WTADJX Example #2

SUDAAN Statements and Results Illustrated

- Nonresponse adjustment and standard-error estimation
- ADJUST = NONRESPONSE
- ADJUST = POST; POSTWGT
- CALVARS
- CLASS; VAR
- VDIFFVAR

Input Data Set(s): DAWN.SAS7bdat

Example

Using manufactured sample and frame data inspired by the Drug Abuse Warning Network (DAWN) survey and its public-use data sethttp://www.samhsa.gov/data/2k11/DAWN/2k9DAWNED/HTML/DAWN2k9ED.htm), we estimate the annual number of drug-related emergency department visits in the U.S. and by census region after adjusting for unit nonresponse by modeling response as a logistic function then using procedures in SUDAAN 11. We also compute standard errors for the resulting estimates assuming the response model is correct.

Solution

RLOGIST (LOGISTIC in stand-alone SUDAAN), WTADJUST, and WTADJX are all used to adjust for the simulated unit nonresponse in our sample data set. The set has a binary variable RESPONDENT, which was generated as a function of the (annual) number of drug-related emergency-department visits – the survey value of interest. That is to say, the sample units denoted as nonrespondents were not missing at random.

Nevertheless, we first estimate the total number of drug-related emergency-department visits and compute their standard errors using RLOGIST, WTADJUST, and WTADJX assuming (incorrectly) that nonrespondents were missing at random. We then use WTADJX correctly assuming nonresponse to be a logistic function of drug-related emergency-department visits and compare the result. We provide the more advanced user a test to compare result across models statistically.

In most of what follows, we assume the SAS-callable version SUDAAN is being used.

The following variables from the DAWN dataset are of interest in this example:

Variable	Definition
RECORD	
STRATUM	
BIG_N	Population size in stratum
N	Sample size in stratum
W	Weight (BIG_N/N)
REGION	East = 1; South = 2; Midwest = 3; West = 4
PUBLIC	Yes = 1 (the alternative is a privately owned hospital)
METRO	Yes = 1 (i.e., located in urban area)

FRAME_VISITS	Number of previous-year emergency-department visits recorded on the frame
ER_VISITS	Annual drug-related emergency-department (room) visits collected on the survey
RESPONDENT	Yes = 1; No = 0

The following step creates some more variables.

In *Exhibit 1*, we download the data set and create a new frame-visit overall size variable Z, by dividing FRAME_VISITS by 1000. It sometimes it helps to reduce the size of relatively large variables when running WTADJX, as we shall see.

We will assume that response is a logistic function of the log of either frame visits (FRAME_VISITS) or drug-related emergency-department visits (ER_VISITS). This sensibly means that a one percent increase in, say frame visits, changes the probability of response by *q* percent. Two variables are created for this purpose, LOG_FRAME and LOG_ER.

We create two frame-visit size calibration variables (PUBLICZ and METROZ) by multiplying corresponding original variables (PUBLIC and METRO) by Z.

Exhibit 1. Downloading the Data and Creating Some Variables

```
LIBNAME IN \\rtints29\sudaan\phil\";
DATA R; SET IN.DAWN;
Z = FRAME_VISITS/1000;
W1 = W - 1;
ZW1 = Z*W1;

LOG_FRAME = LOG(FRAME_VISITS);
LOG_ER = LOG(ER_VISITS);

PUBLICZ = PUBLIC*Z;
PUBLICZW1 = PUBLIC*Z*W1;
METROZ = METRO*Z;
METROZW1 = METRO*Z*W1;
```

Fitting the Same Model with RLOGIST and WTADJUST

We first assume that response is a logistic function of the log of frame visits. There are (at least) two reasonable ways to estimate drug-related emergency-department visits by region under this model, one uses RLOGIST and the other WTADJUST.

Since the design is stratified simple random sampling without replacement (DESIGN = STRWOR) the term VARNONADJ is added to the design statement in both procedures to calculate standard errors correctly under the assumed response model. Had a with-replacement design been specified (i.e., DESIGN = STRWR or WR), this addition would not be necessary. Also needed for proper standard-error computation when using a without-replacement design is the statement: TOTCNT BIG_N.

