Dataset Integrity Check for the DPPOS Phase 2 Data Files

Prepared by Allyson Mateja
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
May 20, 2016

Contents

1 Standard Disclaimer 3

2 Study Background 3

3 Archived Datasets 3

4 Statistical Methods 4

5 Results 4

6 Conclusions 4

7 References 4

Table A: Variables used to replicate Table 1. Characteristics of the Diabetes Prevention Program Outcomes Study Cohort 5

Table B: Comparison of values computed in integrity check to reference article Table 1 values 6

Attachment A: SAS Code 9

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

Effective prevention is needed to combat the worldwide epidemic of type 2 diabetes. The Diabetes Prevention Program (DPP, 1996-2001) was a randomized trail comparing an intensive lifestyle intervention or masked metformin with placebo in a cohort selected to be of very high risk of developing diabetes. 2776 (88%) of the surviving DPP cohort were followed up in the DPP Outcomes Study (DPPOS, Sept 1, 2002, to Jan 2, 2014) and analyzed by intention to treat on the basis of their original DPP assignment. The primary outcomes were development of diabetes and the prevalence of microvascular disease. Lifestyle intervention of metformin significantly reduced diabetes development over 15 years. There were no overall differences in the aggregate microvascular outcome between treatment groups; however, those who did not develop diabetes had a lower prevalence of microvascular complications than those who did develop diabetes. This result supports the importance of diabetes prevention.

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the either the Data folder in either the DPP or DPPOS data package. For DPP baseline and DPP-end (2001) columns in the table, variables were taken from the 'demographic', 'events', 'f02', 's03', and 'lab' datasets in the DPP data package. For DPPOS-end (2013) columns in the table, variables were taken from the 'demographic', 'events', 'f02', and 'lab' datasets in the DPPOS data package.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Diabetes Prevention Program Research Group in Lancet Diabetes Endocrinol in 2015. To verify the integrity of the datasets, descriptive statistics were computed.

5 Results

Note that not all of the participants in DPP and DPPOS gave consent for sharing their data. As a result, of the 2,776 subjects who continued in DPPOS from the DPP, only 2,607 are included in the data package. In addition, given that only age group is included in the data package, age was approximated using the mid-point of each age group for each participant.

For Table 1 in the publication [1], Characteristics of the Diabetes Prevention Program Outcomes Study Cohort, 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 published in Table 2. The results of the replication are within expected results given missing data.

6 Conclusions

The NIDDK repository is confident that the DPPOS data files to be distributed are a true copy of the manuscript data.

7 References

[1] Nathan, D.M., Barrett-Connor, E., Crandall, J.P., Edelstein, S.L., Goldberg, R.B., Horton, E.S., Knowler, W.C., Mather, K.J., Orchard, T.J., Pi-Sunyer, X., Schade, D., Temprosa, M., and the Diabetes Prevention Program Research Group. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. Lancet Diabetes Endocrinol 2015.

Table A: Variables used to replicate Table 1. Characteristics of the Diabetes Prevention Program Outcomes Study Cohort

Characteristic	File.Variable(s)
Age, y, mean (SD)	DEMOGRAPHIC.AGEGROUP, EVENTS.TOTALTIM
Female, n (%)	DEMOGRAPHIC.SEX
Race or Ethnic Origin, n (%)	DEMOGRAPHIC.RACE_ETH
Weight	F02.APWGHT1, F02.APWGHT2, F02.APWGHT3, DEMOGRAPHIC.BMICAT, S03.SOHGHT1, SO3.SOHGHT2, SO3.SOHGHT3
Body mass index, mean (SD)	F02.APWGHT1, F02.APWGHT2, F02.APWGHT3, DEMOGRAPHIC.BMICAT, S03.SOHGHT1, SO3.SOHGHT2, SO3.SOHGHT3
Diabetes Cases	EVENTS.DIABF
Diabetes Duration	EVENTS.DIABT, EVENTS.TOTALTIM
FPG (mmol/L)	LAB.G000/18
HbA1c (%)	LAB.HBA1
Treatment Group	DEMOGRAPHIC.ASSIGN

