Dataset Integrity Check for The Environmental Determinants of Diabetes in the Young (TEDDY) Study: Lernmark Data File

Prepared by

RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709-2194
June 18, 2012

Revision History

Version	Author / Title	Date	Comments
1.0	Norma Pugh	August 31, 2012	Original

Table of Contents

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 on a first (or second) exercise 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

Enrollment of participants in longitudinal research protocols can be particularly difficult when children are the target population. Across the world, there are country-specific studies that provide relevant information on enrollment experiences. However, there may be international differences in these experiences which are not captured by country-specific studies [1].

The TEDDY Study is an international, longitudinal, observational study that identifies young infants at increased genetic risk for type 1 diabetes (T1DM). The study, conducted in the countries of Finland, Germany, Sweden and the United States, aims to identify environmental triggers of T1DM in genetically at-risk children through observation and data collection over a 15 year time period [1].

Lernmark et al. address the importance of country specific estimates for enrollment targets in longitudinal pediatric studies and suggest that enrollment estimates should be lowered when the study involves the general population, painful procedures, or makes multiple demands on families [1].

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the TEDDY Data folder in the Official Archive. For this replication, all variables were taken from the SAS data file M_11_BLERNMARK_NIDDK_FINAL located in the Official Archive.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Lernmark et al [1] in Contemporary Clinical Trials in March 2011.

To verify the integrity of the M_11_BLERNMARK_NIDDK_FINAL data file housed at the repository, descriptive statistics on enrollment were computed, by country (Tables B, C, E, and F). The SAS code for our analysis is included in Attachment 1.

5 Results

Table 1 in the publication [1], Number of all HLA eligible children and number of HLA eligible children excluded from the TEDDY study by country and by general population (GP) versus first degree T1DM relative (FDR) status, reports on the types of children excluded and the primary reasons for exclusion by country. Our Table A lists the variables we used in our replication and Tables B and C compare the results calculated from the archived data file to the results published in Table 1. The results of the replication are similar to published results.

Table 2 in the publication [1], is Number of TEDDY eligible children and number (%) of TEDDY eligible families who refused enrollment by country and by general population (GP) versus first degree T1DM relative (FDR) status, reports on the types of families who refused enrollment and the primary reasons for refusal by country. Our Table D lists the variables we used in our replication and Tables E and F compare the results calculated from the archived data file to the results published in Table 2. Again, the results of the replication are similar to published results.

6 Conclusions

The NIDDK repository is confident that the Lernmark TEDDY data file to be distributed is a true copy of the study data.

7 References

1. Lernmark, B, Bennett Johnson S, Vehik K, Smith L, Ballard L, Baxter J, McLeod W, Roth R, Simellg T, on behalf of the TEDDY Study Group. Enrollment experiences in a pediatric longitudinal observational study: The Environmental Determinants of Diabetes in the Young (TEDDY) study; Contemporary Clinical Trials 32(2011) 517-523.

Table A: Variables Used to Replicate Table 1, Number of all HLA eligible children and number of HLA eligible children excluded from the TEDDY study by country and by general population (GP) versus first degree T1DM relative (FDR) status

Table Variable	Variables Used in Replication
Country	country
Child is eligible	maskid (all unique values)
Child is excluded	excluded=1
GP child	fdr=0
FDR child	fdr=1
Reason for exclusion: No response to calls/messages	inelig_cat5=1
Reason for exclusion: Incorrect contact info	inelig_cat4=1
Reason for exclusion: Appointment not in window	inelig_cat3=1

*All variables taken from dataset M_11_BLERNMARK_NIDDK_FINAL.

Table B: Comparison of Values Computed in Integrity Check to Reference Article Table 1 Values

Country = Finland, Germany, Sweden

Characteristic	Finland			Germany			Sweden		
	Lernmark	Integrity Check	Diff	Lernmark	Integrity Check	Diff	Lernmark	Integrity Check	Diff
Number of HLA eligible children	2959	2959	0	1216	1216	0	2983	2983	0
Number of excluded children (% of HLA eligible children)									
All excluded children	62 (2)	62 (2)	0	211 (18)	211 (17)	0 (1)	107 (4)	107 (4)	0
Excluded GP children	57 (2)	57 (2)	0	200 (17)	200 (17)	0	101 (3)	101 (3)	0
Excluded FDR children	5 (0.2)	5 (0.2)	0	11 (1)	11 (1)	0	6 (0.2)	6 (0.2)	0
Primary reason for exclusion: Number excluded by reason (% of children excluded)									
No response to calls/messages	39 (63)	39 (63)	0	191 (91)	191 (91)	0	76 (71)	76 (71)	0
Incorrect contact info	1 (2)	1 (2)	0	7 (3)	7 (3)	0	6 (6)	6 (6)	0
Appointment not in window	11 (18)	11 (18)	0	5 (2)	5 (2)	0	18 (17)	18 (17)	0

Table C: Comparison of Values Computed in Integrity Check to Reference Article Table 1 Values Country = US, All

