

# Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Pub48 Steck

**Prepared by Allyson Mateja**

**IMS Inc.**

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

**September 30, 2016**

## 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: Characteristics of TEDDY subjects by antibody and diabetes status.....	4
Table B: Comparison of values computed in integrity check to reference article Table 1 values.....	5
Table C: Variables used to replicate Figure 1: Progression to diabetes in children with confirmed autoantibodies .....	6
Figure A: Comparison of values computed in integrity check to reference article Figure 1 values .....	7
Table D: Variables used to replicate Figure 2: Progression to diabetes in children expressing 1 autoantibody, 2 autoantibodies or 3 autoantibodies by family history .....	9
Figure B: Comparison of values computed in integrity check to reference article Figure 2 values .....	9
Attachment A: SAS Code.....	11

## **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\_48\_asteck\_niddk\_30apr2014\_1.sas7bdat” and “m\_48\_asteck\_niddk\_30apr2014\_2.sas7bdat” datasets.

## **4 Statistical Methods**

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

## 5 Results

For Table 1 in the publication [1], Characteristics of TEDDY subjects by antibody and diabetes status, 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. The results of the replication are an exact match to the published results.

For Figure 1 in the publication [1], Progression to diabetes in children with confirmed autoantibodies, Table C lists the variables that were used in the replication and Figure A compares the results calculated from the archived data files to the results published in Figure 1. The results of the replication are an exact match to the published results.

For Figure 2 in the publication [1], Progression to diabetes in children expressing 1 autoantibody, 2 autoantibodies or 3 autoantibodies by family history, Table D lists the variables that were used in the replication and Figure B compares the results calculated from the archived data files to the results published in Figure 2. The results of the replication are an exact match to published results.

## 6 Conclusions

The NIDDK repository is confident that the TEDDY M48 data files to be distributed are a true copy of the study data.

## 7 References

[1] Steck, A.K., Vehik, K., Bonifacio, E., Lernmark, A., Ziegler, A., Hagopian, W.A., She, J., Simell, O., Akolkar, B., Krischer, J., Schatz, D, Rewers, M., and the TEDDY study group. "Predictors of Progression From the Appearance of Islet Autoantibodies to Early Childhood Diabetes: The Environmental Determinants of Diabetes in the Young (TEDDY)". *Diabetes Care* (2015).

**Table A:** Variables used to replicate Table 1: Characteristics of TEDDY subjects by antibody and diabetes status

<b>Table Variable</b>	<b>dataset.variable</b>
Autoantibody status	m_48_asteck_niddk_30apr2014_1.persist_conf_ab, m_48_asteck_niddk_30apr2014_1.ab_status_grps
Type 1 Diabetes	m_48_asteck_niddk_30apr2014_2.T1D
Male sex, n (%)	m_48_asteck_niddk_30apr2014_1.sex
FDR (%)	m_48_asteck_niddk_30apr2014_1fdr
HLA Category	m_48_asteck_niddk_30apr2014_1.hla_5grps
Follow-up duration (years)	m_48_asteck_niddk_30apr2014_1.followup
Age at first Ab positivity	m_48_asteck_niddk_30apr2014_2.agefirstconfposy
First mIAAs (years)	m_48_asteck_niddk_30apr2014_2.ageconfmiaa
First GADs (years)	m_48_asteck_niddk_30apr2014_2.ageconfgad
First IA-2As (years)	m_48_asteck_niddk_30apr2014_2.ageconfia2a
Age at diabetes diagnosis (years)	m_48_asteck_niddk_30apr2014_2.aget1dy

