

Creation of the National Sample

**National Sample Project of the National
Trauma Data Bank (NTDB), the American
College of Surgeons**

Draft March 2007

Contents

Section	Page
1. Introduction	1
2. Sampling Universe	2
3. Sample Design	3
4. Hospital Sample Selection	4
5. Calculation of Final Weights	5

Appendixes

Appendix A: Documentation of data files and Steps Involved in Creating the NTDB National Sample used for Analysis.....	6
Appendix B: Description of data files and data programs used for creating the 2003 National Sample.....	8
Appendix C: SAS source code for selecting the 100 samples	11
Appendix D: SAS Program Specifications for Sample Weight Adjustments.....	14
Appendix E: SAS Source Code	16

Tables

Number	Page
Table 1. Updated TIEP hospital universe	3
Table 2. Hospital sample allocation	4

1. INTRODUCTION

The National Trauma Data Bank (NTDB), which is managed by the American College of Surgeons (ACS), contains more than one million records voluntarily provided by 405 U.S. trauma centers (American College of Surgeons, Committee on Trauma [ACS/COT], 2003). The NTDB represents the largest compilation of traumatic injury data in the United States and contains detailed clinical indicators and other trauma care information on patients admitted into hospitals. The data are useful for research in injury epidemiology and prevention, acute care, and health services policy. However, the NTDB is subject to a limitation of all "convenience samples." The data may not be representative of all trauma hospitals in the nation and thus do not allow statistically valid inferences about national injury incidence and prevalence. This means that findings from analyses on outcome measures such as survival probability, length of hospital stay, or other indicators would have limited relevance to the universe of all patients (Clark and Winchell, 2004). Although nationally representative data such as the National Hospital Discharge Survey data are available, they typically lack the richness of the NTDB data, which contain detailed injury information including a wide array of diagnostic and clinical indicators.

The National Sample Project (NSP) is intended to (1) create a traumatic injury database from a nationally representative sample of trauma hospitals; (2) collect a wide variety of diagnostic and clinical indicators complementary to the NTDB; and (3) produce national baseline estimates of variables and indices associated with hospitalized traumatic injuries such as prehospital diagnosis and management, trauma outcomes, and other variables that characterize the dimensions of trauma treatment. It is also intended to help characterize the circumstances of the injury incident (e.g., external causes of injury) and injured patients treated in trauma centers in the United States. This report describes the following major tasks in the development of National Sample:

1. Determination of the trauma center hospital universe and construction of a sampling frame.
2. Development of sample design.
3. Selection of sample hospitals.

2. SAMPLING UNIVERSE

The Trauma Information Exchange Program (TIEP) of the American Trauma Society maintains the National Inventory of Hospital Trauma Centers, which contains the most complete and up-to-date list of trauma centers in the United States (MacKenzie et al., 2003). This inventory includes hospital information such as location, designated level of care, number of annual emergency room (ER) visits, and other organizational characteristics extracted from the American Hospital Association's (AHA's) Annual Survey of Hospitals (AHA, 2002). For this project, we updated data in the National Inventory of Hospital Trauma Centers and used this updated version to construct the sampling frame.

The designated levels of care of trauma centers are usually made either by the ACS/COT or by states. The TIEP follows a clearly defined procedure to determine the level of care designation for trauma center hospitals included in the National Inventory of Hospital Trauma Centers: (1) in states that have a formal process for designating or certifying trauma centers, the designated level of care is determined through discussions with the appropriate lead agencies; (2) in states that do not have such a formal process, the designated level of care must be verified by the ACS/COT; (3) if there are discrepancies between hospital self-reported, state-reported, and the ACS/COT-verified level of care designation, TIEP gives priority in this order: state, ACS/COT, and hospital report. Self-designated trauma center hospitals without outside verification of resources or capabilities are excluded from the inventory.

There is a direct link between this inventory and the NTDB. The information on the status of the hospitals contributing data to the NTDB was collected when the inventory was created. After obtaining the hospital list, we worked with the NTDB staff to identify hospitals that have been contributing or have agreed to contribute data to the NTDB. Only hospitals with a level I or level II designation were included in the sampling frame. Level III, level IV, and pediatric hospitals were not included during the current phase of the project. The distributions of the hospitals on the sampling frame are shown in Table 1. Of the 453 level I and level II trauma center hospitals, 179 could be identified as NTDB-contributing hospitals, including 15 hospitals that have recently agreed to contribute data in the future. The other 274 hospitals could not be identified as NTDB-contributing hospitals and were designated non-NTDB hospitals.

