

# Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub38a Uusitalo

Prepared by Jane Rideau Demuth

**IMS Inc.**

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

**September 7, 2018**

## Contents

1 Standard Disclaimer .....	2
2 Study Background .....	2
3 Archived Datasets .....	2
4 Statistical Methods .....	3
5 Results .....	3
6 Conclusions .....	3
7 References .....	3
Table A: Variables used to replicate Table 2: Characteristics associated with any islet autoimmunity .....	4
Table B: Comparison of values computed in integrity check to reference article Table 2 values.....	5
Table C: Variables used to replicate Table 3: Association of timing of dietary exposures with the risk of islet autoimmunity (any IA, IAA or GADA)	
Table D: Comparison of values computed in integrity check to reference article Table 3 values	
Table E: Variables used to replicate Table 4: Categorized duration of breastfeeding and timing of introduction of complementary foods and risk of islet autoimmunity (any IA or GADA)	
Table F: Comparison of values computed in integrity check to reference article Table 4 values	
Attachment A: SAS Code.....	7

## 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 folder in the data package. For this replication, variables were taken from the “m\_38a\_uusitalo\_niddk\_31jan2017.sas7bdat” dataset.

## 4 Statistical Methods

Analyses were performed to duplicate results for the data published by Uusitalo et al [1] in the journal Diabetes Care in 2018. To verify the integrity of the dataset, descriptive statistics were computed.

## 5 Results

For Table 2 in the publication [1], Characteristics associated with any islet 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 in Table 2.

For Table 3 in the publication [1], Association of timing of dietary exposures with the risk of islet autoimmunity (any IA, IAA or GADA), Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data files to the results published in Table 3. The results of the replication are an exact match to the published results.

For Table 4 in the publication [1], Categorized duration of breastfeeding and timing of introduction of complementary foods and risk of islet autoimmunity (any IA or GADA), Table E lists the variables that were used in the replication and Table F compares the results calculated from the archived data files to the results published in Table 2.

## 6 Conclusions

The NIDDK repository is confident that the TEDDY M38a data files to be distributed are a true copy of the study data.

## 7 References

[1] Uusitalo U, Lee H-S, Aronsson C A, et al. Early Infant Diet and Islet Autoimmunity in the TEDDY Study. Diabetes Care 2018 Jan; dc171983.

**Table A:** Variables used to replicate Table 2: Characteristics associated with any islet autoimmunity

<b>Table Variable</b>	<b>dataset.variable</b>
Country	m_38a_uuusitalo_niddk_31jan2017.count
High risk HLA genotype (DR3/4)	m_38a_uuusitalo_niddk_31jan2017.dr34
First degree relative with T1D	m_38a_uuusitalo_niddk_31jan2017.fdr
Sex of the child	m_38a_uuusitalo_niddk_31jan2017.female
Probiotics, age at first exposure	m_38a_uuusitalo_niddk_31jan2017.suppl_flag_28d
All/Any IA/IAA/GADA	m_38a_uuusitalo_niddk_31jan2017.compevent

