

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

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

USERS PROGRAM



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

```
/*Import core datasets*/
%MACRO LOOP;
%LOCAL I;
%DO I=2006 %TO 2012; /* Update here when new datasets become available*/

  LIENAME NEDS '\\cdc\project\NCIRD_DVD_EB_DATA_1\hcup\neds\data';

  DATA NEDS_&I._CORE (compress=binary);
    SET NEDS.NEDS_&I._CORE;
  RUN;

  PROC CONTENTS DATA=NEDS_&I._CORE; *** Review contents of dataset;
  RUN;

  %END;
%MEND LOOP;
%LOOP;
QUIT;
```

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:
 - HFMD = A HFMD diagnosis coded in any position
 - HFMD_F = A HFMD diagnosis in the first diagnosis code

```
/* Create a variable indicating whether any diagnosis code includes HFMD*/
%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;
  RUN;
  /*Recode gender variable*/
  IF FEMALE=0 THEN SEX=1;else;
  IF FEMALE=1 THEN SEX=2;
  RUN;
  %END;
%MEND LOOP;
%LOOP;
QUIT;
```

Step 3: Create Subpop

```
/*Combine years and output subpopulation to control size and processing of analysis dataset*/
DATA HFM_06_12;
SET HFM_2006 HFM_2007 HFM_2008 HFM_2009 HFM_2010 HFM_2011 HFM_2012;
IF HFM=1 AND AGE LT 5 THEN OUTPUT;
RUN;
```

Step 4: Analysis Formats and Denominator

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

```
PROC FORMAT;
VALUE AGER 1='LT 1' 2='1-2' 3='2-3';
VALUE SEX 1='MALE' 2='FEMALE';
VALUE REGION 1='NORTHEAST' 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';
VALUE PAY 1='MEDICARE' 2='MEDICAID' 3='PRIVATE, INCLUDING HMO' 4='SELF-PAY' 5='NO CHARGE' 6='OTHER';
RUN;

%LET DENOM=10000;
```

Step 5: Create hospital weights table

```
%MACRO LOOP;
%LOCAL I;
%DO I=2010 %TO 2011;

  PROC SQL;
  CREATE TABLE HOSP&I AS
    SELECT YEAR, HOSPID,HOSP_BEDSIZE,HOSP_CONTROL,
           HOSP_LOCATION,HOSP_REGION,
           HOSP_TEACH, &DB.STRATUM, DISCWT
    FROM adm.&DB&I.WT
    ORDER BY YEAR,HOSPID;
  %END;
%MEND LOOP;

%LOOP;
QUIT;

DATA HOSPWTS;
SET HOSP2010 HOSP2011;
RUN;
```


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

```
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)
HOSFWTS (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;
/*SEX=2; AGE=0; AMONTH=3;*/
END;

NEW_STRAT=NEDS_STRATUM||'-'||LEFT(YEAR);
NEW_STRAT2=YEAR*10000000+NEDS_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;
```

Step 7: Create Census Analysis Dataset

```
DATA CEN LT3;
SET CEN.BR90_10_B (KEEP=YEAR RACE SEX AGE POP STATE);
IF 2006<=YEAR<=2010;
DIED=2;

FORMAT SEX SEX. AGER AGER. YEAR 5. SEX SEX. REGION REGION. ;
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=0 THEN AGER=1;
IF AGE IN (1,.) THEN AGER=2;
IF AGE=2 THEN AGER=3;

IF AGE GT 2 THEN DELETE;

RENAME POP=POP2;
RUN;

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

Step 8: Create Output Datasets with Estimates

```
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,COUNT);
RUN;

$MACRO LOOP;
$LOCAL I;
$DO I=1 $TO &COUNT;|
$let ti=%compress(&&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. AMONTH MONTHF.;
QUIT;
```

Step 9: Calculate Rates

```
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;
```

Step 10: Create Epi Curve

```
%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;

