Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub30 LSmith

Prepared by Jane Wang
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
September 01, 2015

Table of Contents

1 Standard Disclaimer

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

2 Study Background

The TEDDY study was designed to follow children with and without a family history of T1D to understand the environmental factors that contribute to the disease. Newborn children younger than 4 months were screened for high-risk HLA alleles, and those with qualifying haplotypes were eligible for follow-up. Information is collected on medical information (infections, medication, immunizations), exposure to dietary and other environmental factors, negative life events, family history, tap water, and measurements of psychological stress. Biospecimens, including blood, stool, urine, and nail clippings, are taken at baseline and follow-up study visits. The primary outcome measures include two endpoints—the first appearance of one or more islet cell autoantibodies (GADA, IAA, or IA-2A), confirmed at two consecutive visits, and development of T1D. The cohort will be followed for 15 years, or until the occurrence of one of the primary endpoints.

3 Archived Datasets

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the TEDDY folder in the data package. For this replication, variables were taken from “Pub30_LSmith_NIDDK_submission” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Laura Smith et al [1] Diabetes Care 2014; 37:325–331 | DOI: 10.2337/dc13-0449. To verify the integrity of the dataset, descriptive statistics of preventive actions variable (table 1), Socio demographic variables (table 2), and psychosocial variables (table 3) were computed, by 6 months and 15 months visits.

5 Results

Table A, C, and E lists the variables that were used in the replication and Table B, D, and F compares the results calculated from the archived data file to the results published in Table 1, Table 2, and Table 3. The results of the replication are the similar to the published results.

6 Conclusions

The NIDDK repository is confident that the TEDDY Pub30 LSmith data files to be distributed are a true copy of the study data.

7 References

Laura B. Smith, Kristian F. Lynch,Judith Baxter, Barbro Lernmark, Roswith Roth, Tuula Simell, Suzanne Bennett Johnson, and the TEDDY Study Group. Factors Associated With Maternal-Reported Actions to Prevent Type 1 Diabetes in the First Year of the TEDDY Study. Diabetes Care 2014; 37:325–331 | DOI: 10.2337/dc13-0449

Table A: Variables used to replicate Tables 1 in the publication.

Table Variable	Variables Used in Replication from the "Table 1" Dataset for 6 months preventive actions	Variables Used in Replication from the "Table 1" Dataset for 15 months preventive actions
No action reported	prevent_action_yes_6mo	prevent_action_yes_15mo
Any action reported	anything_prevent_six	anything_prevent_fifteen
Breast-feeding	breast_feed_6mo	breast_feed_15mo
Diet	diet_6mo	diet_15mo
Dietary supplements	diet_sup_6mo	diet_sup_15mo
Illness prevention	ill_prevent_6mo	ill_prevent_15mo
Alternative medicine	med_vitamins_6mo	med_vitamins_15mo
Physical activity	physical_activity_6mo	physical_activity_15mo
Stress reduction	stress_reduction_6mo	stress_reduction_15mo
Other action	other_6mo	other_15mo

Table B: Comparison of values computed in integrity check to reference article Table 1 values: Action to prevent type 1 diabetes by category

Preventive actions	**6 Months [Manuscript]**	**6 Months [DSIC]**	**6 Months [Difference]**	**15 Months [Manuscript]**	**15 Months [DSIC]**	**15 Months [Difference]**
No action reported	70.1 (5,337)	70.1(5331)	0 (0)	57.2 (3,720)	57.2 (3715)	0 (0)
Any action reported	29.9 (2,276)	29.9(2276)	0 (0)	42.8 (2,783)	42.8 (2778)	0 (0)
Breast-feeding	6.2 (486)	6.4 (486)	-0.2 (0)	3.6 (286)	4.4 (286)	-0.8 (0)
Diet	20.9 (1,631)	21.4 (1631)	-0.5 (0)	29.2 (2,279)	35.0 (2279)	-3.8 (0)
Dietary supplements	1.3 (102)	1.3(102)	0 (0)	2.2 (167)	2.6 (167)	-0.4 (0)
Illness prevention	1.8 (150)	2.0 (150)	-0.2 (0)	2.1 (165)	2.5 (165)	-0.4 (0)
Alternative medicine	0.1 (8)	0.1(8)	0 (0)	0.2 (15)	0.2(15)	0 (0)
Physical activity	1.3 (102)	1.3 (102)	0 (0)	4.5 (349)	5.4 (349)	-0.9 (0)
Stress reduction	1.3 (102)	1.3(102)	0 (0)	1.5 (118)	1.8(118)	-0.9 (0)
Other action	5.5 (436)	5.7(436)	-0.2 (0)	6.3 (493)	7.6 (493)	-0.7 (0)