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

Characteristic	Persistent confirmed Ab+ Manuscript (N=577)	Persistent confirmed Ab+ DSIC (N=577)	Diff. (N=0)	Ab- Manuscript (N=7,195)	Ab- DSIC (N=7,195)	Diff. (N=0)
Male sex, n (%)	326 (57)	326 (57)	0 (0)	3,621 (50)	3,621 (50)	0 (0)
FDR (%)	124 (21)	124 (21)	0 (0)	586 (8)	586 (8)	0 (0)
HLA-DR3/4	293 (51)	293 (51)	0 (0)	2,747 (38)	2,747 (38)	0 (0)
HLA-DR4/4	100 (17)	100 (17)	0 (0)	1,406 (20)	1,406 (20)	0 (0)
HLA-DR3/3	72 (12)	72 (12)	0 (0)	1,561 (22)	1,561 (22)	0 (0)
HLA-DR4/8	86 (15)	86 (15)	0 (0)	1,294 (18)	1,294 (18)	0 (0)
Other HLA genotypes	26 (5)	26 (5)	0 (0)	187 (3)	187 (3)	0 (0)
Follow-up duration (years)	5.73 (4.4-7.2)	5.73 (4.4-7.2)	0 (0-0)	4.81 (2.1-6.6)	4.81 (2.1-6.6)	0 (0-0)
Age at first Ab positivity	2.06 (1.2-3.4)	2.06 (1.2-3.4)	0 (0-0)	-	-	-
First mIAAs (years)	1.96 (1.1-3.2)	1.96 (1.1-3.2)	0 (0-0)	-	-	-
First GADs (years)	2.48 (1.5-3.8)	2.48 (1.5-3.8)	0 (0-0)	-	-	-
First IA-2As (years)	2.56 (1.8-4.0)	2.56 (1.8-4.0)	0 (0-0)	-	-	-
Age at diabetes diagnosis (years)	-	-	-	-	-	-

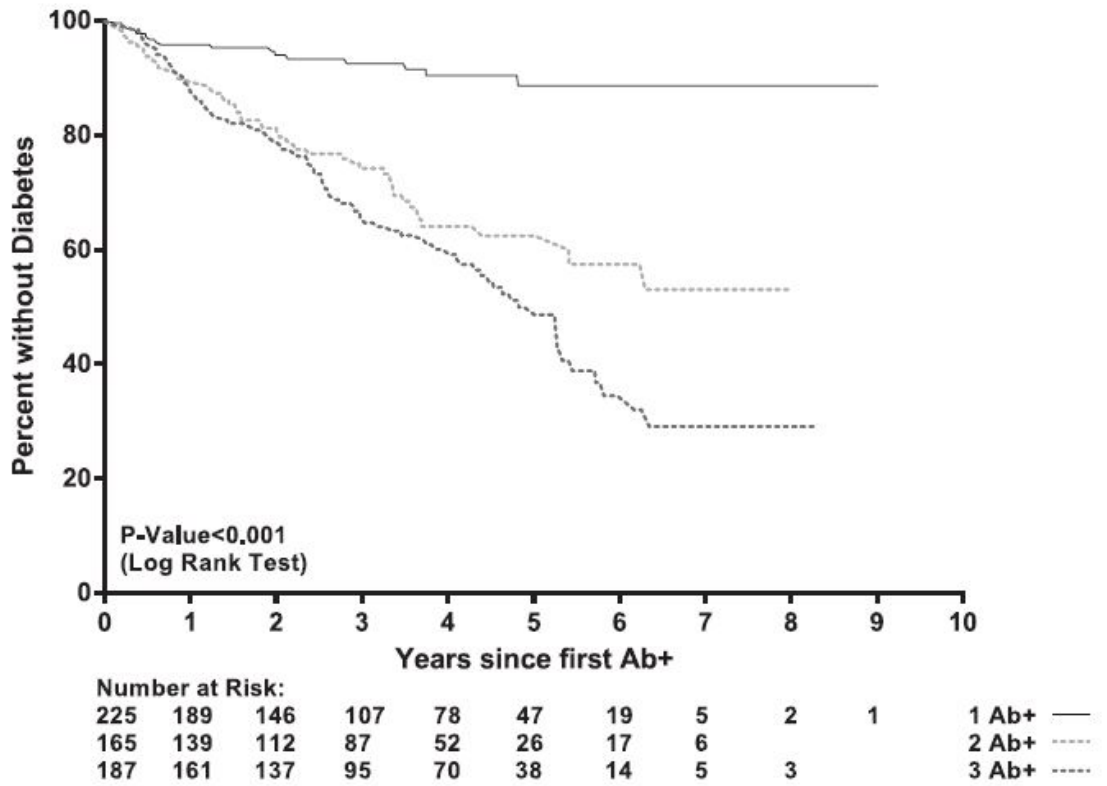
