Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) M90a Larsson

Prepared by Sabrina Chen
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
February 23, 2020

Contents

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/private_orig_data/M_90a_HLarsson_NIDDK_Submission folder in the data package. For this replication, variables were taken from the “m_90a_hlarsson_niddk_31july2016.sas7bdat” dataset.

4 Statistical Methods

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

5 Results

For **Comparison of Data in the publication**, 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.

6 Conclusions

The results of the replication are almost an exact match to the published results. The DCC confirmed that there was a typo in the Total group for the "(2) 24 upto 36 months" category as noted in Table B below.

7 References

[1] Helena Elding Larsson, Kristian F. Lynch, Maria Lönnrot, Michael J. Haller, Åke Lernmark, William A. Hagopian, Jin-Xiong She, Olli Simell, Jorma Toppari, Anette-G. Ziegler, Beena Akolkar, Jeffrey P. Krischer, Marian J. Rewers, Heikki Hyöty, for the TEDDY Study Group. Pandemrix® vaccination is not associated with increased risk of islet autoimmunity or type 1 diabetes in the TEDDY study children. Diabetologia (2018) 61:193–202.

Table A: Variables used to replicate data in the publication.

Table Variable	**dataset.variable**
Country	m_90a_hlarsson_niddk_31july2016.country
Sex	m_90a_hlarsson_niddk_31july2016.sex
Family History	m_90a_hlarsson_niddk_31july2016.fdr
HLA-DR	m_90a_hlarsson_niddk_31july2016.hla_dr_5grps
Age on March 1 2010	m_90a_hlarsson_niddk_31july2016.age_mar2010_5grps
serconverted for any of the three Islet Autoantibodies (GADA, IA2A and IAA) by 10 years of age ?	m_90a_hlarsson_niddk_31july2016.ia_any_10yr
Child was still at risk of Islet Autoantibody seroconverson as of October 1st 2009	m_90a_hlarsson_niddk_31july2016.ia_atrisk_yes
serconverted for second Islet Autoantibodies by 10 years of age ?	m_90a_hlarsson_niddk_31july2016.multiple_ia_10yr
Child was still at risk of multiple Islet Autoantibody seroconverson as of October 1st 2009	m_90a_hlarsson_niddk_31july2016.multi_ia_atrisk_yes
0= type 1 diagnosis not recorded, 1 = diagnosed with type 1 diabetes by 10 years of age	m_90a_hlarsson_niddk_31july2016.t1d_10yr

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

	Total			H1N1 vaccination in relation to risk of islet autoantibodies			H1N1 vaccination in relation to risk of multiple islet autoantibodies			H1N1 vaccination in relation to risk of type 1 diabetes		
	Manuscript	**DSIC**	**Diff**	**Manuscript**	**DSIC**	**Diff**	**Manuscript**	**DSIC**	**Diff**	**Manuscript**	**DSIC**	**Diff**
All Children	3401	3401	0	232	232	0	148	148	0	96	96	0
Finland	1438	1438	0	97	97	0	67	67	0	49	49	0
Sweden	1963	1963	0	135	135	0	81	81	0	47	47	0
Female	1659	1659	0	103	103	0	63	63	0	46	46	0
Male	1742	1742	0	129	129	0	85	85	0	50	50	0
GenPop	3123	3123	0	199	199	0	122	122	0	78	78	0
FDR	278	278	0	33	33	0	26	26	0	18	18	0
(2) HLA-DR-DQ 3-2/X	665	665	0	41	41	0	27	27	0	20	20	0
(3) DR-DQ 3-2/4-8	695	695	0	47	47	0	25	25	0	13	13	0
(1) HLA-DR-DQ 3-2/3-2	1298	1298	0	111	111	0	80	80	0	53	53	0
(5) DR-DQ 4-8/4-8	649	649	0	28	28	0	12	12	0	7	7	0
(0) less than 12 months	752	752	0	71	71	0	44	44	0	26	26	0
(1) 12 upto <24 months	646	646	0	50	50	0	33	33	0	16	16	0
(2) 24 upto 36 months	649	694	-45	43	43	0	25	25	0	23	23	0
(3) 36 upto 48 months	587	587	0	34	34	0	23	23	0	17	17	0
(4) 48 months or older	722	722	0	34	34	0	23	23	0	14	14	0

