Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M107 Johnson

Prepared by NIDDK-CR
February 21, 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 M107 study sought to investigate the association between maternal consumption of gluten-containing foods and other selected foods during late pregnancy, and offspring risk of islet autoimmunity (IA) and T1D.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in the TEDDY M107 folder in the data package. For this replication, variables were taken from the "m_107_rjohnson_niddk_28feb2019.sas7bdat" dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Maternal and participant characteristics for children participating in the TEDDY 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 are an exact match to the published results.

6 Conclusions

The NIDDK Central Repository is confident that the TEDDY M107 data files to be distributed are a true copy of the study data.

7 References

[1] Johnson RK, Tamura R, Frank N, Uusitalo U, Yang J, Niinistö S, Aronsson CA, Ziegler AG, Hagopian W, Rewers M, Toppari J, Akolkar B, Krischer J, Virtanen SM, Norris JM. Maternal Food Consumption during Late Pregnancy and Offspring Risk of Islet Autoimmunity and Type 1 Diabetes. Diabetologia, 64(7), 1604-1612, July 2021. doi: https://doi.org/10.1007/s00125-021-05446-y

Table A: Variables used to replicate Table 1 – Maternal and participant characteristics for children participating in the TEDDY study

Table Variable	**dataset.variable**
Birth year	m_107_rjohnson_niddk_28feb2019.birth_year
Sex	m_107_rjohnson_niddk_28feb2019.sex
FDR with type 1 diabetes	m_107_rjohnson_niddk_28feb2019.fdr
HLA genotype	m_107_rjohnson_niddk_28feb2019.hla_category
Country	m_107_rjohnson_niddk_28feb2019.country
Consuming probiotics before 4 weeks	m_107_rjohnson_niddk_28feb2019.probiotics_within_4_weeks
Age at delivery	m_107_rjohnson_niddk_28feb2019.mom_age_birth
Education	m_107_rjohnson_niddk_28feb2019.education_mom_group3
Smoking during pregnancy	m_107_rjohnson_niddk_28feb2019.any_smoking
Alcohol during pregnancy	m_107_rjohnson_niddk_28feb2019.alcoholfreq1trimester m_107_rjohnson_niddk_28feb2019.alcoholfreq2trimester m_107_rjohnson_niddk_28feb2019.alcoholfreq3trimester

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

Characteristics	Publication: All participants (n=8556)	DSIC: All participants (n=8556)	Diff. (n=0)
Child characteristics			
Birth year			
2004-2005	1415 (17)	1415 (17)	0 (0)
2006	1516 (18)	1516 (18)	0 (0)
2007	1828 (21)	1828 (21)	0 (0)
2008	1726 (20)	1726 (20)	0 (0)
2009-2010	2071 (24)	2071 (24)	0 (0)
Sex (male)	4330 (51)	4330 (51)	0 (0)
FDR with type 1 diabetes	951 (11)	951 (11)	0 (0)
HLA genotype			
HLA-DR3/4	3339 (39)	3339 (39)	0 (0)
HLA-DR4/4	1674 (20)	1674 (20)	0 (0)
HLA-DR4/8	1474 (17)	1474 (17)	0 (0)
HLA-DR3/3	1791 (21)	1791 (21)	0 (0)
Others	278 (3)	278 (3)	0 (0)
Country			
USA	3661 (43)	3661 (43)	0 (0)
Finland	1808 (21)	1808 (21)	0 (0)
Germany	582 (7)	582 (7)	0 (0)
Sweden	2505 (29)	2505 (29)	0 (0)
Consuming probiotics before 4 weeks			
Yes	821 (10)	821 (10)	0 (0)
No	7735 (90)	7735 (90)	0 (0)
Maternal characteristics			
Age at delivery, median years (IQR)	30 (27, 34)	30 (27, 34)	0 (0, 0)
Education			
High school qualification or less	1575 (18)	1575 (18)	0 (0)
Vocational school or some university	1999 (23)	1999 (23)	0 (0)
University degree or more	4118 (48)	4118 (48)	0 (0)
Missing	864 (10)	864 (10)	0 (0)
Smoking during pregnancy			
No	7246 (85)	7246 (85)	0 (0)
Yes	1150 (13)	1150 (13)	0 (0)
Missing	160 (2)	160 (2)	0 (0)
Alcohol during pregnancy			
No	5546 (65)	5546 (65)	0 (0)
Yes	2851 (33)	2851 (33)	0 (0)
Missing	159 (2)	159 (2)	0 (0)

Attachment A: SAS Code

```
libname m107 "X:\NIDDK\niddk-dr_studies6\TEDDY\private_created_data\M107";

/************************/
/* M107 Johnson DSIC */
/************************/

proc freq data=m107.m_107_rjohnson_niddk_28feb2019;
tables birth_year Sex fdr hla_category country probiotics_within_4_weeks education_mom_group3
any_smoking/missing;
where hla_category ^= 0;
run;

proc means data=m107.m_107_rjohnson_niddk_28feb2019 n median q1 q3;
var mom_age_birth;
where hla_category ^= 0;
run;

data one; set m107.m_107_rjohnson_niddk_28feb2019;
any_alc = 0;
if AlcoholFreq1Trimester ^= "Not at all" OR AlcoholFreq2Trimester ^= "Not at all" OR
AlcoholFreq3Trimester ^= "Not at all"
        then any_alc = 1;
if AlcoholFreq1Trimester = "" OR AlcoholFreq2Trimester = "" OR AlcoholFreq3Trimester = ""
        then any_alc = .;
run;

proc freq data=one;
tables any_alc/missing;
where hla_category ^= 0;
run;
```