

Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub19 Aronsson

Prepared by Allyson Mateja

IMS Inc.

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

November 17, 2016

Contents

1 Standard Disclaimer	3
2 Study Background	3
3 Archived Datasets	3
4 Statistical Methods	3
5 Results	4
6 Conclusions	4
7 References	5
Table A: Variables used to replicate Table 1: Nondietary Factors and Risk of CDA (Primary End Point) and CD (Secondary End Point)	6
Table B: Comparison of values computed in integrity check to reference article Table 1 values	6
Table C: Variables used to replicate Table 2: Background Characteristics of Age at Introduction to Gluten and Breastfeeding Duration, by Country	9
Table D: Comparison of values computed in integrity check to reference article Table 2 values	9
Table E: Variables used to replicate Table 3: Time to First Introduction to Gluten-Containing Cereals and Risk for CDA and CD	12
Table F: Comparison of values computed in integrity check to reference article Table 3 values	12
Table G: Variables used to replicate Table 4: Gluten Introduction Under Conditions of Long-Term Continued Breastfeeding (>1 mo), Short-Term Continued Breastfeeding (Continued ≤ 1 mo After Gluten Introduction), and Discontinued Breastfeeding Before Gluten Introduction and the Risk for CDA and CD	12
Table H: Comparison of values computed in integrity check to reference article Table 4 values	13
Table I: Variables used to replicate Supplementary Table 5: Time to First Introduction to Gluten-Containing Cereals and Risk for CDA and CD According to Country	13
Table J: Comparison of values computed in integrity check to reference article Supplementary Table 5 values	14
Table K: Variables used to replicate Supplementary Table 6: Gluten Introduction Under Conditions of Long-Term Continued Breastfeeding (>1 mo), Short-Term Continued Breastfeeding (Continued ≤ 1 mo After Gluten Introduction), and Discontinued Breastfeeding Before Gluten Introduction and the Risk for CDA and CD According to Country	15

Table L: Comparison of values computed in integrity check to reference article Table 6 values 15

Attachment A: SAS Code..... 18

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_19_caronsson_niddk_31july2013.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Aronsson et al [1] in *Pediatrics* in 2015. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Nondietary Factors and Risk of CDA (Primary End Point) and CD (Secondary End Point), Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are almost an exact match to the published results.

For Table 2 in the publication [1], Background Characteristics of Age at Introduction to Gluten and Breastfeeding Duration, by Country, Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data file to the results published in Table 2. The results of the replication are almost an exact match to the published results.

For Table 3 in the publication [1], Time to First Introduction to Gluten-Containing Cereals and Risk for CDA and CD, Table E lists the variables that were used in the replication and Table F compares the results calculated from the archived data file to the results published in Table 3. The results of the replication are almost an exact match to the published results.

For Table 4 in the publication [1], Gluten Introduction Under Conditions of Long-Term Continued Breastfeeding (>1 mo), Short-Term Continued Breastfeeding (Continued \leq 1 mo After Gluten Introduction), and Discontinued Breastfeeding Before Gluten Introduction and the Risk for CDA and CD, Table G lists the variables that were used in the replication and Table H compares the results calculated from the archived data file to the results published in Table 4. The results of the replication are almost an exact match to the published results.

For Supplementary Table 5 in the publication [1], Time to First Introduction to Gluten-Containing Cereals and Risk for CDA and CD According to Country, Table I lists the variables that were used in the replication and Table J compares the results calculated from the archived data file to the results published in Supplementary Table 5. The results of the replication are almost an exact match to the published results.

For Supplementary Table 6 in the publication [1], Gluten Introduction Under Conditions of Long-Term Continued Breastfeeding (>1 mo), Short-Term Continued Breastfeeding (Continued \leq 1 mo After Gluten Introduction), and Discontinued Breastfeeding Before Gluten Introduction and the Risk for CDA and CD According to Country, Table K lists the variables that were used in the replication and Table L compares the results calculated from the archived data file to the results published in Supplementary Table 6. The results of the replication are almost an exact match to the published results.

6 Conclusions

The NIDDK repository is confident that the TEDDY M19 data files to be distributed are a true copy of the study data.

7 References

[1] Aronsson, C.A., Lee, H., Liu, E., Uusitalo, U., Hummel, S., Yang, J., Hummel, M., Rewers, M., She, J., Simell, O., Toppari, J., Ziegler, A., Krischer, J., Virtanen, S., Norris, J.M., Agardh, D., and the TEDDY study group. "Age at Gluten Introduction and Risk of Celiac Disease". *Pediatrics* (2015) 135 (2):239-245.

Table A: Variables used to replicate Table 1: Nondietary Factors and Risk of CDA (Primary End Point) and CD (Secondary End Point)

Table Variable	Variable
CDA	htg_conf
CD	celiac_disease
Age at end point	timetocd, time_conftga
Country	country
Gender	female
HLA Genotype	hla_d
First degree relative with celiac disease	celiac_fdr
Season of birth	birth_season
Maternal smoking during pregnancy	smoker
Maternal education	education_mom_group1
Maternal age at delivery	maternal_age

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

Characteristic	CDA Manuscript (n=773)	CDA DSIC (n=773)	Diff. (n=0)	CD Manuscript (n=307)	CD DSIC (n=307)	Diff. (n=0)
Median (range) age at end point, y	3.0 (0.9-7.5)	3.0 (0.9-7.5)	0 (0-0)	3.8 (1.2-8.8)	3.5 (1.2-8.4)	0.3 (0-0.4)

