Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M123 Aronsson

Prepared by NIDDK-CR
August 16, 2021

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 M123 nested case-control study sought to understand the relationship between 25(OH)D concentrations during early infancy and risk of celiac disease autoimmunity in genetically predisposed children. The manuscript for this study is pending publication.

3 Archived Datasets

All 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_123_carronson_niddk_30jun2019.sas7bdat" dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data found in the manuscript by Aronsson et al. [1] that is pending publication. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the manuscript [1], Descriptive characteristics of subjects in the nested case-control study, 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 match the results to be published.

6 Conclusions

The NIDDK Central Repository is confident that the TEDDY M123 data files to be distributed are a true copy of the study data.

7 References

[1] Aronson CA, Liu X, Norris JM, Uusitalo U, Butterworth MD, Koletzko S, Virtanen SM, Erlund I, Kurppa K, Hagopian W, Rewers MJ, She J, Toppari J, Ziegler AG, Akolkar B, Krischer JP, Agardh D. Associations of 25(OH)D levels in infancy with celiac disease autoimmunity in at-risk children (in press).

Table A: Variables used to replicate Table 1 – Descriptive characteristics of subjects in the nested case-control study

Table Variable	**dataset.variable**
Female sex (yes)	m_123_carronson_niddk_30jun2019.female
Clinical Center	m_123_carronson_niddk_30jun2019.cc
HLA genotype	m_123_carronson_niddk_30jun2019.hla_dr3
Long distance protocol (yes)	m_123_carronson_niddk_30jun2019.subject_ldp
FDR with type 1 diabetes (yes)	m_123_carronson_niddk_30jun2019.fdr
FDR with celiac disease (yes)	m_123_carronson_niddk_30jun2019.celiac_fdr
Age at CDA (years)	m_123_carronson_niddk_30jun2019.case_age_yr
Developed celiac disease during follow-up (yes)	m_123_carronson_niddk_30jun2019.cd
Persistent confirmed islet autoantibodies (yes)	m_123_carronson_niddk_30jun2019.ia
Islet autoantibody positivity prior to CDA (yes)	m_123_carronson_niddk_30jun2019.ia_bf_cda
Season of birth	m_123_carronson_niddk_30jun2019.birth_season
Breastfeeding duration (months)	m_123_carronson_niddk_30jun2019.brstfed_dur_mo m_123_carronson_niddk_30jun2019.exclbr_dur_mo
Age at gluten introduction (months)	m_123_carronson_niddk_30jun2019.gluten_cereals_mo
Maternal education	m_123_carronson_niddk_30jun2019.education_mom_group1
Maternal vitamin D supplementation during pregnancy (yes)	m_123_carronson_niddk_30jun2019.mother_vitD

