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. We describe burden of disease by analyzing the statistical association of various diseases with: hospitalizations, emergency department (ED) visits, ambulatory/outpatient (doctors’ office) visits, and deaths. In this short paper, we discuss the use of 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 (HFMD) associated emergency department (ED) visits in children less than 5 years old.

INTRODUCTION

The Nationwide Emergency Department Sample (NEDS) is an annual database started in 2006 that includes a 20% stratified sample of hospital based emergency departments from 30 states. “The NEDS is the largest all-payer emergency department database in the United States” and accounts for approximately 135 million ED visits per year. The database design has remained relatively unchanged since its inception so multi-year studies are convenient and SAS® programs can be reused year-to-year. Additionally, design variables are already provided with the datasets (HCUP).

SUDAAN® VS. SAS SURVEY PROCEDURES?

Why did I use SAS®-callable Sudaan® to estimate rates of disease rather than the available SAS® survey procedures, SURVEYMEANS? Briefly, Sudaan® is still widely considered the industry standard for analysis of complex survey designs. 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

The following steps create analysis variables and datasets, and produce weighted estimates of HFMD-associated ED visits among children less than 5 years old (see Appendix for full code and output):

1. Import core datasets:
   - Yearly datasets available from AHRQ with a data use agreement
   - Text files are converted to SAS files before analysis
   ```sas
   %MACRO LOOP;
   %LOCAL I;
   %DO I=2006 %TO 2012; /* Update here when new datasets become available*/
   LIBNAME NEDS '\\cdc\project\NCIRD_DVD_EB_DATA_1\hcup\neds\data';
   DATA NEDS_&_I._CORE (compress=binary);
   ```
2. Create analysis variable:

- The KID observations include up to 15 ICD-9-CM codes per record so several different variables were created for different study needs.
- An array is used to create the hand, foot, and mouth disease analysis variables:
  - $HFM = HFMD$ diagnosis coded in any of the 15 ICD-9-CM variable positions
  - $HFM_F = HFMD$ diagnosis coded in the first ICD-9-CM variable

```
%MACRO LOOP;
  %LOCAL I;
  %DO I=2006 %TO 2012;
    DATA HFM_&I (compress=binary);
    SET NEDS_&I._CORE;
    ARRAY DIAGNOsis (15) $ DX1-DX15;
    DO i=1 to 15;
      IF SUBSTR(DIAGNOSIS(i),1,4) IN ('0743') THEN DO;
        HFM=1;
        IF I=1 THEN HFM_F=1; ELSE HFM_F=0;
      END;
    END;
    /*Recode gender variable*/
    IF FEMALE=0 THEN SEX=1;else;
    IF FEMALE=1 THEN SEX=2;
    RUN;
  %END;
%END LOOP;
QUIT;
```

3. Combine years of data and subset on subpopulation of interest in order to control size and processing of data:

```
DATA HFM_06_12;
  IF HFM=1 AND AGE LT 5 THEN OUTPUT;
RUN;
***DELETE WORK DATASETS TO AVOID RUNNING OUT OF RESOURCES IN SAS ***;
%MACRO LOOP;
  %LOCAL I;
  %DO I=2006 %TO 2012;
    PROC DATASETS;
    DELETE HFM_&I NEDS_&I._CORE;
  %END;
%END LOOP;
QUIT;
```
4. Create analysis formats and rate denominator:
   - Format labels for analysis variables
   - Set denominator for rate calculations

   ```sas
   PROC FORMAT;
   VALUE SEX 1='MALE' 2='FEMALE';
   VALUE AGEGRP 1='LT 1' 2='1-4';
   VALUE REGION 1='NORTEAST' 2='MIDWEST' 3='SOUTH' 4='WEST';
   VALUE MONTHF 1='JAN' 2='FEB' 3='MAR' 4='APR' 5='MAY' 6='JUN' 7='JUL' 8='AUG' 9='SEP' 10='OCT' 11='NOV' 12='DEC';
   RUN;
   ```

   ```sas
   %LET DENOM=100000; *** Denominator for rates is 100,000 person visits***;
   ```

5. Create hospital weights table from the hospital weights file included with year NEDS datasets:

   ```sas
   %MACRO LOOP;
   %LOCAL I;
   %DO I=2006 %TO 2012;
   %END;
   ```

   ```sas
   PROC SQL;
   CREATE TABLE HOSP&I AS
   SELECT YEAR,HOSP_ED,HOSP_CONTROL,HOSP_TRAUMA,HOSP_REGION,
   HOSP.UR_TEACH,NEDS.STRATUM,DISCWT FROM NEDS.NEDS. &I. HOSPITAL
   ORDER BY YEAR,HOSP_ED;
   %END;
   %LET DENOM=100000; *** Denominator for rates is 100,000 person visits***;
   ```