Characteristic	CDA Manuscript (n=773) n (%)	CDA DSIC (n=773) n (%)	Diff. (n=0)	CDA Manuscript (n=773) HR (95% CI)	CDA DSIC (n=773) HR (95% CI)	Diff. (n=0)
Country						
United States	228 (9)	228 (9)	0 (0)	1	1	0
Sweden	333 (16)	333 (16)	0 (0)	1.74 (1.47-2.06)	1.74 (1.47-2.06)	0 (0-0)
Finland	171 (12)	171 (12)	0 (0)	1.22 (1.00-1.49)	1.22 (1.00-1.49)	0 (0-0)
Germany	41 (11)	41 (11)	0 (0)	1.21 (0.87-1.69)	1.21 (0.87-1.69)	0 (0-0)
Gender						
Boys	317 (10)	317 (10)	0 (0)	1	1	0
Girls	456 (14)	456 (14)	0 (0)	1.55 (1.34-1.79)	1.55 (1.34-1.79)	0 (0-0)
HLA genotype						
TEDDY other	430 (8)	430 (8)	0 (0)	1	1	0

Characteristic	CDA Manuscript (n=773) n (%)	CDA DSIC (n=773) n (%)	Diff. (n=0)	CDA Manuscript (n=773) HR (95% CI)	CDA DSIC (n=773) HR (95% CI)	Diff. (n=0)
DR3-DQ2/DR3-DQ2	343 (26)	343 (26)	0 (0)	3.56 (3.09-4.11)	3.57 (3.09-4.11)	0.01 (0-0)
First degree relative with celiac disease						
No	732 (12)	732 (12)	0 (0)	1	1	0
Yes	41 (28)	41 (28)	0 (0)	2.60 (1.90-3.57)	2.60 (1.90-3.57)	0 (0-0)
Season of birth						
Winter (December - February)	185 (11)	185 (11)	0 (0)	1	1	0
Spring (March - May)	212 (14)	212 (14)	0 (0)	1.18 (0.97-1.44)	1.18 (0.97-1.44)	0 (0-0)
Summer (June - August)	194 (12)	194 (12)	0 (0)	1.06 (0.87-1.30)	1.06 (0.87-1.30)	0 (0-0)
Fall (September - November)	182 (11)	182 (11)	0 (0)	1.03 (0.84-1.26)	1.03 (0.84-1.27)	0 (0-0.01)
Maternal smoking during pregnancy						
No	710 (12)*	702 (12)	8 (0)	1	1	0
Yes	63 (10)	63 (10)	0 (0)	0.85 (0.65-1.09)	0.85 (0.65-1.10)	0 (0-0.01)
Maternal education						
Basic primary	129 (11)	129 (11)	0 (0)	0.89 (0.74-1.08)	0.89 (0.74-1.08)	0 (0-0)
Higher education	640 (12)	640 (12)	0 (0)	1	1	0
Maternal age at delivery, y						
Mean ± SD	30.9 ± 4.7	30.9 ± 4.7	0 ± 0	1.0 (0.98-1.01)	1.0 (0.99-1.01)	0 (0.01-0)

*Note that the published value of 710 is a typo, as it includes the number of subjects with missing data for this question. The correct value is 702.

Characteristic	CD Manuscript (n=307) n (%)	CD DSIC (n=307) n (%)	Diff. (n=0)	CD Manuscript (n=307) HR (95% CI)	CD DSIC (n=307) HR (95% CI)	Diff. (n=0)
Country						
United States	93 (4)	93 (4)	0 (0)	1	1	0

Characteristic	CD Manuscript (n=307) n (%)	CD DSIC (n=307) n (%)	Diff. (n=0)	CD Manuscript (n=307) HR (95% CI)	CD DSIC (n=307) HR (95% CI)	Diff. (n=0)
Sweden	151 (7)	151 (7)	0 (0)	1.73 (1.34-2.24)	1.73 (1.34-2.24)	0 (0-0)
Finland	52 (4)	52 (4)	0 (0)	0.84 (0.60-1.18)	0.84 (0.60-1.18)	0 (0-0)
Germany	11 (3)	11 (3)	0 (0)	0.75 (0.40-1.40)	0.75 (0.40-1.41)	0 (0-0.01)
Gender						
Boys	107 (3)	107 (3)	0 (0)	1	1	0
Girls	200 (6)	200 (6)	0 (0)	2.02 (1.59-2.55)	2.02 (1.59-2.55)	0 (0-0)
HLA genotype						
TEDDY other	146 (3)	146 (3)	0 (0)	1	1	0
DR3-DQ2/DR3-DQ2	161 (12)	161 (12)	0 (0)	4.58 (3.66-5.74)	4.58 (3.66-5.73)	0 (0-0.01)
First degree relative with celiac disease						
No	282 (4)	282 (4)	0 (0)	1	1	0
Yes	25 (17)	25 (17)	0 (0)	3.83 (2.55-5.77)	3.83 (2.55-5.77)	0 (0-0)
Season of birth						
Winter (December - February)	63 (4)	63 (4)	0 (0)	1	1	0
Spring (March - May)	95 (6)	95 (6)	0 (0)	1.55 (1.13-2.13)	1.55 (1.13-2.13)	0 (0-0)
Summer (June - August)	75 (5)	75 (5)	0 (0)	1.19 (0.85-1.68)	1.19 (0.85-1.67)	0 (0-0.01)
Fall (September - November)	74 (5)	74 (5)	0 (0)	1.16 (0.83-1.63)	1.16 (0.83-1.63)	0 (0-0)
Maternal smoking during pregnancy						
No	275 (5)	275 (5)	0 (0)	1	1	0
Yes	28 (5)	28 (5)	0 (0)	0.98 (0.67-1.45)	0.98 (0.67-1.45)	0 (0-0)
Maternal education						
Basic primary	54 (5)	54 (5)	0 (0)	0.94 (0.70-1.27)	0.94 (0.70-1.27)	0 (0-0)
Higher education	253 (5)	253 (5)	0 (0)	1	1	0
Maternal age at delivery, y						

