

Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub31 Johnson

Prepared by Allyson Mateja

IMS Inc.

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

April 24, 2017

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
Table A: Variables used to replicate Table 1: Predictors of TEDDY Study Withdrawal in Years 2-3 by Block: Demographic, Maternal Lifestyle Behaviors, Stress and Child Illness, Maternal Reactions to Child’s Type 1 Diabetes Risk, and In-Study Behaviors.....	4
Table B: Comparison of values computed in integrity check to reference article Table 1 values.....	5
Attachment A: SAS Code	8

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 the “m_31_sjohnson_niddk_30jun2011.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Johnson et al [1] in The Journal of Pediatric Psychology in 2015. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Predictors of TEDDY Study Withdrawal in Years 2-3 by Block: Demographic, Maternal Lifestyle Behaviors, Stress and Child Illness, Maternal Reactions to Child's Type 1 Diabetes Risk, and In-Study Behaviors, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are almost an exact match to the published results, with some errors due to rounding.

6 Conclusions

The NIDDK repository is confident that the TEDDY M31 data files to be distributed are a true copy of the study data.

7 References

[1] Johnson, S.B., Lynch, K.F., Baxter, J., Lernmark, B., Roth, R., Simell, T., Smith, L., and the TEDDY study group. "Predicting Later Study Withdrawal in Participants Active in a Longitudinal Birth Cohort Study for 1 Year: The TEDDY Study". *The Journal of Pediatric Psychology* (2015) 1-11.

Table A: Variables used to replicate Table 1: Predictors of TEDDY Study Withdrawal in Years 2-3 by Block: Demographic, Maternal Lifestyle Behaviors, Stress and Child Illness, Maternal Reactions to Child’s Type 1 Diabetes Risk, and In-Study Behaviors

Table Variable	Dataset Variable
Country of residence	country
Gender of child	female
Child ethnic/minority status	ethnic_minority
Only child	single_child
Maternal age at child's birth	maternal_age
Parents married or living together	childparentsstatus
Mother's education	education_mom_group3
Crowding (persons/room)	crowdingspace
Working during pregnancy	worked_preg
Alcohol consumption during pregnancy	alc_3rd_trimester
Smoking (pregnancy or during first year of TEDDY)	smoke_pregnancy, mom_smokes
Working outside the home in the first year of TEDDY	mom_works
Serious negative life event in the first year of TEDDY	serious_neg_life_event_count
Number of child illnesses in the first year of TEDDY	count_illness
Risk perception accuracy (15 mo)	anydev_fifteen, anydev
State Anxiety Inventory Score (15 mo)	stai_fifteen, stai_six, stai
Edinburgh Postnatal Depression Score (6 mo)	totalblues
Bradley Depression subscale (15 mo)	depressiontotal
TEDDY recruitment cohort	birth_cohort
Father's participation in first year of TEDDY	dad_participation
Study compliance in first year of TEDDY	missed_visit, missed_blood_draw
Maternal study satisfaction (15 mo)	satis_score_15mo_mom, satis_score_6mo_mom
Withdrawal	late_dropout3yr

Table B: Comparison of values computed in integrity check to reference article Table 1 values

Blocks	Factors	Total N or mean (SD) Manuscript	Total N or mean (SD) DSIC	Diff.	Withdrawals % Manuscript	Withdrawals % DSIC	Diff.
Demographic	Country of residence						
	Finland	845	845	0	11.7	11.7	0
	Sweden	1,230	1,230	0	12.1	12.2	0.1
	Germany	112	112	0	18.8	18.8	0
	United States	1,141	1,141	0	11.7	11.7	0
	Gender of child						
	Boy	1,716	1,716	0	11.9	12.0	0.1
	Girl	1,612	1,612	0	12.3	12.3	0
	Child ethnic/minority status						
	No	2,880	2,880	0	11.6	11.6	0
	Yes	448	448	0	15.4	15.4	0
	Only child						
	No	1,871	1,871	0	11.2	11.2	0
	Yes	1,457	1,457	0	13.2	13.3	0.1
	Maternal age at child's birth	30.7 (5.1)	30.7 (5.1)	0 (0)			
	Parents married or living together						
	No	140	140	0	20.0	20.0	0
	Yes	3,188	3,188	0	11.8	11.8	0
	Mother's education						0
	Primary education	661	661	0	15.7	15.7	0
	Trade school/some college	790	790	0	14.8	14.8	0
	Graduated college	1,877	1,877	0	9.7	9.7	0
	Crowding (persons/room)	2.02 (1.18)	2.02 (1.18)	0 (0)			
Maternal Lifestyle Behaviors	Working during pregnancy						
	Reduced work/did not work at all	1,657	1,657	0	14.3	14.3	0
	Worked all trimesters	1,692	1,692	0	9.9	9.9	0