The addition of the CLASS REGION and VAR ER_VISITS statements in both procedures result in the estimated means and totals for the variable ER_VISITS being computed by region. These estimates are accompanied by their standard errors under the assumed response model. This is a new feature of SUDAAN 11.
Both procedures model response as a function of LOG_FRAME and an intercept since both employ the statement: MODEL RESPONDENT = LOG_FRAME;. The LOWERBD 1 and CENTER 2 statements in WTADJUST are there so that a standard logistic response model will be fit. This is done via calibration as opposed to the quasi-maximum-likelihood routine used in RLOGIST. If a different CENTER were used, estimated means and totals would be unaffected, but the fitted logistic model would be reparameterized. If we let the centering parameter be its default value (which is $(10^{20} + 1)/2$ in this context), the model model coefficients would not converge.

Exhibit 2 contains the code for the two procedures.

Exhibit 2. Using PROC RLOGIST and WTADJUST to Adjust for Unit Nonresponse

```
PROC RLOGIST DATA = R DESIGN = STRWOR VARNONADJ;  NEST STRATUM;
   WEIGHT W; TOTCNT BIG_N;
   CLASS REGION;  VAR ER_VISITS;
   MODEL RESPONDENT = LOG_FRAME;
RUN;

PROC WTADJUST DATA = R DESIGN = STRWOR ADJUST = NONRESPONSE
VARNONADJ;
   NEST STRATUM; WEIGHT W; TOTCNT BIG_N;
   CLASS REGION; VAR ER_VISITS ; LOWERBD 1; CENTER 2;
   MODEL RESPONDENT = LOG_FRAME;
RUN;
```

Slightly edited versions of two of the outputted tables are provided in *Exhibits 3* and *4.* These results are similar but not identical (by contrast, the adjusted estimates for means and totals are over 10% higher than when unadjusted). RLOGIST produces the estimates with lower standard errors in most cases with this data.

Exhibit 3. Two Key RLOGIST Results

```
-------------------------------------------------------------------------------------------------
Independent                                                                        P-value
  Variables and        Beta                     Lower 95%    Upper 95%              T-Test
  Effects              Coeff.          SE Beta   Limit Beta   Limit Beta   T-Test B=0   B=0
-------------------------------------------------------------------------------------------------
Intercept                    -2.80         1.35        -5.46        -0.14        -2.07     0.0393
LOG_FRAME                     0.27         0.14        -0.01         0.55         1.91     0.0575
-------------------------------------------------------------------------------------------------
```

and

```
----------------------------------------------------------------------------------------------------------------
|                 |                |  REGION                                                                    |
| Variable        |                |----------------------------------------------------------------------------|
|                 |                | Total        | 1             | 2             | 3             | 4             |
----------------------------------------------------------------------------------------------------------------
|                 |                |              |               |               |               |               |
| ER_VISITS       | Mean           |       858.02 |       2190.32 |       1159.92 |        426.00 |       1540.93 |
|                 | SE Mean        |        62.93 |        299.67 |        227.26 |         36.20 |        310.60 |
|                 | Total          |   5395035.89 |     775779.95 |    1821447.05 |    1499939.77 |    1297869.12 |
|                 | SE Total       |    404227.38 |     126787.02 |     372572.36 |     171359.14 |     333987.63 |
```

Exhibit 4. The Analogous RLOGIST Results

```
-----------------------------------------------------------------------------------------------------------------
Independent
  Variables and                                                                       Respond-   Nonresp-
  Effects                                                                  P-value    ent        ondent
                Beta                    Lower 95%    Upper 95%             T-Test     Sample     Sample
                Coeff.        SE Beta   Limit Beta   Limit Beta   T-Test B=0   B=0        Size       Size
-----------------------------------------------------------------------------------------------------------------
Intercept             3.13       1.63        -0.07         6.33         1.92     0.0555        154        192
LOG_FRAME            -0.30       0.17        -0.64         0.03        -1.78     0.0760          .          .
```

