

# Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub34 Yang

**Prepared by Jane Rideau Demuth**

**IMS Inc.**

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

**April 9, 2019**

## 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: Maternal and child characteristics of the TEDDY participants by country .....	4
Table B: Comparison of values computed in integrity check to reference article Table 1 values.....	5
Attachment A: SAS Code.....	5

## **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\_34\_jyang\_niddk\_30jun2015.sas7bdat” dataset.

## **4 Statistical Methods**

Analyses were performed to duplicate results for the data published by Yang et al [1] in the journal European Journal of Clinical Nutrition in 2017. To verify the integrity of the dataset, descriptive statistics were computed.

## 5 Results

For Table 1 in the publication [1], Maternal and child characteristics of the TEDDY participants by country, 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.

## 6 Conclusions

The NIDDK repository is confident that the TEDDY M34a data files to be distributed are a true copy of the study data.

## 7 References

[1] Yang J, Tamura R N, Uusitalo U M, et al. Vitamin D and probiotics supplement use in young children with genetic risk for type 1 diabetes. *European Journal of Clinical Nutrition* volume 71, pages 1449–1454 (2017).

**Table A:** Variables used to replicate Table 1: Maternal and child characteristics of the TEDDY participants by country

<b>Table Variable</b>	<b>dataset.variable</b>
Country	m_34_jyang_niddk_30jun2015.country
Maternal age	m_34_jyang_niddk_30jun2015.maternal_age
Prepregnancy BMI	m_34_jyang_niddk_30jun2015.bmi
Maternal diabetes status	m_34_jyang_niddk_30jun2015.diabetes
Maternal education	m_34_jyang_niddk_30jun2015.education_mom_group1
Mother smoking during pregnancy	m_34_jyang_niddk_30jun2015.smoker
Mother drinking during pregnancy	m_34_jyang_niddk_30jun2015.drinker
Birth weight	m_34_jyang_niddk_30jun2015.babysweightgrams
Gestational age (weeks)	m_34_jyang_niddk_30jun2015.gestational_age
Premature birth (<37 weeks)	m_34_jyang_niddk_30jun2015.preterm
Sex	m_34_jyang_niddk_30jun2015.sex
Mode of delivery	m_34_jyang_niddk_30jun2015.delivery
Being the first child	m_34_jyang_niddk_30jun2015.mom_first_child
Having FDR with T1D	m_34_jyang_niddk_30jun2015.fdr
Breastfed during the first year	m_34_jyang_niddk_30jun2015.ever_brstfed
Duration of breastfeeding (months)	m_34_jyang_niddk_30jun2015.breast_milk_end_age_mos
Year of birth	m_34_jyang_niddk_30jun2015.yob

