

Dataset Integrity Check for the Long-Term Investigative Follow-Up in TrialNet Study (TN16 LIFT) Data

Prepared by NIDDK-CR
June 7, 2022

Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	3
6 Conclusions	3
7 References	3
Table A: Variables used to replicate descriptive statistics in the text of the Results section.....	4
Table B: Comparison of values computed in integrity check to reference article descriptive statistics in the text of the Results section values	5
Attachment A: SAS Code	6

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 TN16 LIFT study is a clinical study that follows participants from TrialNet prevention studies after diagnosis with type 1 diabetes (T1D). The study follows individuals who developed diabetes while being monitored in the Pathway to Prevention Study or during participation in any prevention trial. The purpose is to study people early in the disease process. Additionally, the study follows individuals who participated in any new onset studies to learn more about diabetes and if there are any long-term effects from the treatments used in those studies.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in the TN16 folder in the data package. For this replication, variables were taken from the “tn16_registration.sas7bdat” and “tn16_diabeteshealthinfo” datasets. Since all participants included in the TN16 study were originally participants in the TN01 Pathway to Prevention study, some demographic data were used from the TN01 study for this replication.

4 Statistical Methods

Analyses were performed to replicate results for the data published by Bogun et al. [1] for C-Peptide Levels in Subjects Followed Longitudinally Before and After Type 1 Diabetes Diagnosis in TrialNet. To verify the integrity of the data, only descriptive statistics were computed for the provided variables. Participant IDs included in the publication were provided by the DCC for the purposes of this replication.

5 Results

Using the information in the Results section in the publication [1], the descriptive statistics described in the text were replicated. 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 that were published. The results of the replication are within expected variation to the published results.

6 Conclusions

The NIDDK Central Repository is confident that the TN16 data files to be distributed are a true copy of the study data as the results of the replication are within expected variation to the published results.

7 References

[1] Bogun MM, Bundy BN, Goland RS, Greenbaum CJ. C-Peptide Levels in Subjects Followed Longitudinally Before and After Type 1 Diabetes Diagnosis in TrialNet. *Diabetes Care*, 43(8), 1836-1842, August 2020. doi: <https://doi.org/10.2337/dc19-2288>

Table A: Variables used to replicate descriptive statistics in the text of the Results section

Table Variable	dataset.variable
Sex	tn01_nh01_screening.sex tn16_registration.maskid
Age	tn01_nh01_screening.BirthMonth tn01_nh01_screening.BirthYear tn16_registration.Date_of_Registration tn16_registration.maskid
Race	tn01_nh01_screening.Race_AmericanIndianorAlaskaNat tn01_nh01_screening.Race_Asian tn01_nh01_screening.Race_BlackorAfricanAmerican tn01_nh01_screening.Race_NativeHawaiianorOtherPaci tn01_nh01_screening.Race_Unknownornotreported tn01_nh01_screening.Race_White tn01_nh01_screening.Ethnicity tn16_registration.maskid
Post-diagnosis oral glucose tolerance testing (OGTT) performed	tn01_nh08_diabesteset_old.diagnosisday tn01_nh08_diabesteset_old.diagnosismonth tn01_nh08_diabesteset_old.diagnosisyear tn16_diabeteshealthinfo.visit tn16_diabeteshealthinfo.OGTTcollected tn16_diabeteshealthinfo.maskid

Table B: Comparison of values computed in integrity check to reference article descriptive statistics in the text of the Results section values

Variable	Cohort 1 (n=80)	DSIC Cohort 1 (n=80)	Diff. (n=0)	Cohort 2 (n=127)	DSIC Cohort 2 (n=127)	Diff. (n=0)
Sex male, n (%)	41 (51)	41 (51.9)	(0.9)	66 (52)	66 (53)	0 (0)
Age, median (range)	14 (3.9,55.1)	14 (4,55)	0 (0.1, 0.1)	13.5 (4.3,51.1)	14 (5, 51)	0.5 (0.7, 0.1)
Race/Ethnicity						
White	74	75	1	120	119	1
African American	2	2	0	1	1	0
Asian	2	2	0	1	1	0
Unknown	3	3	0	5	2	3
Hispanic/Latino	3	3	0	5	5	0
Post-diagnosis OGTTs performed, median (range) (Cohort 1 only)	2 (1,3)	2 (1,3)	0 (0,0)	-	-	-