and

```
--------------------------------------------------------------------------------------------------------------
|                 |             | REGION                                                                     |
| Variable        |             |----------------------------------------------------------------------------|
|                 |             | Total        | 1             | 2             | 3             | 4             |
--------------------------------------------------------------------------------------------------------------
|                 |             |              |               |               |               |               |
| ER_VISITS       | Mean        |       843.90 |       2170.05 |       1135.54 |        424.14 |       1528.21 |
|                 | SE Mean     |        70.07 |        303.76 |        224.86 |         35.59 |        312.18 |
|                 | Total       |   5316600.95 |     755474.88 |    1786181.15 |    1504703.25 |    1270241.65 |
|                 | SE Total    |    441439.17 |     131260.91 |     377767.89 |     171569.81 |     335005.73 |
```

An advantage of WTADJUST over RLOGIST is that one can directly output the adjusted weights by adding a statement like:

OUTPUT IDVAR WTFINAL ADJFACTOR/FILENAME=OUT REPLACE;

WTFINAL are the adjusted weights, ADJFACTOR the adjustment factors used to produce them, and OUT is the data set containing both as well as other variables listed on a separate IDVAR statement. There is no parallel statement in RLOGIST.

Although we are concentrating here on means and totals, ratios and their standard errors under the assumed response model can also be computed. Adding the variable PUBLIC_VISITS = PUBLIC*ER_VISITS, one could estimate the fraction of drug-related emergency-department visits at public hospitals with RLOGIST using the code in Exhibit 5. The output is in *Exhibit 6*.

Exhibit 5. Estimating a Ratio with Weights Adjusting Using RLOGIST

```
DATA R; SET R; PUBLIC_VISITS = PUBLIC*ER_VISITS;

PROC RLOGIST DATA = R DESIGN = STRWOR VARNONADJ;  NEST STRATUM;
   WEIGHT W; TOTCNT BIG_N;
   CLASS REGION;  NUMER PUBLIC VISITS; DENOM ER_VISITS;
   MODEL RESPONDENT = LOG_FRAME;
    OUTPUT RHAT SE_RHAT/FILENAME = OUT REPLACE;
RUN;
PROC PRINT DATA = OUT; ID REGION; VAR RHAT SE_HAT; RUN;
RUN;
```

Exhibit 6. The Output from Estimating a Ratio with Weights Adjusting Using RLOGIST

```
----------------------------------------------------------------------------------------------
|                |                  | REGION                                                  |
| Variable       |                  |---------------------------------------------------------|
|                |                  | Total       | 1           | 2           | 3           | 4           |
----------------------------------------------------------------------------------------------
|                |                  |             |             |             |             |             |
| PUBLIC_VISITS  | Ratio            |        0.44 |        0.92 |        0.91 |        0.00 |        0.00 |
|                | SE Ratio         |        0.06 |        0.04 |        0.05 |        0.00 |        0.00 |
----------------------------------------------------------------------------------------------
```

WTADJX to Fit the Same Model But in a Different Way

A third method of estimating drug-related emergency-department visits the same logistic model uses WTADJX. *Exhibit 7* again employs LOG_FRAME as the sole covariate in the model statement but adds the statement: CALVARS FRAME_VISITS, which means WTADJX attempts to calibrate on frame visits rather on in its log:

Exhibit 7. Using WTADJX to Adjust for Unit Nonresponse

```
PROC WTADJX DATA = R DESIGN = STRWOR ADJUST = NONRESPONSE  VARNONADJ ;
NEST STRATUM; WEIGHT W; TOTCNT BIG_N;
CLASS REGION; VAR ER_VISITS ; LOWERBD 1; CENTER 2;
MODEL RESPONDENT = LOG_FRAME;
CALVARS   FRAME_VISITS;
RUN;
```

Unfortunately, calibration does not occur even though the “parameters [the coefficient of the model] converge”, as can be seen in the output page displayed on *Exhibit 8.*