Table C: Variables used to replicate Tables 2 in the publication.

Table Variable	**Variables Used in Replication from the "Table 2"**
Married/cohabitating	married_living
Maternal education	education_mom_group3
Sex of child	female
Child's ethnic minority status	ethnic_minority
Only child	single_child
Country of residence	country
FDR with type 1 diabetes	fdr
Maternal age, years	maternal_age
Household crowding,normed score	crowding_norm

Table D: Comparison of values computed in integrity check to reference article Table 2 values: Socio demographic variables associated with actions to prevent diabetes

	6 months Total N[Manuscript]	6 months Total N[DSIC]	6 months Total N[Difference]	6 months yes (%)[Manuscript]	6 months yes (%)[DSIC]	6 months yes (%)[Difference]
Married/cohabitating No	300	300	0	33.3	33.3	0.0
Married/cohabitating Yes	6,915	6,915	0	29.8	29.8	0.0
Maternal education High school/	1,427	1,427	0	24.8	24.8	0.0
Primary Trade school	1,802	1,802	0	29.9	29.9	0.0
Higher education	3,976	3,976	0	32	32.0	0.0
Sex of child Male	3,871	3,871	0	29.9	29.9	0.0
Sex of child Female	3,742	3,742	0	29.9	29.9	0.0
Child's ethnic minority status No	6,070	6,070	0	29.3	29.3	0.0
Child's ethnic minority status Yes	1,142	1,142	0	32.8	32.8	0.0
Only child No	4,185	4,185	0	28.9	28.9	0.0
Only child Yes	3,028	3,028	0	31.5	31.5	0.0
Country of residence U.S.	3,111	3,111	0	31.7	31.7	0.0
Country of residence Finland	1,627	1,627	0	32.5	32.5	0.0
Country of residence Germany	522	522	0	41.8	41.8	0.0
Country of residence Sweden	2,353	2,353	0	23.1	23.1	0.0
FDR with type 1 diabetes No	6,787	6,787	0	28	28.0	0.0
FDR with type 1 diabetes Yes	826	826	0	45.5	45.5	0.0
Maternal age, years	30.4 (5.2)	30.4(5.2)	0(0)	30.9 (5.3)	30.9(5.2)	0(0)
Household crowding, normed score	2.03 (1.18)	2.03 (1.18)	0(0)	2.02 (1.19)	2.02(1.19)	0(0)