Characteristic	US			All		
	Lernmark	Integrity Check	Difference	Lernmark	Integrity Check	Difference
Number of HLA eligible children	9277	9277	0	16,435	16,435	0
Number of excluded children (% of HLA eligible children)						
All excluded children	3457 (38)	3457 (37)	0 (1)	3837 (24)	3837 (23)	0 (1)
Excluded GP children	3364 (37)	3364 (36)	0 (1)	3722 (23)	3722 (23)	0
Excluded FDR children	93 (1)	93 (1)	0	115 (1)	115 (1)	0
Primary reason for exclusion: Number excluded by reason (% of children excluded)						
No response to calls/messages	2512 (73)	2512 (73)	0	2818 (73)	2818 (73)	0
Incorrect contact info	295 (9)	295 (9)	0	309 (8)	309 (8)	0
Appointment not in window	563 (16)	563 (16)	0	597 (16)	597 (16)	0

Table D: Variables Used to Replicate Table 2, Number of TEDDY eligible children and number (%) of TEDDY eligible families who refused enrollment by country and by general population (GP) versus first degree T1DM relative (FDR) status

Table Variable	Variables Used in Replication
Country	country
Child is eligible	maskid (all unique values), excluded=0
Family refused	ref_enr=1
GP family refused	ref_enr=1, fdr=0
FDR family refused	ref_enr=1, fdr=1
Reason for refusal: Protocol characteristics	ref_enr=1, ref_cat4=1
Reason for refusal: Family reasons	ref_enr=1, ref_cat5=1
Reason for refusal: Moving, unavailable	ref_enr=1, ref_cat2=1
Reason for refusal: Wants to wait and see	ref_enr=1, ref_cat3=1
Reason for refusal: No reason given	ref_enr=1, ref_cat1=1

*All variables taken from dataset M_11_BLERNMARK_NIDDK_FINAL.

Table E: Comparison of Values Computed in Integrity Check to Reference Article Table 2 Values

Country = Finland, Germany, Sweden

Characteristic	Finland			Germany			Sweden		
	Lernmark	Integrity Check	Diff	Lernmark	Integrity Check	Diff	Lernmark	Integrity Check	Diff
Number of TEDDY eligible children	2897	2897	0	1005	1005	0	2876	2876	0
Number of TEDDY eligible families who refused (% of TEDDY eligible children)									
All families who refused	1428 (49)	1428 (49)	0	560 (56)	560 (56)	0	833 (29)	833 (29)	0
All GP families who refused	1332 (46)	1332 (46)	0	519 (52)	519 (52)	0	794 (28)	794 (28)	0
All FDR families who refused	96 (3)	96 (3)	0	41 (4)	41 (4)	0	39 (1)	39 (1)	0
Primary reasons for refusal: Number refused by reason (% of children who refused)									
Protocol characteristics	389 (27)	399 (28)	+10 (+1)	300 (54)	300 (54)	0	320 (38)	320 (38)	0
Family reasons	266 (19)	268 (19)	+2 (0)	110 (20)	110 (20)	0	363 (44)	363 (44)	0
Moving, unavailable	20 (1)	20 (1)	0	7 (1)	7 (1)	0	25 (3)	25 (3)	0
Wants to wait and see	21 (1)	22 (2)	+1 (+1)	20 (4)	17 (3)	-3 (-1)	17 (2)	20 (2)	+3 (0)
No reason given	719 (50)	719 (50)	0	126 (23)	126 (23)	0	105 (13)	105 (13)	0

Table F: Comparison of Values Computed in Integrity Check to Reference Article Table 2 Values

Country = US, All

Characteristic	US			All		
	Lernmark	Integrity Check	Difference	Lernmark	Integrity Check	Difference
Number of TEDDY eligible children	5820	5820	0	12,598	12,598	0
Number of TEDDY eligible families who refused (% of TEDDY eligible children)						
All families who refused	3043 (52)	3043 (52)	0	5864 (47)	5864 (47)	0
All GP families who refused	2944 (51)	2944 (51)	0	5589 (44)	5589 (44)	0
All FDR families who refused	99 (2)	99 (2)	0	275 (2)	275 (2)	0
Primary reasons for refusal: Number refused by reason (% of children who refused)						
*Protocol characteristics	1192 (39)	1192 (39)	0	2201 (38)	2211 (38)	+10 (0)
Family reasons	1187 (39)	1187 (39)	0	1926 (33)	1928 (33)	+2 (0)
Moving, unavailable	182 (6)	182 (6)	0	234 (4)	234 (4)	0
Wants to wait and see	150 (5)	150 (5)	0	208 (4)	209 (4)	+1 (0)
No reason given	332 (11)	332 (11)	0	1282 (22)	1282 (22)	0

*Published results for 'protocol characteristics' include a typo. Per DCC, the actual result corresponds with the Integrity Check.

