Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub57 Larsson

Prepared by Dominick Parisi
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
May 1, 2018

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 “m_57_hlarsson.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Helena Elding Larsson et al [1] in Pediatric Diabetes in 2013. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], **Characteristics of the first 100 the Environmental Determinants of Diabetes in the Young (TEDDY) children diagnosed with T1D**, 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 in Table 1. The results of the replication are an exact match to the published results.

For Table 2 in the publication [1], **Symptoms and laboratory data at onset of T1D,** Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data files to the results published in Table 2. The numbers for HbA1c at diagnosis by cut points (≥ 6.5 and < 5.7) were switched in the final publication. The data included in the data package are correct. Otherwise, the results of the replication are an exact match to the published results.

For Figure 2 in the publication [1], Table E lists the variables that were used in the replication and Table F compares the results calculated from the archived data files to the results published in Figure 2. The results of the replication are an exact match to the published results.

6 Conclusions

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

7 References

[1] Elding Larsson H, Vehik K, Gesualdo P, Akolkar B, Hagopian W, Krischer J, Lernmark A° , Rewers M, Simell O, She J-X, Ziegler A, Haller MJ, and the TEDDY Study Group. Children followed in the TEDDY study are diagnosed with type 1 diabetes at an early stage of disease. Pediatric Diabetes 2013.

Table A: Variables used to replicate Table 1**:** Characteristics of the first 100 the Environmental Determinants of Diabetes in the Young (TEDDY) children diagnosed with T1D

Table Variable	**dataset.variable**
Gender	m_57_hlarsson_niddk_30nov2011_1.sex
First degree relative	m_57_hlarsson_niddk_30nov2011_1.fdr
Family member with T1D	m_57_hlarsson_niddk_30nov2011_1.family_mem
Mother's diabetes status	m_57_hlarsson_niddk_30nov2011_1.diabetes
Age at diagnosis	m_57_hlarsson_niddk_30nov2011_1.aget1d
Site	m_57_hlarsson_niddk_30nov2011_1.cc

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

Variable	TEDDY Manuscript (n=100)	TEDDY DSIC (n=100)	Diff. (n=0)
Sex (female)	45	45	0
First-degree relative (yes)	33	33	0
Family member with T1D			
- Mother only	4	4	0
- Father only	18	18	0
- Sibling only	9	9	0
- Mother and father	2	2	0
Mother's diabetes status			
- Gestational	4	4	0
- Type 1	6	6	0
- No diabetes	86	86	0
- Missing	4	4	0
Age at diagnosis (yr)	2.3 (.69-6.27)	2.3 (.69-6.27)	0 (0-0)
N and median, min and max age at diagnosis by site			
- Colorado	14, 1.76 (1.1-3.7)	14, 1.76 (1.1-3.7)	0, 0 (0-0)
- Georgia/Florida	4, 2.45 (1.2-4.3)	4, 2.45 (1.2-4.3)	0, 0 (0-0)
- Washington	7, 2.77 (0.87-4.9)	7, 2.77 (0.87-4.9)	0, 0 (0-0)
- Finland	35, 2.05 (0.69-5.1)	35, 2.05 (0.69-5.1)	0, 0 (0-0)
- Germany	13, 1.96 (1.0-3.2)	13, 1.96 (1.0-3.2)	0, 0 (0-0)
- Sweden	27, 2.98 (0.9-6.3)	27, 2.98 (0.9-6.3)	0, 0 (0-0)

Table C: Variables used to replicate Table 2: Symptoms and laboratory data at onset of T1D