6. Create analysis datasets based on required study criteria:
   - Combine hospital weights datasets and merge with subpopulation dataset
   - Create dummy records for the hospital strata that are not captured in the subpopulation, i.e. strata without records of HFMD in children less than 5

   ```sas
   DATA HOSPWTS;
   RUN;
   ```

   ```sas
   OPTIONS BUFNO=500 BUFSIZE=32k;
   ```
DATA HFM_NEDS_ANALYSIS(SGIO=YES);
MERGE HFM.HFM_06_12 (IN=C SGIO=YES DROP=KEY_ED A WEEKEND DQTR CHRON1-
CHRON15 INTENT_SELF_HARM DXCCS1-DXCCS15 E_CCS1-E_CCS4 I NEDS_STRATUM)
HOSPWT S (IN=B RENAME=(HOSP_REGION=REGION));
BY YEAR HOSP_ED;

*This adds dummy records for any strata that might not be captured in
the subset, to ensure correct variances;
IF B=1 AND (C NE 1) THEN DO;
HFM=0;
DISCWT=0.00000001;
END;

NEW_STRAT=NE DS_STRATUM || '-' || LEFT(YEAR);
NEW_STRAT2=YEAR*10000000+NE DS_STRATUM;

IF AGE=0 THEN AGEGRP=1; ELSE
IF AGE IN (1,2,3,4) THEN AGEGRP=2;
IF HFM_F=. THEN HFM_F=0;

FORMAT YEAR 5. SEX SEX. REGION REGION. AMONTH MONTHF. AGEGRP AGEGRP.;
RUN;

PROC SORT DATA=HFM_NEDS_ANALYSIS;
  BY NEW_STRAT2 HOSP_ED;
RUN;

7. Create census analysis dataset for use with rate calculations:
   - I used the bridged race census data files available from the National Center for Health Statistics
     (CDC, 2016)
LIBNAME CEN '\\cdc\project\NCIRD_DVD_E B_DATA_1\census\bridged_race\data' ;
DATA CEN;
SET CEN.br90_13_B (KEEP=YEAR SEX AGE POP STATE);
  IF 2006<=YEAR<=2012;
  DIED=2;
FORMAT SEX SEX. ;

LABEL POP='AGE GROUP CENSUS';
IF STATE
  IN(23,33,50,25,44,09,36,34,42) THEN REGION=1;
IF STATE IN
  (39,18,17,26,55,27,19,29,38,46,31,20) THEN REGION=2;
IF STATE
  IN(10,24,11,51,54,37,45,13,12,21,47,01,28,05,22,40,48)
  THEN REGION=3;
IF STATE
  IN(30,16,56,08,35,04,49,32,53,41,06,02,15) THEN REGION=4;
IF AGE GT 4 THEN DELETE;
RENAME POP=POP2;
RUN;

DATA CEN;
SET CEN;
IF AGE=0 THEN AGEGRP=1;
IF AGE IN (1,2,3,4) THEN AGEGRP=2;
IF SEX=. THEN SEX=1;
FORMAT AGEGRP AGEGRP .;
RUN;

PROC FREQ DATA=CEN;
TABLES YEAR SEX AGEGRP AGE REGION ;
FORMAT YEAR 5. SEX SEX. REGION REGION . ;
RUN;

PROC SUMMARY NWAY MISSING;
CLASS YEAR SEX AGEGRP REGION DIED;
VAR POP2;
OUTPUT OUT=CENSUS SUM=;
RUN;

8. Create output datasets with weighted estimates of disease:
   - Sudaan® crosstab procedure used to create summary tables of HFM unweighted cases, weighted cases, and standard errors by age group and year, sex and year, region and year

options mprint;
DATA CAT;
INPUT CAT $ @@;
CARDS;
HFM ;
RUN;

DATA _NULL_;
SET CAT END=LAST;
CALL SYMPUT('CAT'||LEFT(PUT(_N_,3.)),CAT);
IF LAST THEN CALL SYMPUT('COUNT'_N_);
RUN;