Exhibit 8. Calibration Fails (Using WTADJX)

```
-------------------------------------------------------------------------------------------------------------
             Maximum
             Adjustment
Calibration  Factor       Sum of        Sum of        Sum of Final                              Original
  Variables  Among        Original      Trimmed       Adjusted                    Final Weight  Unequal
             Responden-   Weights Over  Weights Over  Weights Over  Control       Sum Minus     Weighting
             ts           Respondents   Respondents   Respondents   Totals        Controls      Effect
-------------------------------------------------------------------------------------------------------------
Intercept          2.0078       2728.78       2728.78       5478.79       6300.00       -821.21        2.0094
FRAME_VISITS            .   65053163.78   65053163.78  130612288.42  130612288.39          0.03             .
-------------------------------------------------------------------------------------------------------------
```

It is imperative that this section of the output be checked whenever WTADJUST or WTADJX is used. If the numbers in the second-to-last column are not all zeros (or very close to it), then the estimates for means and totals will be specious. This strong statement does not apply when here are more calibration variables than model variables, which WTADJX allows. That is not the case here, however.

Simply replacing FRAME_VISITS by Z (which is FRAME_VISITS/1000) in the previous code fixes things, as can be seen with the new version of the output in *Exhibit 9.*

Exhibit 9. Calibration Fails No Longer

```
---------------------------------------------------------------------------------------------------------
             Maximum
             Adjustment
Calibration  Factor        Sum of        Sum of        Sum of Final                              Original
  Variables  Among         Original      Trimmed       Adjusted                    Final Weight  Unequal
             Responden-    Weights Over  Weights Over  Weights Over  Control       Sum Minus     Weighting
             ts            Respondents   Respondents   Respondents   Totals        Controls      Effect
---------------------------------------------------------------------------------------------------------
Intercept        3.1605        2728.78       2728.78       6300.00       6300.00          0.00      2.0094
Z                 .           65053.16      65053.16     130612.29     130612.29         -0.00       .
---------------------------------------------------------------------------------------------------------
```

With weighted totals equaling their control targets, we can concentrate on other aspects of the output in *Exhibit 10.*

Exhibit 10. Two Key WTADJX Results

```
-------------------------------------------------------------------------------------------
Independent                                                                     P-value
  Variables and       Beta                      Lower 95%    Upper 95%              T-Test
  Effects             Coeff.          SE Beta   Limit Beta   Limit Beta   T-Test B=0   B=0
-------------------------------------------------------------------------------------------
Intercept                      2.67       1.48        -0.25         5.59        1.80    0.0734
LOG_FRAME                     -0.25       0.15        -0.56         0.05       -1.64    0.1010
-------------------------------------------------------------------------------------------
```

and

```
----------------------------------------------------------------------------------------------------
|                 |             | REGION                                                            |
| Variable        |             |-------------------------------------------------------------------|
|                 |             | Total        | 1            | 2            | 3            | 4          |
----------------------------------------------------------------------------------------------------
|                 |             |              |              |              |              |            |
| ER_VISITS       | Mean        |       863.98 |      2198.93 |      1170.11 |       426.77 |    1546.26 |
|                 | SE Mean     |        55.20 |       299.70 |       219.21 |        37.26 |     307.20 |
|                 | Total       |   5443069.22 |    786582.43 |   1841426.54 |   1501971.84 | 1313088.42 |
|                 | SE Total    |    347760.02 |    131421.13 |    345990.63 |    171716.81 |  332202.08 |
----------------------------------------------------------------------------------------------------
```

The estimated model coefficients displayed in *Exhibit 10* are no more efficient than under RLOGIST in *Exhibit 3* (i.e., their coefficients of variation are not any smaller); in fact, the standard errors of the model coefficient are slightly higher using WTADJX rather than RLOGIST. The standard errors of the means and totals displayed in *Exhibit 10*, by contrast, are generally lower than when using RLOGIST. Recall that our goal is to estimate the total number of drug-related emergency-department visits. The estimated coefficients of the response model are simply means to an end.