The hospitals in the sampling frame are not distributed uniformly across geographic regions. The Midwest has more trauma center hospitals (153/453) than any other region, and the West has the fewest (86/453). The South has the highest proportion of trauma center hospitals contributing to NTDB ($64/109 = 59$ percent), whereas the Northeast has the highest proportion of trauma center hospitals that do not contribute ($80/105 = 76$ percent). By the designated level of care, the South ($58/109 = 53$ percent) and the Northeast

(54/109 = 51 percent) have more level I hospitals but the West (62/86 = 72 percent) and the Midwest (99/153 = 65 percent) have more level II hospitals.

Table 1. Updated TIEP hospital universe

Stratum ^a	U.S. Census Region				Total
	Midwest	Northeast	South	West	
NTDB					
Level I	27	14	39	17	97
Level II	23	11	25	23	82
Subtotal	50	25	64	40	179
Non-NTDB					
Level I	27	40	19	7	93
Level II	76	40	26	39	181
Subtotal	103	80	45	46	274
Total	153	105	109	86	453

^a Sampling strata were formed according to contributing (NTDB) or not contributing (non-NTDB) data to NTDB and designated level of trauma care.

3. SAMPLE DESIGN

A stratified sample design was used, and 100 sample hospitals of level I and level II were to be included. Stratification was based on U.S. Census region (four regions), level of trauma care designation (two categories), and NTDB reporting status (two categories). Thus, there were 16 total strata: 8 NTDB strata and 8 non-NTDB strata. Of the 100 sample hospitals, 90 were allocated to the known NTDB-contributing hospitals and 10 to non-NTDB hospitals. The reason for this dramatic oversampling of NTDB-contributing hospitals was to avoid the extraordinary effort and the expense that would be required to recruit a large number of non-NTDB hospitals during the first year of this project. The sample size of 100 hospitals was chosen on the basis of recent NTDB data that suggested that a sample of 100 hospitals would provide estimates having sufficient precision for most analyses at the national level.

The 90 NTDB-contributing sample hospitals were further allocated to the 8 NTDB strata in proportion to the number of hospitals in each of these strata as indicated in Table 1. The 10 non-NTDB sample hospitals were also proportionally allocated to the 8 strata while each stratum was made sure to contain at least one sample hospital. The resulting sample allocation is shown in Table 2, and it is apparent that the overall distribution of the 100

sample hospitals by region and level of care reflects the distribution of the NTDB-contributing hospitals because of the oversampling of these hospitals.

Table 2. Hospital sample allocation

Stratum ^a	U.S. Census Region				Total
	Midwest	Northeast	South	West	
NTDB					
Level I	14	7	20	8	49
Level II	11	6	12	12	41
Subtotal	25	13	32	21	90
Non-NTDB					
Level I	1	2	1	1	5
Level II	2	1	1	1	5
Subtotal	3	3	2	2	10
Total	28	16	34	23	100

^a Sampling strata were formed according to hospitals contributing (NTDB) or not contributing (non-NTDB) data to NTDB and designated level of trauma care.

4. HOSPITAL SAMPLE SELECTION

Sample hospitals were drawn within strata by using the probability-proportional-to-size (PPS) without-replacement method (Levy and Lemeshow, 1999). When there are large differences in the size of sampling units (e.g., hospitals) with respect to variables of importance (e.g., number of trauma admissions), the PPS method tends to produce estimates with smaller variance compared to an equal probability sampling method. Furthermore, with a PPS sample, one has the flexibility to select a second stage sample of equal size (i.e., patients), which will result in equal overall weighting and may lead to more equitable response burden among the hospitals. Although this is not an issue now, it may be later when patient-level data collection is planned. For example, a study may be needed to validate the patient records submitted by the hospital samples. This may require retrieving and manually reviewing medical records and can be time-consuming and financially costly. A possible study design will be to select a subsample of the hospital samples and a same fixed number of patient records from each selected hospitals. Under the PPS sample design, the selected patient records for the validation study will have equal sample weights.

The size measurement used was the annual number of ER visits in 2003. In the process of selecting the samples, we noted a number of hospitals with very large numbers of ER visits. An investigation by the NTDB staff revealed that these hospitals were members of hospital

systems and provided the aggregated ER visit numbers of their affiliated hospital systems. These hospitals were then contacted to provide their hospital-specific number of ER visits. For a few hospitals that did not provide the data, we used the average of the system-wide number of ER visits per hospital as their estimated size measurement. Also, seven hospitals were included in the sample with certainty due to their large size measurements. For these sample hospitals, the probability of selection is thus one.