Characteristic	CD Manuscript (n=307) n (%)	CD DSIC (n=307) n (%)	Diff. (n=0)	CD Manuscript (n=307) HR (95% CI)	CD DSIC (n=307) HR (95% CI)	Diff. (n=0)
Mean ± SD	30.8 ± 4.5	30.8 ± 4.5	0 ± 0	1.0 (0.97-1.02)	1.0 (0.97-1.02)	0 (0-0)

Table C: Variables used to replicate Table 2: Background Characteristics of Age at Introduction to Gluten and Breastfeeding Duration, by Country

Table Variable	Variable
CDA	htg_conf
CD	celiac_disease
Country	country
Time to gluten introduction	ugluten_no_oat_w
Category of age at gluten introduction	ugluten_no_oat_4
Duration of exclusive breastfeeding	exclbtime_w
Duration of any breastfeeding	brsttime_w

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

United States (n=2542)

Variable	n (%) Manuscript	n (%) DSIC	Diff.	Mean ± SD Manuscript	Mean ± SD DSIC	Diff.	Median (Q1, Q3) Manuscript	Median (Q1, Q3) DSIC	Diff.
Time to gluten introduction, wk	2542	2542	0	30.1 ± 8.4	30.1 ± 8.4	0 ± 0	30.4 (26.1, 34.9)	30.4 (26.1, 34.9)	0 (0, 0)
Category of age at gluten introduction									
Early (< 17 wk)	98 (4)	98 (4)	0 (0)	11.6 ± 3.9	11.6 ± 3.9	0 ± 0	13.0 (11.0, 14.0)	13.0 (11.0, 14.0)	0 (0, 0)
Reference (17-26 wk)	447 (18)	447 (18)	0 (0)	21.0 ± 2.2	21.0 ± 2.2	0 ± 0	21.7 (19.6, 21.7)	21.7 (19.6, 21.7)	0 (0,0)
Late (>26 wk)	1997 (78)	1997 (79)	0 (1)	33.0 ± 6.6	33.0 ± 6.6	0 ± 0	30.4 (26.1, 34.9)	30.4 (26.1, 34.9)	0 (0, 0)
Duration of exclusive breastfeeding, wk	2542	2542	0	5.4 ± 8.3	5.4 ± 8.3	0 ± 0	1.0 (0.07, 8.0)	1.0 (0.07, 8.0)	0 (0, 0)
Duration of any breastfeeding, wk	2542	2542	0	38.4 ± 34.0	38.4 ± 34.0	0 ± 0	34.6 (10.0, 56.6)	34.6 (10.0, 56.6)	0 (0, 0)

Finland (n=1472)

Variable	n (%) Manuscript	n (%) DSIC	Diff.	Mean ± SD Manuscript	Mean ± SD DSIC	Diff.	Median (Q1, Q3) Manuscript	Median (Q1, Q3) DSIC	Diff.
Time to gluten introduction, wk	1472	1472	0	27.8 ± 8.1	27.8 ± 8.1	0 ± 0	26.1 (23.9, 30.4)	26.1 (23.9, 30.4)	0 (0, 0)
Category of age at gluten introduction									
Early (< 17 wk)	47 (3)	47 (3)	0 (0)	7.8 ± 4.1	8.0 ± 4.1	0.2 ± 0	7.0 (4.3, 10.0)	7.0 (4.3, 10.0)	0 (0, 0)
Reference (17-26 wk)	463 (32)	463 (31)	0 (1)	22.1 ± 1.8	22.1 ± 1.8	0 ± 0	21.7 (21.7, 23.9)	21.7 (21.7, 23.9)	0 (0, 0)
Late (>26 wk)	962 (65)	962 (65)	0 (0)	31.5 ± 6.7	31.5 ± 7.0	0 ± 0.3	30.4 (26.1, 34.9)	30.4 (26.1, 34.9)	0 (0, 0)
Duration of exclusive breastfeeding, wk	1472	1472	0	7.2 ± 8.6	7.2 ± 8.6	0 ± 0	3.0 (0.07, 13.0)	3.0 (0.07, 13.0)	0 (0, 0)
Duration of any breastfeeding, wk	1472	1472	0	39.4 ± 24.7	39.4 ± 24.7	0 ± 0	39.1 (21.7, 54.6)	39.1 (21.7, 54.6)	0 (0, 0)

Germany (n=380)

Variable	n (%) Manuscript	n (%) DSIC	Diff.	Mean ± SD Manuscript	Mean ± SD DSIC	Diff.	Median (Q1, Q3) Manuscript	Median (Q1, Q3) DSIC	Diff.
Time to gluten introduction, wk	380	380	0	32.8 ± 10.4	32.8 ± 10.4	0 ± 0	30.4 (26.1, 37.0)	30.4 (26.1, 37.0)	0 (0, 0)
Category of age at gluten introduction									
Early (< 17 wk)	18 (5)	18 (5)	0 (0)	14.5 ± 1.9	14.5 ± 1.9	0 ± 0	14.0 (14.0, 16.0)	14.0 (14.0, 16.0)	0 (0, 0)
Reference (17-26 wk)	53 (14)	53 (14)	0 (0)	22.0 ± 1.9	22.0 ± 1.9	0 ± 0	21.7 (19.6, 23.9)	21.7 (19.6, 23.9)	0 (0, 0)
Late (>26 wk)	309 (81)	309 (81)	0 (0)	35.7 ± 9.2	35.7 ± 9.2	0 ± 0	32.6 (29.3, 39.1)	32.6 (29.3, 39.1)	0 (0, 0)