Another Example of Using WTADJX to Fit the Same Model

In principle, it may be possible to increase the efficiency by adding additional variables to the CALVARS statement. That doesn't happen here. As an example, consider what occurs when we add Z to the CALVARS statement.

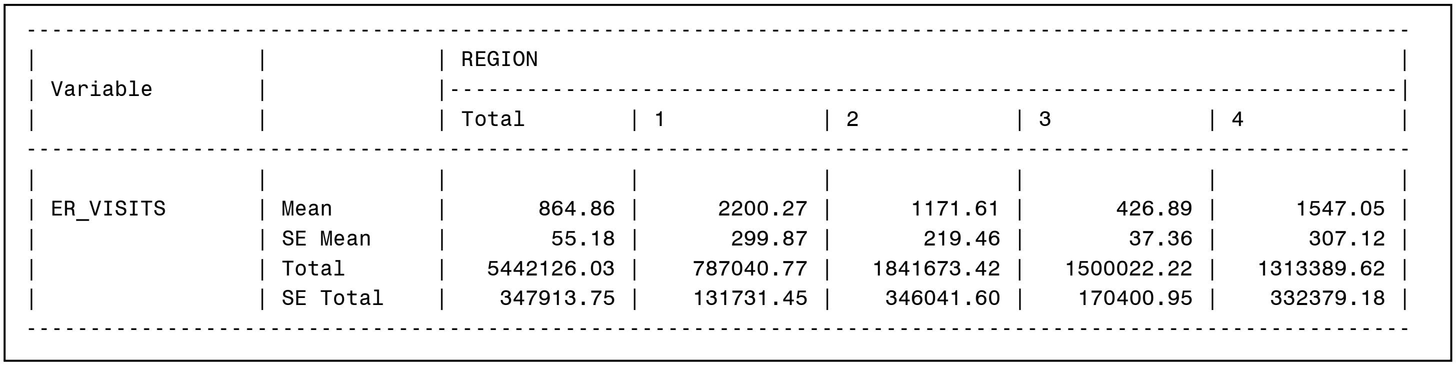
At first, we get the following message in the output:

DATA WARNING: Parameters have not converged in 10 iterations.

We can fix that by adding MAXITER = 20 after VARNONADJ (or by adding BESTIM = REDUCED, which also works in the case).

As can be seen from the results in *Exhibit 11,* the impact or changing the CALVARS statement on the computed standard errors is slight.

Exhibit 11. Standard Errors from WTADJX with a Different CALVARS Statement



```
----------------------------------------------------------------------------------------------------
|                 |             | REGION                                                           |
| Variable        |             |------------------------------------------------------------------|
|                 |             | Total        | 1            | 2            | 3            | 4            |
----------------------------------------------------------------------------------------------------
|                 |             |              |              |              |              |              |
| ER_VISITS       | Mean        |       864.86 |      2200.27 |      1171.61 |       426.89 |      1547.05 |
|                 | SE Mean     |        55.18 |       299.87 |       219.46 |        37.36 |       307.12 |
|                 | Total       |   5442126.03 |    787040.77 |   1841673.42 |   1500022.22 |   1313389.62 |
|                 | SE Total    |    347913.75 |    131731.45 |    346041.60 |    170400.95 |    332379.18 |
----------------------------------------------------------------------------------------------------
```

Adding other variables to the CALVARS statement, like METRO or METROZ results in general *decreases* in efficiency (not shown). There is likewise little to be gained in terms of potential bias from adding variables to the MODEL statement. There is the possibility of efficiency gains, but these can be accomplished using ADJUST =POST, which will be addressed in a later section of this set of examples.

An Example of Using WTADJX When Nonrespondents are Assumed Not Missing at Random
and a Test for Comparing the Results of Fitting Different Response Models

Let us return to the the code used Exhibit 7. Recall we replaced FRAME_VISITS by Z in the CALVARS statement so that the control targets could be reached.

In *Exhibit 12*, we also replace LOG_FRAME in the MODEL statement by LOG_ER. This profoundly alters the response model implicitly fitted by WTADJX from one that assumes response is a logistic function of a frame value, the log the frame number of visits, to one that assumes response is a function of a survey value, the log of the number of drug-related emergency-department visits). Nonrespondents are no longer assumed to be missing at random.