5. CALCULATION OF FINAL WEIGHTS

The final weights for each of the hospital were adjusted by the number of ER visits per month, since the volume of ER visits can vary from month to month. The final weights for hospitals with less than 30 ER visits in one month were classified as a non-responder for that month and the final weight was set to missing. The procedures developed for calculating the final weights can be found in Appendix C and the SAS program implementing those procedures can be found in Appendix D. The SAS program inputs a database file consisting of trauma events by date of arrival at the ER from the sample hospitals along with their sample design weights and then it calculates final weights for each record by month. The date of arrival at the ER was used since it is a required field in NTDB, compared to using injury date which is not consistently reported in NTDB.

**APPENDIX A:
DOCUMENTATION OF DATA FILES AND STEPS INVOLVED IN
CREATING THE NTDB NATIONAL SAMPLE USED FOR ANALYSIS**

Purpose: This document describes major steps and data files required for constructing the NTDB National Sample. The text below describes the various steps taken to create the sample, and a flowchart of these steps can be found in Figure 1.

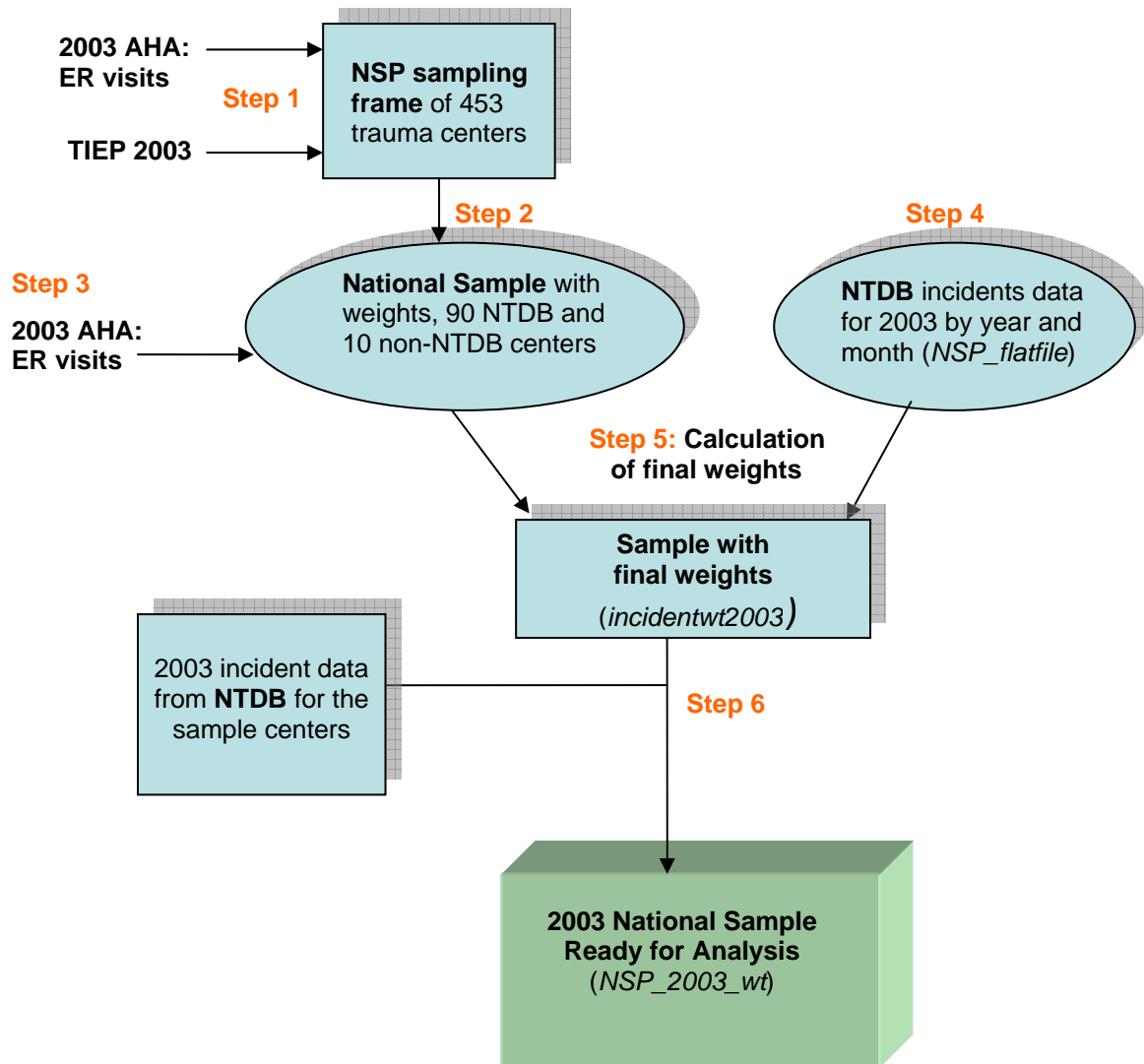
Creation of the 2003 National Sample:

