

Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M105 Mclver

Prepared by NIDDK-CR
September 17, 2021

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 – Participant characteristics	4
Table B1: Comparison of values computed in integrity check to reference article Table 1 values (Total and U.S.)	5
Table B2: Comparison of values computed in integrity check to reference article Table 1 values (Finland, Germany, and Sweden).....	6
Table B3: Comparison of values computed in integrity check to reference article Table 1 values (Colorado, Georgia/Florida, and Washington).....	7
Attachment A: SAS Code	8

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 M105 study sought to identify differences in physical activity among young children within the U.S. and across countries.

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_105_KMcIver_NIDDK_31JAN2017.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data published by McIver et al. [1] for Cross-Country Comparisons of Physical Activity and Sedentary Behavior among 5-Year-Old Children. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Participant Characteristics, Table A lists the variables that were used in the replication, and Tables B1-B3 compare the results calculated from the archived data files to the results published in Table 1. The results of the replication are within expected variation to the published results. There was a discrepancy found in “Mother’s education” for Colorado, GA/FL, and Washington where the values for this variable were transposed in the publication. This information was sent to the Data Coordinating Center and an erratum will be sent to the journal of the publication.

6 Conclusions

The NIDDK repository is confident that the TEDDY M105 data files to be distributed are a true copy of the study data. While there was a discrepancy noted in “Mother’s education”, the TEDDY Data Coordinating Center confirmed that the issue was an entry error in the manuscript and not in the TEDDY M105 data files to be distributed.

7 References

[1] McIver KL, Pate RR, Dowda M, Johnson SB, Yang J, Butterworth M, Liu X. Cross-Country Comparisons of Physical Activity and Sedentary Behavior among 5-Year-Old Children. *International Journal of Pediatrics*, 7912894, 2020. doi: <https://doi.org/10.1155/2020/7912894>

Table A: Variables used to replicate Table 1 – Participant characteristics

Table Variable	dataset.variable
Accelerometer wear time, minutes/day	m_105_kmciver_niddk_31jan2017.Wear_Time
Age, months; mean (SD)	m_105_kmciver_niddk_31jan2017.age
Body weight, kg; mean (SD)	m_105_kmciver_niddk_31jan2017.Wtkg
BMI, kg/m ² ; mean (SD)	m_105_kmciver_niddk_31jan2017.bmi
Males, <i>N</i> (%)	m_105_kmciver_niddk_31jan2017.sex
Race/ethnicity; <i>N</i> (%)	m_105_kmciver_niddk_31jan2017.Race_Ethnicity
Ethnic minority, yes; <i>N</i> (%)	m_105_kmciver_niddk_31jan2017.Ethnic_Minority
Season; <i>N</i> (%)	m_105_kmciver_niddk_31jan2017.Meas_Season
Mother's education; <i>N</i> (%)	m_105_kmciver_niddk_31jan2017.momed

Table B1: Comparison of values computed in integrity check to reference article Table 1 values (Total and U.S.)