Table Variable	**dataset.variable**
Was child symptomatic	m_57_hlarsson_niddk_30nov2011_1.WasChildSymptomatic
Polydipsia	m_57_hlarsson_niddk_30nov2011_1.polydipsia
Polyphagia	m_57_hlarsson_niddk_30nov2011_1.polyphagia
Polyuria	m_57_hlarsson_niddk_30nov2011_1.polyuria
Was child hospitalized	m_57_hlarsson_niddk_30nov2011_1.WasThisChildHospitalized
Was child treated in emergency room	m_57_hlarsson_niddk_30nov2011_1.WasThisChildTreatedinEmergency
Weight at diagnosis (kg)	m_57_hlarsson_niddk_30nov2011_1.CurrentWeightkg
Height at diagnosis (cm)	m_57_hlarsson_niddk_30nov2011_1.HeightCm
Weight loss reported at diagnosis	m_57_hlarsson_niddk_30nov2011_1.WeightLossAmountkg
Diagnostic Test	m_57_hlarsson_niddk_30nov2011_1.GlucoseTestType1_1
Average pH	m_57_hlarsson_niddk_30nov2011_1.pHResult
pH classification	m_57_hlarsson_niddk_30nov2011_1.ph1
Glucose (mmol/l)	m_57_hlarsson_niddk_30nov2011_1.glucose
HbA1C (%, mmol/mol) at diagnosis	m_57_hlarsson_niddk_30nov2011_1.HemoglobinA1cResult
HbA1c at diagnosis by cut points	m_57_hlarsson_niddk_30nov2011_1.hb1
Urine ketones at diagnosis	m_57_hlarsson_niddk_30nov2011_1.UrineKetonesResults
Blood ketones at diagnosis	m_57_hlarsson_niddk_30nov2011_1.bloodket

Table D: Comparison of values computed in integrity check to reference article Table 2 values

Variable	TEDDY Manuscript (n=100)	TEDDY DSIC (n=100)	Diff (n=0)
Was Child symptomatic (yes)	64	64	0
Polydipsia (yes)	53	53	0
Polyphagia (yes)	4	4	0
Polyuria (yes)	51	51	0
Was child hospitalized (yes)	82	82	0
Was child treated in emergency room (yes)	6	6	0
Weight at diagnosis (kg)	13 (6.5-27)	13 (6.5-27)	0 (0-0)
Height at diagnosis (cm)	90 (70-126)	90 (90-126)	0 (0-0)
Weight loss reported at diagnosis (kg)	0.5 (0.03-4.4)	0.5 (0.3-4.4)	0 (0-0)
Diagnostic Test			
- Fasting glucose	12	12	0
- OGTT – 2 hr	13	13	0
- Postprandial glucose	14	14	0
- Random glucose	61	61	0
Average pH (n=80)	7.4 (0.1)	7.4 (0.1)	0 (0)
- ≥ 7.1 and < 7.3	5	5	0
- < 7.1	3	3	0
Average Glucose (mmol/l) (n=98)	19.6 (9.7)	19.6 (9.7)	0 (0)

Variable	TEDDY Manuscript (n=100)	TEDDY DSIC (n=100)	Diff (n=0)
HbA1C (%, mmol/mol) at diagnosis			
- All (n=98)	7.4 (1.9), 57	7.4 (1.9), 57	
- Colorado	7.1 (1.4), 54	7.1 (1.4), 54	
- Georgia/Florida	7.4 (1.3), 57	7.4 (1.3), 57	
- Washington	8.7 (2.3), 72	8.7 (2.3), 72	
- Finland	7.6 (1.7), 60	7.6 (1.7), 60	
- Germany	8.4 (2.3), 68	8.4 (2.3), 68	
- Sweden	6.4 (1.7), 46	7.4 (1.7), 46	0
HbA1c at Diagnosis by Cutpoints (n=98)			
- ≥ 6.5	15	60	45
- ≥ 5.7 and < 6.5	23	23	0
- < 5.7	60	15	45
Urine ketones at diagnosis (n=80)			
- Large	9	9	0
- Moderate	11	11	0
- Small	9	9	0
- Trace	5	5	0
- Negative	46	46	0
Blood ketones at diagnosis (mmol/L)	0.98 (2.2)	0.98 (2.2)	0 (0)

Table E: Variables used to replicate Figure 2**:**

Table Variable	**dataset.variable**
first confirmed positive autoantibodies	m_57_hlarsson_niddk_30nov2011_1.firstpos
confirmed autoantibodies at diagnosis	m_57_hlarsson_niddk_30nov2011_1.confabstatus
persist confirmed autoantibodies at diagnosis	m_57_hlarsson_niddk_30nov2011_1.persist_confstatus

