Abt

Abt Associates Inc.

Introduction to the Public Use Data Set:

The Study of Prenatal and Infant Exposure to Thimerosal and Neuropsychological Outcomes at Ages 7-10 Years

Cambridge, MA
Lexington, MA
Hadley, MA
Bethesda, MD
Chicago, IL

September 19th, 2007

Prepared for
William Thompson
National Immunization Program
Centers for Disease Control and
Prevention
Atlanta, GA 30329

Prepared by
Cristofer Price
Laura Simpson

Abt Associates Inc.
Suite 800 North
4550 Montgomery Avenue
Bethesda, MD 20814-3343

This documentation was prepared by Cristofer Price and Laura Simpson of Abt Associates Inc. for the Immunization Safety Office (ISO) of the Centers for Disease Control and Prevention (CDC) Atlanta, GA 30333. Questions about the documentation, substantive questions, or technical issues regarding the data file should be directed to the CDC ISO, MS D26, 1600 Clifton Road, Atlanta, Georgia 30333 (404-639-8256).

The *Public Use Data Set* for *the Study of Prenatal and Infant Exposure to Thimerosal and Neuropsychological Outcomes at Ages 7-10 Years* is contained on a CD that is comprised of two major partitions. The first of the two major partitions contains the data files, documents, and reports corresponding to the final analysis data set that was used to produce the results in the published manuscript titled "Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years", by Thompson et. al, (2007). The second partition contains historical data and documents, from preliminary analyses. While it is very unusual to include preliminary data and documents on a public use data set, we felt that including them would make the decision that was made to exclude several records from the final analysis data set, more open and transparent to the research community.

The events leading to the decision to exclude seven records from the final data set were as follows. A set of preliminary analyses were run and results were produced using a data set that included records from n=1,054 children that participated in the clinical assessments. A reader of a preliminary draft report noticed an inconsistency that had not been previously identified. The issue concerned a condition called 'fetal growth retardation' that had been an exclusionary criterion for the creation of the sampling frame. Among the 1,054 children in the preliminary data set, there were seven that did not have codes for 'fetal growth retardation' in the electronic records that were used in the creation of the sampling frame, but did have diagnoses of 'intrauterine growth retardation' listed in the medical record abstraction data. It was not until a reader pointed out that 'intrauterine growth retardation' is the same condition as 'fetal growth retardation', that we realized that we had ineligible subjects in the preliminary data set. The decision was made that the ineligible subjects should be removed from the final data set.

Since the results from the preliminary data set exist, we thought that the research community would be best served by having the preliminary data and results posted on this CD, where researchers would have the opportunity to examine them.

Each of the two partitions (final and preliminary) contains three types of data files, each of which is provided in two formats, (e.g, SAS files, ASCII files). The first of the three types is called 'the Main Analysis File'. All results presented in the Technical Report[1] and the published manuscript can be reproduced using the data contained within the Main Analysis File of the final data set partition. The Main Analysis File contains n=1,047 records, corresponding to one record per child. The remaining two data files contained within the set are provided to give researchers a rich source of data on the timing, sources, and amount of prenatal and postnatal exposure to ethylmercury from thimerosal-containing vaccines and immune globulins. Those data sets are called 'the child vaccination histories' and 'prenatal exposures' files and are explained in the documents titled "*Documentation and Codebook for the Prenatal Ethlymercury Exposures File*", and "*Documentation and Codebook for the*

---

[1] Technical Report, Volumes I and II of the Study of Infant Environmental Exposure to Thimerosal and Neuropsycological Outcomes at Ages 7 to 10 Years.

Child Vaccination Histories File". The three data files can be linked by merging on the variable *ChildID*, which is included in each of the three data files.

Exhibit 1.		
Materials included in the Public Use Data Set for the Study of Prenatal and Infant Exposure to Thimerosal and Neuropsychological Outcomes at Ages 7-10 Years		
Introduction		
Introduction to The Public Use Data Set	PubUse_Introduction.doc	* Overview of contents of the Restricted Access Data Set * Terms of Use * Contact information
Technical Report		
Technical Report, Volume I Technical Report, Volume II	Tech Rept Thimerosal Vol 1.pdf Tech Rept Thimerosal Vol 2.pdf	Volume I: * Study background and research questions * Study design * Sample * Data sources * Measures * Analysis approach * Results Volume II: * Full detail model results * Model checking * Post-analysis power
Main Analysis File		
Documentation and Codebook for the **Analysis File**	PubUse_MainAnalysisFile.pdf	* Documentation and codebook
Analysis File (SAS Data Set)	PubUse_MainAnalysisFile.sas7bdat	* SAS file (SAS version 8)
Analysis File (ASCII Text Data Set)	PubUse_MainAnalysisFile.dat	* ASCII Text Data Set
Layout file for **Analysis File** (ASCII Text Data Set)	PubUse_MainAnalysisFile_Layout.txt	* Text file * Layouts of variable names, columns
Child Vaccine Histories File		
Documentation and Codebook for the **Child Vaccine Histories File**	PubUse_ChildVaccinationHistories.pdf	* Documentation and codebook
Child Vaccine Histories File (SAS Data Set)	PubUse_ChildVaccineHistories.sas7bdat	* SAS file (SAS version 8)
Child Vaccine Histories File (ASCII Text Data Set)	PubUse_ChildVaccineHistories.dat	* ASCII Text Data Set
Layout file for **Child Vaccine Histories File** (ASCII Text Data Set)	PubUse_ChildVaccineHistories_Layout.txt	* Text file * Layouts of variable names, columns
Child Vaccine Histories File	PubUse_ChildVaccineHistories.xls	* Excel file is convenient for