1. The NSP Sampling frame was created (*NSP_SAMPLING_FRAME_2003*). This is an excel file with the 453 Trauma Centers of Trauma Level I and II, based on TIEP 2003 data with the corresponding 2003 emergency department visits from AHA data.
2. A stratified sample design was used to create the National Sample (*weightswithreplacements*) of 100 trauma centers from the NSP sampling frame, whereof 90 trauma centers were in NTDB and 10 were not in NTDB. Stratification was based on U.S. Census region (four regions), level of trauma care designation (two categories), and NTDB reporting status (two categories). A SAS program with PROC SURVEYSELECT was used to perform this step.
3. The emergency visits from the AHA 2003 data is merged into the National Sample (*weightswithreplacements*).
4. The 2003 incidents data from NTDB for the trauma centers in the National Sample was created (*NSP flatfile*). This file contains the incident key, incident month, and incident year.
5. The weights are adjusted for nonresponses and changes in the number of Emergency Room Visits. This was done by running the weight adjustment program (*WeightV1_r1.sas*). The National Sample file (*weightswithreplacements*) and the *NSP flatfile* are used as input files for this program. The output file (*incidentwt2003*) contains the adjusted sample weights.
6. The 2003 National Sample with adjusted weights and analysis variables (*NSP_2003_wt*) were created by merging the file with adjusted sample weights (*incidentwt2003*) with 2003 NTDB data by trauma center, year and month.

Creation of the Future National Sample:

The above steps are repeated as necessary, according to the document “**Maintenance of the National Sample**” for weight adjustment, refreshing and replacing (Attachment 3).

Figure 1.: Flow chart of Data files and Major Steps.



**APPENDIX B:
DESCRIPTION OF DATA FILES AND DATA PROGRAMS USED FOR
CREATING THE 2003 NATIONAL SAMPLE.**

Filename	Description	Variables included
NSP_SAMPLING_FRAME_2003	Excel file with centers from TIEP 2003 data that was used as sampling frame. Only level 1 and 2 were included => 453 centers - Merged TIEP and AHA data	Trauma_centers_name_from_Tiep Address City State Zip_code Tiep_trauma_level Submit_data_to_NTDB (Yes/No) Facility_Code_in_NTDB Records_submitted_from_1999 Hospital_Reported_State_Trauma_L Hospital_Reported_ACS_Trauma_Lev In_AHA_ Hospital_name_in_AHA_Database Facility_Admissions Emergency_room_visits Outpatient_visit Total_Surgical_Operations Bed_size_range AHA_Trauma_level Location_rural_or_urban Fips_State_and_Country_Code AHA_number Census_Region Prime_Stra Secnd_Stra
pps_sample.sas	SAS program that selects the 100 hospitals for the sample using probability-proportional-to-size (PPS) method. This file uses the NSP_SAMPLING_FRAME_2003 as an input file	
weightswithreplacements	SAS data set with the 100 sample hospitals using 2003 NSP sampling frame. RTI provides this by running proc surveyselect	Trauma_centers_name_from_Tiep Address City State Zip_code Tiep_trauma_level Submit_data_to_NTDB (Yes/No) Facility_Code_in_NTDB Wght - weight

Filename	Description	Variables included
nonntdb	Excel file with the NTDB facility_key for the 10 centers in NSP sample that are "non-NTDB" centers. If they are not registered dummy number is used for fac_key. This file is used for an input file for the following program: create_inputset_for_LW_p rogram.sas	Trauma_centers_name_from_Tiep Address City State Zip_code Tiep_trauma_level Submit_data_to_NTDB (Yes/No) Facility_Code_in_NTDB Wght
create_NSP_flatfile.sas	SAS program to create the NSP_flatfile from incident data in Oracle Input files needed for this program are: 1)weightswithreplacements 2) nonntdb	Fac_key Inc_key ed_arr_yr ed_arr_mon inc_yr inc_mon
NSP_flatfile	SAS data set with the incident data file with month and year of incident for the 100 NSP sample hospitals	Fac_key Inc_key ed_arr_yr ed_arr_mon inc_yr inc_mon
create_inputset_for_LW_program.sas	Create the input data file for WeightV1_r1.sas This program adds the ervisits to the file (using NSP_frame) to the weightswithreplacements data set. Output file: NSP_sample	
NSP_sample	SAS data set with the 100 sample hospitals and the ervisits & updated ervisits. This file is used as an input file for non-response and post-stratification program: WeightV1_r1.sas	FAC_KEY TRAUMA_CENTER_NAME_FROM_TIEP ADDRESS CITY STATE ZIP_CODE ERVISITS ERVISITSUPD SECND_STRA WGHT SELECTIONPROB - 1/wght

