Dataset Integrity Check for Treatment Options for Type 2 Diabetes in Adolescents & Youth Long Term Follow-Up (TODAY2) Study Data

Prepared by NIDDK-CR
July 26, 2022

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 TODAY2 study was a longitudinal follow-up study to the TODAY (Treatment Options for Type 2 Diabetes in Adolescents and Youth) study that continued the care and observation of the TODAY cohort participants beyond the end of the TODAY intervention trial. The TODAY2 study had two phases: 1) Transition of TODAY participants to non-blinded, non-randomized standard diabetes care and management with monitoring and follow-up for up to 24 months, and 2) Long-term longitudinal follow-up of the TODAY cohort based on findings from TODAY.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in the TODAY and TODAY2 folders in the data packages. For this replication, variables were taken from the TODAY datasets “pat.sas7bdat”, “cbl.sas7bdat”, “baseline.sas7bdat”, and “primout.sas7bdat”, and also from the TODAY2 dataset “visitfinal.sas7bdat”.

4 Statistical Methods

Analyses were performed to replicate results for the data in the publication by Bjornstad et al. [1]. To verify the integrity of the data, descriptive statistics were computed for the baseline characteristics.

5 Results

For Table 1 in the publication [1], Characteristics of the Participants, 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 in Table 1. The results of the replication are within expected variation of the published results.

6 Conclusions

The NIDDK Central Repository is confident that the TODAY2 data files to be distributed are a true copy of the study data.

7 References

[1] Bjornstad P, Drews KL, Caprio S, Gubitosi-Klug R, Nathan DM, Tesfaldet B, Tryggestad J, White NH, Zeitler P. Long-Term Complications in Youth-Onset Type 2 Diabetes. The New England Journal of Medicine, 385(5), 416-426, July 2021. doi: https://doi.org/10.1056/NEJMoa2100165

Table A: Variables used to replicate Table 1 – Characteristics of the Participants

Table Variable	**dataset.variable**
Age - years	visitfinal.pvisit pat.age
Female sex - %	visitfinal.pvisit pat.sex
Race and ethnic group - %	visitfinal.pvisit pat.race
Lived in household with income < $25,000 - %	visitfinal.pvisit pat.houseinc
Time since diagnosis of type 2 diabetes – months	visitfinal.pvisit pat.dxtime
Glycated hemoglobin	visitfinal.pvisit cbl.HbA1c
BMI	visitfinal.pvisit baseline.bmi
Treatment assignment at start of clinical trial - %	visitfinal.pvisit primout.tx

Table B: Comparison of values computed in integrity check to reference article Table 1 values

Characteristic (Baseline)	**TODAY2 Phase 1 Cohort (n=550)**	**TODAY2 Phase 1 Cohort DSIC (n=549)**	**Diff. (n=1)**	**TODAY2 Phase 2 Cohort (n=500)**	**TODAY2 Phase 2 Cohort DSIC (n=474)**	**Diff. (n=26)**
Age - years	13.9 ± 2.0	14.0 ± 0.9	0.1 ± 1.1	13.8 ± 2.0	14.0 ± 0.9	0.2 ± 1.1
Female sex - %	65.1	65.2	0.1	65.4	65.4	0
Race and ethnic group - %						
Hispanic	39.8	39.5	0.3	38.2	38.4	0.2
Non-Hispanic Black	33.6	33.9	0.3	34.8	35.6	0.8
Non-Hispanic White	19.3	19.1	0.2	19.2	18.1	1.1
Other	7.3	7.5	0.2	7.8	7.8	0
Lived in household with income < $25,000 - %	42.8	43.6	0.8	44.5	44.5	0
Time since diagnosis of type 2 diabetes - months	7.7 ± 5.9	N/A	N/A	7.8 ± 5.9	N/A	N/A
Glycated hemoglobin						
Percentage	6.0 ± 0.7	6.0 ± 0.8	0 ± 0.1	6.0 ± 0.8	6.0 ± 0.8	0.0 ± 0.0
Mean level - mmol/mol	42.0 ± 7.7	42.0 ± 8.5	0 ± 0.8	42.0 ± 8.7	42.3 ± 8.6	0.3 ± 0.1
BMI	35.0 ± 7.8	35.0 ± 1.7	0 ± 6.1	35.1 ± 7.8	34.9 ± 1.7	0.2 ± 6.1
Treatment assignment at start of clinical trial - %						
Metformin alone	33.8	33.9	0.1	33.6	34.4	0.8
Metformin plus rosiglitazone	32.9	32.8	0.1	33.0	32.5	0.5
Metformin plus lifestyle intervention	33.3	33.3	0.0	33.4	33.1	0.3

