Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M104 RRoth

Prepared by Sabrina Chen
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705

March 4, 2020

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/private_orig_data/M_104_RRoth_NIDDK_Submission folder in the data package. For this replication, variables were taken from the “m_104_rroth_niddk_31dec2014_1.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Roswith Roth et al [1] in Developmental Psychobiology 2017. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For **Comparison of Data in the publication**, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data files to the results published.

6 Conclusions

The results of the replication are an exact match to the published results.

7 References

[1] Roswith Roth, Judith Baxter, Kendra Vehik, Diane Hopkins, Michael Killian, Patricia Gesualdo, Jessica Melin, Barbara Simell, Elisabeth Strauss, Åke Lernmark, Suzanne Bennett Johnson The TEDDY Study Group. The feasibility of salivary sample collection in an international pediatric cohort: The the TEDDY study. Developmental Psychobiology. 2017;9999:1–10.

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

Table Variable	dataset.variable
Of N3 kids, reasons that we did not get the data-not offered SSP	m_104_rroth_niddk_31dec2014_1.N6B
Of N3 kids, No reason given-not offered SSP	m_104_rroth_niddk_31dec2014_1.N6C
Of N1 kids, the number who had one or more saliva collections=N2	m_104_rroth_niddk_31dec2014_1.N2
Of N3 kids, Number who refused the procedure	m_104_rroth_niddk_31dec2014_1.N6A
Country 1=US, 0=EU	m_104_rroth_niddk_31dec2014_1.country
Time point of saliva collection, 302=42m,555=54m,309=66m	m_104_rroth_niddk_31dec2014_1.TIME_POINT_CD

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

	42 Month						54 Month					
	DSIC	Manuscript	Diff	DSIC	Manuscript	Diff	DSIC	Manuscript	Diff	DSIC	Manuscript	Diff
	COUNT_42			PERCENT_42			COUNT_54			PERCENT_54		
Total												
Completed visit	4307	4,307	0				4545	4,545	0			
Not offered SSP	207	207	0	4.8	4.8	0	345	345	0	7.6	7.6	0
Total eligible for SSP	4100	4,100	0	95.2	95.2	0	4200	4,200	0	92.4	92.4	0
≥1 sample	3948	3,948	0	96.3	96.3	0	4016	4,016	0	95.6	95.6	0
Refused	152	152	0	3.7	3.7	0	184	184	0	4.4	4.4	0
US												
Completed visits	1090	1,090	0				1402	1,402	0			
Not offered SSP	141	141	0	12.9	12.9	0	247	247	0	17.6	17.6	0
Total eligible for SSP	949	949	0	87.1	87.1	0	1155	1,155	0	82.4	82.4	0
≥1 sample	935	935	0	98.5	98.5	0	1128	1,128	0	97.7	97.7	0
Refused	14	14	0	1.5	1.5	0	27	27	0	2.3	2.3	0
EU												
Completed visits	3217	3,217	0				3143	3,143	0			
Not offered SSP	66	66	0	2.1	2.1	0	98	98	0	3.1	3.1	0
Total eligible for SSP	3151	3,151	0	97.9	97.9	0	3045	3,045	0	96.9	96.9	0
≥1 sample	3013	3,013	0	95.6	95.6	0	2888	2,888	0	94.8	94.8	0
Refused	138	138	0	4.4	4.4	0	157	157	0	5.2	5.2	0

	66 Month						Total					
	DSIC	Manuscript	Diff	DSIC	Manuscript	Diff	DSIC	Manuscript	Diff	DSIC	Manuscript	Diff
	COUNT_66			PERCENT_66			COUNT			PERCENT		
Total												
Completed visit	3838	3,838	0				12690	12,690	0			
Not offered SSP	262	262	0	6.8	6.8	0	814	814	0	6.4	6.4	0
Total eligible for SSP	3576	3,576	0	93.2	93.2	0	11876	11,876	0	93.6	93.6	0
≥1 sample	3426	3,426	0	95.8	95.8	0	11390	11,390	0	95.9	95.9	0
Refused	150	150	0	4.2	4.2	0	486	486	0	4.1	4.1	0
US												
Completed visits	1221	1,221	0				3,713	3,713	0			
Not offered SSP	164	164	0	13.4	13.4	0	552	552	0	14.9	14.9	0
Total eligible for SSP	1057	1,057	0	86.6	86.6	0	3161	3,161	0	85.1	85.1	0
≥1 sample	1038	1,038	0	98.2	98.2	0	3101	3,101	0	98.1	98.1	0
Refused	19	19	0	1.8	1.8	0	60	60	0	1.9	1.9	0
EU												
Completed visits	2617	2,617	0				8,977	8,977	0			
Not offered SSP	98	98	0	3.7	3.7	0	262	262	0	2.9	2.9	0
Total eligible for SSP	2519	2,519	0	96.3	96.3	0	8715	8,715	0	97.1	97.1	0
≥1 sample	2388	2,388	0	94.8	94.8	0	8289	8,289	0	95.1	95.1	0
Refused	131	131	0	5.2	5.2	0	426	426	0	4.9	4.9	0