Variable	n (%) Manuscript	n (%) DSIC	Diff.	Mean ± SD Manuscript	Mean ± SD DSIC	Diff.	Median (Q1, Q3) Manuscript	Median (Q1, Q3) DSIC	Diff.
Duration of exclusive breastfeeding, wk	380	380	0	10.1 ± 11.1	10.1 ± 11.1	0 ± 0	4.0 (0.07, 21.7)	4.0 (0.07, 21.7)	0 (0, 0)
Duration of any breastfeeding, wk	380	380	0	36.5 ± 27.2	36.5 ± 27.2	0 ± 0	34.7 (14.1, 49.3)	34.7 (14.1, 49.3)	0 (0, 0)

Sweden (n=2042)

Variable	n (%) Manuscript	n (%) DSIC	Diff.	Mean ± SD Manuscript	Mean ± SD DSIC	Diff.	Median (Q1, Q3) Manuscript	Median (Q1, Q3) DSIC	Diff.
Time to gluten introduction, wk	2042	2042	0	21.6 ± 4.5	21.6 ± 4.5	0 ± 0	21.7 (17.4, 23.9)	21.7 (17.4, 23.9)	0 (0, 0)
Category of age at gluten introduction									
Early (< 17 wk)	233 (11)	233 (11)	0 (0)	15.0 ± 1.6	15.0 ± 1.6	0 ± 0	15.3 (14.0, 16.0)	15.3 (14.0, 16.0)	0 (0, 0)
Reference (17-26 wk)	1330 (65)	1330 (65)	0 (0)	20.6 ± 2.3	20.6 ± 2.3	0 ± 0	21.7 (18.0, 21.7)	21.7 (18.0, 21.7)	0 (0, 0)
Late (>26 wk)	479 (24)	479 (23)	0 (1)	27.7 ± 2.9	27.7 ± 2.9	0 ± 0	26.1 (26.1, 28.3)	26.1 (26.1, 28.3)	0 (0, 0)
Duration of exclusive breastfeeding, wk	2042	2042	0	7.5 ± 7.8	7.6 ± 7.8	0.1 ± 0	4.0 (0.07, 16.0)	4.0 (0.07, 16.0)	0 (0, 0)
Duration of any breastfeeding, wk	2042	2042	0	32.7 ± 22.7	32.7 ± 22.7	0 ± 0	31.9 (17.4, 43.4)	31.9 (17.4, 43.4)	0 (0, 0)

Table E: Variables used to replicate Table 3: Time to First Introduction to Gluten-Containing Cereals and Risk for CDA and CD

Table Variable	Variable
CDA	htg_conf
CD	celiac_disease
Gluten Introduction	ugluten_no_oat_4

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

Gluten Introduction	CDA Manuscript n (%)	CDA DSIC n (%)	Diff.	CDA Manuscript HR (95% CI)	CDA DSIC HR (95% CI)	Diff.
< 17 wk	52 (13)	52 (13)	0 (0)	1.06 (0.79-1.42)	1.05 (0.78-1.41)	0.01 (0.01-0.01)
17-26 wk	315 (14)	315 (14)	0 (0)	1	1	0
> 26 wk	406 (11)	406 (11)	0 (0)	0.97 (0.82-1.15)	0.96 (0.81-1.13)	0.01 (0.01-0.02)

Gluten Introduction	CD Manuscript n (%)	CD DSIC n (%)	Diff.	CD Manuscript HR (95% CI)	CD DSIC HR (95% CI)	Diff.
< 17 wk	13 (3)	13 (3)	0 (0)	0.59 (0.33-1.04)	0.59 (0.34-1.05)	0 (0.01-0.01)
17-26 wk	141 (6)	141 (6)	0 (0)	1	1	0
> 26 wk	153 (4)	153 (4)	0 (0)	0.90 (0.69-1.18)	0.85 (0.66-1.11)	0.05 (0.03-0.07)

Table G: Variables used to replicate Table 4: Gluten Introduction Under Conditions of Long-Term Continued Breastfeeding (>1 mo), Short-Term Continued Breastfeeding (Continued ≤ 1 mo After Gluten Introduction), and Discontinued Breastfeeding Before Gluten Introduction and the Risk for CDA and CD

Table Variable	Variable
CDA	htg_conf
CD	celiac_disease
Continued Breastfeeding at Time of Gluten Introduction	ddd

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

Continued Breastfeeding at Time of Gluten Introduction	CDA Manuscript n (%)	CDA DSIC n (%)	Diff.	CDA Manuscript HR (95% CI)	CDA DSIC HR (95% CI)	Diff.
Long-term (continued > 1 mo)	474 (13)	474 (13)	0 (0)	1.23 (1.05-1.44)	1.22 (1.04-1.43)	0.01 (0.01-0.01)
Short-term (continued ≤ 1 mo)	62 (12)	62 (12)	0 (0)	1.08 (0.82-1.44)	1.08 (0.82-1.43)	0 (0-0.01)
Discontinued previous gluten introduction	231 (10)	231 (10)	0 (0)	1	1	0

Continued Breastfeeding at Time of Gluten Introduction	CD Manuscript n (%)	CD DSIC n (%)	Diff.	CD Manuscript HR (95% CI)	CD DSIC HR (95% CI)	Diff.
Long-term (continued > 1 mo)	186 (5)	186 (5)	0 (0)	1.13 (0.88-1.46)	1.14 (0.88-1.46)	0.01 (0-0)
Short-term (continued ≤ 1 mo)	25 (5)	25 (5)	0 (0)	1.07 (0.69-1.67)	1.06 (0.68-1.65)	0.01 (0.01-0.02)
Discontinued previous gluten introduction	93 (4)	93 (4)	0 (0)	1	1	0