Variable	Publication Total (n=1948)	DSIC Total (n=1948)	Diff. (n=0)	Publication U.S. (n=826)	DSIC U.S. (n=826)	Diff. (n=0)
Accelerometer wear time, minutes/day	1049.5 (160.3)	1049.5 (160.3)	0 (0)	1018.2 (174.8)	1018.2 (174.8)	0 (0)
Age, months; mean (SD)	61.2 (1.5)	61.2 (1.5)	0 (0)	61.2 (1.4)	61.2 (1.4)	0 (0)
Body weight, kg; mean (SD)	19.7 (2.8)	19.7 (2.8)	0 (0)	19.3 (2.7)	19.3 (2.7)	0 (0)
BMI, kg/m ² ; mean (SD)	15.8 (1.4)	15.8 (1.4)	0 (0)	15.7 (1.4)	15.7 (1.4)	0 (0)
Males, <i>N</i> (%)	1019 (51.4)	1019 (51.4)	0 (0)	435 (52.7)	435 (52.7)	0 (0)
Race/ethnicity; <i>N</i> (%)						
Hispanic	--	--	--	--	--	--
White, Non-Hispanic	--	--	--	--	--	--
African American, Non-Hispanic	--	--	--	--	--	--
Other	--	--	--	--	--	--
Missing/unknown	--	--	--	--	--	--
Ethnic minority, yes; <i>N</i> (%)	271 (13.9)	271 (13.9)	0 (0)	211 (26.2)	211 (26.2)	0 (0)
Season; <i>N</i> (%)						
Fall	465 (23.4)	465 (23.4)	0 (0)	173 (20.9)	173 (20.9)	0 (0)
Spring	524 (26.4)	524 (26.4)	0 (0)	208 (25.2)	208 (25.2)	0 (0)
Summer	506 (25.5)	506 (25.5)	0 (0)	234 (28.3)	234 (28.3)	0 (0)
Winter	489 (24.7)	489 (24.7)	0 (0)	211 (25.5)	211 (25.5)	0 (0)
Mother's education; <i>N</i> (%)						
Less than college grad	726 (36.9)	726 (36.9)	0 (0)	272 (33.1)	272 (33.1)	0 (0)
College grad or higher	1243 (63.1)	1243 (63.1)	0 (0)	550 (66.9)	550 (66.9)	0 (0)

Table B2: Comparison of values computed in integrity check to reference article Table 1 values (Finland, Germany, and Sweden)

Variable	Publication Finland (n=368)	DSIC Finland (n=368)	Diff. (n=0)	Publication Germany (n=85)	DSIC Germany (n=85)	Diff. (n=0)	Publication Sweden (n=705)	DSIC Sweden (n=705)	Diff. (n=0)
Accelerometer wear time, minutes/day	1048.2 (153.1)	1048.2 (153.1)	0 (0)	1024.7 (139.0)	1024.7 (139.0)	0 (0)	1089.9 (138.3)	1089.9 (138.3)	0 (0)
Age, months; mean (SD)	61.5 (1.6)	61.5 (1.6)	0 (0)	62.9 (1.8)	62.9 (1.8)	0 (0)	61.0 (1.3)	61.0 (1.3)	0 (0)
Body weight, kg; mean (SD)	19.6 (2.8)	19.6 (2.8)	0 (0)	19.7 (2.6)	19.7 (2.6)	0 (0)	20.1 (2.7)	20.1 (2.7)	0 (0)
BMI, kg/m ² ; mean (SD)	15.9 (1.4)	15.9 (1.4)	0 (0)	15.3 (1.3)	15.3 (1.3)	0 (0)	16.0 (1.4)	16.0 (1.4)	0 (0)
Males, N (%)	186 (50.5)	186 (50.5)	0 (0)	47 (55.3)	47 (55.3)	0 (0)	351 (49.8)	351 (49.8)	0 (0)
Race/ethnicity; N (%)									
Hispanic	--	--	--	--	--	--	--	--	--
White, Non-Hispanic	--	--	--	--	--	--	--	--	--
African American, Non-Hispanic	--	--	--	--	--	--	--	--	--
Other	--	--	--	--	--	--	--	--	--
Missing/unknown	--	--	--	--	--	--	--	--	--
Ethnic minority, yes; N (%)	10 (2.8)	10 (2.8)	0 (0)	13 (15.3)	13 (15.3)	0 (0)	37 (5.3)	37 (5.3)	0 (0)
Season; N (%)									
Fall	94 (25.5)	94 (25.5)	0 (0)	23 (27.1)	23 (27.1)	0 (0)	175 (24.8)	175 (24.8)	0 (0)
Spring	85 (23.1)	85 (23.1)	0 (0)	23 (27.1)	23 (27.1)	0 (0)	208 (29.5)	208 (29.5)	0 (0)
Summer	115 (31.3)	115 (31.3)	0 (0)	17 (20.0)	17 (20.0)	0 (0)	140 (20.0)	140 (19.9)	0 (0.1)
Winter	74 (20.1)	74 (20.1)	0 (0)	22 (25.9)	22 (25.9)	0 (0)	182 (25.8)	182 (25.8)	0 (0)
Mother's education; N (%)									
Less than college grad	113 (31.1)	113 (31.1)	0 (0)	44 (51.8)	44 (51.8)	0 (0)	297 (42.5)	297 (42.5)	0 (0)
College grad or higher	250 (68.9)	250 (68.9)	0 (0)	41 (48.2)	41 (48.2)	0 (0)	402 (57.5)	402 (57.5)	0 (0)