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

	Country:	US			Finland			Germany			Sweden		
		n=3725	n=3724	dif=1	n=1832	n=1832	dif=0	n=595	n=594	dif=1	n=2525	n=2525	dif=0
Variable	Value	MS n (%) or mean (std)	IMS n (%) or mean (std)	dif	MS n (%) or mean (std)	IMS n (%) or mean (std)	dif	MS n (%) or mean (std)	IMS n (%) or mean (std)	dif	MS n (%) or mean (std)	IMS n (%) or mean (std)	dif
Maternal age (years)		30.1 (5.8)	30.1 (5.8)	NULL	29.9 (5.0)	29.9 (5.0)	NULL	31.4 (5.0)	31.3 (5.0)	0.1 (0)	30.8 (4.7)	30.8 (4.7)	NULL
Prepregnancy BMI		25.7 (6.1)	25.7 (6.1)	NULL	24.2 (4.6)	24.2 (4.6)	NULL	24.3 (5.0)	24.3 (5.0)	NULL	24.4 (4.6)	24.3 (4.6)	0.1 (0)
	Data available n=	3571	3570	1	1790	1790	NULL	592	593	1	2472	2472	NULL
Maternal diabetes status	Gestational	191 (5.1)	191 (5.1)	NULL	184 (10.0)	184 (10.0)	NULL	28 (4.7)	28 (4.7)	NULL	74 (2.9)	74 (2.9)	NULL
	Type 1	121 (3.2)	121 (3.2)	NULL	66 (3.6)	66 (3.6)	NULL	106 (17.8)	106 (17.8)	NULL	44 (1.7)	44 (1.7)	NULL
	Type 2	16 (0.4)	16 (0.4)	NULL	0	0	NULL	0	0	NULL	7 (0.3)	7 (0.3)	NULL
	None	3246 (87.1)	3245 (87.1)	1 (0)	1482 (80.9)	1482 (80.9)	NULL	394 (66.3)	395 (66.4)	1 (0.1)	2336 (92.5)	2336 (92.5)	NULL
	Missing	151 (4.1)	151 (4.1)	NULL	100 (5.5)	100 (5.5)	NULL	66 (11.1)	66 (11.1)	NULL	64 (2.5)	64 (2.5)	NULL
Maternal education	Basic	567 (15.2)	567 (15.2)	NULL	181 (9.9)	181 (9.9)	NULL	66 (11.1)	66 (11.1)	NULL	776 (30.7)	776 (30.7)	NULL
	Higher	2723 (73.1)	2722 (73.1)	1 (0)	1521 (83.0)	1521 (83.0)	NULL	449 (75.6)	450 (75.6)	1 (0)	1510 (59.8)	1510 (59.8)	NULL
	Missing	435 (11.7)	435 (11.7)	NULL	130 (7.1)	130 (7.1)	NULL	79 (13.3)	79 (13.3)	NULL	239 (9.5)	239 (9.5)	NULL
Mother smoking during pregnancy	No	3211 (86.2)	3210 (86.2)	1 (0)	1516 (82.8)	1516 (82.8)	NULL	470 (79.1)	471 (79.2)	1 (0.1)	2150 (85.1)	2150 (85.1)	NULL
	Yes	420 (11.3)	420 (11.3)	NULL	275 (15.0)	275 (15.0)	NULL	123 (20.7)	123 (20.7)	NULL	350 (13.9)	350 (13.9)	NULL
	Missing	94 (2.5)	94 (2.5)	NULL	41 (2.2)	41 (2.2)	NULL	1 (0.2)	1 (0.2)	NULL	25 (1.0)	25 (1.0)	NULL
Mother drinking during pregnancy	No	2266 (60.8)	2265 (60.8)	1 (0)	1242 (67.8)	1242 (67.8)	NULL	319 (53.7)	320 (53.8)	1 (0.1)	1803 (71.4)	1803 (71.4)	NULL

