Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub40 Roth

Prepared by Michael Spriggs
IMS Inc.
3901 Calverton Blvd, Suite 200, Calverton MD, 20705
April 25, 2016

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 the "mp40_rroth_niddk" dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Roth et al Pediatric Diabetes 2015: 16: 287–298 doi: 10.1111/pedi.12168 [1]. To verify the integrity of the dataset, sociodemographic and maternal lifestyle factors (Table 1) were computed, by country.

5 Results

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 identical to the published results.

6 Conclusions

The NIDDK repository is confident that the TEDDY mp40_rroth_niddk data files to be distributed are a true copy of the study data.

7 References

[1] Roth, Roswith, et al. "Maternal anxiety about a child's diabetes risk in the TEDDY study: the potential role of life stress, postpartum depression, and risk perception." Pediatric diabetes 16.4 (2015): 287-298.

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

Table Variable	Variables Used in Replication from the "Table 1" dataset
Gender of child	female
Child ethnic minority	ethnic_minority
Child has FDR with T1D	FDR
Child is an only child	single_child
Maternal education	education_mom_group3
Married or living status	married_living
Working status	worked
Smoking during pregnancy:	smoker
Alcohol during pregnancy	alc_pregnancy
Maternal age at child's birth (yr)	maternal_age
Household crowding*	crowding_norm

Table B: Comparison of values computed in integrity check to reference article Table 1 values: Sociodemographic and maternal lifestyle factors by country

USA	Mean /N	Mean /N	Mean /N [Difference]
Gender of child			
Boy	1583(50.9)	1583 (50.9)	0(0)
Girl	1527(49.1)	1527 (49.1)	0(0)
Child ethnic minority			
No	2034(69.9)	2034 (69.9)	0(0)
Yes	874(30.1)	874 (30.1)	0(0)
Child has FDR with T1D			
No	2782(89.5)	2782 (89.5)	0(0)
Yes	328(10.5)	328 (10.5)	0(0)
Child is an only child			
No	1810(62.7)	1810 (62.7)	0(0)
Yes	1075(37.3)	1075 (37.3)	0(0)
Maternal age at child's birth (yr %)	30.5(5.7)	30.5(5.7)	0(0)
Maternal education			
Basic primary	441(15.3)	441 (15.3)	0(0)
Graduated trade school education/college	720(25.0)	720 (25.0)	0(0)
Higher education	1723(59.7)	1723 (59.7)	0(0)
Married or living status			
Married or living together	2711(93.8)	2711 (93.8)	0(0)
Single parent	178(6.2)	178 (6.2)	0(0)
Maternal lifestyle variables during			
Working status			
Did not work or reduced work hours	1368(46.1)	1368 (46.1)	0(0)
Worked all three trimesters	1599(53.9)	1599 (53.9)	0(0)
Smoking during pregnancy			
No	2748(90.5)	2748 (90.5)	0(0)
Yes	287(9.5)	287 (9.5)	0(0)
Alcohol during pregnancy			
No	2345(76.6)	2345 (76.6)	0(0)
Yes	715(23.4)	715 (23.4)	0(0)
Household crowding	1.6(1.3)	1.6(1.3)	0(0)

Finland	Mean /N	Mean /N	Mean /N [Difference]
Gender of child			
Boy	829(51.0)	829 (51.0)	0(0)
Girl	798(49.0)	798 (49.0)	0(0)
Child ethnic minority			
No	1502(97.0)	1502 (97.0)	0(0)
Yes	47(3.0)	47 (3.0)	0(0)
Child has FDR with T1D			
No	1479(90.9)	1479 (90.9)	0(0)
Yes	148(9.1)	148 (9.1)	0(0)
Child is an only child			
No	875(55.7)	875 (55.7)	0(0)
Yes	696(44.3)	696 (44.3)	0(0)
Maternal age at child's birth (yr %)	30.0(5.0)	30.0(5.0)	0(0)
Maternal education			
Basic primary	154(9.8)	154 (9.8)	0(0)
Graduated trade school education/college	447(28.6)	447 (28.6)	0(0)
Higher education	964(61.6)	964 (61.6)	0(0)
Married or living status			
Married or living together	1518(96.7)	1518 (96.7)	0(0)
Single parent	52(3.3)	52 (3.3)	0(0)
Maternal lifestyle variables during			
Working status			
Did not work or reduced work hours	883(56.7)	883 (56.7)	0(0)
Worked all three trimesters	673(43.3)	673 (43.3)	0(0)
Smoking during pregnancy			
No	1393(87.2)	1393 (87.2)	0(0)
Yes	204(12.8)	204 (12.8)	0(0)
Alcohol during pregnancy			
No	1331(82.5)	1331 (82.5)	0(0)
Yes	283(17.5)	283 (17.5)	0(0)
Household crowding	2.5(0.9)	2.5(0.9)	0(0)