**Table B:** Comparison of values computed in integrity check to reference article Table 2 values

Variable	Value	All (ms n)	All (IMS n)	All (Diff n)	Any IA (ms n, %)	Any IA (IMS n, %)	Any IA (Diff n)	IAA (ms n, %)	IAA (IMS n, %)	IAA (Diff n)	GADA (ms n, %)	GADA (IMS n, %)	GADA (Diff n)
Country	Finland	1654	1654	0	180 (25.6)	180 (25.6)	0 (0)	89 (32.7)	89 (32.7)	0 (0)	110 (36.8)	60 (20.1)	50 (16.7)
	Germany	510	510	0	54 (7.7)	54 (7.7)	0 (0)	20 (7.4)	20 (7.4)	0 (0)	60 (20.1)	15 (5.0)	45 (15.1)
	Sweden	2283	2283	0	240 (34.1)	240 (34.1)	0 (0)	83 (30.5)	83 (30.5)	0 (0)	15 (5.0)	114 (38.1)	99 (33.1)
	US	116	3116	3000	229 (32.6)	229 (32.6)	0 (0)	80 (29.4)	80 (29.4)	0 (0)	114 (38.1)	110 (36.8)	4 (1.3)
High risk HLA genotype (DR3/4)	Yes	2956	2956	0	344 (48.9)	344 (48.9)	0 (0)	128 (47.1)	128 (47.1)	0 (0)	150 (50.2)	150 (50.2)	0 (0)
	No	4607	4607	0	359 (51.1)	359 (51.1)	0 (0)	144 (52.9)	144 (52.9)	0 (0)	149 (49.8)	149 (49.8)	0 (0)
First-degree relative with T1D	Yes	857	857	0	136 (19.4)	136 (19.4)	0 (0)	58 (21.3)	58 (21.3)	0 (0)	52 (17.4)	52 (17.4)	0 (0)
	No	6706	6706	0	567 (80.6)	567 (80.6)	0 (0)	214 (78.7)	214 (78.7)	0 (0)	247 (82.6)	247 (82.6)	0 (0)
Sex of the child	Female	3698	3698	0	322 (45.8)	322 (45.8)	0 (0)	152 (44.1)	120 (44.1)	32 (0)	144 (48.2)	144 (48.2)	0 (0)
	Male	3865	3865	0	381 (54.2)	381 (54.2)	0 (0)	120 (55.9)	152 (55.9)	32 (0)	155 (51.8)	155 (51.8)	0 (0)
Probiotics, age at first exposure	<28 days	538	538	0	41 (5.8)	41 (5.8)	0 (0)	15 (5.5)	15 (5.5)	0 (0)	15 (5.0)	15 (5.0)	0 (0)

<b>Variable</b>	<b>Value</b>	<b>All (ms n)</b>	<b>All (IMS n)</b>	<b>All (Diff n)</b>	<b>Any IA (ms n, %)</b>	<b>Any IA (IMS n, %)</b>	<b>Any IA (Diff n)</b>	<b>IAA (ms n, %)</b>	<b>IAA (IMS n, %)</b>	<b>IAA (Diff n)</b>	<b>GADA (ms n, %)</b>	<b>GADA (IMS n, %)</b>	<b>GADA (Diff n)</b>
	>=28 days	7025	7025	0	662 (94.2)	662 (94.2)	0 (0)	257 (94.5)	257 (94.5)	0 (0)	284 (95.0)	284 (95.0)	0 (0)

**Table C:** Variables used to replicate Table 3: Association of timing of dietary exposures with the risk of islet autoimmunity (any IA, IAA or GADA)

<b>Table Variable</b>	<b>dataset.variable</b>
Islet autoimmunity	m_38a_uusitalo_niddk_31jan2017.compevent
Exclusive breastfeeding	m_38a_uusitalo_niddk_31jan2017.EXCLBTIME
Any breastfeeding	m_38a_uusitalo_niddk_31jan2017.BRSTTIME
Any infant formula	m_38a_uusitalo_niddk_31jan2017.FORMULA_ANYTIME
Cow's milk	m_38a_uusitalo_niddk_31jan2017.COW_MILK_ANYTIME
Any solid food	m_38a_uusitalo_niddk_31jan2017.ANYSOLIDTIME
Selected foods	m_38a_uusitalo_niddk_31jan2017.SUMMARYTIME
Cereals, any	m_38a_uusitalo_niddk_31jan2017.NONCORN_CEREALSTIME
Gluten containing cereals	m_38a_uusitalo_niddk_31jan2017.GLUTENTIME
Rice	m_38a_uusitalo_niddk_31jan2017.RICETIME
Root vegetables	m_38a_uusitalo_niddk_31jan2017.TOTAL_ROOTSTIME
Potatoes	m_38a_uusitalo_niddk_31jan2017.POTATOESTIME
Fruits and berries	m_38a_uusitalo_niddk_31jan2017.FRUIT_AND_BERRIESTIME
Meat	m_38a_uusitalo_niddk_31jan2017.MEATIME
Egg	m_38a_uusitalo_niddk_31jan2017.EGGTIME
Fish or other seafood	m_38a_uusitalo_niddk_31jan2017.FISHTIME



