

Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M65 Lundgren

Prepared by Dominick Parisi

IMS Inc.

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

February 10, 2020

Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	3
6 Conclusions	3
The results of the replication are almost an exact match to the published results.	3
7 References	3
Table A: Variables used to replicate data in the publication.	4
Table B: Comparison of values computed in integrity check to reference article data values	5
Attachment A: SAS Code	9

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 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.

3 Archived Datasets

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the TEDDY/private_orig_data/M_65_MLundgren_NIDDK_Submission folder in the data package. For this replication, variables were taken from the “m_65_mlundgren_niddk__30oct2015.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Markus Lundgren et al [1] in BMC Pediatrics in 2017. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For **Appendix B: Characteristics of the subjects used in analysis for analgesic use and islet cell autoimmunity**, 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.

6 Conclusions

The results of the replication are an exact match to the published results.

7 References

[1] Markus Lundgren, Leigh Johnson Steed, Roy Tamura, Berglind Jonsdottir, Patricia Gesualdo, Claire Crouch, Maija Sjöberg, Gertie Hansson, William A. Hagopian, Anette G. Ziegler, Marian J. Rewers, Åke Lernmark, Jorma Toppari, Jin-Xiong She, Beena Akolkar, Jeffrey P. Krischer, Michael J. Haller, Helena Elding Larsson and for the TEDDY Study Group. Analgesic antipyretic use among young children in the TEDDY study: no association with islet autoimmunity. *BMC Pediatrics* (2017) 17:127 DOI 10.1186/s12887-017-0884-y

Table A: Variables used to replicate data in the publication.

Table Variable	dataset.variable
HLA category	m_65_mlundgren_niddk__30oct2015.hla_category
Gender	m_65_mlundgren_niddk__30oct2015.gender
1 st Degree Relative with type 1 diabetes	m_65_mlundgren_niddk__30oct2015.fdr
Ever Breastfed	m_65_mlundgren_niddk__30oct2015.ever_breastfed
Probiotic Use	m_65_mlundgren_niddk__30oct2015.probiotic3
rs1004446	m_65_mlundgren_niddk__30oct2015.rs1004446_A
rs10517086	m_65_mlundgren_niddk__30oct2015.rs10517086_A
rs12708716	m_65_mlundgren_niddk__30oct2015.rs12708716_G
rs2292239	m_65_mlundgren_niddk__30oct2015.rs2292239_A
rs2476601	m_65_mlundgren_niddk__30oct2015.rs2476601_A
rs2816316	m_65_mlundgren_niddk__30oct2015.rs2816316_C
rs3184504	m_65_mlundgren_niddk__30oct2015.rs3184504_A
rs4948088	m_65_mlundgren_niddk__30oct2015.rs4948088_A
Country	m_65_mlundgren_niddk__30oct2015.country

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

	Manuscript Overall (n=6894)	DSIC Overall (n=6894)	Diff Overall (n=0)
HLA			
DR3/DR4	2684 (39)	2684 (39)	0 (0)
DR4/DR4	1358 (20)	1358 (20)	0 (0)
DR4/DR8	1192 (17)	1192 (17)	0 (0)
DR3/DR3	1428 (21)	1428 (21)	0 (0)
All Others	232 (3)	232 (3)	0 (0)
Gender: Male	3513 (51)	3513 (51)	0 (0)
1 st Degree Relative	789 (11)	789 (11)	0 (0)
Ever Breastfed	6640 (96)	6640 (96)	0 (0)
Probiotic Use	1044 (15)	1044 (15)	0 (0)
rs100446			
0	2766 (40)	2766 (40)	0 (0)
1	3180 (46)	3180 (46)	0 (0)
2	948 (14)	948 (14)	0 (0)
rs10517086			
0	3523 (51)	3523 (51)	0 (0)
1	2826 (41)	2826 (41)	0 (0)
2	545 (8)	545 (8)	0 (0)
rs12708716			
0	3042 (44)	3042 (44)	0 (0)
1	3053 (44)	3053 (44)	0 (0)
2	799 (12)	799 (12)	0 (0)

	Manuscript Overall (n=6894)	DSIC Overall (n=6894)	Diff Overall (n=0)
rs2292239			
0	3154 (46)	3154 (46)	0 (0)
1	3045 (44)	3045 (44)	0 (0)
2	695 (10)	695 (10)	0 (0)
rs2476601			
0	5472 (79)	5472 (79)	0 (0)
1	1331 (19)	1331 (19)	0 (0)
2	91 (1)	91 (1)	0 (0)
rs2816316			
0	4612 (67)	4612 (67)	0 (0)
1	2049 (30)	2049 (30)	0 (0)
2	233 (3)	233 (3)	0 (0)
rs3184504			
0	2141 (31)	2141 (31)	0 (0)
1	3372 (49)	3372 (49)	0 (0)
2	1381 (20)	1381 (20)	0 (0)
rs4948088			
0	6283 (91)	6283 (91)	0 (0)
1	590 (9)	590 (9)	0 (0)
2	21 (<1)	21 (<1)	0 (0)