Characteristic	Persistent confirmed Ab+ (no diabetes) Manuscript (N=413)	Persistent confirmed Ab+ (no diabetes) DSIC (N=413)	Diff. (N=0)	Type 1 diabetes Manuscript (N=164)	Type 1 diabetes DSIC (N=164)	Diff. (N=0)
Male sex, n (%)	236 (57)	236 (57)	0 (0)	90 (55)	90 (55)	0 (0)
FDR (%)	73 (18)	73 (18)	0 (0)	51 (31)	51 (31)	0 (0)
HLA-DR3/4	200 (48)	200 (48)	0 (0)	93 (57)	93 (57)	0 (0)
HLA-DR4/4	77 (19)	77 (19)	0 (0)	23 (14)	23 (14)	0 (0)
HLA-DR3/3	57 (14)	57 (14)	0 (0)	15 (9)	15 (9)	0 (0)
HLA-DR4/8	65 (16)	65 (16)	0 (0)	21 (13)	21 (13)	0 (0)
Other HLA genotypes	14 (3)	14 (3)	0 (0)	12 (7)	12 (7)	0 (0)
Follow-up duration (years)	6.36 (5.1-7.5)	6.36 (5.1-7.5)	0 (0-0)	3.35 (2.1-5.2)	3.35 (2.1-5.2)	0 (0-0)
Age at first Ab positivity	2.52 (1.5-4.0)	2.52 (1.5-4.0)	0 (0-0)	1.28 (0.9-2.0)	1.28 (0.9-2.0)	0 (0-0)
First mIAAs (years)	2.46 (1.3-3.6)	2.46 (1.3-3.6)	0 (0-0)	1.29 (1.0-2.0)	1.29 (1.0-2.0)	0 (0-0)
First GADs (years)	3.02 (2.0-4.3)	3.02 (2.0-4.3)	0 (0-0)	1.54 (1.1-2.3)	1.54 (1.1-2.3)	0 (0-0)
First IA-2As (years)	3.58 (2.5-4.9)	3.58 (2.5-4.9)	0 (0-0)	2.00 (1.3-2.7)	2.00 (1.3-2.7)	0 (0-0)
Age at diabetes diagnosis (years)	-	-	-	3.30 (2.0-5.1)	3.30 (2.0-5.1)	0 (0-0)

