

**Addendum to
Healthcare Cost and utilization Project (HCUP)
Methods Series Report #2005-07
Calculating Kids' Inpatient Database (KID) Variances**

HCUP Methods Series Report #2005-07 *Calculating Kids' Inpatient Database (KID)* provides two approaches to calculating variances for subpopulations that ensure that there is at least one observation for each hospital in the Kids' Inpatient Database (KID) when defining the subpopulation. The first example uses the entire KID sample. The second example uses only the subsample of KID records corresponding to the subpopulation of interest. Having at least one observation from each NIS hospital is essential to the accuracy of the variance estimates.

This addendum provides additional guidance on a SAS® option for PROC SURVEYMEANS that will ensure that the variance estimate is calculated correctly if an *analytic data element* is missing for all observations in a hospital.

Consider the following example. If an analysis of the 2019 KID is focused on total hospital charges (HCUP data element TOTCHG), the user might not realize that five hospitals in the 2019 KID are missing data for TOTCHG on all observations. By default, PROC SURVEYMEANS completely excludes an observation from analysis if that observation has a missing value. Unless the option NOMCAR is specified in the PROC SURVEYMEANS, SAS will exclude these hospitals from the variance estimate of TOTCHG and, thereby, underestimate the variance.

The NOMCAR option on the PROC SURVEYMEANS statement became available in SAS 9.2. From the [SAS Version 9.4 documentation for PROC SURVEYMEANS](#):

- This option requests that the procedure treat missing values in the variance computation as not missing completely at random (NOMCAR) for Taylor series variance estimation. When you specify the NOMCAR option, PROC SURVEYMEANS computes variance estimates by analyzing the nonmissing values as a domain or subpopulation, where the entire population includes both nonmissing and missing domains.
- The NOMCAR option applies only to Taylor series variance estimation.
- The Taylor series method is the default for PROC SURVEYMEANS if you do not specify the VARMETHOD= option.

Demonstration of NOMCAR Option

In this section we demonstrate the effect on standard errors when using the NOMCAR option with PROC SURVEYMEANS using the 2019 NIS. These concepts also apply to the KID for all data years.

In SAS Example 1, we use the 2019 NIS and define the following three data elements:

- All observations have a non-missing value (data element ALLCODED)
- Random pattern of missing data (data element ONEMISSING)
 - Only one observation in every NIS hospital has a missing value
 - If a hospital only had 1 record in the NIS, it's value of ONEMISSING is non-missing
- Nonrandom pattern of missing data (data element HOSPMISSING)
 - Five hospitals have data set to missing for all observations
 - All other hospitals have non-missing data for all observations.

Next, PROC SURVEYMEANS was run with and without the NOMCAR option.

SAS Example 1

```
/* Create test file with three new data elements. */

data test;
  set nis2019.NIS_2019_CORE ;
  by hosp_nis key_nis;
  ALLCODED=_n_;
  ONEMISSING=ALLCODED;
  if first.hosp_nis and not last.hosp_nis then onemissing=.;
  HOSPMISSING=ALLCODED;
  if hosp_nis in (10002,30440,30660,90100,90539) then hospmissing=.;
run;

/* Obtain estimates without the use of NOMCAR: The following SAS code produces estimates
of the sums, means, and standard errors for three data elements defined above. */

proc surveymeans data=test sumwgt sum mean ;
  weight DISCWT ;
  cluster HOSP_NIS ;
  strata NIS_STRATUM ;
  var ALLCODED ONEMISSING HOSPMISSING ;
run;

/* Obtain estimates with the use of NOMCAR: The following SAS code produces estimates of
the sums, means, and standard errors for three data elements defined above. */

proc surveymeans data=test sumwgt sum mean NOMCAR;
  weight DISCWT ;
  cluster HOSP_NIS ;
  strata NIS_STRATUM ;
  var ALLCODED ONEMISSING HOSPMISSING ;
run;
```

The data summary table for both SURVEYMEANS had the same values (Table A.1).

Table A.1. Data Summary for PROC SURVEYMEANS from SAS Example 1 with and without NOMCAR

Data Summary	
Number of Strata	201
Number of Clusters	4,568
Number of Observations	7,083,805
Sum of Weights	35,419,023

Source: Agency for Healthcare Research and Quality (AHRQ), Healthcare Cost and Utilization Project (HCUP), National Inpatient Sample (NIS), 2019

Table A.2 provides a comparison of the results of the two PROC SURVEYMEANS.

- The values for the sum of weights, mean, and sum are identical across the three data elements and two PROC SURVEYMEANS.
- The standard errors for the data elements ALLCODED and ONEMISSING are identical across the two approaches because all hospitals in the NIS are represented.
- The standard errors using the NOMCAR option are larger for the data element HOSPMISSING because the NOMCAR option is not excluding the five hospitals with missing data from the standard error calculation.