	Manuscript USA (n=2838)	Manuscript Finland (n=1582)	Manuscript Germany (n=425)	Manuscript Sweden (n=2049)	DSIC USA (n=2838)	DSIC Finland (n=1582)	DSIC Germany (n=425)	DSIC Sweden (n=2049)	Diff USA (n=0)	Diff Finland (n=0)	Diff Germany (n=0)	Diff Sweden (n=0)
HLA												
DR3/DR4	1133 (40)	525 (33)	164 (39)	862 (42)	1133 (40)	525 (33)	164 (39)	862 (42)	0 (0)	0 (0)	0 (0)	0 (0)
DR4/DR4	582 (21)	256 (16)	70 (16)	450 (22)	582 (21)	256 (16)	70 (16)	450 (22)	0 (0)	0 (0)	0 (0)	0 (0)
DR4/DR8	382 (13)	508 (32)	39 (9)	263 (13)	382 (13)	508 (32)	39 (9)	263 (13)	0 (0)	0 (0)	0 (0)	0 (0)
DR3/DR3	665 (23)	231 (15)	89 (21)	443 (22)	665 (23)	231 (15)	89 (21)	443 (22)	0 (0)	0 (0)	0 (0)	0 (0)
All Others	76 (3)	62 (4)	63 (15)	31 (2)	76 (3)	62 (4)	63 (15)	31 (2)	0 (0)	0 (0)	0 (0)	0 (0)
Gender: Male	1458 (51)	805 (51)	218 (51)	1032 (50)	1458 (51)	805 (51)	218 (51)	1032 (50)	0 (0)	0 (0)	0 (0)	0 (0)
1 st Degree Relative	325 (11)	143 (9)	173 (41)	148 (7)	325 (11)	143 (9)	173 (41)	148 (7)	0 (0)	0 (0)	0 (0)	0 (0)
Ever Breastfed	2646 (93)	1573 (99)	397 (93)	2024 (99)	2646 (93)	1573 (99)	397 (93)	2024 (99)	0 (0)	0 (0)	0 (0)	0 (0)
Probiotic Use	75 (3)	609 (39)	109 (26)	251 (12)	75 (3)	609 (39)	109 (26)	251 (12)	0 (0)	0 (0)	0 (0)	0 (0)
rs100446												
0	1151 (41)	663 (42)	174 (41)	778 (38)	1151 (41)	663 (42)	174 (41)	778 (38)	0 (0)	0 (0)	0 (0)	0 (0)
1	1302 (46)	728 (46)	196 (46)	954 (47)	1302 (46)	728 (46)	196 (46)	954 (47)	0 (0)	0 (0)	0 (0)	0 (0)
2	385 (14)	191 (12)	55 (13)	317 (15)	385 (14)	191 (12)	55 (13)	317 (15)	0 (0)	0 (0)	0 (0)	0 (0)
rs10517086												
0	1476 (52)	790 (50)	191 (45)	1066 (52)	1476 (52)	790 (50)	191 (45)	1066 (52)	0 (0)	0 (0)	0 (0)	0 (0)
1	1144 (40)	672 (42)	203 (48)	807 (39)	1144 (40)	672 (42)	203 (48)	807 (39)	0 (0)	0 (0)	0 (0)	0 (0)
2	218 (8)	120 (8)	31 (7)	176 (9)	218 (8)	120 (8)	31 (7)	176 (9)	0 (0)	0 (0)	0 (0)	0 (0)
rs12708716												
0	1248 (44)	724 (46)	174 (41)	896 (44)	1248 (44)	724 (46)	174 (41)	896 (44)	0 (0)	0 (0)	0 (0)	0 (0)
1	1265 (45)	680 (43)	191 (45)	917 (45)	1265 (45)	680 (43)	191 (45)	917 (45)	0 (0)	0 (0)	0 (0)	0 (0)
2	325 (11)	178 (11)	60 (14)	236 (12)	325 (11)	178 (11)	60 (14)	236 (12)	0 (0)	0 (0)	0 (0)	0 (0)