Table I: Variables used to replicate Supplementary Table 5: Time to First Introduction to Gluten-Containing Cereals and Risk for CDA and CD According to Country

Table Variable	Variable
CDA	htg_conf
CD	celiac_disease
Gluten Introduction	ugluten_no_oat_4
Country	country

Table J: Comparison of values computed in integrity check to reference article Supplementary Table 5 values

Gluten Introduction	U.S. Manuscript n (%)	U.S. DSIC n (%)	Diff.	U.S. Manuscript HR (95% CI)	U.S. DSIC HR (95% CI)	Diff.
CDA						
< 17 wk	9 (9)	9 (9)	0 (0)	1.06 (0.51-2.19)	1.06 (0.51-2.19)	0 (0-0)
17-26 wk	37 (8)	37 (8)	0 (0)	1	1	0
> 26 wk	182 (9)	182 (9)	0 (0)	1.0 (0.70-1.43)	1.0 (0.70-1.43)	0 (0-0)
CD						
< 17 wk	2 (2)	2 (2)	0 (0)	0.65 (0.12-2.34)	0.54 (0.13-2.35)	0.09 (0.01-0.01)
17-26 wk	17 (4)	17 (4)	0 (0)	1	1	0
> 26 wk	74 (4)	74 (4)	0 (0)	0.88 (0.52-1.50)	0.88 (0.52-1.50)	0 (0-0)

Gluten Introduction	Finland Manuscript n (%)	Finland DSIC n (%)	Diff.	Finland Manuscript HR (95% CI)	Finland DSIC HR (95% CI)	Diff.
CDA						
< 17 wk	6 (13)	6 (13)	0 (0)	1.26 (0.53-2.95)	1.25 (0.53-2.95)	0.01 (0-0)
17-26 wk	46 (10)	46 (10)	0 (0)	1	1	0
> 26 wk	119 (12)	119 (12)	0 (0)	1.14 (0.80-1.60)	1.13 (0.80-1.60)	0.01 (0-0)
CD						
< 17 wk	1 (2)	1 (2)	0 (0)	0.70 (0.09-5.40)	0.70 (0.09-5.41)	0 (0-0.01)
17-26 wk	12 (3)	12 (3)	0 (0)	1	1	0
> 26 wk	39 (4)	39 (4)	0 (0)	1.21 (0.62-2.36)	1.21 (0.62-2.36)	0 (0-0)

Gluten Introduction	Germany Manuscript n (%)	Germany DSIC n (%)	Diff.	Germany Manuscript HR (95% CI)	Germany DSIC HR (95% CI)	Diff.
CDA						
< 17 wk	1 (6)	1 (6)	0 (0)	1.07 (0.12-9.31)	1.07 (0.12-9.31)	0 (0-0)
17-26 wk	5 (9)	5 (9)	0 (0)	1	1	0
> 26 wk	35 (11)	35 (11)	0 (0)	1.12 (0.44-2.88)	1.12 (0.44-2.88)	0 (0-0)
CD						
< 17 wk	0 (0)	0 (0)	0 (0)	NA	NA	NA
17-26 wk	1 (2)	1 (2)	0 (0)	1	1	0
> 26 wk	10 (3)	10 (3)	0 (0)	1.91 (0.24-15.2)	1.91 (0.24-15.0)	0 (0-0.02)

Gluten Introduction	Sweden Manuscript n (%)	Sweden DSIC n (%)	Diff.	Sweden Manuscript HR (95% CI)	Sweden DSIC HR (95% CI)	Diff.
CDA						
< 17 wk	36 (15)	36 (15)	0 (0)	1.00 (0.70-1.42)	1.00 (0.70-1.42)	0 (0-0)
17-26 wk	227 (17)	227 (17)	0 (0)	1	1	0
> 26 wk	70 (15)	70 (15)	0 (0)	0.85 (0.65-1.11)	0.85 (0.65-1.11)	0 (0-0)
CD						
< 17 wk	10 (4)	10 (4)	0 (0)	0.58 (0.30-1.11)	0.58 (0.30-1.11)	0 (0-0)
17-26 wk	111 (8)	111 (8)	0 (0)	1	1	0
> 26 wk	30 (6)	30 (6)	0 (0)	0.72 (0.48-1.08)	0.72 (0.48-1.08)	0 (0-0)

Table K: Variables used to replicate Supplementary Table 6: Gluten Introduction Under Conditions of Long-Term Continued Breastfeeding (>1 mo), Short-Term Continued Breastfeeding (Continued ≤ 1 mo After Gluten Introduction), and Discontinued Breastfeeding Before Gluten Introduction and the Risk for CDA and CD According to Country

Table Variable	Variable
CDA	htg_conf
CD	celiac_disease
Continued Breastfeeding at Time of Gluten Introduction	ddd
Country	country

Table L: Comparison of values computed in integrity check to reference article Table 6 values

Continued BF at Time of Gluten Introduction	U.S. Manuscript n (%)	U.S. DSIC n (%)	Diff.	U.S. Manuscript HR (95% CI)	U.S. DSIC HR (95% CI)	Diff.
CDA	227 (9.0)	227 (9.1)	0 (0.1)	-	-	-
Long-term (continued > 1 mo)	126 (10.4)	126 (10.4)	0 (0)	1.31 (0.99-1.71)	1.27 (0.96-1.66)	0.04 (0.03-0.05)
Short-term (continued ≤ 1 mo)	13 (7.7)	13 (7.7)	0 (0)	0.90 (0.50-1.62)	0.88 (0.49-1.57)	0.02 (0.01-0.05)
Discontinued previous gluten introduction	88 (7.9)	88 (7.9)	0 (0)	1	1	0
CD	93 (3.7)	93 (3.7)	0 (0)	-	-	-
Long-term (continued > 1 mo)	53 (4.4)	53 (4.4)	0 (0)	1.39 (0.91-2.12)	1.32 (0.86-2.02)	0.07 (0.05-0.1)

