

Dataset Integrity Check for the TEDDY Pub21 BLernmark Data Files

Prepared by Jane Wang

IMS Inc.

3901 Calverton Blvd, Suite 200 Calverton MD 20705

July 11, 2015

Table of Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	3
6 Conclusions	3
7 References	3
Attachment A: SAS Code	12
Table A: Variables used to replicate Table 1	4
Table B: Variables used to replicate Table 2.....	4
Table C: Comparison of values computed in integrity check to reference article Table 1 values: <u>Most common reasons for staying in TEDDY (Reason as “Very Important by country”</u>	5
Table D: Comparison of values computed in integrity check to reference article Table 2 values: <u>Respondent satisfaction with different components of TEDDY study by country</u>	7

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 “pub21_blermark_niddk” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Barbro Lernmark et al [1], published in J Clin Trials in 2012. To verify the integrity of the dataset, descriptive statistics of baseline characteristics were computed, by different country (Table B, Table C).

5 Results

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

6 Conclusions

The NIDDK repository is confident that the TEDDY Pub21 BLernmark data files to be distributed are a true copy of the study data.

7 References

1. Barbro Lernmark, Kristian Lynch, Lori Ballard, Judith Baxter, Roswith Roth, Tuula Simell and Suzanne Bennett Johnson, for the TEDDY Study Group. Reasons for Staying as a Participant in the Environmental Determinants of Diabetes in the Young (TEDDY) Longitudinal Study. Lernmark et al., J Clin Trials 2012, 2:2

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

Table Variable	Variables Used in Replication from the Table 1 Dataset
Knowing someone is watching my child for development of diabetes	watchingchildfordevelopofdiabete
Helping science discover the causes of type 1 diabetes	helpsciencediscovertype1diab
Getting my child's antibody results	gettingchildantibodyresults
Being seen by the same TEDDY staff at each visit	seenbysameteddstaff
Knowing my child might be able to participate in future prevention studies	childparticipatefuturestudies
Total number of questionnaires	parent_answered

Table B: Variables used to replicate Table 2 in the publication.

Table Variable	Variables Used in Replication from the Table 2 Dataset
Working with the TEDDY staff	WorkingWithTheTEDDYStaff
Getting questions answered	GettingMyQuestionsAnswered
Wait before visit starts	WaitBeforeTheTEDDYVisitStarts
Day or time visit is scheduled	DayTimeVisitScheduled
Time to complete a visit	TimeToCompleteVisit
Reminders for the visits	RemindersForTheTEDDYVisits
Transportation to the visit	TransportToTEDDYVisit
Parking for a TEDDY visit	ParkingForATEDDYVisit

Table C: Comparison of values computed in integrity check to reference article Table 1 values: Most common reasons for staying in TEDDY
 (Reason as “Very Important by country”)

Reasons for staying in TEDDY	FINLAND [Manuscript]	FINLAND [DSIC]	FINLAND [Difference]	GERMANY [Manuscript]	GERMANY [DSIC]	GERMANY [Difference]	SWEDEN [Manuscript]	SWEDEN [DSIC]	SWEDEN [Difference]
Knowing someone is watching my child for development of diabetes (N)	330	330	0	93	93	0	667	667	0
Knowing someone is watching my child for development of diabetes (%)	61.5	61.7	-0.2	73.2	73.228	0	81.1	81.4	-0.3
Helping science discover the causes of type 1 diabetes (N)	299	299	0	75	75	0	593	593	0
Helping science discover the causes of type 1 diabetes (%)	55.7	56.1	-0.4	59.1	59.524	-0.4	72.1	72.1	0
Getting my child’s antibody results (N)	404	404	0	78	78	0	443	443	0
Getting my child’s antibody results (%)	75.2	75.5	-0.3	61.4	61.905	-0.5	53.9	54.0	-0.1
Being seen by the same TEDDY staff at each visit (N)	231	231	0	26	26	0	503	503	0
Being seen by the same TEDDY staff at each visit (%)	43	43.3	-0.3	20.5	20.968	-0.5	61.2	61.6	-0.4
Knowing my child might be able to participate in future prevention studies (N)	139	139	0	49	49	0	461	461	0
Knowing my child might be able to participate in future prevention studies (%)	25.9	26.1	-0.2	38.6	39.2	-0.6	56.1	56.2	-0.1
Total number of questionnaires (N)	537	537	0	127	127	0	822	822	0

Reasons for staying in TEDDY	US [Manuscript]	US [DSIC]	US [Difference]	All [Manuscript]	All [DSIC]	All [Difference]
Knowing someone is watching my child for development of diabetes (N)	400	400	0	1490	1490	0
Knowing someone is watching my child for development of diabetes (%)	77.8	78.0	-0.2	74.5	74.7	-0.2
Helping science discover the causes of type 1 diabetes (N)	395	395	0	1362	1362	0
Helping science discover the causes of type 1 diabetes (%)	76.8	77.3	-0.5	68.1	68.4	-0.3
Getting my child's antibody results (N)	340	340	0	1265	1265	0
Getting my child's antibody results (%)	66.1	66.5	-0.4	63.3	63.5	-0.2
Being seen by the same TEDDY staff at each visit (N)	136	136	0	896	896	0
Being seen by the same TEDDY staff at each visit (%)	26.5	26.5	0	44.8	45.1	-0.3
Knowing my child might be able to participate in future prevention studies (N)	190	190	0	839	839	0
Knowing my child might be able to participate in future prevention studies (%)	37	37.3	-0.3	42	42.2	-0.2
Total number of questionnaires (N)	514	514	0	2000	2000	0

Table D: Comparison of values computed in integrity check to reference article Table 2 values: Respondent satisfaction with different components of TEDDY study by country

Reasons for satisfaction with different components of the TEDDY study by country	All [Manuscript]	All [DSIC]	All [Difference]	FINLAND [Manuscript]	FINLAND [DSIC]	FINLAND [DSIC]	SWEDEN [Manuscript]	SWEDEN [DSIC]	SWEDEN [Difference]
Working with the TEDDY staff									
Works Great/ Not a problem (N)	1818	1818	0	514	514	0	806	806	0
Works Great/ Not a problem (%)	97.5	97.5	0	96.1	96.1	0	98.7	98.7	0
Works good most of the time (N)	43	43	0	20	20	0	9	9	0
Works good most of the time (%)	2.3	2.3	0	3.7	3.7	0	1.1	1.1	0
Needs Improvement (N)	3	3	0	1	1	0	2	2	0
Needs Improvement (%)	0.2	0.2	0	0.2	0.2	0	0.2	0.2	0
Getting questions answered									
Works Great/ Not a problem (N)	1757	1757	0	495	495	0	764	764	0
Works Great/ Not a problem (%)	94.4	94.4	0	92.9	92.9	0	93.5	93.5	0
Works good most of the time (N)	99	99	0	35	35	0	52	52	0
Works good most of the time (%)	5.3	5.3	0	6.6	6.6	0	6.4	6.4	0
Needs Improvement (N)	5	5	0	3	3	0	1	1	0
Needs Improvement (%)	0.3	0.3	0	0.6	0.6	0	0.1	0.1	0
Wait before visit starts									
Works Great/ Not a problem (N)	1737	1737	0	487	487	0	794	794	0
Works Great/ Not a problem (%)	93.5	93.5	0	91.4	91.4	0	97.2	97.2	0
Works good most of the time (N)	113	113	0	45	45	0	21	21	0
Works good most of the time (%)	6.1	6.1	0	8.4	8.4	0	2.6	2.6	0
Needs Improvement (N)	8	8	0	1	1	0	2	2	0
Needs Improvement (%)	0.4	0.4	0	0.2	0.2	0	0.2	0.2	0
Day or time visit is scheduled									
Works Great/ Not a problem (N)	1663	1663	0	430	430	0	752	752	0

