-level regression analysis in the distributed linear regression methods project. In this package, the cohort is formed from the PCORnet patient population final specific inclusion/exclusion criteria informed by the Wave 1 and 2 packages. The program then performs an individual-level data pull for the patients who have made it into the cohort. Records from the Demographic, Encounter, Diagnosis, Procedures, Vital, Prescribing, Dispensing, and Harvest table are pulled for each patient in a de-identified manner. Prescribing, Procedure, Dispensing, and Diagnosis record pulls will only be pulled for codes indicated by the look-ups tables in the package. No dates will be returned in the data pull. This code executes against local SAS data stored in PCORnet CDM V3 format.

**NOTE: This query returns patient-level data.**

**Antibiotics Aim1 Distributed Regression Distributed Package.zip**

This is the main package that locally runs at each data-contributing site and that generates the summary-level sums of square and cross products (SSCP) tables (with no patient-level information) required for the distributed linear regression analysis. First, using the datasets created by the all-aims data pull (Antibiotics All-Aims Individual-level Query.zip), the package creates and formats the dependent and independent variables (pc\_rundistreg.sas). Then, each of the 12 linear regressions are executed locally to calculate the SSCP tables (pc\_getsscp.sas). The final output tables consist of the 12 SSCP matrices to be returned to the analysis center.

**NOTE: The distributed regression program will only return summary-level information.**

**Antibiotics Aim 1 Analytic Package Part 1.sas**

This code prepares the confirmatory pooled-individual-level data analysis. After all the participating data-contributing sites have successfully executed the *Antibiotics All-Aims Individual-level Query.zip* package and returned their data pull output individual-level datasets to the analysis center, this program first pools all of them into one analytical dataset and prepares the benchmark pooled patient-level regressions by creating the independent and dependant variables.

**Antibiotics Aim1 Distributed Regression Distributed Package.zip**

This package prepares the pooled summary-level dataset for distributed linear regression. It first creates an all-site SSCP matrix using the SSCP matrices from the participating sites. At the same time, the program also creates site-specific individual fixed effects entries in the all-site SSCP matrix. The program then performs distributed linear regression using only the pooled SSCP matrix. Optionally, this package can also run the benchmark pooled individual-level regression using the dataset created by *Antibiotics Aim 1 Analytic Package Part 1.sas*. If the user opted for also running individual-level regression, results from the distributed linear regression analysis will be compared to those from the individual-level regression analysis.

**Antibiotics Aim1 Analytic Package Part 2.sas**

This program will create Tables 1-4, Supplemental Tables 1-12, and Figure 1 for the manuscript:

Sengwee Toh, ScD, Sheryl L. Rifas-Shiman, MPH, Pi-I Lin, ScD, L. Charles Bailey, MD, PhD, Christopher B. Forrest, MD, PhD, Casie E. Horgan, MPH, Douglas Lunsford, MEd, Erick Moyneur, MA, Jessica L. Sturtevant, MS, Jessica G. Young, PhD, Jason P. Block, MD, MPH, On behalf of the PCORnet Antibiotics and Childhood Growth Study Group. *Privacy-protecting multivariable-adjusted distributed regression analysis for multi-center pediatric study.*

**Antibiotics Maternal-Child Linkage Query.zip**

This query will create and return de-identified individual-level data to address Aim 3 of the PCORnet Antibiotics and Childhood Growth Study, which examines effect modification (treatment effect heterogeneity) according to several pre-specified covariates. These data will be combined to the patient-level dataset created by the *Antibiotics All-Aims Individual-level Query.zip* package. Specifically, the covariates that we will be created and integrated into the analysis are maternal factors, such as maternal demographics, antibiotic use during pregnancy, pre-pregnancy body mass index, and gestational weight gain, among other covariates.

**Note: This query returns de-identified, individual-level data at the datamart-level for the maternity linkage participating sites.**

**Antibiotics Maternal-Child Linkage analytic package part 1.sas**

This package will combine datasets returned from participating sites and will prepare analytic datasets for the maternal-child linkage analysis.

**Antibiotics Maternal-Child Linkage analytic package part 2.sas**

This package will create Tables 1-3 and Supplemental Tables 1-4 for the paper:

Heerman W, Daley M, Boone-Heinonen J, Rifas-Shiman SL, Bailey LC, Forrest C, Young J, Gillman M, Horgan C, Janicke D, Jenter C, Kharbanda E, Lunsford D, Messito MJ, Toh S, Block J.

*Maternal Antibiotic Use During Pregnancy and Childhood Obesity at Age Five Years.*

Int J Obes (Lond). 2019 Jan 22. doi: 10.1038/s41366-018-0316-6.