Blocks	Factors	Total N or mean (SD) Manuscript	Total N or mean (SD) DSIC	Diff.	Withdrawals % Manuscript	Withdrawals % DSIC	Diff.
	Alcohol consumption during pregnancy						
	Total abstinence in third trimester	2,721	1,721	1,000	12.4	12.4	0
	Occasional drink in third trimester	628	628	0	10.7	10.7	0
	Smoking (pregnancy or during first year of TEDDY)						
	No	2,873	1,873	1,000	11.2	11.2	0
	Yes	476	476	0	17.2	17.2	0
	Working outside the home in the first year of TEDDY						
	No	2,218	2,218	0	12.1	12.1	0
	Yes	1,131	1,131	0	12.1	12.1	0
Stress and Child Illness	Serious negative life event in the first year of TEDDY						
	No	1,805	1,805	0	12.4	12.4	0
	Yes	1,651	1,651	0	12.5	12.5	0
	Number of child illnesses in the first year of TEDDY	2.1 (1.2)	2.1 (1.2)	0 (0)			
Maternal Reactions to Child's Type 1 Diabetes Risk	Risk perception accuracy (15 mo)						
	Underestimate	1,279	1,279	0	14.5	14.5	0
	Accurate	1,851	1,851	0	8.5	8.5	0
	State Anxiety Inventory Score (15 mo)	33.6 (9.4)	33.6 (9.4)	0 (0)			
	Edinburgh Postnatal Depression Score (6 mo)	6.2 (4.3)	6.2 (4.3)	0 (0)			

Blocks	Factors	Total N or mean (SD) Manuscript	Total N or mean (SD) DSIC	Diff.	Withdrawals % Manuscript	Withdrawals % DSIC	Diff.
	Bradley Depression subscale (15 mo)	2.24 (1.13)	2.24 (1.13)	0 (0)			
In-Study Behaviors	TEDDY recruitment cohort						
	Year 1	699	699	0	13.6	13.6	0
	Year 2	1,018	1,018	0	13.3	13.3	0
	Year 3	1,189	1,189	0	12.5	12.5	0
	Year 4	516	516	0	8.3	8.3	0
	Father's participation in first year of TEDDY						
	Completed all 3 surveys	2,647	2,647	0	10.2	10.2	0
	Completed 0, 1, or 2 surveys	775	775	0	19.6	19.6	0
	Study compliance in first year of TEDDY						
	No missed study visits or blood draws	2,706	2,706	0	9.4	9.4	0
	Missed ≥ 1 blood draws but not a study visit	545	545	0	19.4	19.5	0.1
	Missed ≥ 1 study visits	171	171	0	35.7	35.7	0
	Maternal study satisfaction (15 mo)						
	Very satisfied (0)	1,496	1,496	0	7.9	7.9	0
	Somewhat satisfied (1-2)	939	939	0	13.2	13.2	0
	Somewhat dissatisfied (3-4)	698	698	0	16.2	16.2	0
	Very dissatisfied (5-6)	289	289	0	23.2	23.2	0