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


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

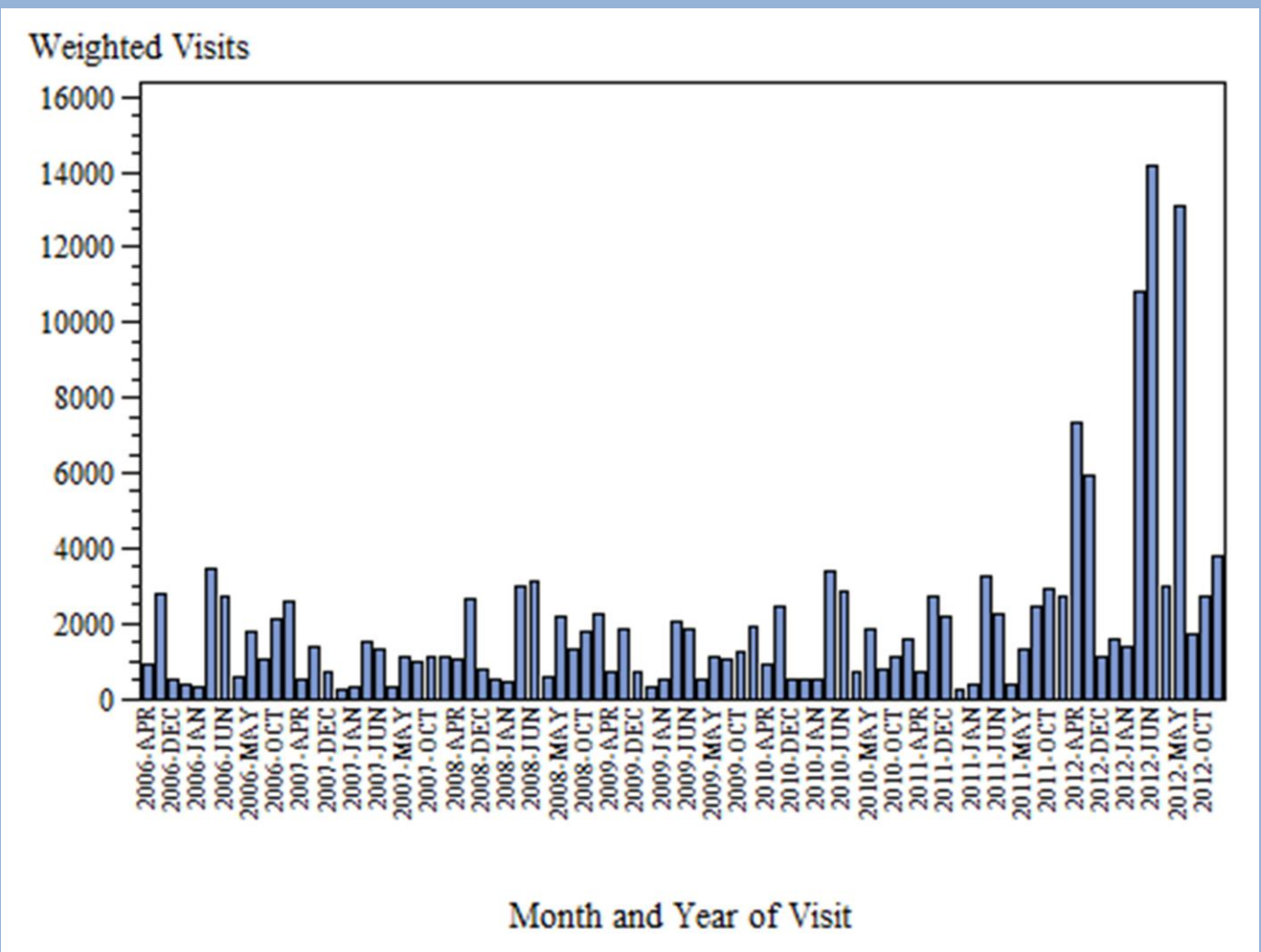
Jessica M. Rudd MPH, PhD Student in Analytics and Data Science

College of Science and Mathematics, Kennesaw State University

Output

HOSPS	STANDARD ERROR	CENSUS TOTAL	HOSPS/100000	raterse1	LOWER 95% CI	UPPER 95% CI
189,815	6,808	140,881,459	134.7	4.83257	125.3	144.2

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



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 2011-2012⁵
- 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

- CDC, National Center for Health Statistics, bridged race data file. (Available at: http://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm#vintage2010) Last accessed 2/21/2013.
- Cody, Ron. 2007. Learning SAS® by Example: A Programmer's Guide. Cary, N.C.: SAS Institute Inc.
- HCUP NIS Database Documentation. Healthcare Cost and Utilization Project (HCUP). November 2011. Agency for Healthcare Research and Quality, Rockville, MD. (Available at: www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp) Last accessed 02/06/2014.



SAS[®] GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL