Dataset Integrity Check for the TrialNet (06) Data Files

Prepared by Jane Wang and Allyson Mateja
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
December 31, 2015

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 objective of this study was to describe a pilot trial of using an omega-3 fatty acid (docosahexaenoic acid [DHA]) to prevent islet cell autoimmunity in infants with an increased risk for developing type 1 diabetes (T1D). Infants from pregnant mothers who either have T1D (or the father or a previous child has T1D) and who entered the study in the third trimester or infants younger than age 5 months having a first-degree family member with T1D were eligible for the study. Infants from either group also had to have an increased genetic (HLA) risk for T1D (or multiple first-degree relatives with T1D) to be eligible. The study is a multicenter, 2-arm, randomized, double-masked clinical trial that will last 4 years (1 year of recruitment and 3 years of treatment). Treatment with DHA (or control) began in the last trimester of pregnancy or in the first 5 months after birth. Inflammatory mediators, including cytokines, chemokines, eicosanoids, and C-reactive protein, are being measured along with fatty acids in maternal and infant blood. Ninety-eight infants were enrolled (41 during pregnancy and 57 in the 5 months after birth). HLA results of the 97 eligible infants (1 infant had a protective 0602 allele and was thus ineligible) showed that 90 have DR3 and/or DR4. Seven infants were enrolled without DR3/4 but who instead had multiple first-degree relatives with T1D. Compliance has been excellent, and no families have discontinued participation. Intervention trials in this high-risk group are feasible but require significant effort to identify potential participants.

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the data folder in the data package. For this replication, variables were taken from the

"tn06_210815_npp04infa.sas7bdat", "tn06_210823_npp05preg.sas7bdat", and "research_labs.sas7bdat" datasets.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Chase, et al in Infant, Child, & Adolescent Nutrition on April 2009 [1]. To verify the integrity of the datasets, descriptive statistics of baseline characteristics were computed, by entry group (Table B).

5 Results

For Table 2 in the publication [1], Infant Baseline information, 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 1. The results of the replication are very similar to the published results.

6 Conclusions

The NIDDK repository is confident that the TrialNet (06) data files to be distributed are a true copy of the manuscript data.

7 References

[1] H. Peter Chase, MD, Ellen Lescheck, MD, Lisa Rafkin-Mervis, MS, CDE, Heidi Krause-Steinrauf, MS, Sonia Chritton, MS, Smita M. Asare,Sara Adams, Jay S. Skyler, MD, Michael Clare-Salzler, MD, and the Type 1 Diabetes TrialNet NIP Study Group. Nutritional Intervention to Prevent (NIP) Type 1 Diabetes. ICAN: Infant, Child, & Adolescent Nutrition April 2009

Table A: Variables used to replicate Table 2: Infant baseline information by entry group.

Table Variable	**dataset.variable**
Breastfeeding	tn06_210815_npp04infa.CurrentlyBreastFeeding
Male	tn06_210815_npp04infa.sex
Race/ethnicity(White)	tn06_210815_npp04infa_white
Gestational age wk	tn06_210815_npp04infa.InfantAgeAtBirth
Weight kg	tn06_210815_npp04infa.WeightEnglish * 0.4535924
Length cm	tn06_210815_npp04infa.LengthEnglish * 2.54
Delivery type	tn06_210823_npp05preg.TypeOfBirth
HLA Genotype allele	research_labs.test_name in ('DQB1_1' 'DQB1_2' 'DR3DR4' 'DRB1_1' 'DRB1_2' 'HLA' 'PROTECTIVE')
Entry (A or B)	tn06_210815_npp04infa.MotherParticipatingInStudyDuri

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

		Entry A (n=41) [Manuscript]	**Entry A (n=41) [DSIC]**	**Entry A (n=0) [Difference]**	**Entry B (n=57) [Manuscript]**	**Entry B (n=57) [DSIC]**	**Entry B (n=0) [Difference]**
Breastfeeding	%	73	73	0	68	63	5
Male	%	46	46	0	51	49	2
Race/ethnicity (White)	%	98	98	0	93	93	0
Gestational age, wk	mean ± SD	38 ± 1.5	38 ± 1.5	0 ± 0	38 ± 1.6	38 ± 1.6	0 ± 0
Weight, kg	mean ± SD	3.6 ± 0.7	3.6 ± 0.7	0 ± 0	3.5 ± 0.7	3.5 ± 0.7	0 ± 0
Length, cm	mean ± SD	50.4 ± 2.5	50.4 ± 2.5	0 ± 0	50.7 ± 3.1	51.0 ± 3.1	0.3 ± 0.0
Delivery type: Vaginal	%	55	58	3	45	48	3
Delivery type: Cesarean section	%	45	43	2	55	52	3
HLA Alleles		n=41	n=38	n=3	n=56	n=56	n=0
DR3/DR4		4	3	1	10	11	1
DR3/DR3		0	1	1	4	2	2
DR4/DR4		6	6	0	5	7	2
DR3/X		11	10	1	11	10	1
DR4/X		16	16	0	23	21	2

