

Dataset Integrity Check for The  
Environmental Determinants of Diabetes  
in the Young (TEDDY) M238 Webb-  
Robertson

Prepared by NIDDK-CR  
March 7, 2022

## Contents

|   |   |
|---|---|
| 1 Standard Disclaimer .....   | 2 |
| 2 Study Background .....  | 2 |
| 3 Archived Datasets .....   | 2 |
| 4 Statistical Methods .....   | 2 |
| 5 Results .....   | 3 |
| 6 Conclusions .....   | 3 |
| 7 References .....  | 3 |
| Table A: Variables used to replicate Table 1 – Characteristics of TEDDY subjects categorized for machine learning based on T1D outcomes at 6 years of age ..... | 4 |
| Table B: Comparison of values computed in integrity check to reference article Table 1 values .....   | 5 |
| Attachment A: SAS Code .....  | 6 |

## 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 manuscript.

## 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 M238 sub study sought to understand if certain genetic, immunologic, and metabolic characteristics, measured at infancy, could be used to predict development of T1D by 6 years of age. 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\_238\_BWebbRobertson\_NIDDK\_30JUN.sas7bdat” dataset.

## 4 Statistical Methods

Analyses were performed to duplicate results for the data in the manuscript by Webb-Robertson 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], Characteristics of TEDDY subjects categorized for machine learning based on T1D outcome at 6 years of age, 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 M238 data files to be distributed are a true copy of the study data.

## 7 References

[1] Webb-Robertson BM, Nakayasu ES, Frohnert BI, Bramer LM, Akers SM, Norris JM, Vehik K, Ziegler AG, Metz TO, Rich SS, Rewers MJ. Integration of Infant Metabolite, Genetic and Islet Autoimmunity Signatures to Predict Type 1 Diabetes by 6 Years of Age (in press).

**Table A:** Variables used to replicate Table 1 – Characteristics of TEDDY subjects categorized for machine learning based on T1D outcomes at 6 years of age

| <b>Table Variable</b>                              | <b>dataset.variable</b>   |
|--|---|
| Number of subjects                                 | m_238_bwebbrobertson_niddk_30jun.outcome                          |
| Female   | m_238_bwebbrobertson_niddk_30jun.sex                              |
| T1D first degree relative                          | m_238_bwebbrobertson_niddk_30jun.fdr                              |
| T1D first degree relative is Mother                | m_238_bwebbrobertson_niddk_30jun.fdr_mother                       |
| T1D first degree relative is Father                | m_238_bwebbrobertson_niddk_30jun.fdr_father                       |
| Gestational age (weeks)                            | m_238_bwebbrobertson_niddk_30jun.gestational_age                  |
| Length at birth                                    | m_238_bwebbrobertson_niddk_30jun.babyslengthcm                    |
| Length at 3 months                                 | m_238_bwebbrobertson_niddk_30jun.height_cm_3m                     |
| Length at 6 months                                 | m_238_bwebbrobertson_niddk_30jun.height_cm_6m                     |
| Length at 9 months                                 | m_238_bwebbrobertson_niddk_30jun.height_cm_9m                     |
| Growth 3 to 6 months                               | m_238_bwebbrobertson_niddk_30jun.height_cm_6m_Height_cm_3m        |
| Growth 3 to 9 months                               | m_238_bwebbrobertson_niddk_30jun.height_cm_9m_height_cm_3m        |
| Growth 6 to 9 months                               | m_238_bwebbrobertson_niddk_30jun.height_cm_9m_height_cm_6m        |
| Weight at birth                                    | m_238_bwebbrobertson_niddk_30jun.babysweightkgrams                |
| Weight at 3 months                                 | m_238_bwebbrobertson_niddk_30jun.weight_kg_3m                     |
| Weight at 6 months                                 | m_238_bwebbrobertson_niddk_30jun.weight_kg_6m                     |
| Weight at 9 months                                 | m_238_bwebbrobertson_niddk_30jun.weight_kg_9m                     |
| Weight gain 3 to 6 months                          | m_238_bwebbrobertson_niddk_30jun.weight_kg_6m_weight_kg_3m        |
| Weight gain 3 to 9 months                          | m_238_bwebbrobertson_niddk_30jun.weight_kg_9m_weight_kg_3m        |
| Weight gain 6 to 9 months                          | m_238_bwebbrobertson_niddk_30jun.weight_kg_9m_weight_kg_6m        |
| Formula (cows milk) before 28 days                 | m_238_bwebbrobertson_niddk_30jun.formula_cows_milk_28days         |
| Formula (cows milk) before 6 months                | m_238_bwebbrobertson_niddk_30jun.formula_cows_milk_6m             |
| Formula (pre- or pro-biotic) before 28 days        | m_238_bwebbrobertson_niddk_30jun.formula_probiotic_or_prebiotic_2 |
| Formula (pre- or pro-biotic) before 6 months       | m_238_bwebbrobertson_niddk_30jun.formula_probiotic_or_prebiotic_6 |
| Gluten before 6 months                             | m_238_bwebbrobertson_niddk_30jun.gluten_6m                        |
| GRS  | m_238_bwebbrobertson_niddk_30jun.TEDDY_GRS_without_DR_DQ          |
| DR3/4  | m_238_bwebbrobertson_niddk_30jun.hla_dr3_4_                       |
| DR4/4  | m_238_bwebbrobertson_niddk_30jun.hla_dr4_4_                       |
| DR3/3  | m_238_bwebbrobertson_niddk_30jun.hla_dr3_3_                       |
| DR4/8  | m_238_bwebbrobertson_niddk_30jun.hla_dr4_8_                       |
| Other  | m_238_bwebbrobertson_niddk_30jun.hla_other_                       |
| Persistent islet autoantibody positive at 9 months | m_238_bwebbrobertson_niddk_30jun.persist_conf_ab_visit_9m         |

