Utilizing SAS® to Estimate Rates of Disease from Nationally Representative Databases
Utilizing SAS® to Estimate Rates of Disease from Nationally Representative Databases

Jessica M. Rudd MPH, PhD Student in Analytics and Data Science
College of Science and Mathematics, Kennesaw State University

ABSTRACT

One of the research goals in public health is to estimate the burden of diseases among the US population by describing their association with:

- Hospitalizations
- ED visits
- Ambulatory/Outpatient visits
- Deaths

We use large, nationally representative databases, such as those offered by the National Center for Health Statistics (NCHS) or the Agency for Healthcare Research and Quality (AHRQ) to produce reliable estimates of disease for studies.

In this example, we use SAS® and SAS®-callable Sudaan to analyze the Nationwide Emergency Department Sample (NEDS), offered by AHRQ, to estimate Hand, Foot, and Mouth Disease Emergency Department (ED) visits in children less than 5 years old.

BACKGROUND

The Nationwide Emergency Department Sample (NEDS) is approximately a 20% clustered sample of hospital based EDs from 30 states.

- Largest all-payer ED database in the US
- Design variables are provided with the datasets!

SUDAAN® VS. SAS® SURVEY PROCEDURES?

Until recent years, SAS® statistical procedures did not take into account design properties of complex samples and would assume a simple random sample design. This can generally lead to underestimation of the variance.

While SAS® has introduced in recent versions the SURVEYMEANS AND SURVEYREG procedures to account for complex designs, adjustments have to be made to several SAS options in these procedures to produce identical results as the SUDAAN procedures.

Using SAS® survey procedures will produce the same weighted estimates.

STEPS FOR PRODUCING DISEASE ESTIMATES

Step 1: Import Datasets

Step 2: Variable Creation

- The NEDS observations include up to 15 ICD-9-CM codes per record
- The following array creates analysis variables for HFMD coded:
  - HFM = A HFMD diagnosis coded in any position
  - HFM_F = A HFMD diagnosis in the first diagnosis code

Step 3: Create Subpop

Step 4: Analysis Formats and Denominator

- Format labels for analysis variables
- Set denominator for rates calculation

Step 5: Create hospital weights table
Utilizing SAS® to Estimate Rates of Disease from Nationally Representative Databases

Jessica M. Rudd MPH, PhD Student in Analytics and Data Science
College of Science and Mathematics, Kennesaw State University

Step 6: Create Analysis Dataset

Step 7: Create Census Analysis Dataset

Step 8: Create Output Datasets with Estimates

Step 9: Calculate Rates

Step 10: Create Epi Curve
• While HFMD is a common childhood infection, hospitalizations associated with the disease are typically rare
• Severe rash and hospitalizations have been associated with HFMD cases reported in the US during 2011-2012
• Without national reporting or surveillance systems for HFMD, it is difficult to know the impact of changes in circulating serotypes on rates of hospitalizations
• This study provides baseline national estimates of HFMD-associated hospitalizations occurring among young children in the United States
• The re-usable SAS code utilized in this project is a readily available resource for researchers to quickly produce estimates of burden of disease as new years of national data become available

References