

Proc SQL and Data Cleaning
Felicita David
Centers for Disease Control and Prevention

Abstract: *Structured Query Language (SQL) is the standardized, widely used language for relational database management systems as defined by American National Standard Institute (ANSI). SQL is a modular type of language and has its own structure and syntax. Proc SQL, a component of base SAS implements SQL for SAS systems. In working with datasets and analytic tools, data cleaning is a necessary task and is an integral part of the process. Fortunately SAS programming language has a lot of 'tools' such as data steps, procedures, and SAS macro facilities. Thus the data analysis does not have to be 'garbage in and garbage out'. Proc SQL provides a comprehensive set of commands for a variety of data related tasks. This paper will focus on features of Proc SQL that could be used as an effective tool for investigating the data for commonly occurring problems like duplication of data, inconsistencies in the reported information with less coding.*

Biography:

Felicita David has been working in SAS for the past 5 years in academic and healthcare organizations. Her current position with Centers for Disease Control and Prevention finds her providing programming support, developing applications in SAS/Web for data management purposes for the various surveys received in the immunization division. Her expertise include programming in Base SAS, SAS/GRAPH, SAS/SQL, ODS, SAS/FSP, SAS/AF, SCL, SAS Macros and other programming languages.

Author Contact:

*Felicita David
Centers for Disease Control and Prevention
1600 Clifton Rd
Atlanta, GA 30329
fcd6@cdc.gov*