%MACRO LOOP;
%LOCAL I;
%DO I=1 %TO &COUNT;
%LET TI=%CMPRES(&&CAT&I);
ODS RTF
FILE="\cdc.gov\private\L317\icj2\Studies\HFM\&ti._rates_&sysdate..rtf"
STYLE=RTF STARTPAGE=NO BODYTITLE SASDATE;
TITLE "Hand, Foot and Mouth Disease ED Visits, Children under 5 yrs, CATEGORY=&TI (All diagnoses), US, 2006-2012, NEDS";
PROC CROSSTAB DATA=HFM_NEDS_ANALYSIS FILETYPE=SAS DESIGN=WR;
WEIGHT DISCWT;
SUBPOPN &TI=1;
NEST NEW_STRAT2 HOSP_ED/ MISSUNIT;
CLASS YEAR SEX AGEGRP REGION /INCLUDE=MISSING;
TABLES SEX AGEGRP REGION year YEAR*(SEX AGEGRP REGION);
SETENV LEFTMGN=1;
RTITLE "Hand, Foot and Mouth ED Visits, Children under 5 yrs, CATEGORY=&TI (All diagnoses), US [NEDS], 2006-2012";
PRINT NSUM WSUM COLPER ROWPER SEWGT SECOL SETOT;
OUTPUT NSUM="UNWEIGHTED CASES" WSUM="WEIGHTED CASES" SEWGT="STANDARD ERROR"/FILENAME="WORK.VAR97" FILETYPE=SAS REPLACE;
FORMAT YEAR 4. SEX SEX. AGEGRP AGEGRP.;
QUIT;

9. Calculate rates of hand, foot, and mouth disease associated hospitalizations:
   • SQL used to combine Sudaan summary tables with census analysis tables

DATA TOTHOSPS;
   SET VAR97;
   IF YEAR=0 AND SEX=-2 AND AGEGRP=0 AND REGION=-2;
   KEEP WSUM SEWGT;
RUN;
PROC SQL;
   CREATE TABLE TOTCEN AS
   SELECT SUM(POP2) AS POP2 LABEL='CENSUS TOTAL' FROM CENSUS;
DATA ALL;
   MERGE TOTHOSPS TOTCEN;
   %MACRO RATECI;
   RATE=(WSUM/POP2)*&DENOM;
ratersel=(RATE*(SEWGT/WSUM));
lci1=(RATE-(1.96*ratersel));
uci1=(RATE+(1.96*ratersel));
   LABEL RATE="HOSPS/&DENOM" WSUM ="HOSPS' LCI1='LOWER 95% CI' UCI1='UPPER 95% CI';
   FORMAT RATE LCI1 UCI1 10.1 WSUM SEWGT COMMA10. POP2 COMMA12.;
   %MEND RATECI;
   %RATECI;
RUN;
PROC PRINT NOOBS LABEL;
   TITLE2 "TOTAL RATE";
RUN;

***************
***** BY YEAR *****
***************;
DATA TOTHOSPS;
SET VAR97;
IF YEAR>0 AND SEX=0 AND AGEGRP=-2 AND REGION=-2;
KEEP YEAR WSUM SEWGT;
RUN;
PROC SQL;
CREATE TABLE TOTCEN AS
SELECT YEAR, SUM(POP2) AS POP2 LABEL='CENSUS TOTAL'
FROM CENSUS
GROUP BY YEAR;

DATA ALL;
MERGE TOTHOSPS TOTCEN;
BY YEAR;
%RATECI;
RUN;

PROC PRINT NOOBS LABEL;
TITLE2 "RATE BY YEAR";
RUN;

ODS RTF CLOSE;
%END;
%MEND LOOP;
%LOOP;
quit;

10. Create epi curve of weighted visits by month and year:

%let gpath='\cdc.gov\private\L317\icj2\Studies\HFM';
%let dpi=200;
ods html close;
ods listing gpath=&gpath image_dpi=&dpi;

/* Sort by onset date */
proc sort data=hfm.hfm_neds_analysis out=dataset;
by amonth hfm;
run;

/* Get onset date range and generate YearMonth variable */
DATA HFM2;
SET HFM.HFM_NEDS_ANALYSIS(KEEP= YEAR AMONTH HFM DISCWT) END=LAST;
IF HFM=1 AND AMONTH NE . ;
MIN=MIN(MIN, AMONTH);
MAX=MAX(MAX, AMONTH);
YEARMONTH= YEAR || '-' || PUT(AMONTH, MONTHF.);
IF LAST THEN DO;
    CALL SYMPUT ("MIN", MIN);
    CALL SYMPUT ("MAX", MAX);
END;
RUN;

