Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub180 Krischer

Prepared by Laura Bowen
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
March 4, 2019

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_180_jkrischer_niddk_30jun2016.sas7bdat" dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Characteristics of children who progressed from multiple autoantibodies to T1D and those who did not, 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.

6 Conclusions

The NIDDK repository is confident that the TEDDY M180 data files to be distributed are a true copy of the study data.

7 References

[1] Jeffrey P. Krischer, Xiang Liu, Åke Lernmark, William A. Hagopian, Marian J. Rewers, Jin-Xiong She, Jorma Toppari, Anette-G. Ziegler, and Beena Akolkar, on behalf of the TEDDY Study Group. "The Influence of Type 1 Diabetes Genetic Susceptibility Regions, Age, Sex, and Family History on the Progression From Multiple Autoantibodies to Type 1 Diabetes: A TEDDY Study Report". Diabetes (2017) 66:3122-3129.

Table A: Variables used to replicate Table 1**:** Characteristics of children who progressed from multiple autoantibodies to T1D and those who did not

Table Variable	**dataset.variable**
Country of residence	m_180_jkrischer_niddk_30jun2016.country
Family history of T1D	m_180_jkrischer_niddk_30jun2016.family
Sex	m_180_jkrischer_niddk_30jun2016.female
HLA-DR-DQ genotypes	m_180_jkrischer_niddk_30jun2016.hla_5grps
Age at multiple persistent confirmed IA (months)	m_180_jkrischer_niddk_30jun2016.multipab_age_mon
Type of first autoantibody	m_180_jkrischer_niddk_30jun2016.firstab

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

	Did not progress to T1D			Progressed to T1D		
Variable	Manuscript (n=222)	DSIC (n=222)	Diff. (n=0)	Manuscript (n=190)	DSIC (n=190)	Diff. (n=0)
Country of residence, n (%)						
U.S.	82 (59)	82 (59)	0 (0)	57 (41)	57 (41)	0 (0)
Finland	53 (48)	53 (48)	0 (0)	57 (52)	57 (52)	0 (0)
Germany	16 (44)	16 (44)	0 (0)	20 (56)	20 (56)	0 (0)
Sweden	71 (56)	71 (56)	0 (0)	56 (44)	56 (44)	0 (0)
Family history of T1D, n (%)						
General population	171 (55)	171 (55)	0 (0)	140 (45)	140 (45)	0 (0)
FDR: mother	12 (50)	12 (50)	0 (0)	12 (50)	12 (50)	0 (0)
FDR: father	27 (50)	27 (50)	0 (0)	27 (50)	27 (50)	0 (0)
FDR: sibling	12 (52)	12 (52)	0 (0)	11 (48)	11 (48)	0 (0)
Sex, n (%)						
Female	88 (49)	88 (49)	0 (0)	92 (51)	92 (51)	0 (0)
Male	134 (58)	134 (58)	0 (0)	98 (42)	98 (42)	0 (0)
HLA-DR-DQ genotypes, n (%)						
DR3/4	121 (52)	121 (52)	0 (0)	110 (48)	110 (48)	0 (0)
DR4/4	46 (61)	46 (61)	0 (0)	30 (39)	30 (39)	0 (0)
DR4/8	31 (56)	31 (56)	0 (0)	24 (44)	24 (44)	0 (0)
DR3/3	16 (53)	16 (53)	0 (0)	14 (47)	14 (47)	0 (0)
FDR Specific	8 (40)	8 (40)	0 (0)	12 (60)	12 (60)	0 (0)
Age at multiple persistent confirmed IA (months), median (IQR)	48(31-74)	48(31-74)	0(0-0)	21(15-31)	21(15-31)	0(0-0)
Type of first autoantibody, n (%)						
GADA only	85 (66)	85 (66)	0 (0)	43 (34)	43 (34)	0 (0)
IAA only	84 (53)	84 (53)	0 (0)	76 (47)	76 (47)	0 (0)
Two or more autoantibodies	49 (42)	49 (42)	0 (0)	68 (58)	68 (58)	0 (0)
IA-2A only	4 (57)	4 (57)	0 (0)	3 (43)	3 (43)	0 (0)

Attachment A: SAS Code

```
*** TEDDY M180 DSIC;
*** Programmer: Laura Bowen;
*** Date: 3/1/2019;


libname origdata "/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_180_JKrischer_NIDDK_Submission/";


data m_180;
  set origdata.m_180_jkrischer_niddk_30jun2016;
run;


proc contents data = m_180;
  title4 'Contents of Raw Data';
run;


proc freq data = m_180;
  table T1D
        (country
         family
         female
         hla_5grps
         firstab) * t1d
         / missing;
  title4 'Table 1';
run;


proc univariate data=m_180;
  class t1d;
  var multipab_age_mon;
run;
```