**Table D:** Comparison of values computed in integrity check to reference article Table 3 values

<b>Variable</b>	<b>Islet autoimmunity</b>	<b>Developed (ms n)</b>	<b>Developed (IMS n)</b>	<b>Developed (Diff n)</b>	<b>Did not develop (ms n)</b>	<b>Did not develop (IMS n)</b>	<b>Did not develop (Diff n)</b>
Exclusive breastfeeding	Any IA	703	703	0	6860	6860	0
	IAA	272	272	0	7291	7291	0
	GADA	299	299	0	7264	7264	0
Any breastfeeding	Any IA	694	694	0	6617	6617	0
	IAA	270	270	0	7041	7041	0
	GADA	294	294	0	7017	7017	0
Any infant formula	Any IA	598	598	0	6006	6006	0
	IAA	235	235	0	6369	6369	0
	GADA	257	257	0	6347	6347	0
Cow's milk	Any IA	701	701	0	6824	6824	0
	IAA	270	270	0	7255	7255	0
	GADA	299	299	0	7226	7226	0
Any solid food	Any IA	701	701	0	6839	6839	0
	IAA	272	272	0	7268	7268	0
	GADA	297	297	0	7243	7243	0
Selected foods	Any IA	701	701	0	6838	6838	0
	IAA	272	272	0	7267	7267	0
	GADA	297	297	0	7242	7242	0
Cereals, any	Any IA	699	699	0	6810	6810	0
	IAA	271	271	0	7238	7238	0
	GADA	296	296	0	7213	7213	0

<b>Variable</b>	<b>Islet autoimmunity</b>	<b>Developed (ms n)</b>	<b>Developed (IMS n)</b>	<b>Developed (Diff n)</b>	<b>Did not develop (ms n)</b>	<b>Did not develop (IMS n)</b>	<b>Did not develop (Diff n)</b>
Gluten containing cereals	Any IA	699	699	0	6708	6708	0
	IAA	271	271	0	7136	7136	0
	GADA	296	296	0	7111	7111	0
Rice	Any IA	699	699	0	6755	6755	0
	IAA	271	271	0	7183	7183	0
	GADA	296	296	0	7158	7158	0
Root vegetables	Any IA	701	701	0	6806	6806	0
	IAA	272	272	0	7235	7235	0
	GADA	297	297	0	7210	7210	0
Potatoes	Any IA	699	699	0	6698	6698	0
	IAA	271	271	0	7126	7126	0
	GADA	296	296	0	7101	7101	0
Fruits and berries	Any IA	700	700	0	6806	6806	0
	IAA	272	272	0	7234	7234	0
	GADA	296	296	0	7210	7210	0
Meat	Any IA	692	692	0	6676	6676	0
	IAA	269	269	0	7099	7099	0
	GADA	292	292	0	7076	7076	0
Egg	Any IA	684	684	0	6458	6458	0
	IAA	266	266	0	6876	6876	0

<b>Variable</b>	<b>Islet autoimmunity</b>	<b>Developed (ms n)</b>	<b>Developed (IMS n)</b>	<b>Developed (Diff n)</b>	<b>Did not develop (ms n)</b>	<b>Did not develop (IMS n)</b>	<b>Did not develop (Diff n)</b>
	GADA	290	290	0	6852	6852	0
Fish or other seafood	Any IA	662	662	0	6210	6210	0
	IAA	256	256	0	6616	6616	0
	GADA	280	280	0	6592	6592	0