Table B3: Comparison of values computed in integrity check to reference article Table 1 values (Colorado, Georgia/Florida, and Washington)

Variable	Publication Colorado (n=367)	DSIC Colorado (n=367)	Diff. (n=0)	Publication GA/FL (n=125)	DSIC GA/FL (n=125)	Diff. (n=0)	Publication Wash. (n=334)	DSIC Wash. (n=334)	Diff. (n=0)
Accelerometer wear time, minutes/day	1084.3 (146.8)	1084.3 (146.8)	0 (0)	1026.4 (154.1)	1026.4 (154.1)	0 (0)	942.5 (180.6)	942.5 (180.6)	0 (0)
Age, months; mean (SD)	60.7 (1.4)	60.7 (1.4)	0 (0)	61.8 (1.5)	61.8 (1.5)	0 (0)	61.4 (1.4)	61.4 (1.4)	0 (0)
Body weight, kg; mean (SD)	18.9 (2.5)	18.9 (2.5)	0 (0)	19.3 (2.3)	19.3 (2.3)	0 (0)	19.7 (3.1)	19.7 (3.1)	0 (0)
BMI, kg/m ² ; mean (SD)	15.6 (1.3)	15.6 (1.4)	0 (0.1)	15.7 (1.3)	15.7 (1.3)	0 (0)	15.9 (1.5)	15.9 (1.5)	0 (0)
Males, N (%)	202 (55.0)	202 (55.0)	0 (0)	60 (48.0)	60 (48.0)	0 (0)	173 (51.8)	173 (51.8)	0 (0)
Race/ethnicity; N (%)									
Hispanic	103 (28.1)	103 (28.1)	0 (0)	8 (6.4)	8 (6.4)	0 (0)	23 (6.9)	23 (6.9)	0 (0)
White, Non-Hispanic	256 (69.8)	256 (69.8)	0 (0)	107 (85.6)	107 (85.6)	0 (0)	276 (82.6)	276 (82.6)	0 (0)
African American, Non-Hispanic	2 (0.5)	2 (0.5)	0 (0)	4 (3.2)	4 (3.2)	0 (0)	5 (1.5)	5 (1.5)	0 (0)
Other	5 (1.4)	5 (1.4)	0 (0)	4 (3.2)	4 (3.2)	0 (0)	19 (5.7)	19 (5.7)	0 (0)
Missing/unknown	1 (0.3)	1 (0.3)	0 (0)	2 (1.6)	2 (1.6)	0 (0)	11 (3.3)	11 (3.3)	0 (0)
Ethnic minority, yes; N (%)	121 (33.2)	121 (33.2)	0 (0)	26 (21.0)	26 (21.0)	0 (0)	64 (20.1)	64 (20.1)	0 (0)
Season; N (%)									
Fall	84 (22.9)	84 (22.9)	0 (0)	19 (15.2)	19 (15.2)	0 (0)	70 (21.0)	70 (21.0)	0 (0)
Spring	91 (24.8)	91 (24.8)	0 (0)	34 (27.2)	34 (27.2)	0 (0)	83 (24.9)	83 (24.9)	0 (0)
Summer	98 (26.7)	98 (26.7)	0 (0)	37 (29.6)	37 (29.6)	0 (0)	99 (29.6)	99 (29.6)	0 (0)
Winter	94 (25.6)	94 (25.6)	0 (0)	35 (28.0)	35 (28.0)	0 (0)	82 (24.6)	82 (24.6)	0 (0)
Mother's education; N (%) ¹									
Less than college grad	235 (64.6)	129 (35.4)	106 (29.2)	97 (78.2)	27 (21.8)	70 (56.4)	218 (65.3)	116 (34.7)	102 (30.6)
College grad or higher	129 (35.4)	235 (64.6)	106 (29.2)	27 (21.8)	97 (78.2)	70 (56.4)	116 (34.7)	218 (65.3)	102 (30.6)

¹The publication has mother's education transposed for Colorado, Georgia/Florida, and Washington.