	15 months Total N[Manuscript]	15 months Total N[DSIC]	6 months Total N[Difference]	15 months yes (%)[Manuscript]	15 months yes (%)[DSIC]	15 months yes (%)[Difference]
Married/cohabitating No	231	231	0	48.1	48.1	0.0
Married/cohabitating Yes	6,188	6,188	0	42.4	42.4	0.0
Maternal education High school/	1,222	1,222	0	38.5	38.5	0.0
Primary Trade school	1,550	1,550	0	43.2	43.2	0.0
Higher education	3,643	3,643	0	43.8	43.8	0.0
Sex of child Male	3,333	3,333	0	42.9	42.9	0.0
Sex of child Female	3,170	3,170	0	42.6	42.6	0.0
Child's ethnic minority status No	5,441	5,441	0	41.6	41.6	0.0
Child's ethnic minority status Yes	926	926	0	48.2	48.2	0.0
Only child No	3,706	3,706	0	40	40.0	0.0
Only child Yes	2,710	2,710	0	46.2	46.2	0.0
Country of residence U.S.	2,585	2,585	0	42.1	42.1	0.0
Country of residence Finland	1,439	1,439	0	44	44.0	0.0
Country of residence Germany	403	403	0	50.4	50.4	0.0
Country of residence Sweden	2,076	2,076	0	41.1	41.1	0.0
FDR with type 1 diabetes No	5,776	5,776	0	41.7	41.7	0.0
FDR with type 1 diabetes Yes	727	727	0	50.9	50.9	0.0
Maternal age, years	30.6 (5.0)	30.6(5.0)	0(0)	31.1 (5.1)	31.1(5.1)	0(0)
Household crowding, normed score	2.00 (1.18)	2.00 (1.18)	0(0)	2.00 (1.18)	2.00(1.18)	0(0)

Table E: Variables used to replicate Tables 3 in the publication.

Table Variable	**Variables Used in Replication from the "Table 3" Dataset for 6 months**	**Variables Used in Replication from the "Table 3" Dataset for 15 months**
Worry about diabetes	worry_six	worry_fifteen
Baby blues (EPDS)c	blues2	blues2
Diabetes risk perception	anydev_six	anydev_fifteen
Diabetes anxiety (STAI) Total Score/Normed Score	stai_six/stai_six_norm	stai_fifteen/stai_fifteen_norm
Baby blues (EPDS) Total Score/Normed Score	totalblues/totalblues_norm	totalblues/totalblues_norm
Depression (WBQ) Total Score/Normed Score	depressiontotal/depress_norm2	depressiontotal/depress_norm2
Belief that diabetes risk can be reduced Total Score/Normed Score	reduce_6mo_agree	reduce_15mo_agree

Table F: Comparison of values computed in integrity check to reference article Table 3 values: Caregiver psychosocial variables associated with actions to prevent diabetes

	6 months Total N[Manuscript]	6 months Total N[DSIC]	6 months Total N[Difference]	6 months yes (%)[Manuscript]	6 months yes (%)[DSIC]	6 months yes (%)[Difference]
Worry about diabetes Never	1,174	1,174	0	17.2	17.2	0
Worry about diabetes Rarely	3,737	3,737	0	27.7	27.7	0
Worry about diabetes Sometimes to very often	2,667	2,667	0	38.6	38.6	0
Baby blues (EPDS) No	6,926	6,926	0	29	29.1	0
Baby blues (EPDS) Yes	681	681	0	38.6	38.6	0
Diabetes risk perception Underestimated	2,941	2,941	0	24.3	24.3	0
Diabetes risk perception Accurate	4,644	4,644	0	33.2	33.2	0

	No, mean (SD)[Manuscript]	**No, mean (SD)[DSIC]**	**No, mean (SD)[Difference]**	**Yes, mean (SD)[Manuscript]**	**Yes, mean (SD)[DSIC]**	**Yes, mean (SD)[Difference]**
Diabetes anxiety (STAI) Total score	35.2 (9.7)	35.2(9.7)	0(0)	37.7 (10.3)	37.7 (10.3)	0(0)
Diabetes anxiety (STAI) Normed score	1.89 (1.17)	1.89(1.17)	0(0)	2.19 (1.19)	2.19 (1.19)	0(0)
Baby blues (EPDS) Total score	5.9 (4.3)	5.91(4.3)	0(0)	6.8 (4.5)	6.8(4.5)	0(0)
Baby blues (EPDS) Normed score	2.49 (1.35)	2.49(1.35)	0(0)	2.76 (1.34)	2.76 (1.34)	0(0)
Depression (WBQ) Score						
Depression (WBQ) Normed score						
Belief that diabetes risk can be reduced	3.30 (1.04)	3.30(1.05)	0(-0.1)	3.91 (0.88)	3.91 (0.88)	0(0)

