Dataset Integrity Check for Diabetes Prevention Program Outcomes Study (DPPOS) Phase 3

Prepared by NIDDK-CR
June 4, 2024

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 Diabetes Prevention Program (DPP) was a multicenter trial examining the ability of an intensive lifestyle program or treatment with metformin to prevent or delay the development of type 2 diabetes in high-risk individuals with prediabetes. The DPP study showed that both interventions reduced the incidence of diabetes in participants, compared with placebo; the lifestyle intervention proved more effective than metformin in preventing the onset of diabetes. The Diabetes Prevention Program Outcomes Study (DPPOS) was the long-term follow-up of the original DPP study. The DPPOS sought to evaluate the effects of the interventions on the further development of diabetes and diabetes complications, including retinopathy, microangiopathy, and cardiovascular disease.

3 Archived Datasets

A full listing of the archived datasets included in the package can be found in the Roadmap document. All data files, as provided by the Data Coordinating Center (DCC), are located in the DPPOS folder in the data package. For this replication, variables were taken from the “demographic.sas7bdat” and “cvd_mort_canc.sas7bdat” datasets.

4 Statistical Methods

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

5 Results

For select sections of Table 2 and Supplemental Figure S1 in the publication [1], Effect of metformin and lifestyle on first major adverse cardiovascular events and the first occurrence of individual

major cardiovascular components and extended cardiovascular outcome events and Consort diagram, respectively, 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 the table and figure. The results of the replication are within expected variation to the published results.

6 Conclusions

The NIDDK Central Repository is confident that the DPPOS Phase 3 data files to be distributed are a true copy of the study data.

7 References

[1] Goldberg RB, Orchard TJ, Crandall JP, Boyko EJ, Budoff M, Dabelea D, Gadde KM, Knowler WC, Lee CG, Nathan DM, Watson K, Temprosa M. Effects of Long-term Metformin and Lifestyle Interventions on Cardiovascular Events in the Diabetes Prevention Program and Its Outcome Study. Circulation, 145(22), 1632-1641, May 2022. doi: https://doi.org/10.1161/CIRCULATIONAHA.121.056756

Table A: Variables used to replicate results – Effect of metformin and lifestyle on first major adverse cardiovascular events and the first occurrence of individual major cardiovascular components and extended cardiovascular outcome events (Table 2) and Consort diagram (Supplemental Figure 1)

Table Variable	**dataset.variable**
Phase 3 participants	demographic.assign
Total major adverse cardiovascular events	cvd_mort_canc.mace
Non-fatal myocardial infarction	cvd_mort_canc.mi
Non-fatal stroke	cvd_mort_canc.stroke
Cardiovascular death	cvd_mort_canc.cvddth
Extended cardiovascular events outcome	cvd_mort_canc.extmace

Table B: Comparison of values computed in integrity check to reference article table and figure

Characteristic	Pub: Placebo	DSIC: Placebo	Diff.	Pub: Metformin	DSIC: Metformin	Diff.	Pub: Lifestyle	DSIC: Lifestyle	Diff.
Supplemental Figure 1									
Phase 3 participants	774	777	3	749	749	0	736	737	1
Table 2 (Number of Events)									
Total major adverse cardiovascular events	98	96	2	101	100	1	111	107	4
Nonfatal myocardial infarction	43	42	1	46	45	1	35	35	0
Nonfatal stroke	28	27	1	16	16	0	39	37	2
Cardiovascular death	27	38	11	39	45	6	37	47	10
Extended cardiovascular events outcome	157	153	4	157	155	2	174	167	7

Attachment A: SAS Code

```
libname p3 "X:\NIDDK\niddk-dr_studies1\DPPOS\private_created_data\DPPOS Phase 3
Submission\Data final" ;
libname dpp "X:\NIDDK\niddk-
dr_studies1\DPP\private_created_data\DPP_V7R\Data\DPP_Data_2008\Non-Form_Data\Data";

/***************************************/
/* DPPOS Phase 3 DSIC - Goldberg et al. */
/***************************************/

*Supp Figure 1 Info;
proc freq data=p3.demographic;
tables assign;
run;

*Pulling baseline data from DPP for full tx assignment for ALL participants;
proc contents data=dpp.basedata;
run;

data base; set dpp.basedata;
keep assign release_id;
run;

proc sort data=base;
by release_id;
run;

*Cardio event data from DPPOS phase 3;
data cardio; set p3.cvd_mort_canc;
run;

proc sort data=cardio;
by release_id;
run;

data one; merge
base (in=a)
cardio (in=b);
by release_id;
if a=b;
run;

*Total MACE, MI, Stroke, CV Death, Extended MACE;
proc freq data=one;
tables (mace mi stroke cvddth extmace)*assign/norow nopercent nocol;
run;
```