Attachment A: SAS Code

```
*** TEDDY M31 DSIC;
*** Programmer: Allyson Mateja;
*** Date: 3/2/2017;

libname sas_data '/prj/niddk/ims_analysis/TEDDY/private_orig_data/m_31_sjohnson_niddk_submission/';

proc format;
  value countryf 1 = 'US'
                2 = 'Finland'
                3 = 'Germany'
                4 = 'Sweden';
  value genderf 0 = 'Boy'
                1 = 'Girl';
  value yesnof 0 = 'No'
               1 = 'Yes';
  value eduf 1 = 'Primary education'
            2 = 'Trade school/some college'
            3 = 'Graduated college';
  value workingf 0 = 'Reduced work/did not work at all'
                 1 = 'Worked all trimesters';
  value alcfc 0 = 'Total abstinence in third trimester'
              1 = 'Occasional drink in third trimester';
  value riskf 0 = 'Underestimate'
              1 = 'Accurate';
  value cohortf 0 = 'Year 1'
                1 = 'Year 2'
                2 = 'Year 3'
                3 = 'Year 4';
  value dadf 0 = 'Completed all 3 surveys'
              1 = 'Completed 0, 1, or 2 surveys';
  value missf 0 = 'No missed study visits or blood draws'
              1 = 'Missed >= 1 blood draws but not a study visit'
              2 = 'Missed >= 1 study visits';
  value satisf 3 = 'Very satisfied'
               4,5 = 'Somewhat satisfied'
               6,7 = 'Somewhat dissatisfied'
               8,9 = 'Very dissatisfied';

data teddym31;
  set sas_data.m_31_sjohnson_niddk_30jun2011;

proc contents data = teddym31;

data table1;
  set teddym31;
  if exclude=0;
```

```

data block1 block2 block3 block4 block5;
  set table1;
  if childparentsstatus in ('Divorced', 'Seperated', 'Unmarried and living apart', 'Widowed') then living_status = 0;
  else if childparentsstatus in ('Married', 'Unmarried but living together') then living_status=1;
  if smoke_pregnancy = 1 or mom_smokes = 1 then smoking=1;
  else smoking=0;
  if serious_neg_life_event_count = 0 then neg_event=0;
  else if serious_neg_life_event_count > 0 then neg_event=1;
  if 0 <= crowdingspace <= 0.49 then crowding = 0;
  else if 0.50 <= crowdingspace <= 0.59 then crowding = 1;
  else if 0.60 <= crowdingspace <= 0.75 then crowding = 2;
  else if 0.76 <= crowdingspace <= 1.00 then crowding = 3;
  else if crowdingspace > 1 then crowding = 4;
  if count_illness in (0,1,2) then illness_num = 0;
  else if count_illness in (3,4) then illness_num = 1;
  else if count_illness in (5,6,7) then illness_num=2;
  else if count_illness in (8,9,10) then illness_num=3;
  else if count_illness > 10 then illness_num=4;
  if country ne . and female ne . and ethnic_minority ne . and single_child ne . and maternal_age ne . and crowdingspace ne . and
  education_mom_group3 ne . and childparentsstatus ne '' then output block1;
  if worked_preg ne . and (smoke_pregnancy ne . or mom_smokes ne .) and alc_3rd_trimester ne . and mom_works ne . then output
  block2;
  if count_illness ne . and serious_neg_life_event_count ne . then output block3;
  if (anydev ne . or anydev_six ne . or anydev_fifteen ne .) and (stai ne . or stai_six ne . or stai_fifteen ne .) and
  depressiontotal ne . and totalblues ne . then output block4;
  if birth_cohort not in (., 4) and (satis_score_6mo_mom ne . or satis_score_15mo_mom ne .) then output block5;

proc freq data = block1;
  tables country country*late_dropout3yr /nocol nopercnt;
  format country countryf.;
  title 'Table 1 - Country of residence';

proc freq data = block1;
  tables female female*late_dropout3yr /nocol nopercnt;
  format female genderf.;
  title 'Table 1 - Gender of child';

proc freq data = block1;
  tables ethnic_minority ethnic_minority*late_dropout3yr /nocol nopercnt;
  format ethnic_minority yesnof.;
  title 'Table 1 - Child ethnic/minority status';

proc freq data = block1;
  tables single_child single_child*late_dropout3yr /nocol nopercnt;
  format single_child yesnof.;
  title 'Table 1 - Only child';

proc means data = block1 n mean std;
  var maternal_age;
  title "Maternal age at child's birth";