	15 months Total N[Manuscript]	**15 months Total N[DSIC]**	**6 months Total N[Difference]**	**15 months yes (%)[Manuscript]**	**15 months yes (%)[DSIC]**	**15 months yes (%)[Difference]**
Worry about diabetes Never (1)	1,163	1,163	0	28.9	28.9	0
Worry about diabetes Rarely (2)	3,212	3,312	0	40.5	40.5	0
Worry about diabetes Sometimes to very often (3)	1,969	1,969	0	54.3	54.3	0
Baby blues (EPDS) No	5,774	5,774	0	41.6	41.6	0
Baby blues (EPDS) Yes	525	525	0	50.5	50.5	0
Diabetes risk perception Underestimated	2,485	2,485	0	37.3	37.3	0
Diabetes risk perception Accurate	3,998	3,998	0	46	46.0	0

	No, mean (SD)[Manuscript]	**No, mean (SD)[DSIC]**	**No, mean (SD)[Difference]**	**Yes, mean (SD)[Manuscript]**	**Yes, mean (SD)[DSIC]**	**Yes, mean (SD)[Difference]**
Diabetes anxiety (STAI) Total score	33.2 (9.4)	33.2 (9.4)	0(0)	35.9 (9.9)	35.9 (9.9)	0(0)
Diabetes anxiety (STAI) Normed score	1.64 (1.17)	1.63 (1.17)	0.01(0)	1.99 (1.17)	1.98 (1.17)	0.01(0)
Baby blues (EPDS) Total score	5.8 (4.2)	5.8 (4.2)	0(0)	6.5 (4.4)	6.5 (4.4)	0(0)
Baby blues (EPDS) Normed score	2.46 (1.33)	2.46 (1.33)	0(0)	2.66 (1.34)	2.66 (1.34)	0(0)
Depression (WBQ) Score	3.15 (2.17)	3.15 (2.17)	0(0)	3.24 (2.18)	3.24 (2.18)	0(0)
Depression (WBQ) Normed score	2.01 (0.72)	2.01 (0.72)	0(0)	2.04 (0.71)	2.04 (0.71)	0(0)
Belief that diabetes risk can be reduced	3.04 (1.04)	3.04 (0.98)	0(0.06)	3.80 (0.83)	3.80 (0.83)	0(0)