	<b>Yes</b>	1372 (36.8)	1372 (36.8)	NULL	549 (30.0)	549 (30.0)	NULL	274 (46.1)	274 (46.1)	NULL	697 (27.6)	697 (27.6)	NULL
	<b>Missing</b>	87 (2.3)	87 (2.3)	NULL	41 (2.2)	41 (2.2)	NULL	1 (0.2)	1 (0.2)	NULL	25 (1.0)	25 (1.0)	NULL
<b>Birth weight (g)</b>		3408 (538)	3408 (538)	NULL	3540 (554)	3540 (554)	NULL	3455 (550)	3455 (549)	0 (1)	3598 (537)	3598 (537)	NULL
	<b>Data available n=</b>	3491	3490	1	1828	1828	NULL	594	595	1	2524	2524	NULL
<b>Gestational age (weeks)</b>		39.2 (1.7)	39.2 (1.7)	NULL	39.7 (1.6)	39.7 (1.6)	NULL	39.3 (1.6)	39.3 (1.6)	NULL	39.8 (1.6)	39.8 (1.6)	NULL
	<b>Data available n=</b>	3717	3716	1	1828	1828	NULL	592	593	1	2524	2524	NULL
<b>Premature birth (&lt;37 weeks)</b>	<b>Yes</b>	245 (6.6)	243 (6.8)	2 (0.2)	84 (4.6)	88 (4.8)	4 (0.2)	32 (5.4)	34 (5.7)	2 (0.3)	122 (4.8)	123 (4.9)	1 (0.1)
	<b>No</b>	3472 (93.4)	3471 (93.2)	1 (0.2)	1748 (95.4)	1744 (95.2)	4 (0.2)	562 (94.6)	561 (94.3)	1 (0.3)	2402 (95.2)	2402 (95.1)	0 (0.1)
<b>Sex</b>	<b>Female</b>	1838 (49.3)	1838 (49.4)	0 (0.1)	900 (49.1)	900 (49.1)	NULL	298 (50.2)	298 (50.1)	0 (0.1)	1249 (49.5)	1250 (49.5)	1 (0)
	<b>Male</b>	1887 (50.7)	1886 (50.6)	1 (0.1)	932 (50.9)	932 (50.9)	NULL	296 (49.8)	297 (49.9)	1 (0.1)	1276 (50.5)	1275 (50.5)	1 (0)
<b>Mode of delivery</b>	<b>Vaginal</b>	2363 (63.4)	2363 (63.5)	0 (0.1)	1511 (82.5)	1511 (82.5)	NULL	386 (65.0)	386 (64.9))	0 (0.1)	2150 (85.2)	2150 (85.1)	0 (0.1)
	<b>Cesarean</b>	1359 (36.5)	1358 (36.5)	1 (0)	317 (17.3)	317 (17.3)	NULL	208 (35.0)	209 (35.1)	1 (0.1)	375 (14.9)	375 (14.9)	NULL
	<b>Missing</b>	3 (<0.01)	3 (0.1)	0 (0.1)	4 (0.2)	4 (0.2)	NULL	0	0	NULL	0	0	NULL
<b>Being the first child</b>	<b>No</b>	1949 (52.3)	1950 (52.4)	1 (0.1)	958 (52.3)	961 (52.5)	3 (0.2)	264 (44.4)	265 (44.5)	1 (0.1)	1229 (48.7)	1229 (48.7)	NULL
	<b>Yes</b>	1373 (36.9)	1375 (36.9)	2 (0)	772 (42.1)	773 (42.2)	1 (0.1)	267 (44.9)	268 (45.0)	1 (0.1)	1096 (43.4)	1096 (43.4)	NULL
	<b>Missing</b>	403 (10.8)	399 (10.7)	4 (0.1)	102 (5.6)	98 (5.4)	4 (0.2)	63 (10.6)	62 (10.4)	1 (0.2)	200 (7.9)	200 (7.9)	NULL
<b>Having FDR with T1D</b>	<b>No</b>	3333 (89.5)	3327 (89.3)	4 (0.2)	1663 (90.8)	1663 (90.8)	NULL	375 (63.1)	376 (63.2)	1 (0.1)	2354 (93.2)	2354 (93.2)	NULL
	<b>Yes</b>	392 (10.5)	397 (10.7)	5 (0.2)	169 (9.2)	169 (9.2)	NULL	219 (36.9)	219 (36.8)	0 (0.1)	171 (6.8)	171 (6.8)	NULL
<b>Breastfed during the first year of year</b>	<b>No</b>	269 (7.2)	273 (7.3)	4 (0.1)	0	0	NULL	26 (4.4)	27 (4.5)	1 (0.1)	27 (1.1)	27 (1.1)	NULL
	<b>Yes</b>	3455 (92.8)	3450 (92.6)	5 (0.2)	1831 (99.9)	1831 (99.9)	NULL	568 (95.6)	568 (95.5)	0 (0.1)	2498 (98.9)	2498 (98.9)	NULL

	<b>Missing</b>	1 (<0.1)	1 (<0.1)	<b>NULL</b>	1 (0.1)	1 (0.1)	<b>NULL</b>	0	0	<b>NULL</b>	0	0	<b>NULL</b>
<b>Duration of breastfeeding (months)</b>		7.5 (7.4)	7.5 (7.6)	0 (0.2)	8.5 (5.6)	8.5 (5.6)	<b>NULL</b>	7.2 (6.0)	7.2 (6.3)	0 (0.3)	7.0 (5.1)	7.0 (5.4)	0 (0.3)
	<b>Data availability n=</b>	3687	3690	3	1817	1817	<b>NULL</b>	571	573	2	2510	2514	4
<b>Year of birth</b>	<b>2004–2005</b>	367 (9.9)	367 (9.9)	<b>NULL</b>	402 (21.9)	402 (21.9)	<b>NULL</b>	80 (13.5)	80 (13.5)	<b>NULL</b>	588 (23.3)	588 (23.3)	<b>NULL</b>
	<b>2006</b>	609 (16.3)	609 (16.4)	0 (0.1)	335 (18.3)	335 (18.3)	<b>NULL</b>	114 (19.2)	114 (19.2)	<b>NULL</b>	483 (19.1)	483 (19.1)	<b>NULL</b>
	<b>2007</b>	870 (23.4)	871 (23.4)	1 (0)	373 (20.4)	373 (20.4)	<b>NULL</b>	107 (18.0)	108 (18.2)	1 (0.2)	500 (19.8)	500 (19.8)	<b>NULL</b>
	<b>2008</b>	867 (23.3)	867 (23.3)	<b>NULL</b>	308 (16.8)	308 (16.8)	<b>NULL</b>	136 (22.9)	136 (22.9)	<b>NULL</b>	439 (17.4)	439 (17.4)	<b>NULL</b>
	<b>2009–2010</b>	1011 (27.1)	1011 (27.1)	<b>NULL</b>	414 (22.6)	414 (22.6)	<b>NULL</b>	157 (26.4)	157 (26.4)	<b>NULL</b>	515 (20.4)	515 (20.4)	<b>NULL</b>



