

Dataset Integrity Check for Joint
modeling of longitudinal autoantibody
patterns and progression to type 1
diabetes: results from the TEDDY study

Prepared by Anne Taylor
IMS Inc.

3901 Calverton Blvd, Suite 200 Calverton, MD 20705
May 08, 2019

Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	2
6 Conclusions	2
7 References	3
Table A: Variables used to replicate Table 1: Description of the study population by type of persistent autoantibody.....	3
Table B: Comparison of values computed in integrity check to reference article Table 1 values.....	5
Attachment A: SAS Code.....	8

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

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_127_MKoehler_NIDDK_Submission/ folder in the data package. For this replication, variables were taken from the “m_127_mkoehler_niddk_31dec2014_1.sas7bdat”, “m_127_mkoehler_niddk_31dec2014_2.sas7bdat”, “m_127_mkoehler_niddk_31dec2014_3.sas7bdat”, and “m_127_mkoehler_niddk_31dec2014_4.sas7bdat” datasets.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Description of the study population by type of persistent autoantibody, 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 data files to be distributed are a true copy of the study data.

7 References

[1] Meike Köhler, Andreas Beyerlein, Kendra Vehik, Sonja Greven, Nikolaus Umlauf, Åke Lernmark, William A. Hagopian, Marian Rewers, Jin-Xiong She, Jorma Toppari¹, Beena Akolkar, Jeffrey P. Krischer, Ezio Bonifacio, Anette-G. Ziegler, TEDDY study group. Joint modeling of longitudinal autoantibody patterns and progression to type 1 diabetes: results from the TEDDY study. *Acta Diabetol.* 2017 Jul 22. DOI 10.1007/s00592-017-1033-7.

Table A: Variables used to replicate Table 1: Description of the study population by type of persistent autoantibody

Table Variable	dataset.variable
Age at respective seroconversion (years)	m_127_mkoehler_niddk_31dec2014_1.ab_age m_127_mkoehler_niddk_31dec2014_2.miaa_age m_127_mkoehler_niddk_31dec2014_3.gad_age m_127_mkoehler_niddk_31dec2014_4.ia2a_age
Girls	m_127_mkoehler_niddk_31dec2014_*.sex
Country	m_127_mkoehler_niddk_31dec2014_*.country
Child having a first-degree relative with T1D	m_127_mkoehler_niddk_31dec2014_*.fdr
HLA-DR genotype – DR3/4	m_127_mkoehler_niddk_31dec2014_*.dr34
HLA-DR genotype – DR4/4	m_127_mkoehler_niddk_31dec2014_*.dr44
HLA-DR genotype – DR4/8	m_127_mkoehler_niddk_31dec2014_*.dr48
HLA-DR genotype – DR3/3	m_127_mkoehler_niddk_31dec2014_*.dr33
HLA-DR genotype – Other	m_127_mkoehler_niddk_31dec2014_*.drrest
Additionally autoantibody positive for IAA	m_127_mkoehler_niddk_31dec2014_3.miaa_pos_t m_127_mkoehler_niddk_31dec2014_4.miaa_pos_t
Additionally autoantibody positive for GADA	m_127_mkoehler_niddk_31dec2014_2.gad_pos_t m_127_mkoehler_niddk_31dec2014_4.gad_pos_t
Additionally autoantibody positive for IA2A	m_127_mkoehler_niddk_31dec2014_2.ia2a_pos_t m_127_mkoehler_niddk_31dec2014_3.ia2a_pos_t
Autoantibody present at first seroconversion	m_127_mkoehler_niddk_31dec2014_2.miaa_first m_127_mkoehler_niddk_31dec2014_3.gad_first m_127_mkoehler_niddk_31dec2014_4.ia2a_first
Number of children who developed T1D	m_127_mkoehler_niddk_31dec2014_*.t1d