**Table E:** Variables used to replicate Table 4: Categorized duration of breastfeeding and timing of introduction of complementary foods and risk of islet autoimmunity (any IA or GADA)

<b>Table Variable</b>	<b>dataset.variable</b>
Any IA/GADA	m_38a_uusitalo_niddk_31jan2017.compevent
Gluten containing cereals	m_38a_uusitalo_niddk_31jan2017.MGLUTENTIME
Any breastfeeding	m_38a_uusitalo_niddk_31jan2017.MBRSTTIME
Cow's milk	m_38a_uusitalo_niddk_31jan2017.MCOW_MILK_ANYTIME
Cereals, any	m_38a_uusitalo_niddk_31jan2017.MNONCORN_CEREALSTIME
Rice	m_38a_uusitalo_niddk_31jan2017.MRICETIME
Potatoes	m_38a_uusitalo_niddk_31jan2017.MPOTATOESTIME
Fruits and berries	m_38a_uusitalo_niddk_31jan2017.MFRUITS_AND_BERRIESTIME
Meat	m_38a_uusitalo_niddk_31jan2017.MMEATIME
Egg	m_38a_uusitalo_niddk_31jan2017.MEGGTIME
Fish or other seafood	m_38a_uusitalo_niddk_31jan2017.MFISHTIME

**Table F:** Comparison of values computed in integrity check to reference article Table 4 values

Islet autoimmunity	Breastfeeding or food	Timing in months	Manuscript n, %	IMS n, %	Diff n, %
Any IA	Gluten containing cereals	<4	28 (6)	31 (6)	3 (0)
		4-9	637 (10)	637 (10)	0 (0)
		>9	31 (14)	31 (14)	0 (0)
	Any breastfeeding	<=7	334 (9)	334 (9)	0 (0)
		>7	360 (11)	360 (11)	0 (0)
	Cow's milk	<=5	584 (10)	584 (9)	0 (1)
		>5	117 (11)	117 (11)	0 (0)
	Cereals	<=4	483 (9)	483 (9)	0 (0)
		>4	216 (10)	216 (10)	0 (0)
	Rice	<=7	618 (9)	624 (9)	6 (0)
		>7	75 (13)	75 (13)	0 (0)
	Fruit and berries	<=4	460 (9)	460 (9)	0 (0)
		>4	240 (10)	240 (10)	0 (0)
	Potato	<=7	578 (10)	578 (10)	0 (0)
		>7	121 (8)	121 (8)	0 (0)
	Meat	<=8	615 (10)	615 (10)	0 (0)
		>8	77 (8)	77 (8)	0 (0)
	Egg	<=9	282 (9)	467 (9)	185 (0)
		>9	402 (10)	217 (10)	185 (0)
GADA	Any breastfeeding	<=6	120 (3)	120 (3)	0 (0)
		>6	174 (5)	174 (5)	0 (0)
	Rice	<=6	264 (4)	268 (4)	4 (0)
		>6	28 (4)	28 (4)	0 (0)
	Fish and seafood	<=9	188 (4)	188 (4)	0 (0)
>9		92 (4)	92 (4)	0 (0)	