**Table C:** Variables used to replicate Figure 1: Progression to diabetes in children with confirmed autoantibodies

Chart Variable	dataset.variable
Autoantibody group	m_48_asteck_niddk_30apr2014_2.pnumab
Years since first Ab+	m_48_asteck_niddk_30apr2014_2.timesincefirst
Type 1 Diabetes	m_48_asteck_niddk_30apr2014_2.T1D

**Figure A:** Comparison of values computed in integrity check to reference article Figure 1 values

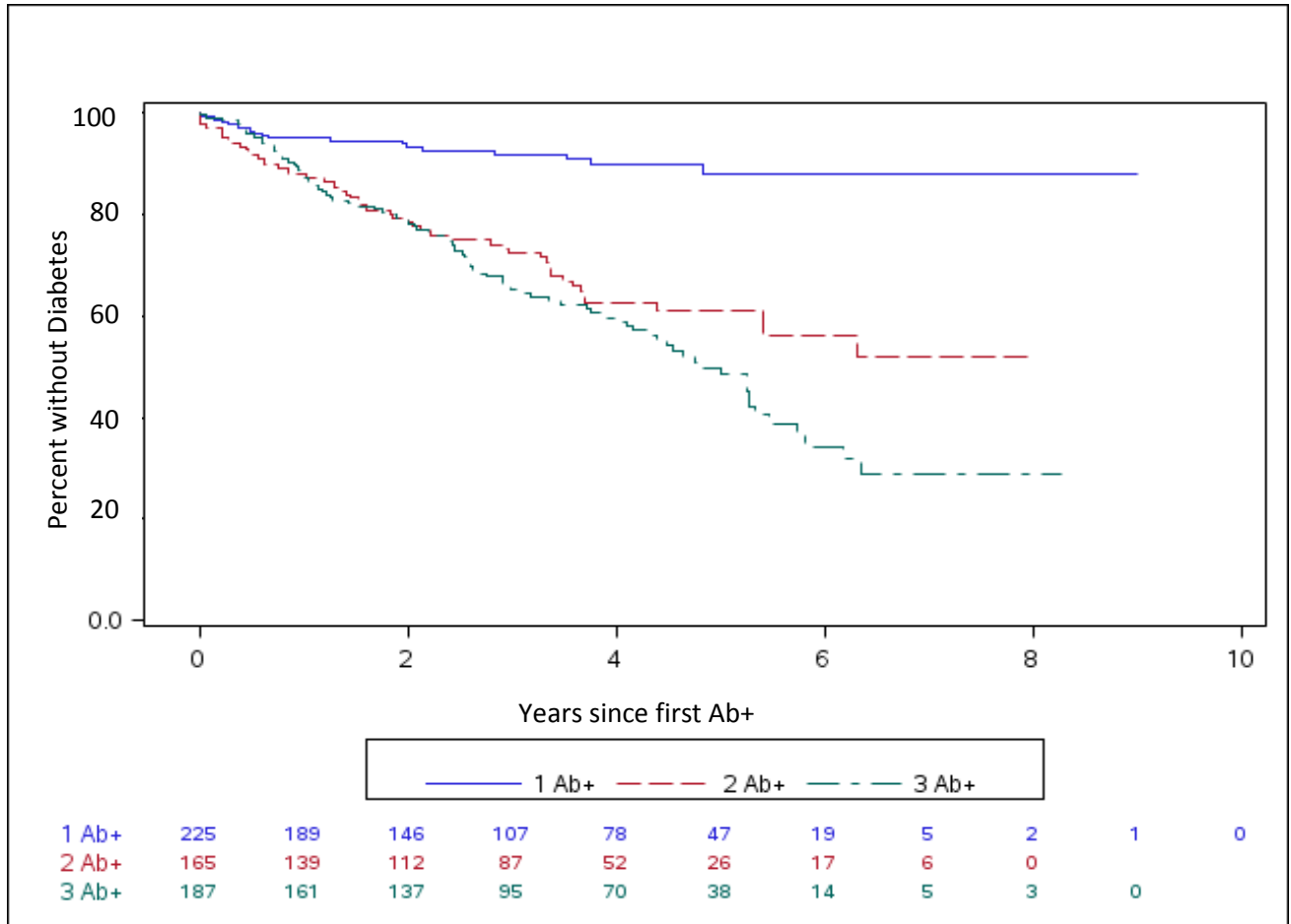
Manuscript