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

	Manuscript Baseline (N = 2776)	DSIC Baseline (N=2607)	DIFF Baseline (N=171)	Manuscript DPP-End Placebo (n=935)	DSIC DPP-End Placebo (N=881)	DIFF DPP-End Placebo (N=34)
Age (years)**	51 (10)	51 (10)	0 (0)	54 (10)	54 (9)	0 (1)
Sex (women)	1887 (68%)	1737 (67%)	150 (1%)	644 (69%)	595 (68%)	49 (1%)
Race or ethnic origin: White	1508 (54%)	1497 (57%)	11 (3%)	501 (54%)	496 (56%)	5 (2%)
Race or ethnic origin: African-American	561 (20%)	549 (21%)	12 (1%)	193 (21%)	191 (22%)	2 (1%)
Race or ethnic origin: Hispanic American	426 (15%)	422 (16%)	4 (1%)	145 (16%)	145 (16%)	0 (0%)
Race or ethnic origin: American Indian	156 (6%)	X	X	55 (6%)	X	X
Race or ethnic origin: Asian American–Pacific Islanders	125 (5%)	X	X	41 (4%)	X	X
Weight (kg)	94 (20)	94 (19)	0 (1)	94 (20)	94 (20)	0 (0)
BMI (kg/m²)	34 (7)	34 (6)	0 (1)	34 (7)	34 (7)	0 (0)
Diabetes cases	0	0	0	278 (30%)	266 (30%)	12 (0%)
Diabetes duration (years): Total cohort	0	0	0	0.5 (0.9)	0.5 (0.9)	0 (0)
Diabetes duration (years): Participants who developed diabetes	0	0	0	1.5 (1.0)	1.7 (1.0)	0.2 (0)
FPG (mmol/L)	5.9 (0.5)	6.0 (0.4)	0.1 (0.1)	6.2 (1.1)	6.3 (1.1)	0.1 (0)
HbA1c (%):Total cohort	5.9% (0.5)	5.9% (0.5)	0% (0)	6.1% (0.7)	6.1% (0.7)	0% (0)
HbA1c (%): Participants who developed diabetes	..	--	--	6.5% (0.9)	6.5% (0.9)	0% (0)

	Manuscript DPP-End Metformin (n=926)	DSIC DPP-End Metformin (n=871)	DIFF DPP-End Metformin (n=55)	Manuscript DPP-End Lifestyle intervention group (n=915)	DSIC DPP-End Lifestyle intervention group (n=855)	DIFF DPP-End Lifestyle intervention group (n=60)
Age (years)**	54 (10)	54 (10)	0 (0)	54 (11)	54 (10)	0 (1)
Sex (women)	619 (67%)	571 (66%)	48 (1%)	624 (68%)	571 (67%)	53 (1%)
Race or ethnic origin: White	516 (56%)	514 (59%)	2 (3%)	491 (54%)	487 (57%)	4 (3%)
Race or ethnic origin: African-American	191 (21%)	185 (21%)	6 (0%)	177 (19%)	173 (20%)	4 (1%)
Race or ethnic origin: Hispanic American	141 (15%)	138 (16%)	3 (1%)	140 (15%)	139 (16%)	1 (1%)
Race or ethnic origin: American Indian	47 (5%)	X	X	54 (6%)	X	X
Race or ethnic origin: Asian American–Pacific Islanders	47 (3%)	X	X	53 (6%)	X	X
Weight (kg)	92 (21)	92 (21)	0 (0)	89 (21)	90 (21)	-1 (0)
BMI (kg/m²)	33 (7)	33 (7)	0 (0)	32 (7)	32 (7)	0 (0)
Diabetes cases	199 (21%)	187 (21%)	12 (0%)	132 (14%)	124 (15%)	8 (1%)
Diabetes duration (years): Total cohort	0.3 (0.8)	0.3 (0.8)	0 (0)	0.2 (0.6)	0.2 (0.7)	0 (0.1)
Diabetes duration (years): Participants who developed diabetes	1.5 (1.0)	1.6 (1.0)	0.1 (0)	1.3 (1.1)	1.4 (1.1)	0.1 (0)
FPG (mmol/L)	5.9 (0.8)	6.0 (0.8)	0.1 (0)	5.9 (0.8)	5.9 (0.8)	0 (0)
HbA1c (%):Total cohort	6.0% (0.5)	6.0% (0.6)	0% (0.1)	5.9% (0.5)	5.9% (0.5)	0% (0)
HbA1c (%): Participants who developed diabetes	6.3% (0.7)	6.3% (0.7)	0% (0)	6.4% (0.7)	6.4% (0.7)	0% (0)

