Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M144 KDriscoll

Prepared by NIDDK-CR
June 6, 2023

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 type 1 diabetes (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.

The M144 study sought to examine adherence to the oral glucose tolerance test (OGTT) in multiple islet autoantibody children in stage 1 of developing type 1 diabetes.

3 Archived Datasets

A full listing of archived datasets included in the package can be found in the Roadmap document. All 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_144_kdriscoll_niddk_30apr2017.sas7bdat" dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data in the publication by Driscoll et al. [1]. To verify the integrity of the data, only descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Participant characteristics at time of OGTT eligibility, 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 in Table 1. The results of the replication are within expected variation of the published results.

6 Conclusions

The NIDDK Central Repository is confident that the TEDDY M144 data files to be distributed are a true copy of the study data.

7 References

[1] Driscoll KA, Tamura R, Johnson SB, Gesualdo P, Clasen J, Smith L, Jacobsen L, Larsson HE, Haller MJ. Adherence to Oral Glucose Tolerance Testing in Children in Stage 1 of Type 1 Diabetes: The TEDDY Study. Pediatric Diabetes, 22(2), 360-368, March 2021. doi: https://doi.org/10.1111/pedi.13149

Table A: Variables used to replicate Table 1 – Participant characteristics at time of OGTT eligibility

Table Variable	**dataset.variable**
Child's sex	m_144_kdriscoll_niddk_30apr2017.sex
Child is a member of a minority group	m_144_kdriscoll_niddk_30apr2017.minority
Mother's age	m_144_kdriscoll_niddk_30apr2017.mom_age
Mother's marital status	m_144_kdriscoll_niddk_30apr2017.two_parent_household
Mother's education	m_144_kdriscoll_niddk_30apr2017.education_mom_group3
Crowding	m_144_kdriscoll_niddk_30apr2017.crowdingspcae
Country	m_144_kdriscoll_niddk_30apr2017.country
Long distance protocol	m_144_kdriscoll_niddk_30apr2017.ldpflag
Maternal study satisfaction	m_144_kdriscoll_niddk_30apr2017.dissatisfied_score_m
First degree relative with T1D	m_144_kdriscoll_niddk_30apr2017.fdr
Child age at second autoantibody positivity	m_144_kdriscoll_niddk_30apr2017.cmp_dys
Child's A1C score	m_144_kdriscoll_niddk_30apr2017.last_hba1c
Child's BMI z-score	m_144_kdriscoll_niddk_30apr2017.bmiz
Mother's smoking status	m_144_kdriscoll_niddk_30apr2017.MotherFemalseCaretakerSmoke
Actions taken to prevent T1D	m_144_kdriscoll_niddk_30apr2017.prevent_count
Monitoring child for T1D	m_144_kdriscoll_niddk_30apr2017.monitor_tm01 m_144_kdriscoll_niddk_30apr2017.monitor_tm02_05 m_144_kdriscoll_niddk_30apr2017.monitor_tm10