**Figure 1**—Progression to diabetes in children with confirmed autoantibodies ( $N = 577$ ). Ab+, antibody positive.



DSIC

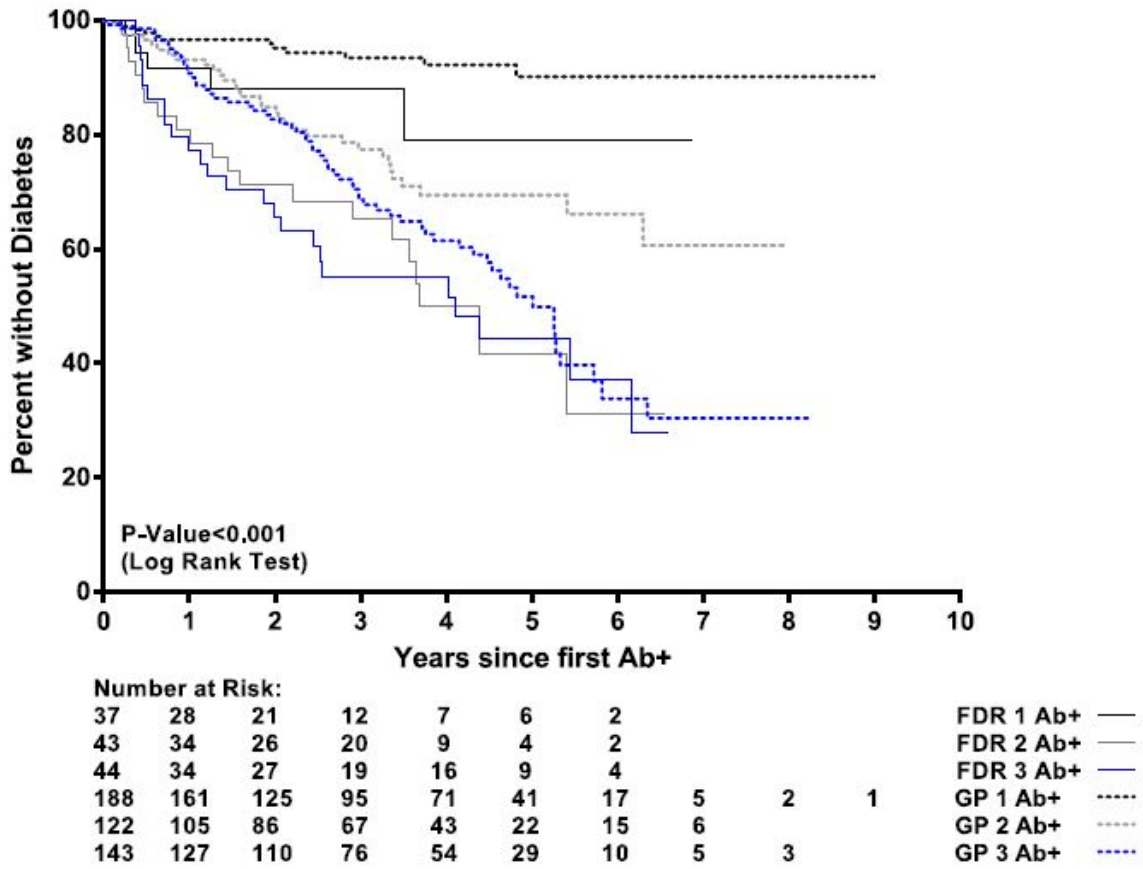


**Table D:** Variables used to replicate Figure 2: Progression to diabetes in children expressing 1 autoantibody, 2 autoantibodies or 3 autoantibodies by family history

Chart Variable	dataset.variable
Autoantibody and family history group	m_48_asteck_niddk_30apr2014_2.fdrab
Years since first Ab+	m_48_asteck_niddk_30apr2014_2.timesincefirst
Type 1 Diabetes	m_48_asteck_niddk_30apr2014_2.T1D