	Manuscript DPPOS-End Placebo group (n=780)	DSIC DPPOS-End Placebo group (n=708)	DIFF DPPOS-End Placebo group (n=72)	Manuscript DPPOS-End Metformin group (n=772)	DSIC DPPOS-End Metformin group (n=707)	DIFF DPPOS-End Metformin group (n=65)
Age (years)**	65 (10)	65 (9)	0 (1)	66 (9)	66 (9)	0 (0)
Sex (women)	538 (69%)	481 (68%)	57 (1%)	525 (68%)	470 (66%)	55 (2%)
Race or ethnic origin: White	411 (53%)	395 (56%)	16 (3%)	420 (54%)	408 (58%)	12 (4%)
Race or ethnic origin: African-American	162 (21%)	156 (22%)	6 (1%)	166 (22%)	160 (23%)	6 (1%)
Race or ethnic origin: Hispanic American	119 (15%)	118 (17%)	1 (2%)	118 (15%)	111 (16%)	7 (1%)
Race or ethnic origin: American Indian	52 (7%)	X	X	43 (6%)	X	X
Race or ethnic origin: Asian American–Pacific Islanders	36 (5%)	X	X	25 (3%)	X	X
Weight (kg)	91 (20)	91 (20)	0 (0)	90 (19)	90 (19)	0 (0)
BMI (kg/m²)	33 (7)	33 (7)	0 (0)	32 (7)	32 (7)	0 (0)
Diabetes cases	564 (60%)	446 (62%)	118 (2%)	506 (55%)	400 (57%)	106 (2%)
Diabetes duration (years): Total cohort	6.5 (6.0)	6.5 (6.0)	0 (0)	5.5 (5.8)	5.5 (5.8)	0 (0)
Diabetes duration (years): Participants who developed diabetes	10.3 (4.3)	10.4 (4.3)	0.1 (0)	9.7 (4.3)	9.7 (4.3)	0 (0)
FPG (mmol/L)	6.8 (1.9)	6.8 (1.9)	0 (0)	6.5 (1.9)	6.5 (1.9)	0 (0)
HbA1c (%):Total cohort	6.3% (1.2)	6.2% (1.2)	0.1% (0)	6.1% (1.1)	6.1% (1.1)	0% (0)
HbA1c (%): Participants who developed diabetes	6.7% (1.4)	6.6% (1.3)	0.1% (0.1)	6.5% (1.3)	6.5% (1.3)	0% (0)