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies6\TrialNet_16\private_created_data\Redacted Data";
libname ids "X:\NIDDK\niddk-dr_studies6\TrialNet_16\private_created_data";
libname tn01 "X:\NIDDK\niddk-dr_studies6\TrialNet_01\private_created_data\TN01_V6\Data";
```

```
/*******/
/* TN16 DSIC Bogun et al. */
/*******/
*calling necessary datasets;
data screen; set tn01.tn01_nh01_screening;
run;
```

```
data cht1; set ids.cohort1;
run;
```

```
data cht2; set ids.cohort2;
run;
```

```
*each cohort;
proc contents data=ids.cohort1;
run;
```

```
proc contents data=ids.cohort2;
run;
```

```
/*******/
/*Cohort 1 Demographics*/
/*******/
*Subset for Cohort1;
proc sort data=screen;
by MaskID;
run;
```

```
proc sort data=cht1;
by maskid;
run;
```

```
data sex;
merge
screen (in=a)
cht1 (in=b);
by maskid;
if b=1;
run;
```

```
*Sex breakdown of Cohort 1;
```

```
proc freq data=sex;  
tables sex;  
run;
```

```
*Age breakdown for cohort 1;  
data reg; set dsic.tn16_registration;  
run;
```

```
*sorting reg by maskid;  
proc sort data=reg;  
by MaskID;  
run;
```

```
*merging reg with Sex dataset;  
data age;  
merge  
sex (in=a)  
reg (in=b);  
by maskid;  
if a=1;  
run;
```

```
*Looking at dates of birth and registration date to gather age of TN16 participants;  
proc freq data=age;  
tables BirthMonth BirthYear Date_of_Registration;  
run;
```

```
data age1; set age;  
keep MaskID BirthMonth BirthYear Date_of_Registration;  
run;
```

```
proc contents data=age1;  
run;
```

```
*creating a new single date variable for DOB;  
data age2; set age1;  
if birthmonth = "Jan" then month = 01;  
if birthmonth = "Feb" then month = 02;  
if birthmonth = "Mar" then month = 03;  
if birthmonth = "Apr" then month = 04;  
if birthmonth = "May" then month = 05;  
if birthmonth = "Jun" then month = 06;  
if birthmonth = "Jul" then month = 07;  
if birthmonth = "Aug" then month = 08;  
if birthmonth = "Sep" then month = 09;  
if birthmonth = "Oct" then month = 10;  
if birthmonth = "Nov" then month = 11;  
if birthmonth = "Dec" then month = 12;
```


run;

data age3; set age2;
birthdate = mdy(month, 1, BirthYear);
format birthdate mmddyy10.;
run;

*creating an age at reg variable;
data age4; set age3;
age = (Date_of_Registration - birthdate)/365;
age1 = round(age, 1.0);
run;

proc means data=age4 median min max;
var age1;
run;

*ethnicity breakdown for cohort 1;
proc freq data=age;
tables Race_AmericanIndianorAlaskaNat Race_Asian Race_BlackorAfricanAmerican
Race_NativeHawaiianorOtherPaci Race_Refused
Race_Unknownornotreported Race_White Ethnicity;
run;

*prediagnosis OGTTs for Cohort 1;
data diag; set tn01.tn01_nh08_diabetesonset_old;
keep DiagnosisDay DiagnosisMonth DiagnosisYear MaskID diagdate;
if DiagnosisMonth = "Jan" then month = 01;
if DiagnosisMonth = "Feb" then month = 02;
if DiagnosisMonth = "Mar" then month = 03;
if DiagnosisMonth = "Apr" then month = 04;
if DiagnosisMonth = "May" then month = 05;
if DiagnosisMonth = "Jun" then month = 06;
if DiagnosisMonth = "Jul" then month = 07;
if DiagnosisMonth = "Aug" then month = 08;
if DiagnosisMonth = "Sep" then month = 09;
if DiagnosisMonth = "Oct" then month = 10;
if DiagnosisMonth = "Nov" then month = 11;
if DiagnosisMonth = "Dec" then month = 12;
diagdate = mdy(Month, DiagnosisDay, DiagnosisYear);
format daigdate mmddyy10.;
run;

proc freq data=diag;
tables diagdate;
run;

proc freq data=dsic.tn16_diabeteshealthinfo;