Filename	Description	Variables included
WeightV1_r1.sas	Weighting program to adjust hospital sampling weight for nonresponse and post-stratification. Input file: NSP_sample and NSP_flatfile. Created the output file: Incidentweight2003.	
incidentwt2003.sas7bdat	2003 National Sample which includes the adjusted weights after adjusting for non-response and post-stratification.	Facility_Code_in_NTDB Inc_yr Month Finalwt
NSP_2003_wt	Final sample for 2003 data which includes the final weights and analysis variables	Facility_Code_in_NTDB Inc_yr Month Finalwt Inc_key Age Gender Race Disstatus LOS D-code E-code

APPENDIX C: SAS SOURCE CODE FOR SELECTING THE 100 SAMPLES

```

/*****
/*
/* Title: pps_sample.sas
/* Author: Lei Li - RTI International
/* Modified by Sandra Goble
/* Project: National Sample Project (NSP)
/*
/* Purpose: Selecting the 100 hospitals using PPS methodology
/*
/*
/* Input data: Hospital sampling frame
/* Name: NSP_SAMPLING_FRAME_2003
/* Variables needed: Name:
/* Facility ID FACILITY_CODE_IN_NTDB
/* Primary strata prime_stra
/* Second strata secnd_stra
/* Trauma level tiep_trauma_level
/* region census_region
/* # ER visits emergency_room_visits
/*
/*
/* Output: DATA FILE WITH 100 HOSPITALS WITH WEIGHTS
/* Name: selectedhospitals
/*
/* LAST Revised: February, 2007
/*
/* Primary strata; 1 NTDB center, 2 non-NTDB center
/* Second strata: There are 8 weighting strata that are combinations
/* of 4 regions (Census regions) and 2 designated
/* levels of care (level I or level II).
/*
/*****

**** READ IN THE SAMPLING FRAME ****;
***** GET THE ERVISITS FOR EACH HOSPITAL FROM THE NSP SAMPLING FRAME ****;
PROC IMPORT
DATAFILE="C:\Sandra\SAS\NSP\NSP_2003\NSP_SAMPLING_FRAME_2003.csv"
OUT=FRAME DBMS=CSV REPLACE;
GETNAMES=YES;
RUN;
DATA SAMPLINGFRAME;
SET FRAME;
FAC_KEY=INPUT(FACILITY_CODE_IN_NTDB, 6.);
RUN;

```

```

proc sort data=samplingframe;
by prime_stra tiep_trauma_level census_region emergency_room_visits;
run;
proc univariate data=samplingframe;
where 1<=prime_stra<=2 and 1<=secnd_stra<=8 ;
by prime_stra tiep_trauma_level census_region ;
var emergency_room_visits;
run;

data frame16strata hospitalswithcertainty;
set samplingframe;
if 1<=prime_stra<=2 and 1<=secnd_stra<=8;

/* exclude certainly strata */
if prime_stra=1 and tiep_trauma_level='I' and census_region='South' and
emergency_room_visits in (158513,174564,195206) then output hospitalswithcertainty;
else if prime_stra=1 and tiep_trauma_level='I' and census_region='West' and
emergency_room_visits=215518 then output hospitalswithcertainty;
else if prime_stra=1 and tiep_trauma_level='II' and census_region='Midwest' and
emergency_room_visits in (85394,95845,114648) then output hospitalswithcertainty;
else output frame16strata;
run;
proc freq data=hospitalswithcertainty;
table prime_stra * tiep_trauma_level * census_region /list ;
run;

proc sort data=frame16strata;
where 1<=prime_stra<=2 and 1<=secnd_stra<=8;
by prime_stra tiep_trauma_level census_region emergency_room_visits;
run;

proc surveyselect data=frame16strata method=pps jtprobs
n=(14 7 17 7
8 6 12 12
1 2 1 1
2 1 1 1) seed=42705 out=chkhit;
size emergency_room_visits;
strata prime_stra tiep_trauma_level census_region ;
run;

libname final 'C:\Sandra\SAS\NSP\NSP_2003\';
data final.selectedhospitals;
set chkhit hospitalswithcertainty(in=certain);
if certain then do; selectionprob=1; samplingweight=1; end;
run;

proc means data=frame16strata min max sum;
where 1<=prime_stra<=2 and 1<=secnd_stra<=8 ;
class prime_stra tiep_trauma_level census_region ;
var emergency_room_visits;
run;
proc means data=chkhit min max sum;
where 1<=prime_stra<=2 and 1<=secnd_stra<=8 ;

