Dataset Integrity Check for the SEARCH for Diabetes in Youth Study (SEARCH)

Prepared by NIDDK-CR
March 25, 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 manuscript.

2 Study Background

SEARCH for Diabetes in Youth was a multi-center study focused on improving the understanding of diabetes among children and young adults in the United States. SEARCH had five study centers and had more than 20,000 study participants. The study has provided insights regarding types of diabetes, its complications, and the affects of diabetes on the lives of youth.

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the SEARCH folder in the data package. For this replication, variables were taken from the “SEARCH.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data in the publication by Shah et al. [1]. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Study cohort at baseline and follow-up, 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 SEARCH data files to be distributed are a true copy of the study data.

7 References

[1] Shah AS, Maahs DM, Stafford JM, Dolan LM, Lang W, Imperatore G, Bell RA, Liese AD, Reynolds K, Pihoker C, Marcovina S, D'Agostino RB, Dabelea D. Predictors of Dyslipidemia Over Time in Youth With Type 1 Diabetes: For the SEARCH for Diabetes in Youth Study. Diabetes Care, 40(4), 607-613, April 2017. doi: https://doi.org/10.2337/dc16-2193

Table A: Variables used to replicate Table 1 – Study cohort at baseline and follow-up

Table Variable	**dataset.variable**
Age (years)	search.age_vis
Race/ethnicity	search.race_pl
Male sex	search.gender
BMI z-score	search.bmi_z20
WHtR	search.height search.nhwaist
Type 1 diabetes duration (years)	search.duration_months
A1C (%)	search.hba1cpcnt
A1C (mmol/mol)	search.hba1cpcnt
TC (mg/dL)	search.tchol_mgdl
LDL-C (mg/dL)	search.ldlchol_mgdl
HDL-C (mg/dL)	search.hdlchol_mgdl
Non-HDL-C (mg/dL)	search.tchol_mgdl search.hdlchol_mgdl
TGs (mg/dL), median (Q1, Q3)	search.trig_mgdl
Systolic blood pressure (mmHg)	search.sbp
Diastolic blood pressure (mmHg)	search.dbp

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

Variable	Baseline – Publication (n=1478)	Baseline – DSIC (n=1478)	Diff. (n=0)	Follow-up – Publication (n=1478)	Follow-up – DSIC (n=1478)	Diff. (n=0)
Age (years)	10.8 ± 3.9	10.8 ± 3.9	0 ± 0	17.9 ± 4.1	17.9 ± 4.1	0 ± 0
Race/ethnicity						
Non-Hispanic White	1141 (77.3)	1141 (77.3)	0 (0)	-	-	-
Non-Hispanic Black	140 (9.5)	140 (9.5)	0 (0)	-	-	-
Hispanic	170 (11.5)	170 (11.5)	0 (0)	-	-	-
Other	26 (1.8)	26 (1.8)	0 (0)	-	-	-
Male sex	743 (50.3)	743 (50.3)	0 (0)	-	-	-
BMI z-score	0.48 ± 1.04	0.48 ± 1.04	0 ± 0	0.59 ± 0.96	0.59 ± 0.96	0 ± 0
WHtR	0.48 ± 0.06	0.48 ± 0.06	0 ± 0	0.51 ± 0.08	0.51 ± 0.08	0 ± 0
Type 1 diabetes duration (years)	0.7 ± 0.5	0.7 ± 0.5	0 ± 0	7.8 ± 1.9	7.8 ± 1.9	0 ± 0
A1C (%)	7.6 ± 1.5	7.6 ± 1.5	0 ± 0	9.2 ± 1.8	9.2 ± 1.8	0 ± 0
A1C (mmol/mol)	59.8 ± 16.1	59.6 ± 16.0	0.2 ± 0.1	76.6 ± 19.9	76.6 ± 19.9	0 ± 0
TC (mg/dL)	159 ± 27	159 ± 27	0 ± 0	169 ± 34	169 ± 34	0 ± 0
LDL-C (mg/dL)	91 ± 22	91 ± 22	0 ± 0	96 ± 28	96 ± 28	0 ± 0
HDL-C (mg/dL)	56 ± 13	56 ± 13	0 ± 0	55 ± 13	55 ± 13	0 ± 0
Non-HDL-C (mg/dL)	103 ± 25	103 ± 25	0 ± 0	114 ± 35	114 ± 35	0 ± 0
TGs (mg/dL), median (Q1, Q3)	55 (42, 71)	55 (42, 70)	0 (0, 1)	75 (56, 105)	75 (56, 105)	0 (0, 0)
Systolic blood pressure (mmHg)	99 ± 12	99 ± 12	0 ± 0	106 ± 11	106 ± 11	0 ± 0
Diastolic blood pressure (mmHg)	63 ± 10	63 ± 10	0 ± 0	69 ± 9	69 ± 9	0 ± 0