Works Great/ Not a problem (%)	89.3	89.3	0	80.4	80.4	0	91.9	91.9	0
Works good most of the time (N)	192	192	0	102	102	0	64	64	0
Works good most of the time (%)	10.3	10.3	0	19.1	19.1	0	7.8	7.8	0
Needs Improvement (N)	7	7	0	3	3	0	2	2	0
Needs Improvement (%)	0.4	0.4	0	0.6	0.6	0	0.2	0.2	0
Time to complete a visit									
Works Great/ Not a problem (N)	1584	1584	0	447	447	0	722	722	0
Works Great/ Not a problem (%)	85.4	85.4	0	84	84.0	0	88.4	88.4	0
Works good most of the time (N)	252	252	0	81	81	0	94	94	0
Works good most of the time (%)	13.6	13.6	0	15.2	15.2	0	11.5	11.5	0
Needs Improvement (N)	18	18	0	4	4	0	1	1	0
Needs Improvement (%)	1	1.0	0	0.8	0.8	0	0.1	0.1	0
Reminders for the visits									
Works Great/ Not a problem (N)	1386	1386	0	413	413	0	524	524	0
Works Great/ Not a problem (%)	76.2	76.2	0	83.6	83.6	0	64.4	64.4	0
Works good most of the time (N)	336	336	0	45	45	0	245	245	0
Works good most of the time (%)	18.5	18.5	0	9.1	9.1	0	30.1	30.1	0
Needs Improvement (N)	97	97	0	36	36	0	45	45	0
Needs Improvement (%)	5.3	5.3	0	7.3	7.3	0	5.5	5.5	0
Transportation to the visit									
Works Great/ Not a problem (N)	1384	1384	0	379	379	0	616	616	0
Works Great/ Not a problem (%)	74.7	74.7	0	70.8	70.8	0	75.5	75.5	0
Works good most of the time (N)	431	431	0	139	139	0	193	193	0
Works good most of the time (%)	23.3	23.3	0	26	26.0	0	23.7	23.7	0
Needs Improvement (N)	37	37	0	17	17	0	7	7	0
Needs Improvement (%)	2	2	0	3.2	3.2	0	0.9	0.9	0
Parking for a TEDDY visit									
Works Great/ Not a problem (N)	994	994	0	168	168	0	417	417	0
Works Great/ Not a problem (%)	54.3	54.3	0	31.5	31.5	0	52.6	52.6	0

Works good most of the time (N)	593	593	0	238	238	0	280	280	0
Works good most of the time (%)	32.4	32.4	0	44.7	44.7	0	35.3	35.3	0
Needs Improvement (N)	245	245	0	127	127	0	96	96	0
Needs Improvement (%)	13.4	13.4	0	23.8	23.8	0	12.1	12.1	0
Reasons for satisfaction with different components of the TEDDY study by country	US [Manuscript]	US [DSIC]	US [Difference]						
Working with the TEDDY staff									
Works Great/ Not a problem (N)	498	498	0						
Works Great/ Not a problem (%)	97.2	97.3	-0.1						
Works good most of the time (N)	14	14	0						
Works good most of the time (%)	2.7	2.7	0						
Needs Improvement (N)	0	0	0						
Needs Improvement (%)	0	0	0						
Getting questions answered									
Works Great/ Not a problem (N)	498	498	0						
Works Great/ Not a problem (%)	97.5	97.5	0						
Works good most of the time (N)	12	12	0						
Works good most of the time (%)	2.3	2.3	0						
Needs Improvement (N)	1	1	0						
Needs Improvement (%)	0.2	0.2	0						
Wait before visit starts									
Works Great/ Not a problem (N)	456	456	0						
Works Great/ Not a problem (%)	89.8	89.8	0						
Works good most of the time (N)	47	47	0						
Works good most of the time (%)	9.3	9.3	0						
Needs Improvement (N)	5	5	0						
Needs Improvement (%)	1	1	0						
Day or time visit is scheduled									

Works Great/ Not a problem (N)	481	481	0
Works Great/ Not a problem (%)	94.5	94.5	0
Works good most of the time (N)	26	26	0
Works good most of the time (%)	5.1	5.1	0
Needs Improvement (N)	2	2	0
Needs Improvement (%)	0.4	0.4	0
Time to complete a visit			
Works Great/ Not a problem (N)	415	415	0
Works Great/ Not a problem (%)	82.2	82.2	0
Works good most of the time (N)	77	77	0
Works good most of the time (%)	15.2	15.2	0
Needs Improvement (N)	13	13	0
Needs Improvement (%)	2.6	2.6	0
Reminders for the visits			
Works Great/ Not a problem (N)	449	449	0
Works Great/ Not a problem (%)	87.9	87.9	0
Works good most of the time (N)	46	46	0
Works good most of the time (%)	9	9.0	0
Needs Improvement (N)	16	16	0
Needs Improvement (%)	3.1	3.1	0
Transportation to the visit			
Works Great/ Not a problem (N)	389	389	0
Works Great/ Not a problem (%)	77.6	77.6	0
Works good most of the time (N)	99	99	0
Works good most of the time (%)	19.8	19.8	0
Needs Improvement (N)	13	13	0
Needs Improvement (%)	2.6	2.6	0
Parking for a TEDDY visit			
Works Great/ Not a problem (N)	409	409	0

Works Great/ Not a problem (%)	80.8	80.8	0
Works good most of the time (N)	75	75	0
Works good most of the time (%)	14.8	14.8	0
Needs Improvement (N)	22	22	0
Needs Improvement (%)	4.3	4.3	0

Attachment A: SAS Code

```
*****  
  
***Program:  
  
***Programmer: Jane Wang  
  
***Date Created: 06/10/2015  
  
***Purpose:  
  
*****;  
  
title1 "%sysfunc(getoption(sysin))";  
  
title2 " ";  
  
options nofmterr;  
  
libname sas_data "/prj/niddk/ims_analysis/TEDDY/private_orig_data/Pub21_BLernmark_niddk_submission/";  
  
*** Reading in the analysis datasets used for the DSIC;  
  
data Pub21_BLernmark          ; set sas_data.pub21_blernmark_niddk          ;  
  
data Pub21_BLernmark;  
  
    set Pub21_BLernmark;  
  
    rename watchingchildfordevelofdiabete = watch_d  
           helpsciencediscovertypeldiab   = help_d  
           gettingchildantibodyresults    = get_result
```

```
seenbysameteddystaff      = see_staff
childparticipatefuturestudies = child_result
parent_answered           = parent_answer
WorkingWithTheTEDDYStaff   = workwithstaff
GettingMyQuestionsAnswered = getqanswer
WaitBeforeTheTEDDYVisitStarts = waitbefore
DayTimeVisitScheduled      = dayschedule
TimeToCompleteVisit        = timecomplete
RemindersForTheTEDDYVisits = reminder
TransportToTEDDYVisit      = transport
ParkingForATEDDYVisit      = parking
;
```

```
*** Data from the Primary outcome 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_pub21_table1.csv';
```

```
FILENAME table2 '/prj/niddk/ims_analysis/TEDDY/private_created_data/teddy_pub21_table2.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_pub21_table1_dsic.csv';
```

```
FILENAME out_t2 '/prj/niddk/ims_analysis/TEDDY/private_created_data/teddy_pub21_table2_dsic.csv';
```

```

%macro baseline_freq1(dataset_name,var_name);

    *** Creating a frequency table in the format of Table 1 in the primary outcome paper;

proc freq data = &dataset_name ;

    table (&var_name.)*country ;

    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._cross;

data &var_name._cross(keep = &var_name country Frequency Percent_keep );

    set &var_name._cross;

    if &var_name = 'Very Important';

    if country = . then Percent_keep = Percent;

    else Percent_keep = colpercent;

proc sort data = &var_name._cross;

    by &var_name;

proc transpose data=&var_name._cross out=&var_name._transpose ;

    by &var_name;

```

```

data &var_name._transpose (drop = _LABEL_ &var_name) ;

  set &var_name._transpose ;

  if _name_ ne 'country';

  length table_name $30.;

  table_name  ="&var_name";

  rename col1 = US

         col2 = FINLAND

         col3 = GERMANY

         col4 = SWEDEN

         col5 = ALL

         _NAME_ = count_pert;

%mend;

%macro baseline_freq2(dataset_name,var_name);

  *** Creating a frequency table in the format of Table 1 in the primary outcome paper;

  proc freq data = &dataset_name ;

    table (&var_name.)*country ;

    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._cross;

data &var_name._cross(keep = &var_name country Frequency );

  set &var_name._cross;

  if &var_name = 1;

proc sort data = &var_name._cross;

  by &var_name;

proc transpose data=&var_name._cross out=&var_name._transpose ;

  by &var_name;

data &var_name._transpose (drop = _LABEL_ &var_name) ;

  set &var_name._transpose ;

  if _name_ ne 'country';

  length table_name $30.;

  table_name = "&var_name";

rename coll = US

         col2 = FINLAND

         col3 = GERMANY

         col4 = SWEDEN

         col5 = ALL

```

```

        _NAME_ = count_pert;
%mend;

%macro baseline_freq3(dataset_name,var_name);

    *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
proc freq data = &dataset_name ;
    table (&var_name.)*country ;
    title3 "Frequency table of the &var_name. variable in the analysis dataset";
    where country ne 3;

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

data &var_name._cross(keep = &var_name country Frequency Percent_keep );
    set &var_name._cross;
    if &var_name ne '';
    if country = . then Percent_keep = Percent;
    else Percent_keep = colpercent;

proc sort data = &var_name._cross;
    by &var_name;

```

```

proc transpose data=&var_name._cross out=&var_name._transpose ;

  by  &var_name;

data &var_name._transpose (drop = _LABEL_ &var_name ) ;

  set &var_name._transpose ;

  if _name_ ne 'country';

  length table_name levels $30.;

  table_name  ="&var_name";

  if &var_name = 'Works Good Most of the Time-Sometimes a problem' then &var_name = 'Works good most of the time';

  levels =&var_name;

  rename coll = US

         col2 = FINLAND

         col3 = SWEDEN

         col4 = ALL

         _NAME_ = count_pert;

%mend;

*****;

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

*****;

```

```
*** Running the baseline_freq on the categorical variables in the Table 1 manuscript file;
```

```
%baseline_freq1(Pub21_BLernmark,watch_d      );  
%baseline_freq1(Pub21_BLernmark,help_d       );  
%baseline_freq1(Pub21_BLernmark,get_result   );  
%baseline_freq1(Pub21_BLernmark,see_staff    );  
%baseline_freq1(Pub21_BLernmark,child_result );  
%baseline_freq2(Pub21_BLernmark,parent_answer );
```

```
data table1_compare;
```

```
  set  watch_d_transpose  
      help_d_transpose  
      get_result_transpose  
      see_staff_transpose  
      child_result_transpose  
      parent_answer_transpose  
;  
;
```

```
proc print data = table1_compare;
```

```
  title3 'table 1 from data set';
```

```

*** Importing the Table 1 Data taken from the primary outcome paper;

data table1_data;

  infile table1 delimiter = ',' MISSOVER DSD firstobs=2 ls=1080;

  length character $100. count_pert $12. table_name $ 30.;

  input character      $ count_pert $ table_name $ FINLAND_ GERMANY_          SWEDEN_US_          ALL_ e $
;

data table1_data;

  set table1_data;

  ordernum = _n_;

proc print data = table1_data;

  title3 'table 1 from paper';

proc sort data = table1_data(drop = e);

  by table_name count_pert;

proc sort data = table1_compare;

  by table_name count_pert;

data table1_combine;

  merge table1_data (in = in2) table1_compare (in = in1);

  by table_name count_pert;

```

```

if in1 and in2;

data table1_combine;

set table1_combine;

diff_all      = round((round(all_,0.1)      - round(all      ,0.1)),0.1);

diff_FINLAND  = round((round(FINLAND_,0.1) - round(FINLAND  ,0.1)),0.1);

diff_GERMANY  = round((round(GERMANY_,0.1) - round(GERMANY  ,0.1)),0.1);

diff_SWEDEN   = round((round(SWEDEN_,0.1) - round(SWEDEN   ,0.1)),0.1);

diff_US       = round((round(US_,0.1)       - round(US       ,0.1)),0.1);

label

character      = "Reasons for staying in TEDDY"

FINLAND_       = "FINLAND [Manuscript]"

FINLAND        = "FINLAND [DSIC]      "

diff_FINLAND   = "FINLAND [Difference]"

GERMANY_       = "GERMANY [Manuscript]"

GERMANY        = "GERMANY [DSIC]      "

diff_GERMANY   = "GERMANY [Difference]"

SWEDEN_       = "SWEDEN [Manuscript]"

SWEDEN        = "SWEDEN [DSIC]      "

diff_SWEDEN   = "SWEDEN [Difference]"

US_           = "US [Manuscript]"

US            = "US [DSIC]      "

```

```
diff_US      = "US [Difference]"
all_         = "all [Manuscript]"
all          = "all [DSIC]      "
diff_all     = "all [Difference]"
;
```

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

```
*** Running the baseline_freq on the categorical variables in the Table 2 manuscript file;
```

```
%baseline_freq3(Pub21_BLernmark,workwithstaff      );
%baseline_freq3(Pub21_BLernmark,getqanswer         );
%baseline_freq3(Pub21_BLernmark,waitbefore         );
%baseline_freq3(Pub21_BLernmark,dayschedule        );
%baseline_freq3(Pub21_BLernmark,timecomplete       );
%baseline_freq3(Pub21_BLernmark,reminder           );
%baseline_freq3(Pub21_BLernmark,transport          );
%baseline_freq3(Pub21_BLernmark,parking            );
```

```

data table2_compare;

  set  workwithstaff_transpose

      getqanswer_transpose

      waitbefore_transpose

      dayschedule_transpose

      timecomplete_transpose

      reminder_transpose

      transport_transpose

      parking_transpose

  ;

proc print data = table2_compare;

  title3 'table 2 from data set';

  *** Importing the Table 2 Data taken from the primary outcome paper;

data table2_data;

  infile table2 delimiter = ',' MISSOVER DSD firstobs=2 ls=1080;

  length character $30. levels $ 30. table_name $ 30. count_pert $12.;

  input character      $ levels $ table_name $ count_pert $ ALL_  FINLAND_          SWEDEN_          US_          e $

;

data table2_data;

  set table2_data;

  ordernum = _n_;

```

```

proc print data = table2_data;

proc sort data = table2_data(drop = e);

  by table_name levels count_pert;

proc sort data = table2_compare;

  by table_name levels count_pert;

data table2_combine;

  merge table2_data (in = in2) table2_compare (in = in1);

  by table_name levels count_pert;

  if in1 and in2 then output table2_combine;

data table2_combine;

  set table2_combine;

  diff_all      = round((round(all_,0.1)      - round(all      ,0.1)),0.1);

  diff_FINLAND  = round((round(FINLAND_,0.1) - round(FINLAND      ,0.1)),0.1);

  diff_SWEDEN   = round((round(SWEDEN_,0.1)   - round(SWEDEN      ,0.1)),0.1);

  diff_US       = round((round(US_,0.1)       - round(US      ,0.1)),0.1);

  label

  character      = "Reasons satisfaction with different components of the TEDDY study by country"

  levels         = "levels"

  FINLAND_       = "FINLAND [Manuscript]"

  FINLAND        = "FINLAND [DSIC]      "

```

```

diff_FINLAND      = "FINLAND [Difference]"
SWEDEN_          = "SWEDEN [Manuscript]"
SWEDEN           = "SWEDEN [DSIC]      "
diff_SWEDEN      = "SWEDEN [Difference]"
US_              = "US [Manuscript]"
US               = "US [DSIC]        "
diff_US          = "US [Difference]"
all_             = "all [Manuscript]"
all              = "all [DSIC]       "
diff_all         = "all [Difference]"
;
proc sort data = table1_combine;
  by ordernum;

*** Outputting the data to a csv format to be added to the DSIC;
ods csv file = out_t1;

run;

proc print data = table1_combine NOOBS label;
  var
character

```

```
FINLAND_  
FINLAND  
diff_FINLAND  
GERMANY_  
GERMANY  
diff_GERMANY  
SWEDEN_  
SWEDEN  
diff_SWEDEN  
US_  
US  
diff_US  
all_  
all  
diff_all  
  
;  
    title "DSIC Check of Table 1 Most common reasons for strying in TEDDY";  
run;  
  
proc sort data = table2_combine;  
    by ordernum;
```

```
*** Outputting the data to a csv format to be added to the DSIC;
```

```
ods csv file = out_t2;
```

```
run;
```

```
proc print data = table2_combine NOOBS label;
```

```
var
```

```
character
```

```
levels
```

```
all_
```

```
all
```

```
diff_all
```

```
FINLAND_
```

```
FINLAND
```

```
diff_FINLAND
```

```
SWEDEN_
```

```
SWEDEN
```

```
diff_SWEDEN
```

```
US_
```

```
US
```

```
diff_US
```

```
;
title "DSIC Check of Table 2 Reasons satisfaction with differnet components of the TEDDY study by country";
run;
```