Attachment A: SAS Code

```
*************************************************************************************************************************
***Program:
***Programmer: Jane Wang
***Date Created: 08/19/2015
***Purpose:
*************************************************************************************************************************;

title1 "%sysfunc(getoption(sysin))";
title2 " ";

options nofmterr;
options nofmterr;
proc format;
 value YESNO
 0=no
 1=yes
;

value COUNTRY
1='US'
2='FIN'
3='GER'
4='SWE'
;
value HLARGC
1='DR3/4' 2='DR4/4' 3='DR4/8' 4='DR3/3' 5='DR4/4b' 6='DR4/1' 7='DR4/13' 9='DR4/9' 10='DR3/9';



/*
*** Data from the Primary hlarg paper that was converted to .csv format so that the DSIC data could be easily compared;
FILENAME table1  '/prj/niddk/ims_analysis/TEDDY/private_created_data/teddy_pub30_table1.csv';

*** Output CSV files that will be converted to .xls before being added to the DSIC document;
FILENAME out_t1  '/prj/niddk/ims_analysis/TEDDY/private_created_data/teddy_pub3_table1_dsic.csv';
*/
*** Reading in the analysis datasets used for the DSIC;
libname sas_data "/prj/niddk/ims_analysis/TEDDY/private_orig_data/Pub30_LSmith_niddk_submission/";
data pub30_lsmith_niddk              ; set sas_data.pub30_lsmith_niddk     ;


%macro baseline_freq1(dataset_name,var_name1,var_name2);

       *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc freq data = &dataset_name ;
```

```
        table &var_name1 /out = &var_name1._cross;
        title3 "Frequency table of the &var_name1. variable in the analysis dataset";

  data &var_name1._cross;
    set &var_name1._cross;
    level=&var_name1;
    rename count = count6
           PERCENT =pert6;


  proc print data = &var_name1._cross;

  proc freq data = &dataset_name ;
        table &var_name2 /out = &var_name2._cross;
        title3 "Frequency table of the &var_name2. variable in the analysis dataset";

  data &var_name2._cross;
    set &var_name2._cross;
    level=&var_name2;
    rename count = count15
           PERCENT = pert15;

*  proc print data = &var_name2._cross;

  data &var_name1._12(drop = level);
    merge &var_name1._cross &var_name2._cross;
    by level;
   length table_name $30.;
    table_name  ="&var_name1";
    level=&var_name1;

  proc print data = &var_name1._12;

%mend;

%macro baseline_freq2(dataset_name,var_name,);
  proc freq data = &dataset_name ;
        table (&var_name.)*prevent_action_yes_6mo ;
        title3 "Frequency table of the &var_name. variable in the analysis dataset";

        *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output CrossTabFreqs = work.&var_name._1cross;
  proc print data = &var_name._1cross;

  data &var_name._1count (keep = table &var_name Frequency  rename = (Frequency = count6));
    set &var_name._1cross;
    if not missing(&var_name) and prevent_action_yes_6mo = .;

  data &var_name._1freq (keep = table &var_name RowPercent  rename = (RowPercent = pert6));
```

```
    set &var_name._1cross;
    if not missing(&var_name) and prevent_action_yes_6mo = 1;

  proc print data = &var_name._1count;
  proc print data = &var_name._1freq;


  proc freq data = &dataset_name ;
       table (&var_name.)*prevent_action_yes_15mo ;
       title3 "Frequency table of the &var_name. variable in the analysis dataset";

       *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output CrossTabFreqs = work.&var_name._2cross;
  proc print data = &var_name._2cross;

  data &var_name._2count (keep = table &var_name Frequency  rename = (Frequency = count15));
    set &var_name._2cross;
    if not missing(&var_name) and prevent_action_yes_15mo = .;

  data &var_name._2freq (keep = table &var_name RowPercent  rename = (RowPercent = pert15));
    set &var_name._2cross;
    if not missing(&var_name) and prevent_action_yes_15mo = 1;

  proc print data = &var_name._2count;
  proc print data = &var_name._2freq;


  data &var_name._12(drop = table &var_name);
    merge &var_name._1count &var_name._1freq &var_name._2count &var_name._2freq;