## Attachment A: SAS Code

```
options mprint nocentre linesize=163 validvarname=upcase;

%let rundate = y2018m08d08;
%let olddate = yYYYYmMMdDD;

title "Program: /prj/niddk/ims_analysis/TEDDY/prog_initial_analysis/M_38a_UUusitalo_NIDDK_Submission/DSIC.paper.review.&rundate..sas";
title2 "This program reviews TEDDY M38a data for level 1 PII";

/*****

programmer: Jane Rideau Demuth

platform: LINUX SASv9.4

date: 30 July 2018

purpose: See title2.

*****/

*****;
*** formats ***;
*****;
proc format;
  value nmsgf
    . = ' '
    low-high = '###'
  ;
  value $cmsgf
    ' ' = ' '
    other = '$$$'
  ;
  value countf
    1='US'
    2='FIN'
    3='GER'
    4='SWE'
  ;
  value dr34f
    0='Others'
    1='HLA DR3/4'
  ;
  value fdrf
    0='GenPop'
    1='FDR'
  ;
```

```

value femalef
  0='Male'
  1='Female'
  ;
value yesnof
  0 = 'No'
  1 = 'Yes'
  ;
value compeventf
  0='no IA'
  1='GAD only in the first positive sample'
  2='MIAA only in the first positive sample'
  3='Others'
  ;
value compevent2f
  0 = 'no IA'
  1,2,3 = 'Any IA'
  ;
value m49f
  0-<4 = '<4'
  4-9 = '4-9'
  9<-high = '>9'
  ;
value m4f
  0-4 = '<=4'
  4<-high = '>4'
  ;
value m5f
  0-5 = '<=5'
  5<-high = '>5'
  ;
value m6f
  0-6 = '<=6'
  6<-high = '>6'
  ;
value m7f
  0-7 = '<=7'
  7<-high = '>7'
  ;
value m8f
  0-8 = '<=8'
  8<-high = '>8'
  ;
value m9f
  0-9 = '<=9'
  9<-high = '>9'
  ;

```

```

*****;
*** input files ***;
*****;

```

```

libname pcsasin "/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_38a_UUusitalo_NIDDK_Submission/";
data m38a;
  set pcsasin.m_38a_uuusitalo_niddk_31jan2017;
title3 "Input file:
/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_38a_UUusitalo_NIDDK_Submission/m_38a_uuusitalo_niddk_31jan2017.sas7bdat";
proc contents data=m38a varnum;

*****;
*** run numbers from select tables ***;
*****;
proc freq data=m38a;
  title4 "Table 2. Characteristics associated with any islet autoimmunity";
  tables (count dr34 fdr female suppl_flag_28d)*compevent / missing;
  format count countf.
         dr34 dr34f.
         fdr fdrf.
         female femalef.
         suppl_flag_28d yesnof.
         compevent compeventf.;

/*proc univariate data=m38a;
  title3 'Univariate of exclusive breast feeding time';
  var exclbtime;*/

%macro table3(varname);
proc freq data=m38a;
  title4 "Table 3. Association of timing of dietary exposures with the risk of islet autoimmunity (any IA, IAA or GADA).";
  title5 "          Where &varname. >= 0";
  where &varname. >= 0;
  tables compevent / missing list;
  format compevent compeventf.;
%mend table3;

%table3(EXCLBTIME);
%table3(BRSTIME);
%table3(FORMULA_ANYTIME);
%table3(COW_MILK_ANYTIME);
%table3(ANYSOLIDTIME);
%table3(SUMMARYTIME);
%table3(NONCORN_CEREALSTIME);
%table3(GLUTENTIME);
%table3(RICETIME);
%table3(TOTAL_ROOTSTIME);
%table3(POTATOESTIME);
%table3(FRUITS_AND_BERRIESTIME);
%table3(MEATIME);
%table3(EGGTIME);
%table3(FISHTIME);

proc freq data=m38a;
  title4 "Table 4. Categorized duration of breastfeeding and timing of introduction of complementary foods";

```



```

title5 "          and risk of islet autoimmunity (any IA or GADA)";
tables compevent*(mglutentime mbrsttime mcow_milk_anytime mnoncorn_cerealstime mricetime mfruits_and_berriestime
                 mpotatoestime mmeatime meggtime) / missing;
format compevent compevent2f.
       mglutentime m49f.
       mbrsttime m7f.
       mcow_milk_anytime m5f.
       mnoncorn_cerealstime mfruits_and_berriestime m4f.
       mricetime mpotatoestime m7f.
       mmeatime m8f.
       meggtime m9f.;

proc freq data=m38a;
  tables compevent*(mbrsttime mricetime mfishtime) / missing;
  format compevent compeventf.
         mbrsttime mricetime m6f.
         mfishtime m9f.;

endsas;

```