Exhibit 12. Using WTADJX to Adjust for Unit Nonresponse that is not Missing at Random

```
PROC WTADJX DATA = R DESIGN = STRWOR ADJUST = NONRESPONSE
VARNONADJ ;
  NEST STRATUM; WEIGHT W; TOTCNT BIG_N;
  CLASS REGION; VAR ER_VISITS ; LOWERBD 1; CENTER 2;
  MODEL RESPONDENT = LOG_ER;
  CALVARS   Z;
  OUTPUT TOTAL SE_TOTAL/FILENAME=OUT0 REPLACE;
RUN;
```

Rather than printing the means and total here, we have output the totals by region and their standard errors under the model into OUT0 for future comparison.

We now construct a test of whether there is a significant difference between estimates computed using MODEL RESPONDENT = LOG_ER in place of MODEL RESPONDENT = LOG_FRAME. The code that follows creates an artificial PSU in a new data set R2 containing two records for each record in R. One is assigned to Domain 1, and weights are effectively adjusted with MODEL RESPONDENT = LOG_ER; the other is assigned to Domain 2, and weights are effective adjusted with MODEL RESPONDENT = LOG_FRAME.

This is done by creating the model variables X1 and X2 and calibration variable Z1 and Z2. X1 and Z1 are zero for records in Domain 2, while X2 and Z2 are zero for records in Domain 1. Dummy variables D1 and D2 are defined analogously and used in place of the intercept.
The statements:

```
CLASS  DOMAIN; VDIFFVAR DOMAIN =(1 2);
```

are used to create an estimate of the difference between the domain means; that is, the estimated total number of drug-related emergency-department visits per hospital computed under each response model. Contrasts of this sort should be done assuming the sampling design is with replacement. In fact, *they cannot be done correctly otherwise* (i.e., when a without-replacement design is specified) *in SUDAAN 11*. Hence, we have DESIGN = WR, no VARNONADJ and no TOTCNT = BIG_N in *Exhibit 13*.

Exhibit 13. Using WTADJX to Test the Difference Between the Outputs of Competing Response Models

```
DATA R2; SET R;
PSU + 1;
 X1 = LOG_ ER; X2 = 0;  Z1 = LOG_FRAME ; Z2 = 0; D1 = 1; D2 = 0; DOMAIN = 1; OUTPUT;
 X1 = 0; X2 = LOG_FRAME;  Z1 = 0; Z2 = LOG_FRAME ; D1 = 0; D2 = 1; DOMAIN = 2; OUTPUT;

PROC WTADJX DATA = R2 DESIGN = WR ADJUST = NONRESPONSE ;  NEST STRATUM PSU;
WEIGHT W;
  CLASS  DOMAIN; VDIFFVAR DOMAIN =(1 2);
  VAR ER_VISITS ; LOWERBD 1; CENTER 2;
  MODEL RESPONDENT = D1 D2  X1 X2/NOINT;
  CALVARS D1 D2 Z1 Z2/NOINT;
RUN;
```

The key output result is displayed in *Exhibit 14.*

Exhibit 14. The Difference Between the Outputs of Competing Response Models

```
----------------------------------------------------
|                 |                  | Contrast     |
| SUDAAN Reserved |                  |--------------|
| Variable One    |                  | CONTRAST_1   |
----------------------------------------------------
|                 |                  |              |
| Total           | Cntrst Mean      |        -9.90 |
|                 | SE Cntrst Mean   |         5.72 |
|                 | T-Test Mean=0    |        -1.73 |
|                 | P-Value Mean=0   |         0.08 |
|                 | Cntrst Total     |    -62385.65 |
|                 | SE Cntrst Total  |     36042.32 |
----------------------------------------------------
```

This exhibit tells us that the mean (and the total since we are calibrating with an intercept) is significantly less when we correctly assume that nonrespondents are not missing at random at the 0.08 level.