```

```

proc freq data = block1;
  tables living_status living_status*late_dropout3yr /nocol nopercnt;
  format living_status yesnof.;
  title 'Table 1 - Parents married or living together';

proc freq data = block1;
  tables education_mom_group3 education_mom_group3*late_dropout3yr /nocol nopercnt;
  format education_mom_group3 eduf.;
  title "Table 1 - Mother's education";

proc means data = block1 n mean std;
  var crowding;
  title 'Crowding (persons/room)';

proc freq data = block2;
  tables worked_preg worked_preg*late_dropout3yr /nocol nopercnt;
  format worked_preg workingf.;
  title 'Working during pregnancy';

proc freq data = block2;
  tables alc_3rd_trimester alc_3rd_trimester*late_dropout3yr /nocol nopercnt;
  format alc_3rd_trimester alcfc.;
  title 'Alcohol consumption during pregnancy';

proc freq data = block2;
  tables smoking smoking*late_dropout3yr /nocol nopercnt;
  format smoking yesnof.;
  title 'Smoking (pregnancy or during first year of TEDDY)';

proc freq data = block2;
  tables mom_works mom_works*late_dropout3yr /nocol nopercnt;
  format mom_works yesnof.;
  title 'Working outside the home in the first year of TEDDY';

proc freq data = block3;
  tables neg_event neg_event*late_dropout3yr /nocol nopercnt;
  format neg_event yesnof.;
  title 'Serious negative life event in the first year of TEDDY';

proc means data = block3 n mean std;
  var illness_num;
  title 'Number of child illnesses in the first year of TEDDY';

data block4;
  set block4;
  if anydev_fifteen = . then anydev_fifteen=anydev;
  if stai_fifteen = . then do;
    if stai_six ne . then stai_fifteen = stai_six;
    else stai_fifteen=stai;
  end;
  if depressiontotal =0 then brad_dep_score = 0;

```

```

        if depressiontotal =1 then brad_dep_score = 1;
        if depressiontotal in (2,3) then brad_dep_score = 2;
        if depressiontotal in (4,5) then brad_dep_score = 3;
        if 6 <= depressiontotal <= 25 then brad_dep_score=4;

proc freq data = block4;
    tables anydev_fifteen anydev_fifteen*late_dropout3yr /nocol nopercnt;
    format anydev_fifteen riskf.;
    title 'Risk perception accuracy (15 mo)';

proc means data = block4 n mean std;
    var stai_fifteen;
    title 'State Anxiety Inventory score (15 mo)';

proc means data = block4 n mean std;
    var totalblues;
    title 'Edinburgh Postnatal Depression score (6 mo)';

proc means data = block4 n mean std;
    var brad_dep_score;
    title 'Bradley Depression subscale (15 mo)';

data block5;
    set block5;
    if dad_participation in (.,1,2) then dad_participation = 0;
    else if dad_participation = 3 then dad_participation = 1;
    if missed_visit = 1 and missed_blood_draw = 1 then missing = 2;
    else if missed_visit = 0 and missed_blood_draw = 0 then missing = 0;
    else missing = 1;
    if satis_score_15mo_mom = . then satis_score_15mo_mom = satis_score_6mo_mom;

proc freq data = block5;
    tables birth_cohort birth_cohort*late_dropout3yr /nocol nopercnt;
    format birth_cohort cohortf.;
    title 'TEDDY recruitment cohort';

proc freq data = block5;
    tables dad_participation dad_participation*late_dropout3yr /nocol nopercnt;
    format dad_participation dadf.;
    title "Father's participation in first year of TEDDY";

proc freq data = block5;
    tables missing missing*late_dropout3yr /nocol nopercnt;
    format missing missf.;
    title 'Study compliance in first year of TEDDY';

proc freq data = block5;
    tables satis_score_15mo_mom satis_score_15mo_mom*late_dropout3yr /nocol nopercnt;
    format satis_score_15mo_mom satisf.;
    title 'Maternal study satisfaction (15 months)';

```