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

Variable	Manuscript Cases (n=281)	DSIC Cases (n=281)	Diff. (n=0)	Manuscript Controls (n=643)	DSIC Controls (n=643)	Diff. (n=0)
N (%) or median (Q1, Q3)						
Female sex (yes)	144 (51.2)	144 (51.2)	0 (0)	320 (49.8)	320 (49.8)	0 (0)
Clinical Center						
Finland	64 (22.8)	64 (22.8)	0 (0)	147 (22.9)	147 (22.9)	0 (0)
Germany	27 (9.6)	27 (9.6)	0 (0)	52 (8.1)	52 (8.1)	0 (0)
Sweden	116 (41.3)	116 (41.3)	0 (0)	275 (42.8)	275 (42.8)	0 (0)
Colorado	36 (12.8)	36 (12.8)	0 (0)	83 (12.9)	83 (12.9)	0 (0)
Washington	16 (5.7)	16 (5.7)	0 (0)	43 (6.7)	43 (6.7)	0 (0)
Georgia	22 (7.8)	22 (7.8)	0 (0)	43 (6.7)	43 (6.7)	0 (0)
HLA genotype						
DR3/3	105 (37.2)	105 (37.4)	0 (0.2)	90 (14.0)	90 (14.0)	0 (0)
DR3/X	121 (43.1)	121 (43.1)	0 (0)	261 (40.6)	261 (40.6)	0 (0)
Other	55 (19.6)	55 (19.6)	0 (0)	292 (45.4)	292 (45.4)	0 (0)
Long distance protocol (yes)	43 (15.3)	43 (15.3)	0 (0)	108 (16.8)	108 (16.8)	0 (0)
FDR with type 1 diabetes (yes)	68 (24.2)	68 (24.2)	0 (0)	152 (23.6)	152 (23.6)	0 (0)
FDR with celiac disease (yes)	27 (9.6)	27 (9.6)	0 (0)	26 (4.0)	26 (4.0)	0 (0)
Age at CDA (years)	2.8 (2.0, 3.9)	2.8 (2.0, 3.9)	0 (0)	NA	NA	NA
Developed celiac disease during follow-up (yes)	102 (36.3)	102 (36.3)	0 (0)	NA	NA	NA
Persistent confirmed islet autoantibodies (yes)	83 (29.5)	83 (29.5)	0 (0)	146 (22.7)	146 (22.7)	0 (0)
Islet autoantibody positivity prior to CDA (yes)	55 (19.6)	55 (19.6)	0 (0)	NA	NA	NA
Season of birth						
Spring (Mar-May)	75 (26.7)	75 (26.7)	0 (0)	149 (23.2)	149 (23.2)	0 (0)
Summer (Jun-Aug)	68 (24.2)	68 (24.2)	0 (0)	171 (26.6)	171 (26.6)	0 (0)
Fall (Sep-Nov)	61 (21.7)	61 (21.7)	0 (0)	173 (26.9)	173 (26.9)	0 (0)
Winter (Dec-Feb)	77 (27.4)	77 (27.4)	0 (0)	150 (23.3)	150 (23.3)	0 (0)
Breastfeeding duration (months)						
Exclusive	0.9 (0.0, 4.0)	0.9 (0.0, 4.0)	0 (0)	0.5 (0.0, 3.2)	0.5 (0.0, 3.2)	0 (0)
Any	8.3 (5.4, 12.1)	8.3 (5.4, 12.1)	0 (0)	8.0 (3.8, 12.0)	8.0 (3.8, 12.0)	0 (0)
Age at gluten introduction (months)	6.0 (5.0, 6.9)	6.0 (5.1, 6.9)	0 (0.1, 0.0)	6.0 (5.0, 6.9)	6.0 (5.1, 6.9)	0 (0.1, 0.0)
Maternal education						
Basic primary	57 (20.3)	57 (20.3)	0 (0)	146 (22.9)	146 (22.9)	0 (0)
Higher education	224 (79.7)	224 (79.7)	0 (0)	492 (77.1)	492 (77.1)	0 (0)
Maternal vitamin D supplementation during pregnancy (yes)	181 (64.4)	181 (64.4)	0 (0)	406 (63.1)	406 (63.1)	0 (0)

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-
dr_studies6\TEDDY\private_orig_data\M_123_CArronson_NIDDK_Submission";

/*****************************/
/*      Calling Dataset      */
/*****************************/

data one; set dsic.m_123_carronson_niddk_30jun2019;
run;

proc contents data=one;
run;

/*****************************/
/*   Replication of Table 1   */
/*****************************/

*Female sex;
proc freq data=one;
tables female*time/ norow nopercent;
run;

*Clinical Center;
proc freq data=one;
tables cc*time/ norow nopercent;
run;

*HLA Genotype;
proc freq data=one;
tables hla_dr3*time/ norow nopercent;
run;

*Long Distance Protocol;
proc freq data=one;
tables subject_ldp*time/ norow nopercent;
run;

*FDR with T1D;
proc freq data=one;
tables fdr*time/ norow nopercent;
run;

*FDR with celiac disease;
proc freq data=one;
tables celiac_fdr*time/ norow nopercent;
run;

*Age at CDA;
```

```
proc means data=one median q1 q3;
var case_age_yr;
where time = 1;
run;

*Developed celiac disease during follow-up;
proc freq data=one;
tables CD*time/ norow nopercent;
run;

*Persistent confirmed islet autoantibodies;
proc freq data=one;
tables ia*time/ norow nopercent;
run;

*Iselt autoantibody positivity prior to CDA;
proc freq data=one;
tables ia_bf_cda*time/ norow nopercent;
run;

*Season of birth;
proc freq data=one;
tables birth_season*time/ norow nopercent;
run;

*Breastfeeding duration;
proc means data=one median q1 q3;
var brstfed_dur_mo exclbf_dur_mo;
class time;
run;

*Age at gluten introduction;
proc means data=one median q1 q3;
var gluten_cereals_mo;
class time;
run;

*maternal education;
proc freq data=one;
tables education_mom_group1*time/ norow nopercent;
run;

*Maternal vitamin D supplementation during pregnancy;
proc freq data=one;
tables mother_vitD*time/ norow nopercent;
run;
```