Germany	Mean /N	Mean /N	Mean /N [Difference]
Gender of child			
Boy	265(50.8)	265 (50.8)	0(0)
Girl	257(49.2)	257 (49.2)	0(0)
Child ethnic minority			
No	429(87.6)	429 (87.6)	0(0)
Yes	61(12.4)	61 (12.4)	0(0)
Child has FDR with T1D			
No	323(61.9)	323 (61.9)	0(0)
Yes	199(38.1)	199 (38.1)	0(0)
Child is an only child			
No	257(52.4)	257 (52.4)	0(0)
Yes	233(47.6)	233 (47.6)	0(0)
Maternal age at child's birth (yr %)	31.6(4.9)	31.6(4.8)	0(0.1)
Maternal education			
Basic primary	65(13.3)	65 (13.3)	0(0)
Graduated trade school education/college	246(50.3)	246 (50.3)	0(0)
Higher education	178(36.4)	178 (36.4)	0(0)
Married or living status			
Married or living together	473(96.5)	473 (96.5)	0(0)
Single parent	17(3.5)	17 (3.5)	0(0)
Maternal lifestyle variables during			
Working status			
Did not work or reduced work hours	302(58.3)	302 (58.3)	0(0)
Worked all three trimesters	216(41.7)	216 (41.7)	0(0)
Smoking during pregnancy			
No	425(82.7)	425 (82.7)	0(0)
Yes	89(17.3)	89 (17.3)	0(0)
Alcohol during pregnancy			
No	381(73.1)	381 (73.1)	0(0)
Yes	140(26.9)	140 (26.9)	0(0)
Household crowding	2.4(1.1)	2.4(1.1)	0(0)

Sweden	Mean /N	Mean /N	Mean /N [Difference]
Gender of child			
Boy	1194 (50.7)	1194 (50.7)	0(0)
Girl	1159 (49.3)	1159 (49.3)	0(0)
Child ethnic minority			
No	2105 (92.9)	2105 (92.9)	0(0)
Yes	160 (7.1)	160 (7.1)	0(0)
Child has FDR with T1D			
No	2202 (93.6)	2202 (93.6)	0(0)
Yes	151 (6.4)	151 (6.4)	0(0)
Child is an only child			
No	1243 (54.8)	1243 (54.8)	0(0)
Yes	1024 (45.2)	1024 (45.2)	0(0)
Maternal age at child's birth (yr %)	30.8 (4.7)	30.8 (4.6)	0(0.1)
Maternal education			
Basic primary	767 (33.8)	767 (33.8)	0(0)
Graduated trade school education/college	389 (17.2)	389 (17.2)	0(0)
Higher education	1111 (49.0)	1111 (49.0)	0(0)
Married or living status			
Married or living together	2213 (97.7)	2213 (97.7)	0(0)
Single parent	53 (2.3)	53 (2.3)	0(0)
Maternal lifestyle variables during			
Working status			
Did not work or reduced work hours	1245 (54.7)	1245 (54.7)	0(0)
Worked all three trimesters	1030 (45.3)	1030 (45.3)	0(0)
Smoking during pregnancy			
No	2038 (87.7)	2038 (87.7)	0(0)
Yes	287 (12.3)	287 (12.3)	0(0)
Alcohol during pregnancy			
No	2131 (90.8)	2131 (90.8)	0(0)
Yes	217 (9.2)	217 (9.2)	0(0)
Household crowding	2.1 (1.1)	2.1 (1.1)	0(0)