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

Type of persistent autoantibody	Variable	Manuscript (n=613)	DSIC (n=613)	Diff. (n=0)
Total	Total number of children	613	613	0
	Age at respective seroconversion (years)	2.2 (1.2, 3.8)	2.3 (1.3, 3.8)	0.1 (0.1, 0.0)
	Girls	268 (44%)	268 (44%)	0 (0%)
	Country – US	206 (34%)	206 (34%)	0 (0%)
	Country – Finland	153 (25%)	153 (25%)	0 (0%)
	Country – Germany	47 (8%)	47 (8%)	0 (0%)
	Country – Sweden	207 (34%)	207 (34%)	0 (0%)
	Child having a first-degree relative with T1D	128 (21%)	128 (21%)	0 (0%)
	HLA-DR genotype – DR3/4	311 (51%)	311 (51%)	0 (0%)
	HLA-DR genotype – DR4/4	106 (17%)	106 (17%)	0 (0%)
	HLA-DR genotype – DR4/8	92 (15%)	92 (15%)	0 (0%)
	HLA-DR genotype – DR3/3	76 (12%)	76 (12%)	0 (0%)
	HLA-DR genotype – Other	28 (5%)	28 (5%)	0 (0%)
	Number of children who developed T1D	175 (29%)	175 (29%)	0 (0%)
IAA	Total number of children	442	442	0
	Age at respective seroconversion (years)	2.0 (1.1, 3.5)	2.0 (1.1, 3.5)	0.0 (0.0, 0.0)
	Girls	200 (45%)	200 (45%)	0 (0%)
	Country – US	136 (31%)	136 (31%)	0 (0%)
	Country – Finland	125 (28%)	125 (28%)	0 (0%)
	Country – Germany	40 (9%)	40 (9%)	0 (0%)
	Country – Sweden	141 (32%)	141 (32%)	0 (0%)
	Child having a first-degree relative with T1D	105 (24%)	105 (24%)	0 (0%)
	HLA-DR genotype – DR3/4	241 (55%)	241 (55%)	0 (0%)

Type of persistent autoantibody	Variable	Manuscript (n=613)	DSIC (n=613)	Diff. (n=0)
	HLA-DR genotype – DR4/4	74 (17%)	74 (17%)	0 (0%)
	HLA-DR genotype – DR4/8	71 (16%)	71 (16%)	0 (0%)
	HLA-DR genotype – DR3/3	30 (7%)	30 (7%)	0 (0%)
	HLA-DR genotype – Other	26 (6%)	26 (6%)	0 (0%)
	Additionally autoantibody positive for GADA	302 (68%)	302 (68%)	0 (0%)
	Additionally autoantibody positive for IA2A	252 (57%)	252 (57%)	0 (0%)
	Autoantibody present at first seroconversion	353 (80%)	353 (80%)	0 (0%)
	Number of children who developed T1D	162 (37%)	162 (37%)	0 (0%)
GADA	Total number of children	466	466	0
	Age at respective seroconversion (years)	2.7 (1.6, 4.2)	2.7 (1.6, 4.2)	0.0 (0.0, 0.0)
	Girls	212 (45%)	212 (45%)	0 (0%)
	Country – US	166 (36%)	166 (36%)	0 (0%)
	Country – Finland	109 (23%)	109 (23%)	0 (0%)
	Country – Germany	32 (7%)	32 (7%)	0 (0%)
	Country – Sweden	159 (34%)	159 (34%)	0 (0%)
	Child having a first-degree relative with T1D	97 (21%)	97 (21%)	0 (0%)
	HLA-DR genotype – DR3/4	251 (54%)	251 (54%)	0 (0%)
	HLA-DR genotype – DR4/4	81 (17%)	81 (17%)	0 (0%)
	HLA-DR genotype – DR4/8	51 (11%)	51 (11%)	0 (0%)
	HLA-DR genotype – DR3/3	64 (14%)	64 (14%)	0 (0%)
	HLA-DR genotype – Other	19 (4%)	19 (4%)	0 (0%)
	Additionally autoantibody positive for IAA	302 (65%)	302 (65%)	0 (0%)
	Additionally autoantibody positive for IA2A	237 (51%)	237 (51%)	0 (0%)
	Autoantibody present at first seroconversion	344 (74%)	344 (74%)	0 (0%)
	Number of children who developed T1D	134 (29%)	134 (29%)	0 (0%)
IA2A	Total number of children	288	288	0
	Age at respective seroconversion (years)	2.8 (1.9, 4.5)	2.8 (1.9, 4.5)	0.0 (0.0, 0.0)

Type of persistent autoantibody	Variable	Manuscript (n=613)	DSIC (n=613)	Diff. (n=0)
	Girls	112 (39%)	112 (39%)	0 (0%)
	Country – US	94 (33%)	94 (33%)	0 (0%)
	Country – Finland	85 (30%)	85 (30%)	0 (0%)
	Country – Germany	22 (8%)	22 (8%)	0 (0%)
	Country – Sweden	87 (30%)	87 (30%)	0 (0%)
	Child having a first-degree relative with T1D	71 (25%)	71 (25%)	0 (0%)
	HLA-DR genotype – DR3/4	148 (51%)	148 (51%)	0 (0%)
	HLA-DR genotype – DR4/4	64 (22%)	64 (22%)	0 (0%)
	HLA-DR genotype – DR4/8	46 (16%)	46 (16%)	0 (0%)
	HLA-DR genotype – DR3/3	16 (6%)	16 (6%)	0 (0%)
	HLA-DR genotype – Other	14 (5%)	14 (5%)	0 (0%)
	Additionally autoantibody positive for IAA	252 (88%)	252 (88%)	0 (0%)
	Additionally autoantibody positive for GADA	237 (83%)	237 (82%)	0 (1%)
	Autoantibody present at first seroconversion	40 (14%)	40 (14%)	0 (0%)
	Number of children who developed T1D	127 (44%)	127 (44%)	0 (0%)