	Manuscript USA (n=2838)	Manuscript Finland (n=1582)	Manuscript Germany (n=425)	Manuscript Sweden (n=2049)	DSIC USA (n=2838)	DSIC Finland (n=1582)	DSIC Germany (n=425)	DSIC Sweden (n=2049)	Diff USA (n=0)	Diff Finland (n=0)	Diff Germany (n=0)	Diff Sweden (n=0)
rs2292239												
0	1363 (48)	743 (47)	196 (46)	852 (42)	1363 (48)	743 (47)	196 (46)	852 (42)	0 (0)	0 (0)	0 (0)	0 (0)
1	1228 (43)	679 (43)	181 (43)	957 (47)	1228 (43)	679 (43)	181 (43)	957 (47)	0 (0)	0 (0)	0 (0)	0 (0)
2	247 (9)	160 (10)	48 (11)	240 (12)	247 (9)	160 (10)	48 (11)	240 (12)	0 (0)	0 (0)	0 (0)	0 (0)
rs2476601												
0	2348 (83)	1158 (73)	327 (77)	1639 (80)	2348 (83)	1158 (73)	327 (77)	1639 (80)	0 (0)	0 (0)	0 (0)	0 (0)
1	465 (16)	389 (29)	90 (21)	387 (19)	465 (16)	389 (29)	90 (21)	387 (19)	0 (0)	0 (0)	0 (0)	0 (0)
2	25 (1)	35 (2)	8 (2)	23 (1)	25 (1)	35 (2)	8 (2)	23 (1)	0 (0)	0 (0)	0 (0)	0 (0)
rs2816316												
0	1839 (65)	1145 (72)	280 (66)	1348 (66)	1839 (65)	1145 (72)	280 (66)	1348 (66)	0 (0)	0 (0)	0 (0)	0 (0)
1	886 (31)	399 (25)	128 (30)	636 (31)	886 (31)	399 (25)	128 (30)	636 (31)	0 (0)	0 (0)	0 (0)	0 (0)
2	113 (4)	38 (2)	17 (4)	65 (3)	113 (4)	38 (2)	17 (4)	65 (3)	0 (0)	0 (0)	0 (0)	0 (0)
rs3184504												
0	959 (34)	545 (34)	94 (22)	543 (27)	959 (34)	545 (34)	94 (22)	543 (27)	0 (0)	0 (0)	0 (0)	0 (0)
1	1332 (47)	767 (48)	223 (52)	1050 (51)	1332 (47)	767 (48)	223 (52)	1050 (51)	0 (0)	0 (0)	0 (0)	0 (0)
2	547 (19)	270 (17)	108 (25)	456 (22)	547 (19)	270 (17)	108 (25)	456 (22)	0 (0)	0 (0)	0 (0)	0 (0)
rs4948088												
0	2561 (90)	1490 (94)	390 (92)	1842 (90)	2561 (90)	1490 (94)	390 (92)	1842 (90)	0 (0)	0 (0)	0 (0)	0 (0)
1	268 (9)	89 (6)	33 (8)	200 (10)	268 (9)	89 (6)	33 (8)	200 (10)	0 (0)	0 (0)	0 (0)	0 (0)
2	9 (<1)	3 (<1)	2 (<1)	7 (<1)	9 (<1)	3 (<1)	2 (<1)	7 (<1)	0 (0)	0 (0)	0 (0)	0 (0)

Attachment A: SAS Code

```

/*****/
/* Formats */
/*****/
PROC FORMAT;
  VALUE COUNTRY
    1 = 'USA'
    2 = 'Finland'
    3 = 'Germany'
    4 = 'Sweden';
RUN;

/*****/
/* Library statements */
/*****/
LIBNAME SASDATA '/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_65_MLundgren_NIDDK_Submission/';

/*****/
/* Prepare dataset for manuscript check */
/*****/

PROC SORT DATA=SASDATA.m_65_mlundgren_niddk__30oct2015 OUT=m_65_mlundgren_niddk__30oct2015;
  BY MASKID;
RUN;

DATA COHORT;
  SET m_65_mlundgren_niddk__30oct2015;
  BY MASKID;
  IF FIRST.MASKID;

  IF COUNTRY='USA' THEN COUNTRY_NUM=1;
  ELSE IF COUNTRY='FIN' THEN COUNTRY_NUM=2;
  ELSE IF COUNTRY='GER' THEN COUNTRY_NUM=3;
  ELSE IF COUNTRY='SWE' THEN COUNTRY_NUM=4;

  if hla_category = 3 or
     hla_category = 5 or
     hla_category = 6 or
     hla_category = 7 or
     hla_category = 8 or
     hla_category = 10 then p_hla_category = 99;

```

```

else p_hla_category = hla_category;

IF HLA_CATEGORY=. OR COUNTRY='' OR FDR=. OR GENDER='' OR PROBIOTIC3=. OR RS1004446_a=. OR rs10517086_a=. OR
rs12708716_g=. OR rs2292239_a=. OR rs2476601_a=. OR rs2816316_c=. OR rs3184504_a=. OR rs4948088_a=. THEN COMPLETE=0;
ELSE COMPLETE=1;

IF HLA_CATEGORY IN(-1,0) THEN DELETE;
IF COMPLETE=0 THEN DELETE;
RUN;

/*****
/* Frequency tables */
*****/
TITLE2 'Frequency tables to replicate Appendix B (Additional file 2)';
PROC FREQ DATA=COHORT;
  TABLE (p_hla_category
         gender
         fdr
         ever_breastfed
         probiotic3
         rs1004446_A
         rs10517086_A
         rs12708716_G
         rs2292239_A
         rs2476601_A
         rs2816316_C
         rs3184504_A
         rs4948088_A) * country_num /norow;
  FORMAT COUNTRY_NUM COUNTRY.;
RUN;

```