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies6\SEARCH\private_created_data\Redacted Data";

/*********************************************/
/* DSIC for SEARCH - Shah et al. */
/*********************************************/

*Subsetting SEARCH to appropriate study population for the publication;
data paper1; set dsic.search;
where paper1 = 1;
run;

*double checking the number of participants included in the publication - 1478;
ods select nlevels;
proc freq data=paper1 nlevels;
tables id;
run;

*Age;
proc freq data=paper1;
tables wave;
run;

proc sort data=paper1;
by wave;
run;

proc means data=paper1;
var age_vis;
by wave;
run;

*Race/ethnicity;
proc freq data=paper1;
tables (race_pl race_v2)*wave/norow nopercent;
run;

data paper1_2; set paper1;
if race_pl = "aspi" or race_pl = "natam" or race_pl = "other" then race_pl = "other";
if race_pl = "unknown" then race_pl = "";
run;

proc freq data=paper1_2;
tables race_pl;
run;
```

```
*Male sex;
proc freq data=paper1;
tables gender;
run;

*BMI z score;
proc means data=paper1 n mean std;
var bmi_z20;
by wave;
run;

*WHtR;
proc contents data=paper1;
run;

proc means data=paper1;
var nawaist nhwaist height;
run;

*converting height from m to cm;
data paper1_3; set paper1;
height1 = height*100;
run;

*creating ratio variable for waist-to-height;
data paper1_4; set paper1_3;
naratio = nawaist/height1;
nhratio = nhwaist/height1;
run;

proc means data=paper1_4;
var naratio nhratio;
by wave;
run;

*type 1 diabetes duration (years);
proc freq data=paper1;
tables DAAIRgrp11_BL Type_DAAsens_2 type_prov type_provider;
run;

data paper1_5; set paper1;
duration_years = (duration_months/12);
run;

proc means data=paper1_5;
var duration_years;
by wave;
run;
```

```
*A1C (%);
proc means data=paper1;
var HbA1cPcnt;
by wave;
run;

*A1C mmol/mol;
data paper1_6; set paper1;
Hba1cmmol = ((HbA1cPcnt-2.15)*10.929);
run;

proc means data=paper1_6;
var Hba1cmmol;
by wave;
run;

*TC;
proc means data=paper1;
var TChol_mgdL;
by wave;
run;

*LDL-C;
proc means data=paper1;
var LDLChol_mgdL;
by wave;
run;

*HDL-C;
proc means data=paper1;
var HDLChol_mgdL;
by wave;
run;

*Non-HDL-C;
data paper1_7; set paper1;
nonHDL = TChol_mgdl - HDLChol_mgdl;
run;

proc means data=paper1_7;
var nonHDL;
by wave;
run;

*TGs;
proc means data=paper1 n median q1 q3;
var Trig_mgdL;
```

```
by wave;
run;

*Systolic and diastolic blood pressure;
proc means data=paper1;
var sbp dbp;
by wave;
run;
```