		Entry A (n=41) [Manuscript]	Entry A (n=41) [DSIC]	Entry A (n=0) [Difference]	Entry B (n=57) [Manuscript]	Entry B (n=57) [DSIC]	Entry B (n=0) [Difference]
X/X		4	2	2	3	5	2

Attachment A: SAS Code

```
*************************************************************************************************************************
***Program:
***Programmer: Corey Del Vecchio
***Date Created: 7/8/2013
***Purpose:
***
***
***
***Source of Request:
***Input Files:
***
***Output Files:
***
***History
***Updated by: Jane Wang
***Date Modified: 10/30/2015
***Updated by: Allyson Mateja 12/31/15
*************************************************************************************************************************;

title1 "%sysfunc(getoption(sysin))";
title2 " ";

proc format;
       value $typef 'cesarean section' = 'Cesarean section'
                    'spontaneous', 'vaginal' = 'Vaginal';

options nofmterr;

 options nofmterr source2 mprint symbolgen spool;
libname sas_data "/prj/niddk/ims_analysis/TrialNet/private_orig_data/Data.Extraction.6.Version151218.PW/6/sasv9/";

data inf_screening              ; set sas_data.tn06_210815_npp04infa        ;
data preg_hx                    ; set sas_data.tn06_210823_npp05preg        ;
data infant_enr_med_hx          ; set sas_data.tn06_210729_npp06infa        ;
data research_labs              ; set sas_data.research_labs                ;

proc sort data = inf_screening(keep = maskid sex Race_White MotherParticipatingInStudyDuri MotherRecievedAnyImmunoglobuli Date_of_Visit InfantAgeAtBirth
LengthEnglish
CurrentlyBreastFeeding MotherBreastfeedingWillingToDi    LengthMetric LengthInfantIN  LengthInfantCM  WeightEnglish WeightInfantKG WeightInfantLB WeightMetric
HLATypingCollected HLAComments HLAHeelStick);
  by maskid;

proc sort data = preg_hx(keep = maskid enterthisstudyduringpregnancy TypeOfBirth);
  by maskid;

proc sort data = infant_enr_med_hx(keep = maskid   EntryAInfantScreeningWithInfan EntryAInfantScreeningWithInfan2);
  by maskid;
```

```
data infantboth;
  merge infant_enr_med_hx (in = in1)  inf_screening (in = in2) preg_hx (in = in3);
  by maskid;
  if in1;

proc sort data = infantboth;
  by maskid;

data dups_infantboth uniq_infantboth;
  set infantboth;
  by maskid;
  if not (first.maskid and last.maskid) then output dups_infantboth;
  else output uniq_infantboth;

proc freq data = uniq_infantboth;
  tables    enterthisstudyduringpregnancy  * MotherParticipatingInStudyDuri/list missing;

proc print data = dups_infantboth;

data dups_infantboth_keep;
  set dups_infantboth;
  if TypeOfBirth in ('cesarean section' 'vaginal');

proc sort data =  dups_infantboth_keep nodupkey;
  by maskid;

proc print data = dups_infantboth_keep;

data infantboth_final;
  set dups_infantboth_keep uniq_infantboth;
  if      Date_of_Visit <= input('11/30/2006' ,mmddyy10.) then visit_cat = 'a 12/2006';
  else if Date_of_Visit <= input('01/26/2007' ,mmddyy10.) then visit_cat = 'b 01/2007';
  else if Date_of_Visit <= input('02/16/2007' ,mmddyy10.) then visit_cat = 'c 02/2007';
  else if Date_of_Visit <= input('03/28/2007' ,mmddyy10.) then visit_cat = 'd 03/2007';
  else if Date_of_Visit <= input('04/23/2007' ,mmddyy10.) then visit_cat = 'e 04/2007';
  else if Date_of_Visit <= input('05/30/2007' ,mmddyy10.) then visit_cat = 'f 05/2007';
  else if Date_of_Visit <= input('06/07/2007' ,mmddyy10.) then visit_cat = 'g 06/2007';
  else if Date_of_Visit <= input('07/17/2007' ,mmddyy10.) then visit_cat = 'h 07/2007';
  else if Date_of_Visit <= input('08/30/2007' ,mmddyy10.) then visit_cat = 'i 08/2007';
  else if Date_of_Visit <= input('09/17/2007' ,mmddyy10.) then visit_cat = 'j 09/2007' ;
  else if Date_of_Visit <= input('10/26/2007' ,mmddyy10.) then visit_cat = 'k 10/2007' ;
  else if Date_of_Visit <= input('11/28/2007' ,mmddyy10.) then visit_cat = 'l 11/2007' ;
  else if Date_of_Visit <= input('12/21/2007' ,mmddyy10.) then visit_cat = 'm 12/2007' ;
  else if Date_of_Visit <= input('01/31/2008' ,mmddyy10.) then visit_cat = 'n 01/2008' ;