Attachment A: SAS Code

```
options errorabend nofmterr;
/*************************************************************************
/*
/* Program: R:\05_Users\Norma\TEDDY\LernmarkPaper\table1.sas
/* Author:  Norma Pugh
/* Date:    August 2012
/* Purpose: Replicate table 1 results.
/*
/*************************************************************************
/* DATA SOURCE */
libname data
'\\samba1.rtp.rti.org\NIDDK\03_Data_And_Tools\Studies\TEDDY\Delivery_from_DCC
\20120511_from_Steven_Fiske\final_upload';

/**********************/
/* ADDITIONAL FORMATS */
/**********************/
proc format;

 value country 1 = "1 = U.S."
               2 = "2 = Finland"
               3 = "3 = Germany"
               4 = "4 = Sweden"
              99 = "Total";
run;

/*************/
/* GET DATA */
/*************/
/* Keep one (last) visit for all HLA eligible children */
data hla; set data.m_11_blernmark_niddk_final; run;


proc sort data=hla; by maskid; run;

data hla; set hla; by maskid; if last.maskid; output; country=99; output;
format country country.; run;


/*****************************/
/* REPLICATE ANALYSIS RESULTS */
/*****************************/
proc freq data=hla; tables country / list nopct nocum out=denom(keep=country
count rename=(count=denom)); title'Country Counts'; run;

proc freq data=hla(where=(excluded=1)) noprint; tables country /
out=denom2(keep=country count rename=(count=denom)); run;

%macro frq(where,title,denom);
proc freq data=hla(where=(&where)) noprint; tables country /
out=frqstats(drop=percent) list nopct nocum; run;
data frqstats; merge frqstats &denom; by country; pct=(count/denom)*100; run;
proc print data=frqstats; title"&title"; run;
```

```
%mend frq;

%frq(%str(excluded=1),%str(All Excluded Children),denom);
%frq(%str(excluded=1 & fdr=0),%str(Excluded GP Children),denom);
%frq(%str(excluded=1 & fdr=1),%str(Excluded FDR Children),denom);
%frq(%str(excluded=1 & inelig_cat5=1),%str(Exclusion: No response),denom2);
%frq(%str(excluded=1 & inelig_cat4=1),%str(Exclusion: Incorrect contact
info),denom2);
%frq(%str(excluded=1 & inelig_cat3=1),%str(Exclusion: Appt. not in
window),denom2);
```

```
options errorabend nofmterr;
/*************************************************************************
/*
/* Program: R:\05_Users\Norma\TEDDY\LernmarkPaper\table2.sas
/* Author:  Norma Pugh
/* Date:    August 2012
/* Purpose: Replicate table 2 results.
/*
/*************************************************************************
/* DATA SOURCE */
libname data
'\\rcdubuntu01.rtp.rti.org\NIDDK\03_Data_And_Tools\Studies\TEDDY\Delivery_fro
m_DCC\20120511_from_Steven_Fiske\final_upload';

/**********************/
/* ADDITIONAL FORMATS */
/**********************/
proc format;
      value country 1 = "1 = U.S."
                    2 = "2 = Finland"
                    3 = "3 = Germany"
                    4 = "4 = Sweden"
                   99 = "Total";
run;

/*************/
/* GET DATA */
/*************/
/* Keep one (last) visit for all HLA eligible children */
data hla; set data.m_11_blernmark_niddk_final; run;

proc sort data=hla; by maskid; run;

data hla; set hla; by maskid; if last.maskid; if excluded=0; output;
country=99; output; format country country.; run;


/*******************************/
/* REPLICATE ANALYSIS RESULTS */
/*******************************/
proc freq data=hla; tables country / list nopct nocum out=denom(keep=country
count rename=(count=denom)); title'Country Counts'; run;

proc freq data=hla(where=(ref_enr=1)) noprint; tables country /
out=denom2(keep=country count rename=(count=denom)); run;

%macro frq(where,title,denom);
proc freq data=hla(where=(&where)) noprint; tables country /
out=frqstats(drop=percent) list nopct nocum; run;
data frqstats; merge frqstats &denom; by country; pct=(count/denom)*100; run;
proc print data=frqstats; title"&title"; run;
%mend frq;

%frq(%str(ref_enr=1),%str(All families who refused),denom);
%frq(%str(ref_enr=1 & fdr=0),%str(All GP families who refused),denom);
%frq(%str(ref_enr=1 & fdr=1),%str(All FDR families who refused),denom);
%frq(%str(ref_enr=1 & ref_cat4=1),%str(Refusal: Protocol),denom2);
```

```
%frq(%str(ref_enr=1 & ref_cat5=1),%str(Refusal: Family reasons),denom2);
%frq(%str(ref_enr=1 & ref_cat2=1),%str(Refusal: Moving),denom2);
%frq(%str(ref_enr=1 & ref_cat3=1),%str(Refusal: Wants to wait and
see),denom2);
%frq(%str(ref_enr=1 & ref_cat1=1),%str(Refusal: No reason given),denom2);
```