ALL	Mean /N	Mean /N	Mean /N [Difference]
Gender of child			
Boy	3871(50.9)	3871(50.9)	0(0)
Girl	3741(49.1)	3741(49.1)	0(0)
Child ethnic minority			
No	6070(84.2)	6070(84.2)	0(0)
Yes	1142(15.8)	1142(15.8)	0(0)
Child has FDR with T1D			
No	6786(89.1)	6786(89.1)	0(0)
Yes	826(10.9)	826(10.9)	0(0)
Child is an only child			
No	4185(58.0)	4185(58.0)	0(0)
Yes	3028(42.0)	3028(42.0)	0(0)
Maternal age at child's birth (yr %)	30.6(5.2)	30.6(5.2)	0(0)
Maternal education			
Basic primary	1427(19.9)	1427(19.9)	0(0)
Graduated trade school education/college	1802(25.0)	1802(25.0)	0(0)
Higher education	3976(55.2)	3976(55.2)	0(0)
Married or living status			
Married or living together	6914(95.8)	6915(95.8)	1(0)
Single parent	300(4.2)	300(4.2)	0(0)
Maternal lifestyle variables during			
Working status			
Did not work or reduced work hours	3798(51.9)	3798(51.9)	0(0)
Worked all three trimesters	3518(48.1)	3518(48.1)	0(0)
Smoking during pregnancy			
No	6604(88.4)	6604(88.4)	0(0)
Yes	867(11.6)	867(11.6)	0(0)
Alcohol during pregnancy			
No	6188(82.0)	6188(82.0)	0(0)
Yes	1355(18.0)	1355(18.0)	0(0)
Household crowding	2.0(1.2)	2.0(1.2)	0(0)

Attachment A: SAS Code

```
*********************************************************************************************************************
***Program:
***Programmer: Michael Spriggs
***Date Created: 04/25/2016
***Purpose:
********************************************************************************************************************;

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

options nofmterr;

libname sas_data "/prj/niddk/ims_analysis/TEDDY/private_orig_data/MP40/";
%include '/prj/niddk/ims_analysis/sas_macros/redaction_data_summary.sas';

*** Reading in the analysis datasets used for the DSIC;
data mp40_rroth_niddk; set sas_data.mp40_rroth_niddk;

data analysis;
        set mp40_rroth_niddk;
        if exclude in(1 2) then delete;

%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";

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

 *proc print data = &var_name._cross;

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

   proc sort data = &var_name._cross;
     by &var_name  country;
 *proc print data = &var_name._cross;

  data &var_name._cross(drop =&var_name i country  Frequency colpercent);;
    set &var_name._cross;
    by &var_name country;
```

```
    length table_name $30.;
    table_name  ="&var_name";
    levels =&var_name;
    array temp1(4) us_count finland_count germany_count sweden_count  ;
    array temp2(4) us_pert finland_pert germany_pert sweden_pert ;
    retain us_count finland_count germany_count sweden_count  us_pert finland_pert germany_pert sweden_pert;
    if first.&var_name then do i = 1 to 4;
      temp1(i) = .;
      temp2(i) = .;
    end;
      temp1(country) = Frequency;
      temp2(country) = round(colpercent,.1);
    if last.&var_name;

 *proc print data = &var_name._cross;

%mend;


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

data analysis;
      set analysis;
      label crowding_norm='Normalized crowding score(0=0-0.49;1=0.50-0.59;2=0.60-0.75;3=0.76-1.00;4=>1.00)';

*** Running the baseline_freq on the  categorical variables in the Table 1 manuscript file;
%baseline_freq3(analysis,female             );
%baseline_freq3(analysis,ethnic_minority    );
%baseline_freq3(analysis,FDR                );
%baseline_freq3(analysis,single_child       );
%baseline_freq3(analysis,education_mom_group3);
%baseline_freq3(analysis,married_living     );
%baseline_freq3(analysis,worked             );
%baseline_freq3(analysis,smoker             );
%baseline_freq3(analysis,alc_pregnancy      );
%baseline_freq3(analysis,alcohol_3rd_trim   );

proc means data=analysis maxdec=1 mean stddev;
      var maternal_age crowding_norm;
      title3 'Means: All';

proc means data=analysis maxdec=1 mean stddev;
      var maternal_age crowding_norm;
      class country;
      title3 'Means: By Country';
```

11

```
data table1_compare(keep=table_name levels USA FIN GERMANY SWEDEN);
  set female_cross
                      ethnic_minority_cross
                      FDR_cross
                      single_child_cross
                      education_mom_group3_cross
                      married_living_cross
                      worked_cross
                      smoker_cross
                      alc_pregnancy_cross
                      alcohol_3rd_trim_cross
                  ;
      length USA $13. FIN $13. GERMANY $13. SWEDEN $13.;
      USA=strip(put(us_count,8.))||' ('||strip(put(us_pert,8.1))||')';
      FIN=strip(put(finland_count,8.))||' ('||strip(put(finland_pert,8.1))||')';
      GERMANY=strip(put(germany_count,8.))||' ('||strip(put(germany_pert,8.1))||')';
      SWEDEN=strip(put(sweden_count,8.))||' ('||strip(put(sweden_pert,8.1))||')';

proc print data =  table1_compare noobs;
    title3 'table 1 from data set';
```