Exhibit 1.		
Materials included in the Public Use Data Set for the Study of Prenatal and Infant Exposure to Thimerosal and Neuropsychological Outcomes at Ages 7-10 Years		
(Excel Spreadsheet)		visual inspection of the data.
Prenatal Ethlymercury Exposures File		
Documentation and Codebook for the **Prenatal Ethylmercury Exposures File**	PubUse_ PrenatalExposures.pdf	* Documentation and codebook
Prenatal Ethylmercury Exposures File (SAS Data Set)	PubUse_PrenatalExposures.sas7bdat	* SAS file (SAS version 8)
Prenatal Ethylmercury Exposures File (ASCII Text Data Set)	PubUse_ PrenatalExposures.dat	* ASCII Text Data Set
Layout file for **Prenatal Ethylmercury Exposures File** (ASCII Text Data Set)	PubUse_ PrenatalExposures _Layout.txt	* Text file * Layouts of variable names, columns
Prenatal Ethylmercury Exposures File (Excel Spreadsheet)	PubUse_ PrenatalExposures.xls	* Excel file is convenient for visual inspection of the data.
Published Manuscript		
Study Manuscript	CDC_Thimerosal_Neurodevelopment.pdf	
Historical Documents and Files (Preliminary Documents and Data)		
Technical Report from Preliminary Analyses: (Based on n=1054 records, including 7 records that were omitted from final data set because of exclusionary birth conditions.)	Preliminary TechReport Thimerosal.pdf	
Main analysis file data for 7 records that were omitted from final data set because of exclusionary birth conditions.	Excluded7_MainAnalysisFile.sas7bdat Excluded7_MainAnalysisFile.dat Excluded7_MainAnalysisFile_Layout.txt	
Child vaccination histories file data for 7 records that were omitted from final data set because of exclusionary birth conditions.	Excluded7_ChdVacHistories.sas7bdat Excluded7_ChdVacHistories.dat Excluded7_ChdVacHistories_Layout.txt Excluded7_ChdVacHistories.xls	
Prenatal ethylmercury exposures data for 7 records that were omitted from final data set because of exclusionary birth conditions.	Excluded7_PrenatalExposures.sas7bdat Excluded7_ PrenatalExposures.dat Excluded7_ PrenatalExposures _Layout.txt Excluded7_ PrenatalExposures.xls	

Date and Place of Publication

The citation for this manuscript is presented below.

Thompson WW, Price C, Goodson B, Shay D, Benson P, Hinrichsen VL, Lewis E, Eriksen E, Ray P, Marcy M, Dunn J, Jackson LA, Lieu T, Black S, Stewart G, Weintraub ES, Davis RL, and DeStefano F. *Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years*. N Engl J M; 357:1281-92.

Data Modification

Not applicable

Confidentiality Procedures

Investigators requested and received an Assurance of Confidentiality under Section 308(d) of the Public Health Service Act for *the Study of Prenatal and Infant Exposure to Thimerosal and Neuropsychological Outcomes at Ages 7-10 Years*. The Assurance of Confidentiality allows for the collection of sensitive information with protection against the dangers of release of the information under the usual conditions of the Privacy Act and the Freedom of Information Act.

To further ensure the privacy of the study participants, the data set provided here has been de-identified or stripped of any individual identifiers, using primarily the safe-harbor method. According to the guidance from CDC and the U.S. Department of Health and Human Services regarding the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, the safe-harbor method is defined as the method used when a covered entity or its business associate de-identifies information by removing 18 identifiers, and the covered entity does not have actual knowledge that the remaining information can be used alone or in combination with other data to identify the subject (45 CFR § 164.514 and CDC MMWR, April 11, 2003). The relevant identifiers of the 18 variables listed under the Privacy Rule [45 CFR § 164.514] were removed from this data set. These identifiers included 1) names (child and parents/guardian); 2) all geographic subdivisions including county, city, street address, precinct, zip codes, and schools/school systems; 3) all elements of dates (except year) including birth dates and vaccination dates; and 4) birth certificate numbers. These de-identification procedures are believed to reduce the risk of disclosure without reducing the usefulness of the data for public health practice and research.

Constraints on Data Usage

The data collected for this study are protected under an Assurance of Confidentiality and the Data Sharing and Use Agreement (DSA) for this study. This agreement is available from http://www.cdc.gov/od/science/iso/research_activities/dataset.htm.

References

CDC/ATSDR Policy on Releasing and Sharing Data. Manual Guide: General Administration, CDC-102, April 16, 2003. Available from http://www.cdc.gov/od/foia/policies/sharing.htm.