Dataset Integrity Check for the TEDDY Pub17 HLarsson Data Files

Prepared by Michael Curry
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
July 8, 2015

Table of 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 SAS data files, as provided by the Data Coordinating Center (DCC), are located in the "\TEDDY\private_orig_data\Pub17_HLarsson_niddk_submission\" file in the data package. For this replication, variables were taken from pub17_hlarsson_niddk_final_2.sas7bdat only.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Helena Larsson, November 2011. To verify the integrity of the datasets, Table 3 from the manuscript was checked (Table B).

5 Results

Replication of Table 3 in the publication [1], 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 3. The results of the replication are complete.

6 Conclusions

The NIDDK repository are confident that the Pub17 data file to be distributed is a copy of the manuscript data with only inconsequential discrepancies.

7 References

1. Elding Larsson H, Vehik K, Bell R, Dabelea D, Dolan L, Pihoker C, et al. Reduced prevalence of diabetic ketoacidosis at diagnosis of type 1 diabetes in young children participating in longitudinal follow-up. *Diabetes Care*. 2011;34:2347–52. doi: 10.2337/dc11-1026.

Table A: Variables used to replicate Table 3: DKA rates in children aged <5 years and <2 years at diagnosis

Characteristic	Variable(s)
Study Name	Study
Age Group	Age
DKA severity	Level
DKA count	Count
Total Diabetes Diagnoses	Total

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

Characteristics from Table 3: DKA rates in children aged <5 years and <2 years at diagnosis	Manuscript (N = 805)	DSIC (N=805)	DIFF (N=0)
Teddy (strict DKA), Total DKA, <2 years	5/31	5/31	0,0
Teddy (strict DKA), Total DKA, <5 years	8/61	8/61	0,0
Teddy (strict DKA), Mild DKA, <2 years	2/31	2/31	0,0
Teddy (strict DKA), Mild DKA, <5 years	5/61	5/61	0,0
Teddy (strict DKA), Severe DKA, <2 years	3/31	3/31	0,0
Teddy (strict DKA), Severe DKA, <5 years	3/61	3/61	0,0
Teddy (broad DKA), Total DKA, <2 years	6/40	6/40	0,0
Teddy (broad DKA), Total DKA, <5 years	9/79	9/79	0,0
Teddy (broad DKA), Mild DKA, <2 years	3/40	3/40	0,0
Teddy (broad DKA), Mild DKA, <5 years	6/79	6/79	0,0
Teddy (broad DKA), Severe DKA, <2 years	3/40	3/40	0,0
Teddy (broad DKA), Severe DKA, <5 years	3/79	3/79	0,0
Sweden, Total DKA, <2 years	51/129	51/129	0,0
Sweden, Total DKA, <5 years	102/604	102/604	0,0
Sweden, Mild DKA, <2 years	39/129	39/129	0,0
Sweden, Mild DKA, <5 years	78/604	78/604	0,0
Sweden, Severe DKA, <2 years	12/129	12/129	0,0
Sweden, Severe DKA, <5 years	24/604	24/604	0,0
SEARCH, Total DKA, <2 years	29/58	29/58	0,0
SEARCH, Total DKA, <5 years	100/275	100/275	0,0
Finland, Total DKA, <2 years	82/183	82/183	0,0
Finland , Total DKA, <5 years	138/737	138/737	0,0
Finland , Mild DKA, <2 years	64/183	64/183	0,0
Finland , Mild DKA, <5 years	114/737	114/737	0,0
Finland , Severe DKA, <2 years	18/183	18/183	0,0
Finland , Severe DKA, <5 years	27/737	24/737	-3,0
Germany, Total DKA, <2 years	235/435	265/435	30,0
Germany, Total DKA, <5 years	583/1812	583/1812	0,0
Germany, Mild DKA, <2 years	171/435	171/435	0,0
Germany, Mild DKA, <5 years	458/1812	458/1812	0,0
Germany, Severe DKA, <2 years	64/435	64/435	0,0
Germany, Severe DKA, <5 years	125/1812	125/1812	0,0

```
libname pub17 '/prj/niddk/ims_analysis/TEDDY/private_orig_data/Pub17_HLarsson_niddk_submission';


proc format;
  value level
  1 = "Total"
  2 = "Mild"
  3 = "Severe"
  ;

  value age2f
  1 = "<2"
  2 = "<5"
  ;
  value $snam_long
  "FIN"    = "Finland Registry"
  "GER"    = "Germany Registry"
  "SWE"    = "Sweden Registry"
  "teddys" = "Teddy (strict)"
  "teddyb" = "Teddy (broad)"
  "SEARCH" = "SEARCH"
  ;
run;



data dka2;
       set pub17.pub17_hlarsson_niddk_final_2;
       format age age2f. study $snam_long.;
       prop = Count/Total;

run;


title2 "dka2";
proc print data = dka2;
run;

title3 "Table 3: DKA rates in children aged <5 years and <2 years at diagnosis";
proc tabulate data = dka2;
       where response = "Y";
       class level age ;
       class study / descending;
       var count total prop;
       table study=''*(count="Cases"*f=4.0 total="Pop"*f=4.0 prop="Percent"*f=PercentN7.1)*sum,level*age / misstext = "NA";
       KEYLABEL SUM=' ';
run;
```