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-  
dr_studies6\TEDDY\private_orig_data\M_105_KMciver_NIDDK_Submission";
```

```
/*  
/* *****  
/* Dataset Integrity Check (DSIC) for the TEDDY M105 */  
/* Data submission */  
/* "Cross-Country Comparisons of Physical Activity */  
/* and Sedentary Behavior among 5-Year-Old Children" */  
/* McIver et al. */  
/* *****  
/*
```

```
*Temp dataset for analysis;  
data m105; set dsic.m_105_kmciver_niddk_31jan2017;  
run;
```

```
*Checking contents;  
proc contents data=m105;  
run;
```

```
*Accelerometer wear time, minutes/day: Total;  
proc means data=m105 mean std;  
var Wear_Time;  
run;
```

```
*Accelerometer wear time, minutes/day by Country;  
proc means data=m105 mean std;  
var Wear_Time;  
class Country;  
run;
```

```
*Accelerometer wear time, minutes/day by Clinical Center;  
proc means data=m105 mean std;  
var Wear_Time;  
class cc;  
run;
```

```
*Age, months total;  
proc means data=m105 mean std;  
var age;  
run;
```

```
*Age, months by country;  
proc means data=m105 mean std;  
var age;  
class country;
```

```
run;
```

```
*Age, month by clinical center;  
proc means data=m105 mean std;  
var age;  
class cc;  
run;
```

```
*Body weight, kg Total;  
proc means data=m105 mean std;  
var Wtkg;  
run;
```

```
*Body weight, kg by country;  
proc means data=m105 mean std;  
var Wtkg;  
class country;  
run;
```

```
*Body weight, kg by clinical center;  
proc means data=m105 mean std;  
var Wtkg;  
class Cc;  
run;
```

```
*BMI, total;  
proc means data=m105 mean std;  
var bmi;  
run;
```

```
*BMI, by country;  
proc means data=m105 mean std;  
var bmi;  
class country;  
run;
```

```
*BMI, by clinical center;  
proc means data=m105 mean std;  
var bmi;  
class cc;  
run;
```

```
*Males, total;  
proc freq data=m105;  
tables Sex;  
run;
```

```
*Males, by country;  
proc freq data=m105;
```

```
tables sex*country/norow nopercnt;  
run;
```

```
*Males, by clinical center;  
proc freq data=m105;  
tables sex*cc/norow nopercnt;  
run;
```

```
*Race/ethnicity by clinical center;  
proc freq data=m105;  
tables Race_Ethnicity*cc/norow nopercnt;  
run;
```

```
*Ethnic minority, yes Total;  
proc freq data=m105;  
tables Ethnic_Minority;  
run;
```

```
*Ethnic minority, yes by country;  
proc freq data=m105;  
tables Ethnic_Minority*country/norow nopercnt;  
run;
```

```
*Ethnic minority, yes by clinical center;  
proc freq data=m105;  
tables Ethnic_Minority*cc/ norow nopercnt;  
run;
```

```
*Season, Total;  
proc freq data=m105;  
tables Meas_Season;  
run;
```

```
*Season, by country;  
proc freq data=m105;  
tables Meas_Season*country/norow nopercnt;  
run;
```

```
*Season, by clinical center;  
proc freq data=m105;  
tables Meas_Season*cc/norow nopercnt;  
run;
```

```
*Mother's education, Total;  
proc freq data=m105;  
tables momed;  
run;
```

```
*Mother's education, by country;  
proc freq data=m105;
```

```
tables momed*country/norow nopercent;  
run;
```

```
*Mother's education, by clinical center;  
proc freq data=m105;  
tables momed*cc/norow nopercent;  
run;
```