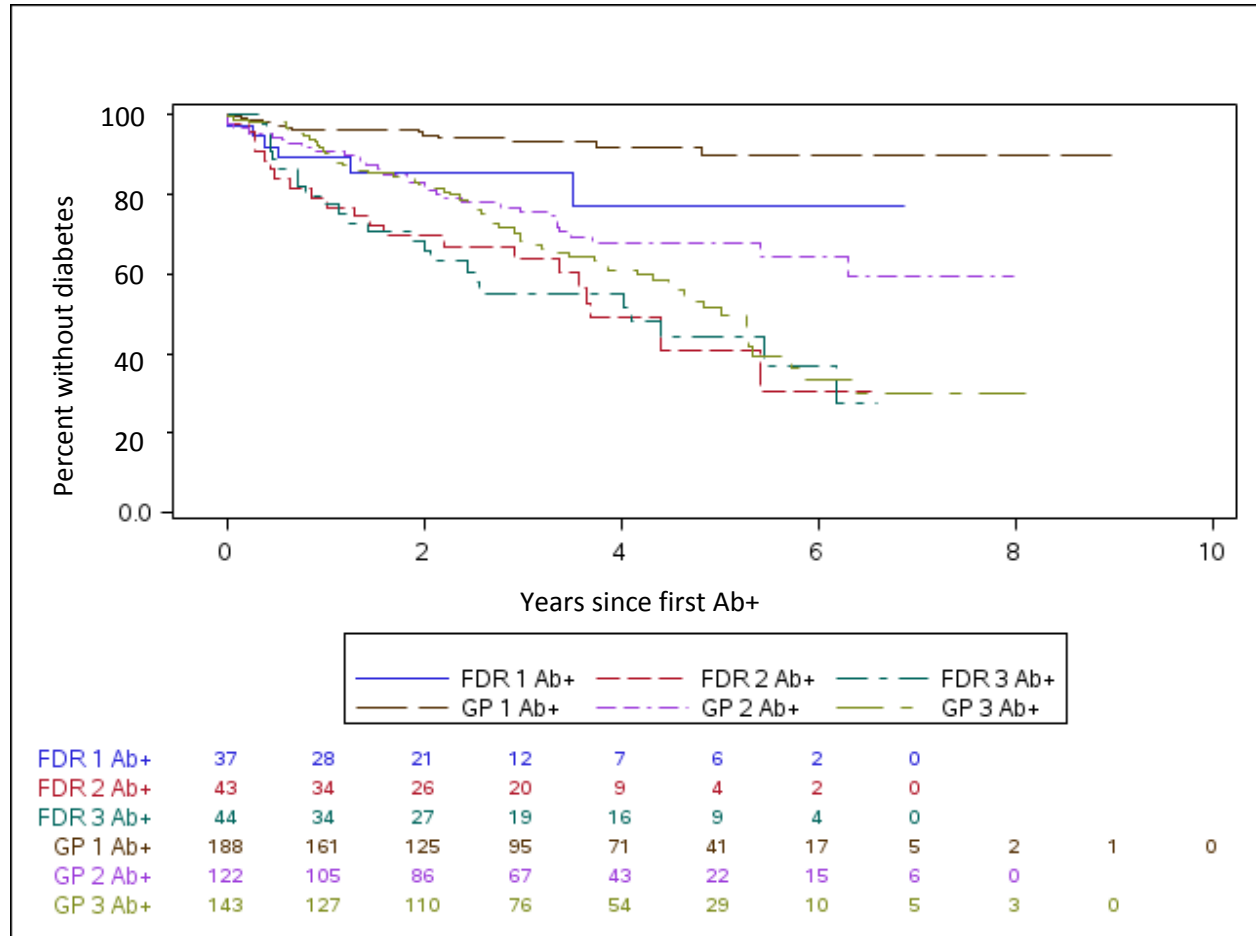
**Figure B:** Comparison of values computed in integrity check to reference article Figure 2 values

Manuscript



**Figure 2—**Progression to diabetes in children expressing 1 autoantibody, 2 autoantibodies or 3 autoantibodies by family history. Ab+, antibody positive; GP, general population.

DSIC



## Attachment A: SAS Code

```
*** TEDDY M48 Analysis DSIC;
*** Date: 9/29/16;
*** Programmer: Allyson Mateja;

libname sas_data '/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_48_ASteck_NIDDK_Submission/';

proc format;
  value groupf 1 = 'Persistent confirmed Ab+'
    2 = 'Ab-';

  value hlaf 1 = 'HLA-DR3/4'
    2 = 'HLA-DR4/4'
    3 = 'HLA-DR4/8'
    4 = 'Other HLA genotypes'
    5 = 'HLA-DR3/3';

  value abnumf 1 = '1 Ab+'
    2 = '2 Ab+'
    3 = '3 Ab+';

  value fdrabnumf 1 = 'FDR 1 Ab+'
    2 = 'FDR 2 Ab+'
    3 = 'FDR 3 Ab+'
    4 = 'GP 1 Ab+'
    5 = 'GP 2 Ab+'
    6 = 'GP 3 Ab+';

data m48_data1;
  set sas_data.m_48_asteck_niddk_30apr2014_1;

data m48_data2;
  set sas_data.m_48_asteck_niddk_30apr2014_2;

data m48_data3;
  set sas_data.m_48_asteck_niddk_30apr2014_3;

proc contents data = m48_data1;
proc contents data = m48_data2;
proc contents data = m48_data3;

proc freq data = m48_data1;
  tables persist_conf_ab*ab_status_grps /list;

data m48_data1;
  set m48_data1;
  if persist_conf_ab = 0 and ab_status_grps = 7 then group = 2;
  else if persist_conf_ab = 1 then group = 1;
```

```

        else group = .;
        if group ne .;

proc sort data = m48_data1;
    by maskid;

proc sort data = m48_data2;
    by maskid;

data m48_data1;
    merge m48_data1 (in=vall)
          m48_data2 (keep=maskid T1D ageconfgad ageconfia2a ageconfmiaa agefirstconfposy agetldy);
    by maskid;
    if vall then output;

proc freq data = m48_data1;
    tables group;
    format group groupf.;
    title 'Table 1 - Group';

proc sort data = m48_data1;
    by group;

proc freq data = m48_data1;
    tables sex;
    by group;
    format group groupf.;
    title 'Table 1 - Sex';

proc freq data = m48_data1;
    tables fdr;
    by group;
    format group groupf.;
    title 'Table 1 - FDR';

proc freq data = m48_data1;
    tables hla_5grps;
    by group;
    format hla_5grps hlaf. group groupf.;
    title 'Table 1 - HLA Category';

proc means data = m48_data1 n median p25 p75;
    var followup;
    class group;
    format group groupf.;
    title 'Table 1 - Follow-up duration (years)';

proc means data = m48_data1 n median p25 p75;
    var agefirstconfposy;
    where group = 1;
    format group groupf.;