    by &var_name;
    length table_name $30.;
    table_name  =compress("&var_name" || &var_name);

  proc print data = &var_name._12;

%mend;

%macro baseline_means(dataset_name,var_name);

proc sort data = pub30_lsmith_niddk;
  by prevent_action_yes_6mo;
       *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc means data = &dataset_name mean Std ;
       var &var_name.;
       by prevent_action_yes_6mo;
       title3 "Frequency table of the &var_name. variable in the analysis dataset";

       *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output Summary = work.&var_name._means1;
```

```
  run;


  data &var_name._means1;
    set &var_name._means1;
    length table_name $30.;
    table_name  ="&var_name";
    if prevent_action_yes_6mo ne .;

  proc print data = &var_name._means1;

   proc sort data = &var_name._means1;
     by table_name prevent_action_yes_6mo;
  data &var_name._means_1(drop = prevent_action_yes_6mo &var_name._Mean &var_name._StdDev i);
    set &var_name._means1;
    by table_name;
    array temp1(2) no_mean6 yes_mean6 ;
    array temp2(2) no_std6 yes_std6 ;
    retain yes_mean6 no_mean6 yes_std6 no_std6;
    if first.table_name then do i = 1 to 2;
      temp1(i) = .;
      temp2(i) = .;
    end;
      temp1(_n_) = round(&var_name._Mean,0.01);
      temp2(_n_) = round(&var_name._StdDev,0.01);
    if last.table_name;

  proc print data = &var_name._means_1;



proc sort data = pub30_lsmith_niddk;
  by prevent_action_yes_15mo;
        *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc means data = &dataset_name mean Std ;
        var &var_name.;
        by prevent_action_yes_15mo;
        title3 "Frequency table of the &var_name. variable in the analysis dataset";

        *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output Summary = work.&var_name._means2;
  run;


  data &var_name._means2;
    set &var_name._means2;
    length table_name $30.;
    table_name  ="&var_name";
```

```
    if prevent_action_yes_15mo ne .;

  proc print data = &var_name._means2;

   proc sort data = &var_name._means2;
     by table_name prevent_action_yes_15mo;
  data &var_name._means_2(drop = prevent_action_yes_15mo &var_name._Mean &var_name._StdDev i);
    set &var_name._means2;
    by table_name;
    array temp1(2) no_mean15 yes_mean15 ;
    array temp2(2) no_std15 yes_std15 ;
    retain yes_mean15 no_mean15 yes_std15 no_std15;
    if first.table_name then do i = 1 to 2;
      temp1(i) = .;
      temp2(i) = .;
    end;
      temp1(_n_) = round(&var_name._Mean,0.01);
      temp2(_n_) = round(&var_name._StdDev,0.01);
    if last.table_name;

  proc print data = &var_name._means_2;

  data &var_name._mean12;
    merge &var_name._means_1 &var_name._means_2;
    by table_name;

  proc print data = &var_name._mean12;

%mend;

%macro baseline_freq3(dataset_name,var_name,grounp_name);
  proc freq data = &dataset_name ;
        table (&var_name.)*&grounp_name ;
        title3 "Frequency table of the &var_name. variable in the analysis dataset";

        *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output CrossTabFreqs = work.&var_name._1cross;
  proc print data = &var_name._1cross;

  data &var_name._1count (keep = table &var_name Frequency );
    set &var_name._1cross;
    if not missing(&var_name) and &grounp_name = .;

  data &var_name._1freq (keep = table &var_name RowPercent );
    set &var_name._1cross;
    if not missing(&var_name) and &grounp_name = 1;

  proc print data = &var_name._1count;
  proc print data = &var_name._1freq;
```

```
  data &var_name._12(drop = table &var_name);
    merge &var_name._1count &var_name._1freq ;
    by &var_name;
    length table_name $30.;
    table_name  =compress("&var_name" || &var_name);
    merge_var = substr(table_name,1,index(table_name,'_')-1);

  proc print data = &var_name._12;

%mend;

%macro baseline_means2(dataset_name,var_name1,var_name2);

proc sort data = pub30_1smith_niddk;
  by prevent_action_yes_6mo;
       *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc means data = &dataset_name mean Std ;
