Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M190 Hummel

Prepared by NIDDK-CR
September 16, 2022

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. Data are 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 M190 study sought to assess whether the duration of breastfeeding is associated with the development of early childhood autoimmunity, allergies, or obesity in the TEDDY cohort.

3 Archived Datasets

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_190_SHummel_NIDDK_31DEC2017.sas7bdat" dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Table 1 – Description of the study population, 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 publication.

6 Conclusions

The NIDDK Central Repository is confident that the TEDDY M190 data files to be distributed are a true copy of the study data.

7 References

[1] Hummel S, Weiß A, Bonifacio E, Agardh D, Akolkar B, Aronsson CA, Hagopian WA, Koletzko S, Krischer JP, Lernmark Å, Lynch K, Norris JM, Rewers MJ, She JX, Toppari J, Uusitalo U, Vehik K, Virtanen SM, Beyerlein A, Ziegler AG. Associations of Breastfeeding with Childhood Autoimmunity, Allergies, and Overweight: The Environmental Determinants of Diabetes in the Young (TEDDY) Study. The American Journal of Clinical Nutrition, 114(1), 134-142, July 2021. doi: https://doi.org/10.1093/ajcn/nqab065

Table A: Variables used to replicate Table 1 – Description of the study population

Table Variable	**dataset.variable**
Duration of any breastfeeding (months)	M_190_SHummel_NIDDK_31DEC2017.any_bf_binary
Duration of exclusive breastfeeding (months)	M_190_SHummel_NIDDK_31DEC2017.excl_bf_binary
Females	M_190_SHummel_NIDDK_31DEC2017.sex
Country	M_190_SHummel_NIDDK_31DEC2017.country
Maternal age at delivery (years)	M_190_SHummel_NIDDK_31DEC2017.maternal_age
Maternal pre-pregnancy BMI (kg/m^2)	M_190_SHummel_NIDDK_31DEC2017.maternal_BMI
Excessive weight gain during pregnancy (yes)	M_190_SHummel_NIDDK_31DEC2017.excessive_weight_gain
Maternal diabetes during pregnancy (yes)	M_190_SHummel_NIDDK_31DEC2017.maternal_diabetes
Maternal smoking during pregnancy (yes)	M_190_SHummel_NIDDK_31DEC2017.smoker
High maternal education (yes)	M_190_SHummel_NIDDK_31DEC2017.education_mom_group2
Preterm delivery (gestational age < 37 weeks)	M_190_SHummel_NIDDK_31DEC2017.preterm
Birth weight (grams)	M_190_SHummel_NIDDK_31DEC2017.babysweightgrams
Cesarean delivery (yes)	M_190_SHummel_NIDDK_31DEC2017.sectio
Firstborn child (yes)	M_190_SHummel_NIDDK_31DEC2017.mom_first_child

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

Variable	**Pub: n (out of 8615)**	**DSIC: n (out of 8615)**	**Diff. (n=0)**	**Publication: n (%)/ median (IQR)**	**DSIC: n (%)/ median (IQR)**	**Diff.**
Duration of any breastfeeding	8576	8576	0			
≤ 6 months				3982 (46.4%)	3982 (46.4%)	0 (0)
> 6 months				4594 (53.6%)	4594 (53.6%)	0 (0)
Duration of exclusive breastfeeding	8611	8611	0			
≤ 3 months				6699 (77.8%)	6699 (77.8%)	0 (0)
> 3 months				1912 (22.2%)	1912 (22.2%)	0 (0)
Females	8615	8615	0	4255 (49.4 %)	4255 (49.4%)	0 (0)
Country	8615	8615	0			
USA				3694 (42.9%)	3694 (42.9%)	0 (0)
Finland				1825 (21.2%)	1825 (21.2%)	0 (0)
Germany				580 (6.7%)	580 (6.7%)	0 (0)
Sweden				2516 (29.2%)	2516 (29.2%)	0 (0)
Maternal age at delivery (years)	8615	8615	0	30 (27-34)	30 (27-34)	0 (0)
Maternal pre-pregnancy BMI (kg/m²)	8372	8372	0	23.6 (21.3-27.0)	23.6 (21.3-27.0)	0 (0)
Excessive weight gain during pregnancy (yes)	8300	8300	0	3985 (48.0%)	3985 (48.0%)	0 (0)
Maternal diabetes during pregnancy (yes)	8252	8252	0	836 (10.1%)	836 (10.1%)	0 (0)
Maternal smoking during pregnancy (yes)	8373	8373	0	1068 (12.8%)	1068 (12.8%)	0 (0)
High maternal education (yes)	7501	7501	0	4110 (54.8%)	4110 (54.8%)	0 (0)
Preterm delivery (gestational age < 37 weeks)	8601	8601	0	477 (5.5%)	477 (5.5%)	0 (0)
Birth weight (grams)	8381	8381	0	3500 (3156-3840)	3500 (3156-3840)	0 (0)
Cesarean delivery (yes)	8610	8610	0	2235 (26.0%)	2235 (26.0%)	0 (0)
Firstborn child (yes)	7455	7455	0	3335 (44.7%)	3335 (44.7%)	0 (0)

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-
dr_studies6\TEDDY\private_orig_data\M_190_SHummel_NIDDK_Submission";

/**************************/
/* DSIC for Hummel et al. */
/**************************/

*temp datasets;
data dsic; set dsic.m_190_shummel_niddk_31dec2017;
run;

proc contents data=dsic;
run;

**duration of any breastfeeding;
proc freq data=dsic;
tables any_bf_binary;
run;

*duration of exclusive breastfeeding;
proc freq data=dsic;
tables excl_bf_binary;
run;

*Females;
proc freq data=dsic;
tables sex;
 run ;

 *Country;
 proc freq data=dsic;
 tables Country;
 run;

 *maternal age at delivery;
 proc means data=dsic n median q1 q3;
 var Maternal_Age;
 run;

 *maternal pre-pregnancy BMI;
 proc means data=dsic n median q1 q3;
 var maternal_BMI;
 run;
```

```
*excessive weight gain during pregnancy;
proc freq data=dsic;
tables excessive_weight_gain;
where excessive_weight_gain ^= "NA";
run;

*maternal diabetes during pregnancy;
proc freq data=dsic;
tables maternal_diabetes;
run;

*maternal smoking during pregnancy;
proc freq data=dsic;
tables Smoker/missing;
run ;

*high maternal education;
proc freq data=dsic;
tables Education_Mom_Group2;
run;

*Preterm delivery (gestational age <37 weeks);
proc freq data=dsic;
tables preterm;
run;

*BIRTH WEIGHT;
proc means data=dsic n median q1 q3;
var Babysweightgrams;
run;

*Cesarean section;
proc freq data=dsic ;
tables sectio;
where sectio ^= "NA";
run;

*First-born child;
proc freq data=dsic;
tables Mom_First_Child;
run;
```