In this example, we know the true response model. In practice, that will usually not be the case. Nevertheless, code similar to that developed above can be used to assess the impact of assuming one response model, which includes not only the variables in the MODEL statement but the choices for the centering and bounding parameters, when another is true. Differences in variables that appear in CALVARS statements are *not* part of the model assumption.

Examples Using ADJUST = POST

One can often remove the (large-sample) bias from our estimated totals while at the same time reducing their standard errors by using ADJUST = POST in place of ADJUST = NONRESPONSE with the code in *Exhibit 15*. Notice that in the newly added POSTWGT statement in the following code, there is a total for both the intercept (6300) and for Z (134000). The results are output into OUT1 for future comparison.

Exhibit 15. Using WTADJX with Frame Data to Adjust for Unit Nonresponse that is not Missing at Random

```
PROC WTADJX DATA = R DESIGN = STRWOR ADJUST = POST
VARNONADJ;
  NEST STRATUM; WEIGHT W; TOTCNT BIG_N;
  CLASS REGION; VAR ER_VISITS ; LOWERBD 1; CENTER 2;
  MODEL RESPONDENT = LOG_ER ;
  CALVARS   Z ;
  POSTWGT 6300  134000 ;
  OUTPUT TOTAL SE_TOTAL/FILENAME=OUT1 REPLACE;
RUN;
```

We can drive standard errors down further with the code in *Exhibit 16,* which combines adjusting for bias with standard-error reduction. Since REGION*Z is in the MODEL and CALVARS statements, LOG_FRAME replaces Z in the CALVARS statement to avoid a singularity.

Exhibit 16. Using WTADJX with Frame Data to Adjust for Unit Nonresponse that is not Missing at Random and Reduce Standard Errors Simultaneously

```
PROC WTADJX DATA = R DESIGN = STRWOR ADJUST = POST
VARNONADJ ;
   NEST STRATUM; WEIGHT W; TOTCNT BIG_N;
   CLASS REGION; VAR ER_VISITS ; LOWERBD 1; CENTER 2;
   MODEL RESPONDENT = LOG_ER PUBLICZ METROZ  REGION*Z;
   CALVARS  LOG_FRAME  PUBLICZ METROZ ; REGION*Z;
   POSTWGT 6300 59500 58000   44000 22000   43000   33000 36000;
   OUTPUT TOTAL SE_TOTAL/FILENAME=OUT2 REPLACE;
RUN;
```

One runs the risk when employing LOWERBD 1 that calibration will fail. Although that doesn't happen here, a better approach is to split bias adjustment and standard-error reduction up into two steps.

After removing the "RUN;" from Exhibit 12, we add a few lines of code before the second WTADJX so that we can conduct a second WTADJX procedure using nearly pseudo-optimal calibration. In those added lines, the adjusted weight from the first run of WTADJX is renamed A. The new code is shown *Exhibit 17.*

Exhibit 17. Using WTADJX in Two Steps: One for Nonresponse Adjustment and for Nearly Quasi-optimal Calibration

```
[continues from Exhibit 12 without the last RUN;]

   IDVAR  RECORD;
   OUTPUT IDVAR WTFINAL ADJFACTOR/FILENAME=STEP1 REPLACE;
RUN:

DATA R2; MERGE R STEP1; BY RECORD;  A = WTFINAL;

PROC WTADJX DATA = R2 DESIGN = STRWR ADJUST = POST;
   NEST _ONE_; WEIGHT A;
  CLASS  REGION;  VAR ER_VISITS  ;
  MODEL RESPONDENT =   W1 PUBLICZW1 METROZW1
REGION*ZW1/NOINT;
  CALVARS   _ONE_ PUBLICZ METROZ  REGION*Z/NOINT;
  POSTWGT 6300  58000   44000  22000   43000   33000 36000;
  OUTPUT TOTAL SE_TOTAL/FILENAME=OUT3 REPLACE;
RUN;
```

Unfortunately, SUDAAN 11 can not correctly compute large-sample standard errors when there are more than one calibration steps. Instead, in the code in Exhibit 17 we compute an upwardly biased measure by using DESIGN = WR (which means dropping VARNONADJ and TOTCNT BIG_N) and NEST _ONE_ (otherwise Stratum 4 would have only a single member, and WTADJX would fail). We also removed ZWI from the MODEL statement and LOG_FRAME from the CALVARS statement, since the latter was not part of a roughly linear model for drug-related emergency-department visits, unlike the other variables in the CALVARS statement.