Attachment A: SAS Code

```
options nocenter validvarname=upcase;

title '/prj/niddk/ims_analysis/TEDDY/prog_initial_analysis/m_104_dsic.sas';
run;

libname dat '/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_104_RRoth_NIDDK_Submission';


proc format;
 value val
 .     = "no value"
 other = "   value"
 ;

 value oneplus
 . = "no value"
 0 = "0"
 1-high = "1+"
 ;

 value zerohi
 . = "no value"
 0-high = "0-high"
 ;

 value sitef
 1 = '(1) US-Colorado'
 2 = '(2) US-Georgia/Florida'
 3 = '(3) US-Washington State'
 4 = '(4) Finland'
 5 = '(5) Germany'
 6 = '(6) Sweden'
 ;

 value malef
 0 = '(B) Female'
 1 = '(A) Male'
 ;

 value fdr
 1='(B) FDR'
 0='(A) GenPop'
 ;

 value ageislet
 1 = '    <12 months'
 2 = '12 =<24 months'
 3 = '24 =<36 months'
 4 = '36 =<48 months'
 5 = '48 = 72 months'
 ;

run;



* produce n and %;
%macro t1(tp, subset, subsetname);

proc freq data=analy noprint;
* where n1=1 and TIME_POINT_CD=&tp;
  where n1=1 and TIME_POINT_CD=&tp and &subset=1 ;
  tables N6B_N6C/list missing out=&subsetname.12;
run;

proc freq data=analy noprint;
  where n1=1 and &subset=1 and (N2=1 or N6A=1);
  tables N2* N6A /missing list out=&subsetname.34;
run;

data &subsetname;
  set &subsetname.12 &subsetname.34;
run;

proc print data=&subsetname;
run;

%mend;


%macro t1total(subset, subsetname);

proc freq data=analy noprint;
  where n1=1 and country in(&subset);
  tables N6B_N6C/list missing out=&subsetname.12;
run;
```

```
proc freq data=analy noprint;
  where n1=1 and country in(&subset) and (N2=1 or N6A=1);
  tables N2* N6A /missing list out=&subsetname.34;
run;

data &subsetname;
  set &subsetname.12 &subsetname.34;
run;

proc print data=&subsetname;
run;

%mend;



data analy;
  set dat.m_104_rroth_niddk_31dec2014_1;

  * Create var for 'not offered SSP' vs 'total eligible for SSP';
  N6B_N6C = sum(N6B, N6C);

  * create subsets;
  if country in(1 0) then do;
    if TIME_POINT_CD =302 then tot_42 = 1;
    else if TIME_POINT_CD = 555 then tot_54 = 1;
    else if TIME_POINT_CD = 309 then tot_66 = 1;
  end;

  if country in(1) then do;
    if TIME_POINT_CD =302 then us_42 = 1;
    else if TIME_POINT_CD = 555 then us_54 = 1;
    else if TIME_POINT_CD = 309 then us_66 = 1;
  end;

  if country in(0) then do;
    if TIME_POINT_CD =302 then eu_42 = 1;
    else if TIME_POINT_CD = 555 then eu_54 = 1;
    else if TIME_POINT_CD = 309 then eu_66 = 1;
  end;

run;

proc contents data=analy;
run;
```

```
proc freq data=analy;
  where n1=1;
  tables TIME_POINT_CD/missing;
  tables time_point_cd*COUNTRY/list missing;
  title3 "confirm subset counts";
run;

proc freq data=analy;
  where n1=1;
  tables  country*time_point_cd*tot_42* tot_54 * tot_66 * us_42 * us_54    * us_66 * eu_42 * eu_54 * eu_66/list missing;
  title3 "check subset flags";
run;


* TOTAL;
%t1(302, tot_42, tot42);
%t1(555, tot_54, tot54);
%t1(309, tot_66, tot66);
%t1total(1 0, us_eu);

proc sort data=tot42 (rename=(count=count_42 percent=percent_42));
  by  N6B_N6C N2 N6A;
run;