	Manuscript DPPOS-End Lifestyle intervention group (n=751)	DSIC DPPOS-End Lifestyle intervention group (n=674)	DIFF DPPOS-End Lifestyle intervention group (n=77)
Age (years)**	66 (11)	66 (10)	0 (1)
Sex (women)	513 (68%)	452 (67%)	61 (1%)
Race or ethnic origin: White	396 (53%)	378 (56%)	8 (3%)
Race or ethnic origin: African-American	148 (20%)	141 (21%)	7 (1%)
Race or ethnic origin: Hispanic American	109 (15%)	107 (16%)	2 (1%)
Race or ethnic origin: American Indian	50 (7%)	X	X
Race or ethnic origin: Asian American–Pacific Islanders	48 (6%)	X	X
Weight (kg)	89 (19)	89 (19)	0 (0)
BMI (kg/m²)	32 (6)	32 (6)	0 (0)
Diabetes cases	480 (52%)	380 (56%)	100 (4%)
Diabetes duration (years): Total cohort	4.8 (5.3)	4.9 (5.3)	0.1 (0)
Diabetes duration (years): Participants who developed diabetes	8.6 (4.3)	8.6 (4.3)	0 (0)
FPG (mmol/L)	6.8 (2.0)	6.7 (1.9)	0.1 (0.1)
HbA1c (%):Total cohort	6.2% (1.2)	6.2% (1.1)	0% (0.1)
HbA1c (%): Participants who developed diabetes	6.7% (1.4)	6.6% (1.3)	0.1% (0.1)

	Manuscript DPPOS-End Non-diabetes (n=1226)	DSIC DPPOS-End Non-diabetes (n=1154)	DIFF DPPOS-End Non-diabetes (n=72)	Manuscript DPPOS-End Diabetes (n=1550)	DSIC DPPOS-End Diabetes (n=1453)	DIFF DPPOS-End Diabetes (n=97)
Age (years)**	67 (10)	66 (10)	1 (0)	65 (10)	65 (10)	0 (0)
Sex (women)	838 (68%)	775 (67%)	63 (1%)	1049 (68%)	962 (66%)	87 (2%)
Race or ethnic origin: White	713 (58%)	708 (61%)	5 (3%)	795 (51%)	789 (54%)	6 (3%)
Race or ethnic origin: African-American	208 (17%)	203 (18%)	5 (1%)	353 (23%)	346 (23%)	7 (0%)
Race or ethnic origin: Hispanic American	183 (15%)	183 (16%)	0 (1%)	243 (16%)	239 (16%)	4 (0%)
Race or ethnic origin: American Indian	70 (6%)	X	X	86 (6%)	X	X
Race or ethnic origin: Asian American–Pacific Islanders	52 (4%)	X	X	73 (5%)	X	X
Weight (kg)	87 (18)	87 (18)	0 (0)	92 (20)	92 (20)	0 (0)
BMI (kg/m^2)	31 (6)	31 (6)	0 (0)	33 (7)	33 (7)	0 (0)
Diabetes cases	0	0	0	100%	100%	0%
Diabetes duration (years): Total cohort	0	0	0	5.6 (5.8)*	5.1 (5.6)*	0.5 (0.2)*
Diabetes duration (years): Participants who developed diabetes	0	0	0	9.6 (4.4)	9.1 (4.4)	0.5 (0)
FPG (mmol/L)	5.7 (0.5)	5.7 (0.5)	0 (0)	7.4 (2.3)	7.4 (2.2)	0 (0.1)
HbA1c (%):Total cohort	5.6% (0.4)	5.6% (0.4)	0% (0)	6.6% (1.4)	6.6% (1.3)	0 (0.1)
HbA1c (%): Participants who developed diabetes	..	--	--	6.6% (1.4)	6.6% (1.3)	0 (0.1)

*Note that this is inclusive of all participants in DPPOS.

**Age was approximated using the mid-point of the age group variable.

Data are mean (SD) of continuous variables or n (%). DPP-end is the last annual visit of the Diabetes Prevention Program (DPP). DPPOS-end includes data from the final annual visit of the DPP Outcomes Study (DPPOS) except for the diabetes cases which include patients who developed diabetes at any point during the study. Diabetes duration for participants who remain non-diabetic calculated as 0 years.

