Maintenance of NTDB 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. Overview **1**

3. Adjustment of Sample Design Weights **1**

4. Sample Replacement and Sample Refreshment **2**

Appendixes

Appendix A: SAS Program Specifications for Sample Weight Adjustments 4

Appendix B: SAS Source Code 6

Appendix C: Dealing with Changes in sample Eligibility/Operational Status.....13

1. INTRODUCTION

The current national sample consists of a stratified sample of 100 hospitals. Stratification was based on U.S. Census region (four regions), level of trauma care designation (two categories), and NTDB reporting status (two categories). The sample consists of 90 hospitals that have contributed data to NTDB and 10 that have not contributed data to NTDB before year 2003.

The national sample is intended to reflect the universe of all trauma level I and II hospitals that provides trauma care. As the hospital universe changes inevitably over time due to closing of the existing hospitals, opening of the new hospitals, and other reasons, the national sample must be evaluated from time to time and revised so that it continues to be representative. In the meantime, the sample should provide certain longitudinal comparability, and thus requires some stability. The various maintenance steps for the national sample is described in this document.

2. OVERVIEW

Annually: The eligibility and operational status of all sample hospitals are evaluated appropriated adjustments will be made according to appendix C. The weights will be adjusted for non-response when the AHA data base is released with all the ER visits for corresponding year.

Every 5 - 10 years: Sample Refreshment.

3. ADJUSTMENT OF SAMPLE DESIGN WEIGHTS

The 100 hospitals in the sample were randomly drawn with probability proportional to the annual number of ER visits. Their design weights of a selected hospital are the inverse of their selection probability and add up to the total number of ER visits. The sample design weight is calculated as follows:

$$w_{hi} = \left(\sum_{i=1}^{N_h} z_{hi} \right) / \left(n_h \, z_{hi} \right).$$

Where z_{hi} is the number of ER visits (the size measurement for sampling) for the ith hospital in stratum h, and N_h and n_h be the total and selected numbers of hospitals in stratum h. One exception is for the seven hospitals selected with certainty whose sample weights are set to equal one.

As usual, it is anticipated that some of the sample hospitals may participate in this project but not be able to contribute complete data throughout the project period for a variety of reasons such as technical difficulty or concerns over confidentiality. Furthermore, the volume of ER visits may change from year to year, which may deviate substantially from the size measurements used for sample selection. The adjustment of the sample weights can be an effective short-term solution for changes in response of the hospitals in the sample. The procedures developed for adjusting the sample design weights for non-respondent hospitals (contributes 30 or fewer injury cases in a month) or change in volume of ER visits can be found in Appendix A and Appendix B contains a SAS program for implementing these procedures.

The weight adjustments for will be done once a year when the AHA data base is released with all the ER visits for corresponding year.

4. SAMPLE REPLACEMENT AND SAMPLE REFRESHMENT

In addition to changes in the volume of ER visits, the status of a few selected hospitals may change due to a variety of reasons such as hospital closure, merging of hospitals, and change in the designated level of trauma care, or even a movement across geographical regions. These changes vary on a case-by-case basis and appendix C lists some scenarios that are likely to occur and the proposed adjustment methods for recalculating the sample design weights. However, given that the National Sample intends to maintain 100 sample hospitals, it may be necessary to draw additional hospitals when changes like these occur and procedures for Sample Replacement and Sample Refreshment is described in further detail below.

Sample Replacement

The term sample replacement refers to those situations when a hospital refuses to participate or decides to drop out, and thus must be replaced to keep 100 hospitals in the sample. A hospital will be replaced after their status participation/eligibility status change is known to NTDB right before the annual adjustment of the weights (section 3). The following steps needs to be completed for the sample replacement:

1. randomly draw a replacement hospital from the same sampling stratum
2. re-calculate the sample weight (section 3)

Sample Refreshment

The term sample refreshment refers to the situations where the sampling frame has changed substantially and some hospitals in the sample must be removed so that new

hospitals previously not in the sampling frame or new NTDB hospitals previously classified as non-NTDB can be represented in the sample. Sample refreshing may be done every 5 - 10 years, or coincide with the release of the National Inventory of Hospital Trauma Centers. The following steps needs to be completed for the sample refreshment:

1. identify and randomly select hospitals in the sample to be removed
2. randomly select new hospitals to be included in the sample
3. re-calculate the sample weights (section 3)

APPENDIX A:
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

α_{1ht} = ($\Sigma w_{hi}z_{hi}$ over all eligible hospitals in stratum h) / ($\Sigma w_{hi}z_{hi}$ 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 B:
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 Incident     inc_mon            */
/*                        Year of Incident      inc_yr             */
/*                                                                 */
/*    Output:   Incident level data set with final weight          */
/*              Name: IncidentWT (by year)                         */
/*                                                                 */
/*    Created: August, 2005                                        */
/*    Revised: September, 2005 - to add further documentation      */
/*                                                                 */
/*  **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 'K:\Lisa\Data\final\';
```

```
*libname dat ‘\yourpathname\’; /*folder for saving input and output data
sets*/

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


/***STEP 1: DATA PREPARATION***/

/*code to convert text file to a sas data set (.dat file has been saved as
.txt file)*/
proc import out=dat.nsp datafile=‘k:\lisa\data\final\NSP_flatfile.txt’
dbms=tab replace;
*proc import out=dat.nsp datafile=‘yourpathname\NSP_flatfile.txt’ dbms=tab
replace;
getnames=yes;
datarow=2;
run;