  else if Date_of_Visit <= input('02/29/2008' ,mmddyy10.) then visit_cat = 'o 02/2008' ;
  else if Date_of_Visit <= input('03/28/2008' ,mmddyy10.) then visit_cat = 'p 03/2008' ;
  else if Date_of_Visit <= input('04/30/2008' ,mmddyy10.) then visit_cat = 'q 04/2008' ;
  else if Date_of_Visit <= input('05/14/2008' ,mmddyy10.) then visit_cat = 'r 05/2008' ;
  weight_t2 = WeightEnglish * 0.4535924 ; * from ib to kg;
  length_t2 = LengthEnglish * 2.54; *form inch to cm;

proc sort data = infantboth_final;
  by MotherParticipatingInStudyDuri;

proc freq data = infantboth_final;
  by MotherParticipatingInStudyDuri;
```

```
  tables CurrentlyBreastFeeding;
  title3 'Table 2 - Breastfeeding, %';

proc freq data = infantboth_final;
  by MotherParticipatingInStudyDuri;
  tables Sex;
  title3 'Table 2 - Male, %';

proc freq data = infantboth_final;
  by MotherParticipatingInStudyDuri;
  tables Race_White /missing;
  title3 'Table 2 - Race/ethnicity, % White';

proc means data = infantboth_final mean std;
  by MotherParticipatingInStudyDuri;
  var InfantAgeAtBirth;
  title3 'Table 2 - Gestational age, wk';

proc means data = infantboth_final mean std;
  by MotherParticipatingInStudyDuri;
  var weight_t2;
  title3 'Table 2 - Weight, KG';

proc means data = infantboth_final mean std;
  by MotherParticipatingInStudyDuri;
  var length_t2;
  title3 'Table 2 - Length, cm';

proc freq data = infantboth_final;
  by MotherParticipatingInStudyDuri;
  tables TypeOfBirth;
  format TypeOfBirth $typef.;
  title3 'Table 2 - Delivery type, %';

/*proc freq data = RESEARCH_LABS;
  tables TEST_NAME visit resulttype;*/

proc freq data = RESEARCH_LABS;
  tables TEST_NAME*visit /list missing;
  where test_name in ('DR3DR4', 'DQB1_1', 'DQB1_2', 'DRB1_1', 'DRB1_2', 'HLA', 'PROTECTIVE');
 /* where visit in ('1 - Infant Screening Visit (TN06)', '2 - Infant Enrollment Visit (TN06)');  */

data RESEARCH_LABS_hla;
  set RESEARCH_LABS;
  if test_name in ('DR3DR4', 'DQB1_1', 'DQB1_2', 'DRB1_1', 'DRB1_2', 'HLA', 'PROTECTIVE');

proc sort data= RESEARCH_LABS_hla;
  by maskid;

proc sort data= infantboth_final;
  by maskid;

data research_labs_hla;
       set research_labs_hla;
       by maskid;
       array results (0:6) $ results0-results6;
```

```
        retain results;
        if first.maskid then do i = 0 to 6;
                results[i] = '';
        end;
        if test_name = 'DR3DR4' then results[0] = result;
        if test_name = 'DQB1_1' then results[1] = result;
        if test_name = 'DQB1_2' then results[2] = result;
        if test_name = 'DRB1_1' then results[3] = result;
  if test_name = 'DRB1_2' then results[4] = result;
  if test_name = 'HLA' then results[5] = result;
  if test_name = 'PROTECTIVE' then results[6] = result;
  if last.maskid then output research_labs_hla;

data RESEARCH_LABS_hla prob;
  merge infantboth_final (in = in1)  RESEARCH_LABS_hla (in = in2);
  by maskid;
  if in1 and in2 then output RESEARCH_LABS_hla;
  else if not in1 and in2 then output prob;

proc sort data=research_labs_hla;
        by MotherParticipatingInStudyDuri;

data research_labs_hla;
        set research_labs_hla;
        if results3='3' and results4 = '3' then allele = 'DR3/DR3';
        else if results3 = '4' and results4 = '4' then allele = 'DR4/DR4';
        else if (results3 = '3' and results4 = '4') or (results3 = '4' and results4 = '3') then allele = 'DR3/DR4';
        else if results3 = '3' or results4 = '3' then allele = 'DR3/X';
        else if results3 = '4' or results4 = '4' then allele = 'DR4/X';
        else allele = 'X/X';

proc freq data=research_labs_hla;
        by MotherParticipatingInStudyDuri;
        tables allele /list missing;

proc freq data=research_labs_hla;
        by MotherParticipatingInStudyDuri;
        tables results0*results1*results2*results3*results4*results5*results6*allele /list missing;
endsas;

proc freq data = RESEARCH_LABS_hla;
  by MotherParticipatingInStudyDuri;
  tables test_name*result /list missing;
```