Attachment A: SAS Code

```
*** DPPOS Phase 2 DSIC;
*** Programmer: Allyson Mateja;
*** Date: May 18, 2016;

title 'DPPOS Phase 2 DSIC';
title2 ' ';

libname dppos2 '/prj/niddk/ims_analysis/DPPOS/private_created_data/SASDATA';
libname dppos1 '/prj/niddk/public_orig_data/DPPOS_V1/Official_Archive/DPPOS_PHASE1_Data/non-form-based/';
libname dppos1fo '/prj/niddk/public_orig_data/DPPOS_V1/Official_Archive/DPPOS_PHASE1_Data/form-based/';

libname dpp '/prj/niddk/public_orig_data/DPP_V2.1/v2.1/Data/DPP_Data_2008/Non_Form_Data/Data/';
libname dpp_form '/prj/niddk/public_orig_data/DPP_V2.1/v2.1/Data/DPP_Data_2008/Form_Data/Data/';
libname bridge '/prj/niddk/public_orig_data/DPPOS_V1/Official_Archive/DPP_Bridge_Official_Archive/DPP_BRIDGE_Data/non_form_based';

proc format;
       value sexf 1 = 'M'
                  2 = 'F';

       value racef 1 = 'White'
                   2 = 'African-American'
                   3 = 'Hispanic'
                   4 = 'All other';

data dppos1_demographic;
       set dppos1.demographic;

data dppos2_demographic;
       set dppos2.demographic;

data dpp_demographic;
       set dpp.basedata;

proc sort data = dpp_demographic;
       by release_id;

proc sort data = dppos1_demographic;
       by release_id;

proc sort data = dppos2_demographic;
       by release_id;

proc contents data = dpp_demographic;
proc contents data = dppos1_demographic;
proc contents data = dppos2_demographic;
```

```
data all_subjects;
       merge dpp_demographic    (keep=release_id bmi_cat rename = (bmi_cat = dpp_bmi_cat))
             dppos1_demographic (keep=release_id bmi_cat rename = (bmi_cat = dppos1_bmi_cat))
             dppos2_demographic (in=val3);
       by release_id;
       if val3 then output all_subjects;

data dpp_labs;
       set dpp.lab;

proc freq data = dpp_labs;
       tables visit;
       title3 'dpp labs';

proc freq data = all_subjects;
       tables dpp_bmi_cat*dppos1_bmi_cat*bmi_cat /list missing;
       title3 'Compare BMI';

proc freq data = all_subjects;
       tables sex;
       format sex sexf.;
       title3 ' Table 1, DPP Baseline, Sex';

proc freq data = all_subjects;
       tables race_eth;
       format race_eth racef.;
       title3 'Table 1, DPP Baseline, Race';

data dpp_events;
       set dpp.events;

data s03;
       set dpp_form.s03;

data dpp_baseline_labs;
       set dpp_labs;
       if visit = 'BAS';

proc sort data = dpp_baseline_labs;
       by release_id;

proc sort data = all_subjects;
       by release_id;

proc sort data = s03;
       by release_id;

proc sort data = dpp_events;
       by release_id;
```

```
data dpp_baseline;
       merge all_subjects (in=val1)
             dpp_baseline_labs
             s03;
       by release_id;
       fpg = g000/18;
       if dpp_bmi_cat = 1 then bmi = 25;
       if dpp_bmi_cat = 2 then bmi = 27;
       if dpp_bmi_cat = 3 then bmi = 29;
       if dpp_bmi_cat = 4 then bmi = 31;
       if dpp_bmi_cat = 5 then bmi = 33;
       if dpp_bmi_cat = 6 then bmi = 35;
       if dpp_bmi_cat = 7 then bmi = 37;
       if dpp_bmi_cat = 8 then bmi = 39;
       if dpp_bmi_cat = 9 then bmi = 41;
       if dpp_bmi_cat = 10 then bmi = 45;
       if sohght1 ne . then height = mean (sohght1, sohght2, sohght3);
       base_weight = (height/100)*(height/100)*bmi;
       if agegroup = 1 then age = 35;
       if agegroup = 2 then age = 42;
       if agegroup = 3 then age = 47;
       if agegroup = 4 then age = 52;
       if agegroup = 5 then age = 57;
       if agegroup = 6 then age = 62;
       if agegroup = 7 then age = 67;
       if val1 then output;

proc means data = dpp_baseline;
       var age;
       title3 'Table 1, DPP Baseline, Age';

proc means data = dpp_baseline;
       var base_weight;