       var &var_name1.;
       by prevent_action_yes_6mo;
       title3 "Frequency table of the &var_name1. variable in the analysis dataset";

       *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output Summary = work.&var_name1._means1;
  run;


  data &var_name1._means1;
    set &var_name1._means1;
    length table_name $30.;
    table_name  ="&var_name1";
    if prevent_action_yes_6mo ne .;
    merge_var = substr(table_name,1,index(table_name,'_')-1);

  proc print data = &var_name1._means1;

   proc sort data = &var_name1._means1;
     by table_name prevent_action_yes_6mo;
  data &var_name1._means_1(drop = prevent_action_yes_6mo &var_name1._Mean &var_name1._StdDev i);
    set &var_name1._means1;
    by table_name;
    array temp1(2) no_mean6 yes_mean6 ;
    array temp2(2) no_std6 yes_std6 ;
    retain yes_mean6 no_mean6 yes_std6 no_std6;
    if first.table_name then do i = 1 to 2;
      temp1(i) = .;
      temp2(i) = .;
    end;
```

```
      temp1(_n_) = round(&var_name1._Mean,0.01);
      temp2(_n_) = round(&var_name1._StdDev,0.01);
    if last.table_name;

  proc print data = &var_name1._means_1;



proc sort data = pub30_lsmith_niddk;
  by prevent_action_yes_15mo;
       *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc means data = &dataset_name mean Std ;
       var &var_name2.;
       by prevent_action_yes_15mo;
       title3 "Frequency table of the &var_name2. variable in the analysis dataset";

       *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output Summary = work.&var_name2._means2;
  run;


  data &var_name2._means2;
    set &var_name2._means2;
    length table_name $30.;
    table_name  ="&var_name2";
    if prevent_action_yes_15mo ne .;
     merge_var = substr(table_name,1,index(table_name,'_')-1);

  proc print data = &var_name2._means2;

   proc sort data = &var_name2._means2;
     by table_name prevent_action_yes_15mo;

  data &var_name2._means_2(drop = prevent_action_yes_15mo &var_name2._Mean &var_name2._StdDev i);
    set &var_name2._means2;
    by table_name;
    array temp1(2) no_mean15 yes_mean15 ;
    array temp2(2) no_std15 yes_std15 ;
    retain yes_mean15 no_mean15 yes_std15 no_std15;
    if first.table_name then do i = 1 to 2;
      temp1(i) = .;
      temp2(i) = .;
    end;
      temp1(_n_) = round(&var_name2._Mean,0.01);
      temp2(_n_) = round(&var_name2._StdDev,0.01);
    if last.table_name;

  proc print data = &var_name2._means_2;
```

```
  data &var_name1._mean12;
    merge &var_name1._means_1 &var_name2._means_2;
    by merge_var;

  proc print data = &var_name1._mean12;

%mend;


*******************************************************************************;
*************** Check Table 1 *************************************************;
*******************************************************************************;

*** Running the baseline_freq on the  categorical variables in the Table 1 manuscript file;
%baseline_freq1(pub30_lsmith_niddk,prevent_action_yes_6mo,prevent_action_yes_15mo        );
%baseline_freq1(pub30_lsmith_niddk,anything_prevent_six , anything_prevent_fifteen );
%baseline_freq1(pub30_lsmith_niddk,breast_feed_6mo      , breast_feed_15mo  );
%baseline_freq1(pub30_lsmith_niddk,diet_6mo             , diet_15mo  );
%baseline_freq1(pub30_lsmith_niddk,diet_sup_6mo         , diet_sup_15mo  );
%baseline_freq1(pub30_lsmith_niddk,ill_prevent_6mo      , ill_prevent_15mo );
%baseline_freq1(pub30_lsmith_niddk,med_vitamins_6mo     , med_vitamins_15mo  );
%baseline_freq1(pub30_lsmith_niddk,physical_activity_6mo, physical_activity_15mo  );
%baseline_freq1(pub30_lsmith_niddk,stress_reduction_6mo , stress_reduction_15mo );
%baseline_freq1(pub30_lsmith_niddk,other_6mo            , other_15mo );

data anything_prevent_six_12;
  set anything_prevent_six_12;
  table_name  = compress(table_name || anything_prevent_six);

data table1_compare (keep = table_name count6 count15 pert6 pert15 );
  set
anything_prevent_six_12 (where = (compress(anything_prevent_six) in ('Yes', 'No')) )