Table F: Comparison of values computed in integrity check to reference article Figure 2

	TEDDY Manuscript (n)	TEDDY DSIC (n)	Diff (n)
(A) First Confirmed positive autoantibodies			
- GAD only	13	13	0
- IA2A only	0	0	0
- IAA only	49	49	0
- GAD and IAA only	28	28	0
- GAD and IA2A only	0	0	0
- IA2A and IAA only	1	1	0
- GAD, IAA and IA2A	3	3	0
(B) Confirmed autoantibodies at diagnosis			
- GAD only	2	2	0
- IAA only	10	10	0
- IA2A only	0	0	0
- GAD and IAA only	24	24	0
- GAD and IA2A only	1	1	0
- IA2A and IAA only	11	11	0
- GAD, IAA and IA2A	46	46	0
(C) Persist confirmed autoantibodies at diagnosis			
- GAD only	2	2	0
- IAA only	14	14	0
- IA2A only	0	0	0
- GAD and IAA only	20	20	0
- GAD and IA2A only	2	2	0
- IA2A and IAA only	12	12	0
- GAD, IAA and IA2A	33	33	0

Attachment A: SAS Code

```
 OPTIONS NOFMTERR;
/*********************/
/* Library Statements */
/*********************/
LIBNAME SASDATA '/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_57_HLarsson_NIDDK_Submission';

/*******************/
/* Import datasets */
/*******************/
DATA ANALYSIS1;
  SET SASDATA.m_57_hlarsson_niddk_30nov2011_1;
RUN;

/***********/
/* Table 1 */
/***********/
TITLE2 'Table 1';
PROC FREQ DATA=ANALYSIS1;
  TABLE Sex
        fdr
        family_mem
        diabetes  ;
RUN;

PROC MEANS DATA=ANALYSIS1 MEDIAN MIN MAX;
  VAR aget1d;
RUN;

PROC MEANS DATA=ANALYSIS1 MEDIAN MIN MAX;
  VAR aget1d;
  CLASS CC;
RUN;

/***********/
/* Table 2 */
/***********/
TITLE2 'Table 2';
PROC FREQ DATA=ANALYSIS1;
  TABLE WasChildSymptomatic
        Polydipsia
        Polyphagia
        Polyuria
        WasThisChildHospitalized
        WasThisChildTreatedInEmergency;
RUN;

PROC MEANS DATA=ANALYSIS1 MEDIAN MIN MAX;
```

```
  VAR CurrentWeightkg
      HeightCm
      WeightLossAmountkg;
RUN;

PROC FREQ DATA=ANALYSIS1;
  TABLE GlucoseTestType1_1;
RUN;

PROC MEANS DATA=ANALYSIS1;
  VAR pHResult;
RUN;

PROC FREQ DATA=ANALYSIS1;
  TABLE ph1;
RUN;

PROC MEANS DATA=ANALYSIS1;
  VAR glucose;
RUN;

PROC MEANS DATA=ANALYSIS1;
  VAR HemoglobinA1cResult;
RUN;

PROC MEANS DATA=ANALYSIS1;
  VAR HemoglobinA1cResult;
  CLASS CC;
  OUTPUT OUT=HBA1C_MEANS MEAN=MEAN STD=STD;
RUN;

PROC MEANS DATA=ANALYSIS1;
  VAR HemoglobinA1cResult;
  CLASS HB1;
RUN;

DATA HBA1C_MEANS;
  SET HBA1C_MEANS;

  MMOL=(MEAN-2.152)/.09148;
RUN;

PROC PRINT DATA=HBA1C_MEANS;
RUN;

PROC FREQ DATA=ANALYSIS1;
  TABLE hb1
        UrineKetonesResult;
RUN;

PROC MEANS DATA=ANALYSIS1;
```

```
  VAR BLOODKET;
RUN;

/*************************/
/* Figure 2: Venn diagrams */
/*************************/
TITLE2 'Venn Diagrams';
PROC FREQ DATA=ANALYSIS1;
  TABLE firstpos
        confabstatus
        persist_confstatus /LIST MISSING;
RUN;
```