Attachment A: SAS Code

```
libname t2 "X:\NIDDK\niddk-dr_studies6\TODAY2\private_orig_data\Datasets extracted from XPT";
libname t1 "X:\NIDDK\niddk-dr_studies6\TODAY\private_created_data\TODAY_V4\Data\sas7bdat";

proc contents data=t1.visit;
run;

proc contents data=t2.visitfinal;
run;

proc freq data=t2.visitfinal;
tables pvisit;
run;

proc contents data=t1.baseline;
run;

proc contents data=t1.pat;
run;

*merging the different dataset;
data t2_2; set t2.visitfinal;
keep releaseid pvisit;
run;

data t1_2; set t1.pat;
run;

proc sort data=t2_2;
by releaseid;
run;

proc sort data=t1_2;
by releaseid;
run;

data demo; merge
t2_2 (in=a)
t1_2 (in=b);
by releaseid;
if a=1;
run;

*splitting the demo datasets into cohorts 1 and 2;
data demo_2; set demo;
array visits pvisit;
```

```
if "P120" in visits OR "P132" in visits OR "P144" in visits OR "P156" in visits OR "P168" in visits OR "P180"
in visits;
run;

data demo_1; set demo;
array visits pvisit;
if "P027" in visits OR "P081" in visits OR "P030" in visits OR "P084" in visits OR "P033" in visits OR "P087"
in visits OR
"P036" in visits OR "P090" in visits OR "P039" in visits OR "P093" in visits OR "P042" in visits OR  "P096"
in visits OR
"P045" in visits OR "P099" in visits OR "P048" in visits OR "P102" in visits OR "P051" in visits OR "P105" in
visits OR
"P054" in visits OR "P108" in visits OR "P057" in visits OR "P111" in visits OR "P060" in visits OR "P114" in
visits OR
"P063" in visits OR "P066" in visits OR "P069" in visits OR "P072" in visits OR "P075" in visits OR "P078" in
visits;
run;

*assessing demographics TDOAY2 phase 1 cohort;
*age;
proc sort data=demo_1 nodupkey out=demo_1_2;
by releaseid;
run;

proc freq data=demo_1_2;
tables age;
run;

*need to change age since it is categorical;
data demo_1_3; set demo_1_2;
if age = 1 then age_1 = 13;
if age = 3 then age_1 = 15;
if age = 14 then age_1 = 14;
if age = 15 then age_1 = 15;
run;

proc means data=demo_1_3 n mean std;
var age_1;
run;

*female sex;
proc freq data=demo_1_3;
tables sex;
run;

*race and ethnic group;
proc freq data=demo_1_3;
tables race;
```

```
run;

*household income;
proc freq data=demo_1_3;
tables houseinc;
run;

*glycated hemoglobin (a1c);
proc freq data=t1.cbl;
tables mvisit;
run;

data cbl; set t1.cbl;
where mvisit = "M00";
keep releaseid mvisit HbA1c;
run;

proc sort data=cbl;
by releaseid;
run;

proc sort data=demo_1_3;
by releaseid;

data demo_1_4; merge
demo_1_3 (in=a)
cbl (in=b);
by releaseid;
if a=1;
run;

*%;
proc means data=demo_1_4 n mean std;
var HbA1c;
run;

*mmol/mol;
data demo_1_4; set demo_1_4;
glyc_hem = 10.929*(HbA1c-2.15);
run;

proc means data=demo_1_4 n mean std;
var glyc_hem;
run;

*BMI;
data base; set t1.baseline;
keep bmi releaseid;
```

```
run;

proc sort data=demo_1_3;
by releaseid;
run;

proc sort data=base;
by releaseid;
run;

data demo_1_4; merge
demo_1_3 (in=a)
base (in=b);
by releaseid;
if a=1;
run;

proc freq data=demo_1_4;
tables bmi;
run;

proc means data=demo_1_4 n mean std;
var bmi;
where bmi ^= 1 AND bmi^= 3;
run;

*treatment assignment;
data prim; set t1.primout;
keep releaseid tx;
run;

proc sort data=prim;
by releaseid;
run;

proc sort data=demo_1_4;
by releaseid;
run;

data demo_1_5; merge
demo_1_4 (in=a)
prim (in=b);
by releaseid;
if a=1;
run;

proc freq data=demo_1_5;
tables tx;
```

```
run;

*TODAY2 phase 2 cohort;
proc sort data=demo_2 nodupkey out=demo_2_2;
by releaseid;
run;

*age;
data demo_2_3; set demo_2_2;
if age = 1 then age_1 = 13;
if age = 3 then age_1 = 15;
if age = 14 then age_1 = 14;
if age = 15 then age_1 = 15;
run;

proc means data=demo_2_3 n mean std;
var age_1;
run;

*Sex;
proc freq data=demo_2_3;
tables sex;
run;

*race and ethnicity;
proc freq data=demo_2_3;
tables race;
run;

*household income;
proc freq data=demo_2_3;
tables houseinc;
run;

*glycated hemoglobin (HbA1c);
proc sort data=cbl;
by releaseid;
run;

proc sort data=demo_2_3;
by releaseid;

data demo_2_4; merge
demo_2_3 (in=a)
cbl (in=b);
by releaseid;
if a=1;
run;
```

```
*%;
proc means data=demo_2_4 n mean std;
var HbA1c;
run;

*mmol/mol;
data demo_2_4; set demo_2_4;
glyc_hem = 10.929*(HbA1c-2.15);
run;

proc means data=demo_2_4 n mean std;
var glyc_hem;
run;

*BMI;
proc sort data=demo_2_4;
by releaseid;
run;

proc sort data=base;
by releaseid;
run;

data demo_2_5; merge
demo_2_4 (in=a)
base (in=b);
by releaseid;
if a=1;
run;

proc freq data=demo_2_5;
tables bmi;
run;

proc means data=demo_2_5 n mean std;
var bmi;
where bmi ^= 1 AND bmi^= 3;
run;

*treatment assignment;
proc sort data=prim;
by releaseid;
run;

proc sort data=demo_2_5;
by releaseid;
run;
```

```
data demo_2_6; merge
demo_2_5 (in=a)
prim (in=b);
by releaseid;
if a=1;
run;

proc freq data=demo_2_6;
tables tx;
run;
```