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

Characteristic	Publication: N (% or M ± SD (min, max) (n=437)	DSIC: N (% or M ± SD (min, max) (n=439)	Diff. (n=2)
Child's sex			
Female	180 (41%)	182 (41.5%)	2 (0.5)
Male	257 (59%)	257 (58.5%)	0 (0.5)
Child is a member of a minority group			
Yes	31 (7%)	33 (8%)	2 (1)
No	397 (93%)	397 (92%)	0 (1)
Mother's age	31.4 ± 4.9 (18, 45)	31.4 ± 4.9 (18, 45)	0 ± 0 (0, 0)
Mother's marital status			
Married or living together	411 (95%)	413 (94.5%)	2 (0.5)
Single	23 (5%)	23 (5%)	0 (0)
Mother's education			
Primary or high school	64 (15%)	65 (15%)	1 (0)
Trade school or some college	95 (22%)	95 (22%)	0 (0)
College	272 (63%)	273 (63%)	1 (0)
Crowding	0.7 ± 0.3 (0.3, 3.0)	0.7 ± 0.3 (0.3, 3.0)	0 ± 0 (0, 0)
Country			
United States	152 (35%)	153 (35%)	1 (0)
Finland	106 (24%)	106 (24%)	0 (0)
Germany	31 (7%)	32 (7%)	1 (0)
Sweden	148 (34%)	148 (34%)	0 (0)
Long distance protocol			
Yes	42 (10%)	42 (10%)	0 (0)
No	395 (90%)	397 (90%)	2 (0)
Maternal study satisfaction	1.4 ± 1.7 (0, 6)	1.4 ± 1.7 (0, 6)	0 ± 0 (0, 0)
First degree relative with T1D			
Yes	94 (22%)	95 (22%)	1 (0)
No	343 (78%)	344 (78%)	1 (0)
Child age at second autoantibody positivity	4.0 ± 2.5	4.0 ± 2.5	0 ± 0
Child's A1C score	5.3 ± 0.3 (4.6, 6.2)	5.3 ± 0.3 (4.6, 6.2)	0 ± 0 (0, 0)
Child's BMI z-score	0.2 ± 1.0 (-2.5, 2.6)	0.2 ± 1.0 (-2.5, 2.6)	0 ± 0 (0, 0)
Mother's smoking status			
Yes	31 (7%)	31 (7%)	0 (0)
No	403 (93%)	406 (93%)	3 (0)
Actions taken to prevent T1D			
Yes	184 (42%)	185 (42%)	1 (0)
No	252 (58%)	253 (58%)	1 (0)
Monitoring child for T1D			
Participation in TEDDY	201 (46%)	202 (46%)	1 (0)
Glucose checking	78 (18%)	78 (18%)	0 (0)
Other	44 (10%)	44 (10%)	0 (0)

Attachment A: SAS Code

```
libname m144 "X:\NIDDK\niddk-dr_studies6\TEDDY\private_created_data\M144";

/*******************/
/* TEDDY M144 DSIC */
/* Driscoll et al */
/*******************/

*temp dataset;
data m144; set m144.m_144_kdriscoll_niddk_30apr2017;
run;

*Table 1;
proc contents data=m144;
run;

proc freq data=m144;
tables cmp_visit;
run;

proc sql;
select count(distinct maskid) as distinct_var1
from m144;
quit;

*demographics;
data dem; set m144;
run;

proc sort data=dem nodupkey;
by maskid;
run;

*sex;
proc freq data=dem;
tables Sex;
run;

*Minority;
proc freq data=dem;
tables minority;
run;

*Mother's age;
proc means data=dem n mean std min max;
var mom_age;
```

```
run;

*Mother's marital status;
proc freq data=dem;
tables two_parent_household;
run;

*Mother's education;
proc freq data=dem;
tables education_mom_group3;
run;

*crowding;
proc means data=dem n mean std min max;
var crowdingspace;
run;

*Teddy specific study variables;

*Country;
proc freq data=dem;
tables country;
run;

*long distance protocol;
proc freq data=dem ;
tables ldp_flag;
run;

*satisfaction;
proc means data=dem n mean std min max;
var dissatisfied_score_m;
run;

*Child medical variables;

*FDR;
proc freq data=dem;
tables fdr;
run;

*child age at second autoantibody positivity;
data dem_1; set dem;
cmp_age_yr = (cmp_dys/365.25);
run;

proc means data=dem_1 n mean std min max;
var cmp_age_yr;
```

```
run;

*A1C;
proc means data=dem n mean std min max;
var last_hba1c;
run;

*BMI z-score;
proc means data=dem n mean std min max;
var bmiz;
run;

*Maternal psychosocial and lifestyle variables;

*smoking status;
proc freq data=dem;
tables MotherFemaleCaretakerSmoke;
run;

*actions taken to prevent;
proc freq data=dem;
tables DoneAnythingStopDiabetes prevent_count;
run;

*monitoring child for t1d;
proc freq data=dem;
tables monitor_tm01 monitor_tm02_05 monitor_tm10 monitor_code_other;
run;
```