       title3 'Table 1, DPP Baseline, Weight';

proc means data = dpp_baseline;
       var bmi;
       title3 'Table 1, DPP Baseline, BMI';

proc means data = dpp_baseline;
       var fpg;
       title3 'Table 1, DPP Baseline, FPG';

proc means data = dpp_baseline;
       var hba1;
       title3 'Table 1, DPP Baseline, HbA1c';

proc sort data = dpp_labs;
```

```
        by release_id visit;

data dpp_labs_end;
        set dpp_labs;
        by release_id visit;
        if last.release_id and visit in ('Y01', 'Y02', 'Y03', 'Y04') then output;

data dpp_f02;
        set dpp_form.f02;

proc sort data = dpp_f02;
        by release_id visit;

data dpp_f02_end;
        set dpp_f02;
        by release_id visit;
        if last.release_id then output;

data dpp_end;
        merge dpp_baseline (in=val1 keep = release_id assign sex race_eth height age)
              dpp_events
              dpp_labs_end
              dpp_f02_end;
        by release_id;
        fpg = g000/18;
        diabdur = totaltim - diabt;
        if diabf = 0 then diabdur = 0;
        if diabdur < 0 then diabdur = 0;
        if apwght1 ne . then weight = mean(apwght1, apwght2, apwght3);
        bmi = weight/((height/100)**2);
        age = age + totaltim;
        if val1 then output;

proc freq data = dpp_end;
        tables assign;
        title3 'Table 1, DPP End, Treatment group';

proc sort data = dpp_end;
        by assign;

proc means data = dpp_end;
        var age;
        class assign;
        types () assign;
        title3 'Table 1, DPP End, Age';

proc freq data = dpp_end;
        tables sex;
        format sex sexf.;
```

```
        by assign;
        title3 'Table 1, DPP End, Sex';

proc freq data = dpp_end;
        tables race_eth;
        format race_eth racef.;
        by assign;
        title3 'Table 1, DPP End, Race/ethnicity';

proc means data = dpp_end;
        var weight;
        class assign;
        types () assign;
        title3 'Table 1, DPP End, Weight';

proc means data = dpp_end;
        var bmi;
        class assign;
        types () assign;
        title3 'Table 1, DPP End, BMI';

proc freq data = dpp_end;
        tables diabf;
        by assign;
        title3 'Table 1, DPP End, Diabetes cases';

proc means data = dpp_end;
        var diabdur;
        class assign;
        types () assign;
        title3 'Table 1, DPP End, Diabetes duration, Total cohort';

proc means data = dpp_end;
        var diabdur;
        where diabf = 1;
        class assign;
        types () assign;
        title3 'Table 1, DPP End, Diabetes duration, Subjects who developed diabetes';

proc means data = dpp_end;
        var fpg;
        class assign;
        types () assign;
        title3 'Table 1, DPP End, FPG';

proc means data = dpp_end;
        var hba1;
        class assign;
        types () assign;
```

```
        title3 'Table 1, DPP End, HbA1c, Total cohort';

proc means data = dpp_end;
        var hba1;
        where diabf = 1;
        class assign;
        types () assign;
        title3 'Table 1, DPP End, HbA1c, Participants who developed diabetes';

data dppos_labs;
        set dppos2.lab;

proc freq data = dppos_labs;
        tables visit;

proc sort data = dppos_labs;
        by release_id;

data dppos_labs_end;
        set dppos_labs;
        if visit = '11A';

data dppos_events;
        set dppos2.events;

proc sort data = dppos_events;
        by release_id;

data dppos_f02;
        set dppos2.f02;

data dppos_f02_end;
        set dppos_f02;
        if visit = '11A';

proc sort data = dppos_f02_end;
        by release_id;

data dppos_end;
        merge dpp_baseline   (in=val1 keep = release_id assign sex race_eth height age)
              dppos_labs_end (in=val2)
              dppos_events