Continued BF at Time of Gluten Introduction	U.S. Manuscript n (%)	U.S. DSIC n (%)	Diff.	U.S. Manuscript HR (95% CI)	U.S. DSIC HR (95% CI)	Diff.
Short-term (continued ≤ 1 mo)	4 (2.4)	4 (2.4)	0 (0)	0.67 (0.24-1.88)	0.64 (0.23-1.80)	0.03 (0.01-0.08)
Discontinued previous gluten introduction	36 (3.2)	36 (3.2)	0 (0)	1	1	0

Continued BF at Time of Gluten Introduction	Finland Manuscript n (%)	Finland DSIC n (%)	Diff.	Finland Manuscript HR (95% CI)	Finland DSIC HR (95% CI)	Diff.
CDA	170 (11.6)	170 (11.6)	0 (0)	-	-	-
Long-term (continued > 1 mo)	106 (12.0)	106 (12.0)	0 (0)	1.24 (0.87-1.75)	1.26 (0.88-1.78)	0.02 (0.01-0.03)
Short-term (continued ≤ 1 mo)	19 (14.0)	19 (14.0)	0 (0)	1.31 (0.77-2.24)	1.33 (0.78-2.28)	0.02 (0.01-0.04)
Discontinued previous gluten introduction	45 (10.0)	45 (10.0)	0 (0)	1	1	0
CD	52 (3.5)	52 (3.5)	0 (0)	-	-	-
Long-term (continued > 1 mo)	31 (3.5)	31 (3.5)	0 (0)	1.05 (0.56-1.99)	1.05 (0.55-1.98)	0 (0.01-0.01)
Short-term (continued ≤ 1 mo)	7 (5.2)	7 (5.2)	0 (0)	1.29 (0.51-3.22)	1.28 (0.51-3.20)	0.01 (0-0.02)
Discontinued previous gluten introduction	14 (3.1)	14 (3.1)	0 (0)	1	1	0

Continued BF at Time of Gluten Introduction	Germany Manuscript n (%)	Germany DSIC n (%)	Diff.	Germany Manuscript HR (95% CI)	Germany DSIC HR (95% CI)	Diff.
CDA	40 (11.8)	40 (11.1)	0 (0.7)	-	-	-
Long-term (continued > 1 mo)	16 (9.8)	16 (9.8)	0 (0)	0.84 (0.44-1.61)	0.79 (0.41-1.52)	0.05 (0.03-0.09)
Short-term (continued ≤ 1 mo)	3 (8.6)	3 (8.6)	0 (0)	0.83 (0.24-2.86)	0.79 (0.23-2.77)	0.04 (0.01-0.09)
Discontinued previous gluten introduction	21 (13.1)	21 (13.1)	0 (0)	1	1	0
CD	11 (2.9)	11 (2.9)	0 (0)	-	-	-
Long-term (continued > 1 mo)	5 (3.1)	5 (3.1)	0 (0)	0.97 (0.29-3.20)	1.06 (0.30-3.79)	0.09 (0.01-0.59)

Continued BF at Time of Gluten Introduction	Germany Manuscript n (%)	Germany DSIC n (%)	Diff.	Germany Manuscript HR (95% CI)	Germany DSIC HR (95% CI)	Diff.
Short-term (continued ≤ 1 mo)	0 (0)	0 (0)	0 (0)	0	0	0
Discontinued previous gluten introduction	5 (3.1)	5 (3.1)	0 (0)	1	1	0

Continued BF at Time of Gluten Introduction	Sweden Manuscript n (%)	Sweden DSIC n (%)	Diff.	Sweden Manuscript HR (95% CI)	Sweden DSIC HR (95% CI)	Diff.
CDA	330 (16.0)	330 (16.4)	0 (0.4)	-	-	-
Long-term (continued > 1 mo)	226 (17.5)	226 (17.5)	0 (0)	1.23 (0.95-1.59)	1.23 (0.95-1.59)	0 (0-0)
Short-term (continued ≤ 1 mo)	27 (17.0)	27 (17.0)	0 (0)	1.11 (0.72-1.72)	1.11 (0.71-1.72)	0 (0.01-0)
Discontinued previous gluten introduction	77 (13.6)	77 (13.6)	0 (0)	1	1	0
CD	151 (7.4)	151 (7.4)	0 (0)	-	-	-
Long-term (continued > 1 mo)	97 (7.5)	97 (7.5)	0 (0)	0.99 (0.69-1.44)	1.01 (0.69-1.47)	0.02 (0-0.03)
Short-term (continued ≤ 1 mo)	14 (8.8)	14 (8.8)	0 (0)	1.17 (0.64-2.16)	1.19 (0.64-2.19)	0.02 (0-0.03)
Discontinued previous gluten introduction	38 (6.7)	38 (6.7)	0 (0)	1	1	0