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

| Variable   | Positive T1D at 6 Years – Manuscript (n=73) | Positive T1D at 6 Years – DSIC (n=73) | Diff. (n=0) | Negative T1D at 6 Years – Manuscript (n=582) | Negative T1D at 6 Years – DSIC (n=582) | Diff. (n=0) |
|--|---|---------------------------------------|-------------|--|--|-------------|
| Female   | 40 (54.8%)                                  | 40 (54.8%)                            | 0 (0)       | 263 (45.2%)                                  | 263 (45.2%)                            | 0 (0)       |
| T1D first degree relative                          | 24 (32.9%)                                  | 24 (32.9%)                            | 0 (0)       | 124 (21.3%)                                  | 124 (21.3%)                            | 0 (0)       |
| T1D first degree relative is Mother                | 5 (6.8%)                                    | 5 (6.8%)                              | 0 (0)       | 41 (7.0%)                                    | 41 (7.0%)                              | 0 (0)       |
| T1D first degree relative is Father                | 14 (19.2%)                                  | 14 (19.2%)                            | 0 (0)       | 64 (11.0%)                                   | 64 (11.0%)                             | 0 (0)       |
| Gestational Age (wks)                              | 39.2  | 39.3                                  | 0.1         | 39.5   | 39.5                                   | 0           |
| Length at birth                                    | 50.8  | 50.8                                  | 0           | 50.7   | 50.7                                   | 0           |
| Length at 3M                                       | 62.8  | 62.8                                  | 0           | 62.5   | 62.5                                   | 0           |
| Length at 6M                                       | 68.4  | 68.4                                  | 0           | 68.1   | 68.1                                   | 0           |
| Length at 9M                                       | 72.8  | 72.8                                  | 0           | 72.6   | 72.6                                   | 0           |
| Growth 3 to 6M                                     | 5.6   | 5.6                                   | 0           | 5.6  | 5.6                                    | 0           |
| Growth 3 to 9 M                                    | 10.0  | 10.0                                  | 0           | 10.0   | 10.0                                   | 0           |
| Growth 6 to 9 M                                    | 4.4   | 4.4                                   | 0           | 4.4  | 4.4                                    | 0           |
| Weight at birth                                    | 3.6   | 3.6                                   | 0           | 3.5  | 3.5                                    | 0           |
| Weight at 3M                                       | 6.7   | 6.7                                   | 0           | 6.5  | 6.5                                    | 0           |
| Weight at 6M                                       | 8.2   | 8.3                                   | 0.1         | 8.1  | 8.1                                    | 0           |
| Weight at 9M                                       | 9.4   | 9.4                                   | 0           | 9.1  | 9.1                                    | 0           |
| Weight gain 3 to 6 M                               | 1.6   | 1.6                                   | 0           | 1.5  | 1.5                                    | 0           |
| Weight gain 3 to 9 M                               | 2.7   | 2.7                                   | 0           | 2.6  | 2.6                                    | 0           |
| Weight gain 6 to 9 M                               | 1.1   | 1.1                                   | 0           | 1.1  | 1.1                                    | 0           |
| Formula (cows milk) before 28 days                 | 33 (42.2%)                                  | 33 (45.2%)                            | 0 (3.0)     | 272 (46.7%)                                  | 272 (46.7%)                            | 0 (0)       |
| Formula (cows milk) before 6 months                | 48 (65.8%)                                  | 48 (65.7%)                            | 0 (0.1)     | 451 (77.5%)                                  | 451 (77.5%)                            | 0 (0)       |
| Formula (pre- or pro-biotic) before 28 days        | 16 (21.9%)                                  | 16 (21.9%)                            | 0 (0)       | 120 (20.6%)                                  | 120 (20.6%)                            | 0 (0)       |
| Formula (pre- or pro-biotic) before 6 months       | 25 (34.2%)                                  | 25 (34.2%)                            | 0 (0)       | 203 (34.9%)                                  | 203 (34.9%)                            | 0 (0)       |
| Gluten before 6 months                             | 30 (41.1%)                                  | 30 (41.1%)                            | 0 (0)       | 265 (45.5%)                                  | 265 (45.5%)                            | 0 (0)       |
| GRS  | 10.5  | 10.5                                  | 0           | 10.2   | 10.2                                   | 0           |
| DR3/4  | 43 (58.9%)                                  | 43 (58.9%)                            | 0 (0)       | 217 (37.3%)                                  | 217 (37.3%)                            | 0 (0)       |
| DR4/4  | 8 (11.0%)                                   | 8 (11.0%)                             | 0 (0)       | 117 (20.1%)                                  | 117 (20.1%)                            | 0 (0)       |
| DR3/3  | 8 (11.0%)                                   | 8 (11.0%)                             | 0 (0)       | 90 (15.5%)                                   | 90 (15.5%)                             | 0 (0)       |
| DR4/8  | 8 (11.0%)                                   | 8 (11.0%)                             | 0 (0)       | 106 (18.2%)                                  | 106 (18.2%)                            | 0 (0)       |
| Other  | 6 (8.2%)                                    | 6 (8.2%)                              | 0 (0)       | 52 (8.9%)                                    | 52 (8.9%)                              | 0 (0)       |
| Persistent islet autoantibody positive at 9 months | 23 (31.5%)                                  | 23 (31.5%)                            | 0 (0)       | 11 (1.9%)                                    | 11 (1.9%)                              | 0 (0)       |

## Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies6\TEDDY\private_created_data\M238";
```

```
/******  
/* Dataset Integrity Check (DSIC) */  
/* For TEDDY M_238 BWebbRobertson */  
/******
```

```
*Replicating Table 1 from manuscript;
```

```
*Subjects divided by +T1D at 6 years old = Outcome variable;  
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables Outcome;  
run;
```

```
*Female;  
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables sex*outcome / norow nopercnt;  
run;
```

```
*T1d first degree relative;  
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables fdr*outcome/norow nopercnt;  
run;
```

```
*T1d FDR is Mother;  
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables fdr_mother*outcome/norow nopercnt;  
run;
```

```
*T1d FDR is father;  
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables fdr_father*outcome/norow nopercnt;  
run;
```

```
*gestational age (wks);  
proc sort data=dsic.m_238_bwebbrobertson_niddk_30jun;  
by outcome;  
run;
```

```
proc means data=dsic.m_238_bwebbrobertson_niddk_30jun n mean;  
var Gestational_age;  
by outcome;  
run;
```

```
*length at birth;
```

```
proc means data=dsic.m_238_bwebbrobertson_niddk_30jun n mean;  
var Babyslengthcm;  
by outcome;  
run;
```

\*length at 3m to 9m;

```
proc means data=dsic.m_238_bwebbrobertson_niddk_30jun n mean;  
var Height_cm_3m Height_cm_6m Height_cm_9m;  
by outcome;  
run;
```

\*growth 3 to 6, 3 to 9, and 6 to 9m;

```
proc means data=dsic.m_238_bwebbrobertson_niddk_30jun n mean;  
var Height_cm_6m_Height_cm_3m Height_cm_9m_Height_cm_3m Height_cm_9m_Height_cm_6m;  
by outcome;  
run;
```

\*weight at birth, 3m, 6m, and 9m;

```
proc means data=dsic.m_238_bwebbrobertson_niddk_30jun n mean;  
var Babysweightkgrams Weight_kg_3m Weight_kg_6m Weight_kg_9m;  
by outcome;  
run;
```

\*Weight gain 3 to 6, 3 to 9, and 6 to 9m;

```
proc means data=dsic.m_238_bwebbrobertson_niddk_30jun n mean;  
var Weight_kg_6m_Weight_kg_3m Weight_kg_9m_Weight_kg_3m Weight_kg_9m_Weight_kg_6m;  
by outcome;  
run;
```

\*Formula (cows milk) before 28 days;

```
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables formula_cows_milk_28days*outcome/norow nopercnt;  
run;
```

\*Formula (cows milk) before 6 months;

```
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables formula_cows_milk_6m*outcome/norow nopercnt;  
run;
```

\*Formula (pre- or pro-biotic) before 28 days and 6 months;

```
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables (formula_probiotic_or_prebiotic_2 formula_probiotic_or_prebiotic_6)*outcome/norow  
nopercnt;  
run;
```

\*gluten before 6 months;

```
proc freq data=dsic.m_238_bwebbrobertson_niddk_30jun;  
tables gluten_6m*outcome/norow nopercnt;
```



**run;**

\*GRS;

**proc means** data=dsic.m\_238\_bwebbrobertson\_niddk\_30jun n mean;

var TEDDY\_GRS\_without\_DR\_DQ;

by outcome;

**run;**

\*DR 3/4, 4/4, 3/3, 4/8 and other;

**proc freq** data=dsic.m\_238\_bwebbrobertson\_niddk\_30jun;

tables (hla\_dr3\_4\_hla\_dr4\_4\_hla\_dr3\_3\_hla\_dr4\_8\_HLA\_Other\_)\*outcome/norow nopercnt;

**run;**

\*persistent islet autoantibody;

**proc freq** data=dsic.m\_238\_bwebbrobertson\_niddk\_30jun;

tables Persist\_conf\_ab\_visit\_9M\*outcome/norow nopercnt;

**run;**