```

```
class prime_stra tiep_trauma_level census_region ;
var emergency_room_visits;
weight samplingweight;
run;
proc means data=samplingframe min max sum;
where 1<=prime_stra<=2 and 1<=secnd_stra<=8 ;
class prime_stra tiep_trauma_level census_region ;
var emergency_room_visits;
run;
proc means data=final.selectedhospitals min max sum;
where 1<=prime_stra<=2 and 1<=secnd_stra<=8 ;
class prime_stra tiep_trauma_level census_region ;
var emergency_room_visits;
weight samplingweight;
run;
```

APPENDIX D: SAS PROGRAM SPECIFICATIONS FOR SAMPLE WEIGHT ADJUSTMENTS

The sample design weights are the inverse of the probability of selection within each hospital stratum. These sample design weights within each hospital stratum is then adjusted for nonresponses on monthly basis, and will be post-stratified according to the updated number of emergency room (ER) visits in the reporting year. This appendix specifies the procedure to calculate and adjust the sample design weights. Appendix B contains a SAS program for implementing the procedure.

B.1 Loading the Input Data Files

Two input data files are needed.

The hospital sample file contains hospital ID, sample selection stratum indicator, size measurements (number of ER visits used for sampling and updated number of ER visits), and selection probability. From these variables, two additional variables are derived. The sample design weight is calculated as the inverse of the selection probability, and the product of the sample weight and the number of ER visits that was used for sampling is also calculated.

The patient (incident) data file contains hospital ID, year and month of ER admission, and year and month of Injury.

B.2 Determining Hospital Nonresponse Status in a Given Month

A respondent hospital should have a minimum of 30 cases per month. A hospital with fewer than the minimum number of cases in a month is defined as a nonrespondent hospital in this month, even if the hospital submitted data.

B.3 Calculating Monthly Adjustment Factors

$$\alpha_{1ht} = (\sum w_{hi} z_{hi} \text{ over all eligible hospitals in stratum } h) / (\sum w_{hi} z_{hi} \text{ over all responding hospitals in month } t)$$

where w_{hi} is the design weight for hospital i in stratum h , and z_{hi} is the number of ER visits used as size measurement for hospital sampling. This adjustment factor essentially rescales the total sum of sample weights among respondent hospitals to the annual number of ER visits.

This will be done within each stratum formed by region and care level. Note that when calculating this adjustment factor, the seven self-representative hospitals (those selected with certainty) will not be excluded from other sample hospitals in the same stratum. In

SUDAAN, a hospital selected with certainty is typically treated as a separate stratum and does not contribute to variance calculation under the with-replacement sample design option.¹ Including self-representative hospitals will lead to increased variance estimates. However, excluding a nonrespondent, self-representative sample hospital will substantially reduce the sampling universe and can lead to great underestimation of the population total, as these hospitals have a very large number of the ER visits and tend to be large in size. The chance of nonresponse among these hospitals is expected to be very small, and special efforts may be made to ensure responses by these hospitals. Thus, inclusion of these hospitals in the nonresponse adjustment may not be necessary. Should it become necessary in the future, some imputation methods may be explored to produce records for any nonrespondent, self-representative sample hospitals.

B.4 Calculating Post-Stratification Factors

The sample weights are brought up to the updated total number of ER visits in the reporting year. This is done by inflating the sample weight by a ratio

$$\alpha_{2h} = Z'_h / Z_h$$

where the numerator is the updated total number of ER visits in stratum h , and the denominator is the total number of ER visits currently used for hospital sampling.

For now, this factor is assumed to be 1.

B.5 Calculating the Final Monthly Weights

$$\text{fnlwt}_{hi} = W_{hi} \alpha_{1ht} \alpha_{2hi}$$

B.6 Merging the Weights Back to the Patient-Level Data File

A dataset with detailed patient and injury information (variables included in the National Data Elements Project) will be created for those sample hospitals currently contributing data to the NTDB.