Notice that there are no bounds in the code, which leaves us with the defaults: a lower bound of 0, a center of 1, and a virtually infinite upper bound. As long as we keep the center at 1, we can adjust the lower and upper bounds without affecting the large-sample unbiasedness of the resulting estimated total.

If the original design specification were with replacement (i.e., DESIGN = STRWR or WR), and the NEST statement didn't have to change to avoid nonconvergence, then the standard error measure produced by the above two-step approach may not be conservative. It fully captures the impact of the second WTADJX on the standard error but not the first, capturing its impact on the weights but not its other effects. These effects are likely to be small in practice.

Comparing the Results

Exhibit 18 combines the output, totals and coefficients of variation (CVs) from each set of estimates.
The estimates using ADJUST=NONRESPONSE are labeled NONTOTAL and NONCV.
The estimates using ADJUST = POST and CALVARS Z are labeled ZTOTAL and ZCV.
The estimates using ADJUST = POST and CALVARS Z PUBLICZ METROZ REGION * Z are labeled ZPMRTOTAL and ZPMRCV.
The estimates from the second WTADJX, which used nearly pseudo-optimal calibration are labeled QOTOTAL and QOCV.

The estimated totals and their CV's are displayed in *Exhibit 19.*

Exhibit 18. Comparing the Results (the Code)

```
DATA OUT0; SET OUT0;
NONCV = SETOTAL/TOTAL;
NONTOTAL = TOTAL;
RUN;

DATA OUT1; SET OUT1;
ZCV = SE_TOTAL/TOTAL;
ZTOTAL = TOTAL;
RUN;

DATA OUT2; SET OUT2;
ZPMRCV = SE_TOTAL/TOTAL;
ZPMRTOTAL = TOTAL;
RUN;
DATA OUT3; SET OUT3;
QOCV = SE_TOTAL/TOTAL;
QOTOTAL = TOTAL;
RUN;

DATA C; MERGE  OUT0 OUT1 OUT2 OUT3; BY VARIABLE REGION;

NONCV = ROUND(NONCV * 100, .01);
ZCV=  ROUND(ZCV * 100, .01);
ZPMCV=  ROUND(ZPMCV * 100, .01);
QOCV = ROUND(QOCV * 100, .01);


PROC PRINT; ID  REGION; VAR NONTOTAL ZTOTAL ZPMRTOTAL QOTOTAL;
PROC PRINT; ID  REGION; VAR  NONCV ZCV ZPMRCV QOCV; RUN;
```

Exhibit 19. Comparing the Results (the Output)

```
           REGION     NONTOTAL       ZTOTAL       ZPMTOTAL       QOTOTAL

              0      5380683.57    5525234.55    5532923.19    5630018.44
              1       778121.80     813097.48     775358.20     842008.47
              2      1815199.62    1878727.90    1844560.88    1867757.33
              3      1490449.25    1489152.34    1399812.14    1410678.05
              4      1296912.89    1344256.82    1513191.97    1509574.59

                      REGION    NONCV     ZCV     ZPMCV    QOCV

                         0       6.50     3.65     3.17    2.32
                         1      16.74    19.07     9.74    5.65
                         2      18.54    16.97     3.21    3.12
                         3      11.34    11.54     4.32    4.02
                         4      25.07    23.67     5.93    6.04
```

Not surprisingly, the three set of final weights using ADJUST = POST generally produce estimated totals with more efficiency than using ADJUST = NONRESPONSE with one model variable did. Calibrating on more variables increases efficiency, and calibrating in two steps, which allows us to use nearly pseudo-optimal calibration, increases efficiency even further except in the West (Region 4).