```

```

        title 'Table 1 - Age at first Ab positivity';

proc means data = m48_data1 n median p25 p75;
    var ageconfmiaa;
    where group = 1;
    format group groupf.;
    title 'Table 1 - First mIAAs (years)';

proc means data = m48_data1 n median p25 p75;
    var ageconfgad;
    where group = 1;
    format group groupf.;
    title 'Table 1 - First GADs (years)';

proc means data = m48_data1 n median p25 p75;
    var ageconfia2a;
    where group = 1;
    format group groupf.;
    title 'Table 1 - First IA-2As (years)';

proc freq data = m48_data1;
    tables t1d;
    where group = 1;
    title 'Table 1 - Diabetes status for Ab+';

data ab_positive;
    set m48_data1;
    if group = 1;

proc sort data = ab_positive;
    by t1d;

proc freq data = ab_positive;
    tables sex;
    by t1d;
    title 'Table 1 - Sex, Ab+ by diabetes status';

proc freq data = ab_positive;
    tables fdr;
    by t1d;
    title 'Table 1 - FDR, Ab+ by diabetes status';

proc freq data = ab_positive;
    tables hla_5grps;
    by t1d;
    format hla_5grps hlaf.;
    title 'Table 1 - HLA Category, Ab+ by diabetes status';

proc means data = ab_positive n median p25 p75;
    var followup;
    class t1d;

```

```

        title 'Table 1 - Follow-up duration (years), Ab+ by diabetes status';

proc means data = ab_positive n median p25 p75;
    var agefirstconfposy;
    class tld;
    title 'Table 1 - Age at first Ab positivity, Ab+ by diabetes status';

proc means data = ab_positive n median p25 p75;
    var ageconfmiaa;
    class tld;
    title 'Table 1 - First mIAAs (years), Ab+ by diabetes status';

proc means data = ab_positive n median p25 p75;
    var ageconfgad;
    class tld;
    title 'Table 1 - First GADs (years), Ab+ by diabetes status';

proc means data = ab_positive n median p25 p75;
    var ageconfia2a;
    class tld;
    title 'Table 1 - First IA-2As (years), Ab+ by diabetes status';

proc means data = ab_positive n median p25 p75;
    var agetldy;
    where tld=1;
    title 'Table 1 - Age at diabetes diagnosis (years)';

data m48_data2;
    set m48_data2;
    time_years = timesincefirst/365.25;
    if fdrab = 1 then fdr_ab=1;
    if fdrab = 3 then fdr_ab=2;
    if fdrab = 5 then fdr_ab=3;
    if fdrab = 2 then fdr_ab=4;
    if fdrab = 4 then fdr_ab=5;
    if fdrab = 6 then fdr_ab = 6;

ods graphics on;

PROC LIFETEST DATA=m48_data2 plots=survival(atrisk(outside)=0 to 10 by 1 nocensor );
TIME time_years * T1D (0);
strata pnumab;
test pnumab;
format pnumab abnumf.;
RUN;

PROC LIFETEST DATA=m48_data2 plots=survival(atrisk(outside)=0 to 10 by 1 nocensor );
TIME time_years * T1D (0);
strata fdr_ab;
test fdr_ab;

```

```
format fdr_ab fdrabnumf.;  
RUN;
```