Table A.2. Results for PROC SURVEYMEANS for SAS Example 1 with and without NOMCAR

Variable	Sum of Weights	Mean	Std Error of Mean	Sum	Std Dev
PROC SURVEYMEANS without NOMCAR					
ALLCODED	35,419,023	3,541,902	18,905	125,450,710,000,000	1,261,970,300,000
ONEMISSING	35,396,223	3,541,847	18,917	125,367,990,000,000	1,261,954,000,000
HOSPMISSING	35,379,103	3,543,792	18,914	125,376,200,000,000	1,261,848,100,000
PROC SURVEYMEANS with NOMCAR					
ALLCODED	35,419,023	3,541,902	18,905	125,450,710,000,000	1,261,970,300,000
ONEMISSING	35,396,223	3,541,847	18,917	125,367,990,000,000	1,261,954,000,000
HOSPMISSING	35,379,103	3,543,792	18,988	125,376,200,000,000	1,263,334,200,000

Source: Agency for Healthcare Research and Quality (AHRQ), Healthcare Cost and Utilization Project (HCUP), National Inpatient Sample (NIS), 2019

Recommended Modification to the SAS Example on Page 10 of this Report

To guard against instances in which an analytic data element, such as TOTCHG, is missing for all observations in a KID hospital, we recommend adding the NOMCAR option to PROC SURVEYMEANS. SAS Example 2 is the recommended modification to the SAS code provided on page 10 of HCUP Methods Series Report #2015-09 *Calculating Nationwide Inpatient Sample (NIS) Variances for Data Years 2011 and Earlier*. The recommendations also apply to the KID for all data years.

SAS Example 2. Recommended Modification to the SAS Code on Page 10 of this Report

```
/* Create analysis file. */

libname IN "location of NIS file" ;
data DIABETES ;
  set IN.NIS_2012_CORE ;
  if DXCCS1 = 50 then DIABETES = 1; else DIABETES = 0 ;
  DISCHGS = 1 ;
run ;

/* Obtain estimates: The following SAS code produces estimates of the sums, means, and
standard errors for the number of discharges, length of stay, and total hospital charges. */

proc surveymeans data= DIABETES sum std mean stderr NOMCAR;
  weight DISCWT ;
  cluster HOSP_NIS ;
  strata NIS_STRATUM ;
  var DISCHGS LOS DIED TOTCHG ;
  domain DIABETES ;
run ;
```

Table A.3 is a copy of Table 1 in the report showing the mean, sum, and the associated variance estimates for the domain of DIABETES=1. The addition on the NOMCAR option to the PROC SURVEYMEANS made *no* difference in the variance estimates.

Table A.3 SAS Output for DIABETES = 1, Domain Statistics

Variable	Label	Mean	Std Error of Mean	Sum	Std Dev
DISCHGS		1.000000	0 <same value using NOMCAR>	528,030	5,451.714493 <same value using NOMCAR>
LOS	Length of stay (cleaned)	4.601903	0.025091 <same value using NOMCAR>	2,429,116	29,480 <same value using NOMCAR>
DIED	Died during hospitalization	0.005531	0.000231 <same value using NOMCAR>	2919.999322	124.540093 <same value using NOMCAR>
TOTCHG	Total charges (cleaned)	33,914	422.923358 <same value using NOMCAR>	17,532,625,874	290,339,323 <same value using NOMCAR>

Source: Agency for Healthcare Research and Quality (AHRQ), Healthcare Cost and Utilization Project (HCUP), National Inpatient Sample (NIS), 2019

In contrast, we see a difference in the variance estimates for total charges for the domain of DIABETES=0 using the NOMCAR option. Table A.4 shows the mean, sum, and the associated variance estimates for the domain of DIABETES=0. It is noted in red when the variance estimate is different using NOMCAR. Evidently, the hospitals with TOTCHG missing on all observations do not have any discharges with a principal diagnosis of diabetes.

Table A.4 SAS Output for DIABETES = 0, Domain Statistics

Variable	Label	Mean	Std Error of Mean	Sum	Std Dev
DISCHGS		1.000000	0 <same value using NOMCAR>	35,956,816	287,436 <same value using NOMCAR>
LOS	Length of stay (cleaned)	4.508937	0.016994 <same value using NOMCAR>	162,077,482	1,415,764 <same value using NOMCAR>
DIED	Died during hospitalization	0.018642	0.000156 <same value using NOMCAR>	670,130.1141	7,721.9791 <same value using NOMCAR>
TOTCHG	Total charges (cleaned)	36,745	451.194131 <451.197386>	1,293,816,265,570	20,425,183,381 <20,426,853,606>

Source: Agency for Healthcare Research and Quality (AHRQ), Healthcare Cost and Utilization Project (HCUP), National Inpatient Sample (NIS), 2019

Please note that the MISSING option would be needed if the data element DIABETES was coded as 1 and missing (.) instead of 1 and 0. The MISSING option treats missing values as a valid (nonmissing) category for all categorical variables, which include CLASS, STRATA, CLUSTER, and DOMAIN variables.

Additional information on the NOMCAR option is available in the paper [Survey Analysis: Options for Missing Data](#) presented at the 2009 NorthEast SAS Users Group (NESUG).