proc sort data=tot54 (rename=(count=count_54 percent=percent_54));
  by  N6B_N6C N2 N6A;
run;

proc sort data=tot66 (rename=(count=count_66 percent=percent_66));
  by  N6B_N6C N2 N6A;
run;

proc sort data=us_eu;
  by  N6B_N6C N2 N6A;
run;

data row_totals;
  merge tot42 tot54 tot66 us_eu;
  by N6B_N6C N2 N6A;
  if N6B_N6C = 1 then row = 1;
  else if N6B_N6C = 0 then row = 2;
  else if N2=1 and N6A = 0 then row = 3;
  else if N2=0 and N6A = 1 then row = 4;
  group = "Total";

run;

proc sort data=row_totals;
  by row;
run;

proc print data=row_totals;
  var N6B_N6C N2 N6A count_42 percent_42 count_54 percent_54 count_66 percent_66 count percent;
  title3 "Total";
run;


** US;
%t1(302, us_42, us42);
%t1(555, us_54, us54);
%t1(309, us_66, us66);
%t1total(1  , us);


proc sort data=us42 (rename=(count=count_42 percent=percent_42));
  by  N6B_N6C N2 N6A;
run;

proc sort data=us54 (rename=(count=count_54 percent=percent_54));
  by  N6B_N6C N2 N6A;
run;

proc sort data=us66 (rename=(count=count_66 percent=percent_66));
  by  N6B_N6C N2 N6A;
run;

proc sort data=us;
  by  N6B_N6C N2 N6A;
run;

data row_us;
  merge us42 us54 us66 us;
  by N6B_N6C N2 N6A;
  if N6B_N6C = 1 then row = 1;
  else if N6B_N6C = 0 then row = 2;
  else if N2=1 and N6A = 0 then row = 3;
  else if N2=0 and N6A = 1 then row = 4;
  group = "US";

run;

proc sort data=row_us;
  by row;
run;

proc print data=row_us;
  var N6B_N6C N2 N6A count_42 percent_42 count_54 percent_54 count_66 percent_66 count percent;
  title3 "US";
run;



** EU;
%t1(302, eu_42, eu42);
%t1(555, eu_54, eu54);
%t1(309, eu_66, eu66);
%t1total(0  , eu);

proc sort data=eu42 (rename=(count=count_42 percent=percent_42));
  by  N6B_N6C N2 N6A;
run;

proc sort data=eu54 (rename=(count=count_54 percent=percent_54));
  by  N6B_N6C N2 N6A;
run;

proc sort data=eu66 (rename=(count=count_66 percent=percent_66));
  by  N6B_N6C N2 N6A;
run;

proc sort data=eu;
  by  N6B_N6C N2 N6A;
run;

data row_eu;
  merge eu42 eu54 eu66 eu;
  by N6B_N6C N2 N6A;
  if N6B_N6C = 1 then row = 1;
  else if N6B_N6C = 0 then row = 2;
  else if N2=1 and N6A = 0 then row = 3;
  else if N2=0 and N6A = 1 then row = 4;
  group = "EU";

run;

proc sort data=row_eu;
  by row;
run;

proc print data=row_eu;
  var N6B_N6C N2 N6A count_42 percent_42 count_54 percent_54 count_66 percent_66 count percent;
  title3 "EU";
run;


* Combine;
data table1;
  set row_totals row_us row_eu;

  percent_42 = put(percent_42,8.1);
  percent_54 = put(percent_54,8.1);
  percent_66 = put(percent_66,8.1);
  percent    = put(percent,8.1);

run;

proc print data=table1;
  var N6B_N6C N2 N6A group count_42 percent_42 count_54 percent_54 count_66 percent_66 count percent;
title3 "Table 1";
run;


ods listing close;
ods phtml file="/prj/niddk/ims_analysis/TEDDY/private_created_data/TEDDY.m104.Table1.xls";

proc print data=table1;
  var N6B_N6C N2 N6A group count_42 percent_42 count_54 percent_54 count_66 percent_66 count percent;
title3 "M104 Table 1";
run;

ods phtml close;
ods listing;
```