data sample;                  /*read in hospital sample data*/
set dat.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 month inc_yr count); /*read in incident data*/
set dat.nsp;
if inc_yr=&year;  /*macro loop selects a specific year of data*/
month=inc_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);
*%year(year=2004);
%year(year=2005);
```

APPENDIX C:
DEALING WITH CHANGES IN SAMPLE PARTICIPATION/OPERATIONAL STATUS

The operation or eligibility status of hospitals in the sampling frame will inevitably change over time. To maintain the desired hospital sample size and comparability of temporal trends in injury estimates, we propose to address changes in hospital eligibility or operational status by the adjustment of sample design weight, the use of replacement hospitals and the update of the sampling frame. This document outlines some strategies for addressing a number of situations that we anticipate are likely to occur (see Table C-1).

Table C-1. A Summary of Rules to deal with Changes in Hospital Eligibility/Operational Status

	Type of Scenario	Replace Sample Hospital	Adjust Sample Weight
(1)	A sample hospital becomes ineligible	No	No
(2)	A sample hospital merges with another sample hospital in the same sampling stratum	No	Yes
(3)	A sample hospital merges with another sample hospital in a different sampling stratum	Yes	Yes
(4)	A sample hospital merges with another hospital not in the sample but in the same sampling stratum	No	Yes
(5)	A sample hospital merges with another hospital not in the sample and not in the same sampling stratum, and the merged hospital stays in the stratum of the sample hospital	No	Yes
(6)	A sample hospital merges with another hospital not in the sample and not in the same sampling stratum, and the merged hospital moves into the stratum of the hospital not in sample	No	Yes
(7)	A sample hospital moves into a different sampling stratum	No	Yes

(1) A sample hospital becomes ineligible. This applies to the closure of a sample hospital or its ER department, or when a hospital loses its designation as a level I or level II trauma care center.

- Do not select a replacement sample hospital. Select additional sample(s) to achieve the target sample size when updating the sample frame.
- Do not adjust hospital sample weight.

(2) A sample hospital merges with another sample hospital in the same sampling stratum.

- Do not replace the sample hospital.
- The sample weight for the merged hospital is $w'_{hi} = (\frac{w_{hi}w_{hj}}{w_{hi} + w_{hj}})(\frac{n_h}{n_h - 1})$, where hospital *i* merges with hospital *j* in stratum *h*, and w_{hi} is the sample design weight and n_h is the sample size.
- The adjusted sample weight for a sample hospital other than merged hospitals is $w'_{hi} = w_{hi}(\frac{n_h}{n_h - 1})$.

(3) A sample hospital merges with another sample hospital in a different sampling stratum.

- Select a replacement hospital from the stratum that lost a sample hospital due to merging with a hospital outside the stratum.
- The adjusted sample weight for the merged hospital is $w'_{lj} = w_{lj} \frac{(1 + \frac{\sum_i z_{hi}}{n_h \sum_j z_{lj}})}{(1 + \frac{\sum_i z_{hi}}{\sum_j z_{lj}} \frac{w_{lj}}{w_{hi}} \frac{n_l}{n_h})}$, where hospital *i* in stratum *h* merges into hospital *j* in stratum *l*, and z_{hi} and z_{lj} are numbers of ER visits.
- The adjusted sample weight for a sample hospital other than merged hospitals in stratum *l* is $w'_{lj} = w_{lj}(1 + \frac{\sum_i z_{hi}}{\sum_j z_{lj}} \frac{1}{n_h})$.
- The sample weight for a sample hospital in stratum *h* is $w'_{hi} = w_{hi}(1 - \frac{1}{n_h})$.

(4) A sample hospital merges with another hospital not in the sample but in the same sampling stratum.

- Do not replace the sample hospital.

- The sample weight for the merged hospital is $w'_{hi} = \frac{w_{hi}}{(1 + \frac{n_h w_{hi} z_{hj}}{\sum_i z_{hi}})}$, where in stratum *h*, sample hospital *i* merges with hospital *j* that is not in the sample.
- The sample weight for a sample hospital other than a merged hospital remains same.

(5) A sample hospital merges with another hospital not in the sample and not in the same sampling stratum, and the merged hospital stays in the stratum of the sample hospital.

- Do not replace the sample hospital.
- The sample weight for the merged hospital is $w'_{hi} = w_{hi} \frac{1 + \frac{z_{lj}}{\sum_i z_{hi}}}{1 + \frac{n_h w_{hi} z_{lj}}{\sum_i z_{hi}}}$, where sample hospital *i* in stratum *h* merges with hospital *j* in stratum *l* but not in the sample for stratum *l*.
- The adjusted sample weight for a sample hospital other than the merged hospital in stratum *h* is $w'_{hi} = w_{hi}(1 + \frac{z_{lj}}{\sum_i z_{hi}})$, and $w'_{lj} = w_{lj}(1 - \frac{z_{lj}}{\sum_j z_{lj}})$.

(6) A sample hospital merges with another hospital not in the sample and not in the same sampling stratum, and the merged hospital moves into the stratum of the hospital not in sample.

- Do not replace the sample hospital.
- The sample weight for the merged hospital is $w'_{lj} = w_{hi} \frac{(1 + \frac{n_h \sum_j z_{lj}}{\sum_i z_{hi}})}{(1 + \frac{n_h w_{hi} z_{lj}}{\sum_i z_{hi}})} \frac{1}{n_l + 1}$, where sample hospital *i* in stratum *h* merges with hospital *j* in stratum *l* but not in the sample for stratum *l*.

- The adjusted sample weight for a sample hospital other than the merged hospital in stratum h is $w'_{hi} = w_{hi}$, and $w'_{lj} = w_{lj}(1 + \frac{\sum_i z_{hi}}{\sum_j z_{lj}} \frac{1}{n_l})(\frac{n_l}{n_l + 1})$.

(7) A sample hospital moves into a different sampling stratum. This applies to changes in the designated level of care or in the geographic region.

- Do not replace the sample hospital.
- The adjusted weight for the moving hospital is $w'_{hi} = w_{hi}(1 + \frac{n_h \sum_j z_{lj}}{\sum_i z_{hi}})(\frac{1}{n_l + 1})$, where sample hospital i in stratum h moves to stratum l.
- The adjusted weight for a sample hospital in stratum l other than the new hospital i is $w'_{lj} = w_{lj}(1 + \frac{\sum_i z_{hi}}{\sum_j z_{lj}} \frac{1}{n_h})(\frac{n_l}{n_l + 1})$.
- The sample weight for a sample hospital in stratum h remains same.

These formulas were developed to take into account the most likely situations that may arise after recruitment of the initial 100 sample hospitals. They will be needed and should be applied in the maintenance of the data system, but may need modifications. We anticipate that the planned activities will include year-end verification of sample hospital status and sample weight adjustment each year. A hospital's status may have changed before the year-end verification, and its design weights will need to be adjusted. The operational meaning of a hospital merging can be variable and requires caution in the use of the above rules. Mathematically, we have implied that the numbers of ER visits in two hospitals are added together when the two hospitals merge. In reality, the number of ER visits of the resulting merged hospital may be quite different. If the ER department of one hospital is merged into the ER department of another hospital due to small patient volume, the potential patients of the first hospital may not necessarily go to the merged hospital's ER.