Attachment A: SAS Code

```
*** TEDDY M19 DSIC;
*** Programmer: Allyson Mateja;
*** Date: 10/20/16;

proc format;
  value countryf 1 = 'US'
                2 = 'Finland'
                3 = 'Germany'
                4 = 'Sweden';

  value sexf 0 = 'M'
             1 = 'F';

  value hlaf 0 = 'TEDDY other'
             1 = 'DR3-DQ2/DR3-DQ2';

  value birthf 1 = 'Spring (March-May)'
               2 = 'Summer (June-August)'
               3 = 'Fall (September-November)'
               4 = 'Winter (December-February)';

  value eduf 1 = 'Basic primary'
             2 = 'Higher education';

  value glutenf 1 = 'Early (<17 wk)'
                 2 = 'Reference (17-26 wk)'
                 3 = 'Late (>26 wk)';

  value feedf 0 = 'Discontinued previous gluten introduction'
                1 = 'Short-term (continued <= 1 mo)'
                2 = 'Long-term (continued > 1 mo)';

libname m19data '/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_19_Caronsson_NIDDK_Submission/';

data m19_data;
  set m19data.m_19_caronsson_niddk_31july2013;

proc contents data = m19_data;

data m19_data;
  set m19_data;
  time_conftga_yr = time_conftga/365.25;
  timetocd_yr = timetocd/365.25;

proc freq data = m19_data;
  tables celiac_disease htg_conf /list missing;
  title 'Table 1 - CDA and CD numbers';
```

```

proc means data = m19_data n median min max;
  var time_conftga_yr;
  where htg_conf=1;
  title 'Table 1 - Age at end point, CDA';

proc means data = m19_data n median min max;
  var timetocd_yr;
  where celiac_disease=1;
  title 'Table 1 - Age at end point, CD';

proc freq data = m19_data;
  tables country*htg_conf /nocol nopercent;
  format country countryf.;
  title 'Table 1 - Country';

proc phreg data=m19_data;
  class country (ref = 'US');
  model time_conftga*htg_conf(0) = country /risklimits;
  format country countryf.;

proc freq data = m19_data;
  tables country*celiac_disease /nocol nopercent;
  format country countryf.;

proc phreg data=m19_data;
  class country (ref = 'US');
  model timetocd*celiac_disease(0) = country /risklimits;
  format country countryf.;

proc freq data = m19_data;
  tables female*htg_conf /nocol nopercent;
  format female sexf.;
  title 'Table 1 - Gender';

proc phreg data=m19_data;
  class female (ref = 'M');
  model time_conftga*htg_conf(0) = female /risklimits;
  format female sexf.;

proc freq data = m19_data;
  tables female*celiac_disease /nocol nopercent;
  format female sexf.;

proc phreg data=m19_data;
  class female (ref = 'M');
  model timetocd*celiac_disease(0) = female /risklimits;
  format female sexf.;

proc freq data = m19_data;
  tables hla_d*htg_conf /nocol nopercent;

```

```

format hla_d hlaf.;
title 'Table 1 - HLA Genotype';

proc phreg data=m19_data;
class hla_d (ref = 'TEDDY other');
model time_conftga*htg_conf(0) = hla_d /risklimits;
format hla_d hlaf.;

proc freq data = m19_data;
tables hla_d*celiac_disease /nocol nopercnt;
format hla_d hlaf.;

proc phreg data=m19_data;
class hla_d (ref = 'TEDDY other');
model timetocd*celiac_disease(0) = hla_d /risklimits;
format hla_d hlaf.;

proc freq data = m19_data;
tables celiac_fdr*htg_conf /nocol nopercnt;
title 'Table 1 - First degree relative with celiac disease';

proc phreg data=m19_data;
class celiac_fdr (ref = '0');
model time_conftga*htg_conf(0) = celiac_fdr /risklimits;

proc freq data = m19_data;
tables celiac_fdr*celiac_disease /nocol nopercnt;

proc phreg data=m19_data;
class celiac_fdr (ref = '0');
model timetocd*celiac_disease(0) = celiac_fdr /risklimits;

proc freq data = m19_data;
tables birth_season*htg_conf /nocol nopercnt;
format birth_season birthf.;
title 'Table 1 - Season of birth';

proc phreg data=m19_data;
class birth_season (ref = 'Winter (December-February)');
model time_conftga*htg_conf(0) = birth_season /risklimits;
format birth_season birthf.;

proc freq data = m19_data;
tables birth_season*celiac_disease /nocol nopercnt;
format birth_season birthf.;

proc phreg data=m19_data;
class birth_season (ref = 'Winter (December-February)');
model timetocd*celiac_disease(0) = birth_season /risklimits;
format birth_season birthf.;

```

```

proc freq data = m19_data;
  tables smoker*htg_conf /nocol nopercnt;
  title 'Table 1 - Maternal smoking during pregnancy';

proc phreg data=m19_data;
  class smoker (ref = '0');
  model time_conftga*htg_conf(0) = smoker /risklimits;

proc freq data = m19_data;
  tables smoker*celiac_disease /nocol nopercnt;

proc phreg data=m19_data;
  class smoker (ref = '0');
  model timetocd*celiac_disease(0) = smoker /risklimits;

proc freq data = m19_data;
  tables education_mom_group1*htg_conf /nocol nopercnt;
  format education_mom_group1 eduf.;
  title 'Table 1 - Maternal education';

proc phreg data=m19_data;
  class education_mom_group1 (ref = 'Higher education');
  model time_conftga*htg_conf(0) = education_mom_group1 /risklimits;
  format education_mom_group1 eduf.;

proc freq data = m19_data;
  tables education_mom_group1*celiac_disease /nocol nopercnt;
  format education_mom_group1 eduf.;

proc phreg data=m19_data;
  class education_mom_group1 (ref = 'Higher education');
  model timetocd*celiac_disease(0) = education_mom_group1 /risklimits;
  format education_mom_group1 eduf.;

proc means data = m19_data mean std;
  var maternal_age;
  where htg_conf=1;
  title3 'Table 1 - Maternal age at delivery';

proc phreg data=m19_data;
  model time_conftga*htg_conf(0) = maternal_age /risklimits;

proc means data = m19_data mean std;
  var maternal_age;
  where celiac_disease=1;

proc phreg data=m19_data;
  model timetocd*celiac_disease(0) = maternal_age /risklimits;

proc sort data = m19_data;
  by country;