Attachment A: SAS Code

```
options nocenter validvarname=upcase;

title '/prj/niddk/ims_analysis/TEDDY/prog_initial_analysis/m_90a_dsic.sas';
run;

libname dat '/prj/niddk/ims_analysis/TEDDY/private_orig_data/M_90a_HLarsson_NIDDK_Submission';


proc format;
 value val
 .     = "no value"
 other = "   value"
 ;

 value oneplus
 . = "no value"
 0 = "0"
 1-high = "1+"
 ;

 value zerohi
 . = "no value"
 0-high = "0-high"
 ;

 value country
 2 = 'Finland'
 4 = 'Sweden'
 ;

 value sexf
 0 = 'Male'
 1 = 'Female'
 ;

 value fdrf
 1='FDR'
 0='GenPop'
 ;

 value hladrf
```

```
 1 = '(1) HLA-DR-DQ 3-2/3-2'
 2 = '(2) HLA-DR-DQ 3-2/X'
 3 = '(3) DR-DQ 3-2/4-8'
 4 = '(4) DR-DQ 4-8/X'
 5 = '(5) DR-DQ 4-8/4-8'
 ;

 value agemarch
 0 = '(0) less than 12 months'
 1 = '(1) 12 upto <24 months'
 2 = '(2) 24 upto 36 months'
 3 = '(3) 36 upto 48 months'
 4 = '(4) 48 months or older'
 ;
run;



* produce n and %;
%macro npercent(rownum, var, varf, subset,  subsetname);
  proc freq data=analy noprint;
    where &subset = 1;
    tables &var/list missing out=tbl1&subsetname;
    format &var &varf..;
  run;

  data tbl1&subsetname;
    length covar covarf $100;
    set tbl1&subsetname;
    covar = "&var";
    covarf = put(&var,&varf..);
    rownum = &rownum;
  run;

  data prnt&subsetname;
    set prnt&subsetname tbl1&subsetname;
  run;

%mend;


%macro univ(rownum, var, subset, supplement, subsetname);

  proc univariate data=analy outtable= univ&subsetname noprint;
    where &subset=1 and &supplement=1 and &var not in(.,0);
```

```
    var &var
       ;
  run;

  data univ&subsetname;
    length covarf $100;
    set univ&subsetname;
        covarf = "&subset";
        rownum = &rownum;
  run;

  data prntuniv&subsetname;
    set prntuniv&subsetname univ&subsetname;
  run;

%mend;


data analy;
  set dat.m_90a_hlarsson_niddk_31july2016;
run;

proc contents data=analy;
title3 "m_90a_hlarsson_niddk_31july2016";
run;

proc freq data=analy;
  tables country sex fdr HLA_DR_5GRPS AGE_MAR2010_5GRPS/list missing;
  tables ia_any_10yr
        ia_atrisk_yes*ia_iaa_only_10yr
        ia_atrisk_yes*ia_gada_only_10yr
        ia_any_10yr*ia_iaa_only_10yr    *ia_gada_only_10yr
        ia_any_10yr*ia_atrisk_yes* ia_iaa_only_10yr    *ia_gada_only_10yr
        ia_any_10yr*ia_atrisk_yes
        multiple_ia_10yr * multi_ia_atrisk_yes
        t1d_10yr/list missing;
title3 "prelim freqs";
run;


data analy;
  set analy;

  if ia_any_10yr=1 and ia_atrisk_yes=1 then ia_event=1;
  else ia_event=0;
```

```
  if multiple_ia_10yr=1 and multi_ia_atrisk_yes=1 then multi_ia_event=1;
  else multi_ia_event=0;

  all=1;

run;

proc freq data=analy;
  tables ia_event*ia_any_10yr*ia_atrisk_yes/list missing;
  tables multi_ia_event*multiple_ia_10yr*multi_ia_atrisk_yes/list missing;
  tables t1d_10yr/missing;
title3 "Event vars";
run;


* check for dups;
proc sort data=analy;
  by MASKID;
run;

data dups;
  set analy;
  by MASKID;
  if not (first.MASKID and last.MASKID);
run;

data _null_;
  set dups;
  abort;
run;



* n and percent;
data prntall;