Attachment A: SAS Code

```

/*****
TEDDY M_127 MKoehler
Saved As: /prj/niddk/ims_analysis/TEDDY/prog_initial_analysis/m_127_mkoehler.check.table1.sas
Programmer: Anne Taylor
Purpose: To check Table 1 of Joint modeling of longitudinal autoantibody patterns and progression
         to type 1 diabetes: results from the TEDDY study.
*****/

options validvarname=upcase mprint;

title 'TEDDY M_127 MKoehler';
title2 "Program Saved As: %sysfunc(getoption(sysin))";

libname m_127 '/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_127_MKoehler_NIDDK_Submission';

proc format;
  value country
    1='US'
    2='Finland'
    3='Germany'
    4='Sweden'
  ;
  value yesno
    0='No'
    1='Yes'
  ;

%macro check(dsn,pop,ab,addab1,addab2);

  proc sort data=m_127.m_127_mkoehler_niddk_31dec2014_&dsn nodupkey out=&pop;
    by mask_id &ab._age sex country fdr dr34 dr44 dr48 dr33 drrest tld;

  data &pop.2;
    set &pop;
    by mask_id;
    if not first.mask_id then abort;

    &ab._age_yrs=&ab._age/12; ***convert age at seroconversion from months to years***;

  keep mask_id &ab._age_yrs sex country fdr dr34 dr44 dr48 dr33 drrest tld;

  label &ab._age_yrs='Age at respective seroconversion (years)'
    fdr='Child having a first-degree relative with T1D'
    dr34='HLA-DR genotype - DR3/4'
    dr44='HLA-DR genotype - DR4/4'
    dr48='HLA-DR genotype - DR4/8'
    dr33='HLA-DR genotype - DR3/3'
    drrest='HLA-DR genotype - Other'

```

```

        tld='Developed T1D'
        ;

proc means data=&pop.2 n nmiss median p25 p75 maxdec=1;
  var &ab._age_yrs;
  title4 'Table 1: Description of the study population by type of persistent autoantibody';
  title5 "&pop";

proc freq data=&pop.2;
  table sex country fdr dr34 dr44 dr48 dr33 drrest/missprint nocum;
  format country country.
         fdr dr34 dr44 dr48 dr33 drrest yesno.
  ;

%if &pop ne Total %then %do;

  ***pull additional autoantibody flags from last record***;
  data &pop.3;
  set m_127.m_127_mkoehler_niddk_31dec2014_&dsn;
  by mask_id &ab._first obstime2;
  if last.&ab._first and not last.mask_id then abort;
  if not first.obstime2 then abort;
  retain &addab1 &addab2;
  if first.mask_id then do;
    &addab1=0;
    &addab2=0;
  end;
  if &addab1._pos_t=1 then &addab1=1;
  else if &addab1=1 then abort;
  if &addab2._pos_t=1 then &addab2=1;
  else if &addab2=1 then abort;
  if last.mask_id then output;
  keep mask_id &addab1 &addab2 &ab._first;
  %if &pop ne IAA %then %do;
    label miaa='Additionally autoantibody positive for IAA';
  %end;
  %if &pop ne GADA %then %do;
    label gad='Additionally autoantibody positive for GADA';
  %end;
  %if &pop ne IA2A %then %do;
    label ia2a='Additionally autoantibody positive for IA2A';
  %end;
  label &ab._first='Autoantibody present at first seroconversion';

proc freq data=&pop.3;
  table &addab1 &addab2 &ab._first/missprint nocum;
  format &addab1 &addab2 &ab._first yesno.;
run;

%end;

```

```
proc freq data=&pop.2;
  table t1d/missprint nocum;
  format t1d yesno.;
run;

%mend check;

%check(1,Total,ab);
%check(2,IAA,miaa,gad,ia2a);
%check(3,GADA,gad,miaa,ia2a);
%check(4,IA2A,ia2a,miaa,gad);
```