breast_feed_6mo_12      (where = (breast_feed_6mo in (1)) )
diet_6mo_12             (where = (diet_6mo in (1))  )
diet_sup_6mo_12         (where = (diet_sup_6mo in ( 1)))
ill_prevent_6mo_12      (where = (ill_prevent_6mo in ( 1)))
med_vitamins_6mo_12     (where = (med_vitamins_6mo in (1)))
physical_activity_6mo_12 (where = (physical_activity_6mo in ( 1)))
stress_reduction_6mo_12 (where = (stress_reduction_6mo in (1)) )
other_6mo_12            (where = (other_6mo in (1)) )
;

proc print data = table1_compare;


*******************************************************************************;
*************** Check Table 2 *************************************************;
*******************************************************************************;
```

```
*** Running the baseline_freq on the  categorical variables in the Table 2 manuscript file;
%baseline_freq2(pub30_lsmith_niddk,married_living        );
%baseline_freq2(pub30_lsmith_niddk,education_mom_group3        );
%baseline_freq2(pub30_lsmith_niddk,female                    );
%baseline_freq2(pub30_lsmith_niddk,ethnic_minority        );
%baseline_freq2(pub30_lsmith_niddk,single_child        );
%baseline_freq2(pub30_lsmith_niddk,country        );
%baseline_freq2(pub30_lsmith_niddk,fdr        );
%baseline_means(pub30_lsmith_niddk, maternal_age      );
%baseline_means(pub30_lsmith_niddk, crowding_norm      );

data table2_freqcompare ;
  set
married_living_12
education_mom_group3_12
female_12
ethnic_minority_12
single_child_12
country_12
fdr_12
;


data table2_meancompare ;
  set  maternal_age_mean12 crowding_norm_mean12
;

proc print data = table2_freqcompare;
proc print data = table2_meancompare;



*******************************************************************************;
*************** Check Table 3 *************************************************;
*******************************************************************************;

*** Running the baseline_freq on the  categorical variables in the Table 3 manuscript file;
%baseline_means2(pub30_lsmith_niddk, stai_six,stai_fifteen      );
%baseline_means2(pub30_lsmith_niddk, stai_six_norm,stai_fifteen_norm      );
%baseline_means2(pub30_lsmith_niddk, reduce_6mo_agree,reduce_15mo_agree      );

%baseline_means(pub30_lsmith_niddk, totalblues      );
%baseline_means(pub30_lsmith_niddk, totalblues_norm      );
%baseline_means(pub30_lsmith_niddk, depressiontotal      );
%baseline_means(pub30_lsmith_niddk, depress_norm2      );

data table3_meancompare(drop= merge_var) ;
  set
```

```
stai_six_mean12
stai_six_norm_mean12
totalblues_mean12
totalblues_norm_mean12
depressiontotal_mean12(drop =no_mean6 no_std6 yes_mean6 yes_std6)
depress_norm2_mean12(drop =no_mean6 no_std6 yes_mean6 yes_std6)
reduce_6mo_agree_mean12
;

%baseline_freq2(pub30_lsmith_niddk,blues2        );
%baseline_freq3(pub30_lsmith_niddk,worry_six, prevent_action_yes_6mo       );
%baseline_freq3(pub30_lsmith_niddk,worry_fifteen, prevent_action_yes_15mo       );

data worry_c;
  merge WORRY_SIX_12 (rename = (Frequency = count6 RowPercent = pert6))
        worry_fifteen_12(rename = (Frequency = count15 RowPercent = pert15))  ;
  by merge_var;
proc print data = worry_c;

%baseline_freq3(pub30_lsmith_niddk,anydev_six, prevent_action_yes_6mo       );
%baseline_freq3(pub30_lsmith_niddk,anydev_fifteen, prevent_action_yes_15mo       );

data anydev_c;
  merge anydev_SIX_12 (rename = (Frequency = count6 RowPercent = pert6))
        anydev_fifteen_12(rename = (Frequency = count15 RowPercent = pert15))  ;
  by merge_var;
proc print data = anydev_c;

data table3_freqcompare(drop= merge_var) ;
  set
worry_c
blues2_12
anydev_c
;



proc print data = table1_compare;
  title3 'table 1 DSIC results';

proc print data = table2_freqcompare;
  title3 'table 2 DSIC freq results';
proc print data = table2_meancompare;
  title3 'table 2 DSIC mean results';


proc print data = table3_freqcompare;
  title3 'table 3 DSIC freq results';
proc print data = table3_meancompare;
```

```
title3 'table 3 DSIC mean results';
```