¹ RTI International. 2004. SUDAAN Language Manual, Release 9.0. Research Triangle Park, NC: RTI.

APPENDIX E: SAS SOURCE CODE

```

/*****
/*
/* Title: WeightV1_r1.sas
/* Author: L. Wrage - RTI International
/* Project: National Sample Project (NSP)
/*
/* Purpose: Weighting program to adjust hospital sampling
/* weight for nonresponse and post-stratification.
/*
/* Input data: 1. Hospital sample data set
/* Name: Sample
/* Variables needed: Name:
/* Facility ID fac_key
/* **Selection probability selectionprob
/* **Weighting strata wtstrata
/* # ER visits ervisits
/* # ER visits updated ervisitsupd
/*
/* 2. Incident data set
/* Name: NSP_flatfile.txt
/* Variables needed: Name:
/* Facility ID fac_key
/* Month of ER visit ED_ARR_MON
/* Year of ER visit ED_ARR_YR
/*
/* Output: Incident level data set with final weight
/* Name: IncidentWT (by year)
/*
/* Created: August, 2005
/* Revised: Feb 2007, Er visit date used instead of incident date
/*
/* **The selection probability is derived...
/*
/* **There are 8 weighting strata that are combinations of
/* 4 regions (Census regions) and 2 designated levels of care
/* (level I or level II).
*****/

options nocenter linesize=163 pagesize=90;
*options youroptions;
libname dat 'D:\data\SAS\NSP\NSP_2003';
*libname dat '\yourpathname\'; /*folder for saving input and output data
sets*/

/****MACRO LOOP FOR YEAR****/
%macro year (year=);

/****STEP 1: DATA PREPARATION****/
/**** NSP FLAT FILE *****/
DATA NSP_flatfile;
set DAT.nsp_flatfile;

```

```

run;

data sample;                                /*read in hospital sample data*/
set DAT.Nsp_sample;
sampwt=1/selectionprob;                    /*calculate sample weight*/
wtervisits=sampwt*ervisits;                /*calculate weighted number of er visits*/
if fac_key=2204 then fac_key=29;          /***fac_key #2204 is #29 in nsp data set*/
proc sort;
by fac_key;
run;

data incident(keep=fac_key ED_ARR_MON MONTH ED_ARR_YR count); /*read in
incident data*/
set dat.NSP_flatfile ;
if ED_ARR_yr=&year;                          /*macro loop selects a specific year of data*/
month=ED_ARR_MON;
count=1;
proc sort; by fac_key month;
run;

/**STEP 2: DETERMINE HOSPITAL RESPONSE STATUS FOR EACH MONTH***/
proc summary data=incident; /*count incidents per month for each facility*/
class fac_key month;
var count;
output out=sum sum=;
data sum(keep=fac_key month count respond);
set sum;
if _TYPE_=3;
if count ge 30 then respond=1;              /*response indicator=1 if incident count
at least 30 in a month*/
*proc print;
run;

%macro month(num=); /*create monthly response indicators for each facility*/
data dat&num (keep=fac_key resp&num);
set sum;
if month=&num;
resp&num=respond;
run;
%mend month;

%month(num=1);
%month(num=2);
%month(num=3);
%month(num=4);
%month(num=5);
%month(num=6);
%month(num=7);
%month(num=8);
%month(num=9);
%month(num=10);
%month(num=11);
%month(num=12);

data sample2;                               /*merge monthly response indicators to sample file*/

```

```

merge sample (in=insamp) dat1 dat2 dat3 dat4 dat5 dat6 dat7 dat8 dat9 dat10
dat11 dat12;
by fac_key;
if insamp;
run;

/**STEP 3: CALCULATE MONTHLY NON-RESPONSE ADJUSTMENT FACTORS WITHIN
WEIGHTING STRATA***/
*proc freq data=sample2;      /*check that there is at least one responding
facility per strata per month*/
*tables wtstrata*(resp1--resp12) / list;
*run;

data sample3 (drop=i);
set sample2;
array resp{12} resp1-resp12;
array wtvis{12} wtervisits1-wtervisits12;
do i=1 to 12;
wtvis{i}=resp{i}*wtervisits;      /*weighted # er visits for respondents*/
end;
run;

%macro month2(num=);
proc summary data=sample3;
class wtstrata;
var resp&num wtervisits wtervisits&num;
output out=sum sum=;
data sum&num;
set sum;
if _TYPE_;
adj&num=(wtervisits/wtervisits&num);      /*nonresponse adjustment factor
(adj&num) = the sum of the weighted er visits for all eligible hospitals in
the stratum / the sum of the weighted er visits for all responding hospitals
in the stratum; calculated within month*/
run;
%mend month2;

%month2(num=1);
%month2(num=2);
%month2(num=3);
%month2(num=4);
%month2(num=5);
%month2(num=6);
%month2(num=7);
%month2(num=8);
%month2(num=9);
%month2(num=10);
%month2(num=11);
%month2(num=12);

data all(keep=wtstrata adj1-adj12);
merge sum1 sum2 sum3 sum4 sum5 sum6 sum7 sum8 sum9 sum10 sum11 sum12;
by wtstrata;
run;

proc sort data=sample3;