## Attachment A: SAS Code

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

%let rundate = y2018m09d14;
%let olddate = YYYYmMMdDD;

title "Program: /prj/niddk/ims_analysis/TEDDY/prog_initial_analysis/M_34_JYang_NIDDK_Submission/DSIC.paper.review.&rundate..sas";
title2 "This program reviews TEDDY M34-JYang paper (Vitamin D and probiotics supplement use in young children with genetic risk for type 1 diabetes)";
/*****

programmer: Jane Rideau Demuth

platform: LINUX SASv9.4

date: 14 September 2018

purpose: See title2.

*****/

*****;
*** formats ***;
*****;
proc format;
  value nmsgf
    . = ' '
    low-high = '###'
  ;
  value $cmsgf
    ' ' = ' '
    other = '$$$'
  ;
  value yesnof
    0 = 'no'
    1 = 'YES'
  ;
  value countryf
    1 = 'USA'
    2 = 'FIN'
    3 = 'DEU'
    4 = 'SWE'
  ;
  value diabetesf
    1 = 'Gestational'
    2 = 'Type 1'
    3 = 'Type 2'
```

```

4 = 'None'
. = '<missing>'
;
value edulf
1 = 'Basic'
2 = 'Higher'
. = '<missing>'
;
value deliveryf
1,3 = 'Vaginal'
2,4 = 'Cesarean'
;
value yobf
2004-2005 = '2004/5'
2006 = '2006'
2007 = '2007'
2008 = '2008'
2009-2010 = '2009/10'
;

*****;
*** input files ***;
*****;
libname pcsasin "/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_34_JYang_NIDDK_Submission/";
data m34;
set pcsasin.m_34_jyang_niddk_30jun2015;
title3 "Input file: /prj/niddk/ims_analysis/TEDDY/private_orig_data/M_34_JYang_NIDDK_Submission/m_34_jyang_niddk_30jun2015.sas7bdat";
*proc contents data=m34 varnum;

*****;
*** check Table 1 ***;
*****;
proc sort data=m34;
by country;

proc univariate data=m34 noprint;
by country;
var maternal_age bmi babysweightgrams gestational_age breast_milk_end_age_mos;
output out=m34uni
n=momage_n bmi_n babywt_n gestage_n brstfedm_n
mean=momage_mean bmi_mean babywt_mean gestage_mean brstfedm_mean
std=momage_std bmi_std babywt_std gestage_std brstfedm_std;

proc print data=m34uni noobs;
by country;
title3 'Check Table 1 (Univariates for MEAN, STD)';
var momage_n momage_mean momage_std bmi_n bmi_mean bmi_std babywt_n babywt_mean babywt_std
gestage_n gestage_mean gestage_std brstfedm_n brstfedm_mean brstfedm_std;
format country countryf.;

proc freq data=m34;

```

```
title3 'Check Table 1';
tables country
      (diabetes education_mom_group1 smoker drinker preterm sex delivery mom_first_child fdr ever_brstfed yob)*country / missing;
format country countryf.
      diabetes diabetesf.
      education_mom_group1 edulf.
      smoker drinker preterm mom_first_child fdr ever_brstfed yesnof.
      delivery deliveryf.
      yob yobf.;

endsas;
```