/* Get weighted visits my month and year */
PROC SUMMARY DATA=HFM2;
CLASS YEARMONTH;
VAR HFM;
WEIGHT DISCWT;
OUTPUT OUT=OUTBREAK SUM=VISITS;
RUN;

DATA OUTBREAK;
SET OUTBREAK;
IF _TYPE_ = 0 THEN DELETE;
LABEL YEARMONTH = 'MONTH AND YEAR OF VISIT' VISITS='WEIGHTED VISITS';
RUN;

GOPTIONS RESET=ALL DEVICE=ACTXIMG;
ODS HTML FILE="OUTBREAK.HTML";
TITLE 'HAND, FOOT, AND MOUTH DISEASE EMERGENCY DEPARTMENT VISITS* AMONG
CHILDREN AGE < 5 YEARS, UNITED STATES, 2006-2012';
PROC GCHART DATA=OUTBREAK;
VBAR YEARMONTH /SUMVAR= VISITS DISCRETE;
RUN;
GOPTIONS RESET=ALL DEVICE=ACTXIMG;
QUIT;

OUTPUT

The overall average rate of Hand, Foot, and Mouth Diseases related hospitalizations is shown in Table 1.

<table>
<thead>
<tr>
<th>HOSPS</th>
<th>STANDARD ERROR</th>
<th>CENSUS TOTAL</th>
<th>HOSPS/100000</th>
<th>raterse1</th>
<th>LOWER 95% CI</th>
<th>UPPER 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>189,815</td>
<td>6,808</td>
<td>140,881,459</td>
<td>134.7</td>
<td>4.83257</td>
<td>125.3</td>
<td>144.2</td>
</tr>
</tbody>
</table>

Table 1. Hand, Foot and Mouth Disease ED Visits, Children under 5 yrs, CATEGORY=HFM (All diagnoses), US, 2006-2012, NEDS, Average Rate

The rate of Hand, Foot, and Mouth Disease related hospitalizations by year is show in Table 2. Bi-annual peaks of HFMD are typical through 2010. However, in late 2011 through spring 2012, increased clusters of disease were noted by medical care providers and the uncharacteristic number of cases those years is reflected in NEDS, as seen in Table 2 and Figure 1.

<table>
<thead>
<tr>
<th>Calendar year</th>
<th>HOSPS</th>
<th>STANDARD ERROR</th>
<th>CENSUS TOTAL</th>
<th>HOSPS/100000</th>
<th>raterse1</th>
<th>LOWER 95% CI</th>
<th>UPPER 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>20,773</td>
<td>1,549</td>
<td>19,938,883</td>
<td>104.2</td>
<td>7.7699</td>
<td>89.0</td>
<td>119.4</td>
</tr>
<tr>
<td>2007</td>
<td>12,320</td>
<td>1,087</td>
<td>20,125,962</td>
<td>61.2</td>
<td>5.4030</td>
<td>50.6</td>
<td>71.8</td>
</tr>
<tr>
<td>2008</td>
<td>22,201</td>
<td>1,698</td>
<td>20,271,127</td>
<td>109.5</td>
<td>8.3741</td>
<td>93.1</td>
<td>125.9</td>
</tr>
<tr>
<td>2009</td>
<td>16,095</td>
<td>910</td>
<td>20,244,518</td>
<td>79.5</td>
<td>4.4940</td>
<td>70.7</td>
<td>88.3</td>
</tr>
<tr>
<td>2010</td>
<td>19,620</td>
<td>1,256</td>
<td>20,189,075</td>
<td>97.2</td>
<td>6.2205</td>
<td>85.0</td>
<td>109.4</td>
</tr>
<tr>
<td>2011</td>
<td>25,967</td>
<td>1,730</td>
<td>20,122,198</td>
<td>129.0</td>
<td>8.5975</td>
<td>112.2</td>
<td>145.9</td>
</tr>
<tr>
<td>2012</td>
<td>72,839</td>
<td>5,873</td>
<td>19,989,696</td>
<td>364.4</td>
<td>29.3790</td>
<td>306.8</td>
<td>422.0</td>
</tr>
</tbody>
</table>

Table 2. Hand, Foot and Mouth Disease ED Visits, Children under 5 yrs, CATEGORY=HFM (All diagnoses), US, 2006-2012, NEDS, Rate by Year
CONCLUSION

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\(^2\). 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


RECOMMENDED READING


CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jessica M. Rudd, MPH
PhD Candidate, Analytics and Data Science
College of Science and Mathematics
Kennesaw State University
1000 Chastain Road
Kennesaw, GA 30144
Phone: 631.275.6698
Email: jrudd1@students.kennesaw.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.