```

```

by wtstrata;
data sample4 (drop=i);
merge sample3 all;
by wtstrata;

array adj{12} adj1-adj12;          /**adj factor***/
array resp{12} resp1-resp12;      /**resp indicator***/
array adj1f{12} adj1_1-adj1_12;
array wt{12} wt1_1-wt1_12;
array wtvis{12} adj_wtervis1-adj_wtervis12;
do i=1 to 12;
adj1f{i}=adj{i}*resp{i};          /**adj factor set to missing for nonresp***/
wt{i}=sampwt*adj1f{i};           /**monthly nonresp adjusted wt***/
wtvis{i}=wt{i}*ervisits;        /**monthly nonresp adj wt # ervisits (for
check)***/
end;
run;

/**STEP 4: CALCULATE POST-STRATIFICATION ADJUSTMENT FACTOR BASED ON UPDATED
# ER VISITS***/
/*NOTE: the updated # er visits is currently set to equal # er visits thus
post-strat factor will=1*/
proc summary data=sample4;
class wtstrata;
var ervisits ervisitsupd;
output out=sum sum=;
data sum;
set sum;
if _TYPE_;
adj_pstrat=ervisitsupd/ervisits;  /*post-stratification adjustment factor
(adj_pstrat) = the sum of the updated number of er visits / the sum of er
visits; calculated within strata*/
run;
data sum(keep=wtstrata adj_pstrat);
set sum;
run;

/**STEP 5: CALCULATE FINAL MONTHLY WEIGHTS***/
proc sort data=sample4; by wtstrata;      /*merge post-stratification
adjustment factor back to file*/
data sample5;
merge sample4 sum;
by wtstrata;
run;

data sample5 (drop=i);
set sample5;
array finalwt{12} finalwt1-finalwt12;
array wt{12} wt1_1-wt1_12;
array ckfnlwt{12} ckfnlwt1-ckfnlwt12;
do i=1 to 12;
finalwt{i}=wt{i}*adj_pstrat;      /*calculate final weight*/
ckfnlwt{i}=finalwt{i}*ervisits;  /*calculate final weighted # ervisits
(for check)*/

```

```

wtupdvisits=sampwt*ervisitsupd;      /*calculate weighted # updated ervisits
(for check)*/
end;
*proc print;
run;

/**CHECKS**/
proc summary data=sample5;
class wtstrata;
var wtervisits adj_wtervis1-adj_wtervis12
wtupdvisits ckfnlwt1-ckfnlwt12;
output out=sum(drop=_TYPE_) sum=;
proc print data=sum;
title4 'Checks for 1) nonresponse adjustment, and 2) post-stratification';
title5;
title6 'Check 1: wtervisits = adj_wtervis1-adj_wtervis12';
title7 'The sum of the weighted #ervisits should equal the sum of each of the
monthly nonresp adjusted weighted #ervisits';
title8;
title9 'Check 2: wtupdvisits = ckfnlwt1-ckfnlwt12';
title10 'The sum of the weighted updated #ervisits should equal the sum of
each of the final monthly weighted #ervisits';
run;

*data dat.sample5&year;      /*hospital sample data set with all variables
created in program*/
*set sample5;
*run;

/**STEP 6: MERGE FINAL WEIGHTS TO INCIDENT DATA SET***/
data wt(keep=fac_key finalwt1-finalwt12);
set sample5;
proc sort data=wt; by fac_key;

data incidentWT (drop=i count);
merge incident (in=ininc) wt;
by fac_key;
if ininc;
array fwt{12} finalwt1-finalwt12;
do i=1 to 12;
if month=i then finalwt=fwt{i};
end;
run;

data dat.incidentWT&year(drop=finalwt1-finalwt12);      /*output incident data
set with final weight*/
set incidentWT;
run;

%mend year;
*%year(year=2001);
*%year(year=2002);
%year(year=2003);

```