```

tables visit OGTTcollected;
run;

data diabhth; set dsic.tn16_diabeteshealthinfo;
run;

proc sort data=diag;
by maskid;
run;

proc sort data=diabhth;
by maskid;

proc sort data=cht1;
by maskid;
run;

data ogtt;
merge
diag (in=a)
diabhth (in=b)
cht1 (in=c);
by maskid;
if c=1;
run;

*assessing the number of OGTTs post-diagnosis;
data ogtt1; set ogtt;
keep MaskID diagdate OGTTCollected Visit_Dt flag ;
if visit_dt > diagdate AND OGTTcollected = "Yes" then flag = 1;
run;

proc freq data=ogtt1;
tables flag;
run;

data ogtt2; set ogtt1;
keep MaskID flag;
run;

proc transpose data=ogtt2 out=ogtt3 prefix=OGTT_;
by maskid;
run;

data ogtt4; set ogtt3;
keep MaskID ogtt_1 ogtt_2 ogtt_3 ogttTot;
ogttTot = Sum(ogtt_1, ogtt_2, ogtt_3);
run;

```

```
proc means data=ogtt4 median min max;  
var ogttTot;  
run;
```

```
/******  
/*Cohort 2 demographics*/  
/******
```

```
proc sort data=screen;  
by MaskID;  
run;
```

```
proc sort data=cht2;  
by maskid;  
run;
```

```
data sex;  
merge  
screen (in=a)  
cht2 (in=b);  
by maskid;  
if b=1;  
run;
```

```
*Sex breakdown of Cohort 2;  
proc freq data=sex;  
tables sex;  
run;
```

```
*Age breakdown for cohort 2;  
data reg; set dsic.tn16_registration;  
run;
```

```
*sorting reg by maskid;  
proc sort data=reg;  
by MaskID;  
run;
```

```
*merging reg with Sex dataset;  
data age;  
merge  
sex (in=a)  
reg (in=b);  
by maskid;  
if a=1;  
run;
```

```
*Looking at dates of birth and registration date to gather age of TN16 participants;
```

```

proc freq data=age;
tables BirthMonth BirthYear Date_of_Registration;
run;

data age1; set age;
keep MaskID BirthMonth BirthYear Date_of_Registration;
run;

proc contents data=age1;
run;

```

*creating a new single date variable for DOB;

```

data age2; set age1;
if birthmonth = "Jan" then month = 01;
if birthmonth = "Feb" then month = 02;
if birthmonth = "Mar" then month = 03;
if birthmonth = "Apr" then month = 04;
if birthmonth = "May" then month = 05;
if birthmonth = "Jun" then month = 06;
if birthmonth = "Jul" then month = 07;
if birthmonth = "Aug" then month = 08;
if birthmonth = "Sep" then month = 09;
if birthmonth = "Oct" then month = 10;
if birthmonth = "Nov" then month = 11;
if birthmonth = "Dec" then month = 12;
run;

```

```

data age3; set age2;
birthdate = mdy(month, 1, BirthYear);
format birthdate mmddyy10.;
run;

```

*creating an age at reg variable;

```

data age4; set age3;
age = (Date_of_Registration - birthdate)/365;
age1 = round(age, 1.0);
run;

```

```

proc means data=age4 median min max;
var age1;
run;

```

*ethnicity breakdown for cohort 2;

```

proc freq data=age;
tables Race_AmericanIndianorAlaskaNat Race_Asian Race_BlackorAfricanAmerican
Race_NativeHawaiianorOtherPaci Race_Refused
Race_Unknownornotreported Race_White Ethnicity;
run;

```