  set _null_;
run;

%npercent(1, country           , country      , all       , all);
%npercent(2, sex               , sexf  , all       , all);
%npercent(3, fdr               , fdrf  , all       , all);
%npercent(4, HLA_DR_5GRPS      , hladrf       , all       , all);
%npercent(5, AGE_MAR2010_5GRPS, agemarch , all       , all);
```

```
proc sort data=prntall;
  by rownum covarf;
run;

proc print data=prntall;
  var rownum covarf count;
title3 "Table 1 - Total";
run;


* n and percent;
data prntia_event;
  set _null_;
run;

%npercent(1, country          , country       , ia_event       , ia_event);
%npercent(2, sex             , sexf    , ia_event       , ia_event);
%npercent(3, fdr             , fdrf    , ia_event       , ia_event);
%npercent(4, HLA_DR_5GRPS     , hladrf        , ia_event       , ia_event);
%npercent(5, AGE_MAR2010_5GRPS, agemarch , ia_event       , ia_event);

proc sort data=prntia_event;
  by rownum covarf;
run;

proc print data=prntia_event;
  var rownum covarf count;
title3 "Table 1 - H1N1 vaccination in relation to risk of islet autoantibodies";
run;


* n and percent;
data prntmulti;
  set _null_;
run;

%npercent(1, country          , country       , multi_ia_event       , multi);
%npercent(2, sex             , sexf    , multi_ia_event       , multi);
%npercent(3, fdr             , fdrf    , multi_ia_event       , multi);
%npercent(4, HLA_DR_5GRPS     , hladrf        , multi_ia_event       , multi);
%npercent(5, AGE_MAR2010_5GRPS, agemarch , multi_ia_event       , multi);

proc sort data=prntmulti;
  by rownum covarf;
run;
```

```
proc print data=prntmulti;
  var rownum covarf count;
title3 "Table 1 - H1N1 vaccination in relation to risk of multiple islet autoantibodies";
run;


* n and percent;
data prntt1d;
  set _null_;
run;

%npercent(1, country          , country       , t1d_10yr       , t1d);
%npercent(2, sex             , sexf    , t1d_10yr       , t1d);
%npercent(3, fdr             , fdrf    , t1d_10yr       , t1d);
%npercent(4, HLA_DR_5GRPS     , hladrf        , t1d_10yr       , t1d);
%npercent(5, AGE_MAR2010_5GRPS, agemarch , t1d_10yr       , t1d);

proc sort data=prntt1d;
  by rownum covarf;
run;

proc print data=prntt1d;
  var rownum covarf count;
title3 "Table 1 - H1N1 vaccination in relation to risk of type 1 diabetes";
run;




* Table 1;
data npercent;
  merge prntall       (in=in1 rename=(count=countall  ))
        prntia_event  (in=in2 rename=(count=countia   ))
        prntmulti     (in=in3 rename=(count=countmulti))
        prntt1d       (in=in4 rename=(counttt1d  ))
        ;
  by rownum covarf;

run;

proc sort data=npercent;
  by rownum covarf;
run;
```

```
proc print data=npercent;
  var rownum  covarf countall countia countmulti counttld;
  title3 "Table 1 - n ";
run;


ods listing close;
ods phtml file="/prj/niddk/ims_analysis/TEDDY/private_created_data/TEDDY.m90a.Table1.xls";

proc print data=npercent;
  var rownum  covarf countall countia countmulti counttld;
  title3 "Table 1 - n ";
run;

ods phtml close;
ods listing;
```