```

```

proc freq data = m19_data;
  tables country;
  format country countryf.;
  title 'Table 2 - Country';

proc means data = m19_data n mean std median p25 p75;
  var gluten_no_oat_w;
  class country;
  format country countryf.;
  title 'Table 2 - Time to gluten introduction';

proc freq data = m19_data;
  tables ugluten_no_oat_4;
  format country countryf. ugluten_no_oat_4 glutenf.;
  by country;
  title 'Table 2 - Category of age of gluten introduction';

proc means data = m19_data n mean std median p25 p75;
  var gluten_no_oat_w;
  class country;
  format country countryf. ugluten_no_oat_4 glutenf.;
  where ugluten_no_oat_4 = 1;

proc means data = m19_data n mean std median p25 p75;
  var gluten_no_oat_w;
  class country;
  format country countryf. ugluten_no_oat_4 glutenf.;
  where ugluten_no_oat_4 = 2;

proc means data = m19_data n mean std median p25 p75;
  var gluten_no_oat_w;
  class country;
  format country countryf. ugluten_no_oat_4 glutenf.;
  where ugluten_no_oat_4 = 3;

proc means data = m19_data n mean std median p25 p75;
  var gluten_no_oat_w;
  class country;
  format country countryf. ugluten_no_oat_4 glutenf.;
  where ugluten_no_oat_4 = 4;

proc means data = m19_data n mean std median p25 p75;
  var exclbtime_w;
  class country;
  format country countryf.;
  title 'Table 2 - Duration of exclusive breastfeeding';

proc means data = m19_data mean std median p25 p75;
  var brsttime_w;
  class country;

```

```

format country countryf.;
title 'Table 2 - Duration of any breastfeeding';

proc freq data = m19_data;
tables u gluten_no_oat_4*htg_conf /nocol nopercnt;
format u gluten_no_oat_4 glutenf.;
title 'Table 3 - Gluten Introduction CDA';

proc phreg data=m19_data;
class u gluten_no_oat_4 (ref = 'Reference (17-26 wk)');
model time_conf tga*htg_conf(0) = u gluten_no_oat_4 country hla_d female celiac_fdr /risklimits;
format u gluten_no_oat_4 glutenf.;

proc freq data = m19_data;
tables u gluten_no_oat_4*celiac_disease /nocol nopercnt;
format u gluten_no_oat_4 glutenf.;
title 'Table 3 - Gluten Introduction CD';

proc phreg data=m19_data;
class u gluten_no_oat_4 (ref = 'Reference (17-26 wk)');
model time_tocd*celiac_disease(0) = u gluten_no_oat_4 country hla_d female celiac_fdr /risklimits;
format u gluten_no_oat_4 glutenf.;

proc freq data = m19_data;
tables ddd*htg_conf /nocol nopercnt;
format ddd feedf.;
title 'Table 4 - Continued Breastfeeding at time of Gluten Introduction CDA';

proc phreg data=m19_data;
class ddd (ref = 'Discontinued previous gluten introduction');
model time_conf tga*htg_conf(0) = ddd country hla_d female celiac_fdr d_gluten_no_oat /risklimits;
format ddd feedf.;

proc freq data = m19_data;
tables ddd*celiac_disease /nocol nopercnt;
format ddd feedf.;
title 'Table 4 - Continued Breastfeeding at time of Gluten Introduction CD';

proc phreg data=m19_data;
class ddd (ref = 'Discontinued previous gluten introduction');
model time_tocd*celiac_disease(0) = ddd country hla_d female celiac_fdr d_gluten_no_oat /risklimits;
format ddd feedf.;

proc sort data = m19_data;
by country;

proc freq data = m19_data;
tables u gluten_no_oat_4*htg_conf /nocol nopercnt;
format u gluten_no_oat_4 glutenf. country countryf.;
by country;
title 'Table 5 - Gluten Introduction CDA';

```



```

proc phreg data=m19_data;
  class u gluten_no_oat_4 (ref = 'Reference (17-26 wk)');
  model time_conftga*htg_conf(0) = u gluten_no_oat_4 hla_d female celiac_fdr /risklimits;
  by country;
  format u glutenf. country countryf.;

proc freq data = m19_data;
  tables u gluten_no_oat_4*celiac_disease /nocol nopercnt;
  format u glutenf. country countryf.;
  by country;
  title 'Table 5 - Gluten Introduction CD';

proc phreg data=m19_data;
  class u gluten_no_oat_4 (ref = 'Reference (17-26 wk)');
  model timetocd*celiac_disease(0) = u gluten_no_oat_4 hla_d female celiac_fdr /risklimits;
  by country;
  format u glutenf. country countryf.;

proc freq data = m19_data;
  tables ddd*htg_conf /nocol;
  by country;
  format ddd feedf. country countryf.;
  title 'Table 6 - Continued Breastfeeding at time of Gluten Introduction CDA';

proc phreg data=m19_data;
  class ddd (ref = 'Discontinued previous gluten introduction');
  model time_conftga*htg_conf(0) = ddd hla_d female celiac_fdr gluten_no_oat_w /risklimits;
  by country;
  format ddd feedf. country countryf.;

proc freq data = m19_data;
  tables ddd*celiac_disease /nocol missing;
  by country;
  format ddd feedf. country countryf.;
  title 'Table 6 - Continued Breastfeeding at time of Gluten Introduction CD';

proc phreg data=m19_data;
  class ddd (ref = 'Discontinued previous gluten introduction');
  model timetocd*celiac_disease(0) = ddd hla_d female celiac_fdr gluten_no_oat_w /risklimits;
  by country;
  format ddd feedf. country countryf.;

```