              dppos_f02_end;
        by release_id;
        diab_dur = totaltim-diabt;
        if diabf = 0 then diab_dur = 0;
        if apwght1 ne . then weight = mean(apwght1, apwght2, apwght3);
        bmi = weight/((height/100)**2);
        fpg = g000/18;
```

```
        age = age+ totaltim;
        if val1 and val2 then output;

proc freq data = dppos_end;
        tables assign;
        title3 'DPPOS End, treatment groups';

proc sort data = dppos_end;
        by assign;

proc means data = dppos_end;
        var age;
        class assign;
        types () assign;
        title3 'Table 1, DPPOS End, Age';

proc freq data = dppos_end;
        tables sex;
        format sex sexf.;
        by assign;
        title3 'Table 1, DPPOS End, Sex';

proc freq data = dppos_end;
        tables race_eth;
        format race_eth racef.;
        by assign;
        by assign;
        title3 'Table 1, DPPOS End, Race/ethnicity';

proc means data = dppos_end;
        var weight;
        class assign;
        types () assign;
        title3 'Table 1, DPPOS End, Weight';

proc means data = dppos_end;
        var bmi;
        class assign;
        types () assign;
        title3 'Table 1, DPPOS End, BMI';

proc freq data = dppos_end;
        tables diabf;
        by assign;
        title3 'Table 1, DPPOS End, Diabetes Cases';

proc means data = dppos_end;
        var diab_dur;
        class assign;
        types () assign;
```

```
        title3 'Table 1, DPPOS End, Diabetes duration, Total cohort';

proc means data = dppos_end;
        var diab_dur;
        where diabf = 1;
        class assign;
        types () assign;
        title3 'Table 1, DPPOS End, Diabetes duration, Subjects who developed diabetes';

proc means data = dppos_end;
        var fpg;
        class assign;
        types () assign;
        title3 'Table 1, DPPOS End, FPG';

proc means data = dppos_end;
        var hba1;
        class assign;
        types () assign;
        title3 'Table 1, DPPOS End, HbA1c, Total Cohort';

proc means data = dppos_end;
        var hba1;
        where diabf = 1;
        class assign;
        types () assign;
        title3 'Table 1, DPPOS End, HbA1c, Subjects who developed diabetes';

data dppos_end_diabetes;
        merge dpp_baseline   (in=val1 keep = release_id assign sex race_eth height age)
              dppos_labs_end
              dppos_events
              dppos_f02_end;
        by release_id;
        diab_dur = totaltim-diabt;
        if diabf = 0 then diab_dur = 0;
        if apwght1 ne . then weight = mean(apwght1, apwght2, apwght3);
        bmi = weight/((height/100)**2);
        fpg = g000/18;
        age = age+totaltim;
        if val1 then output;

proc freq data = dppos_end_diabetes;
        tables diabf;
        title3 'DPPOS End, Diabetes cases';

proc sort data = dppos_end_diabetes;
        by diabf;
```

```
proc means data = dppos_end_diabetes;
       var age;
       class diabf;
       types () diabf;
       title3 'Table 1, DPPOS End, Age';

proc freq data = dppos_end_diabetes;
       tables sex;
       format sex sexf.;
       by diabf;
       title3 'Table 1, DPPOS End, Sex';

proc freq data = dppos_end_diabetes;
       tables race_eth;
       format race_eth racef.;
       by diabf;
       title3 'Table 1, DPPOS End, Race/ethnicity';

proc means data = dppos_end_diabetes;
       var weight;
       class diabf;
       types () diabf;
       title3 'Table 1, DPPOS End, Weight';

proc means data = dppos_end_diabetes;
       var bmi;
       class diabf;
       types () diabf;
       title3 'Table 1, DPPOS End, BMI';

proc freq data = dppos_end_diabetes;
       tables diabf;
       by diabf;
       title3 'Table 1, DPPOS End, Diabetes Cases';

proc means data = dppos_end_diabetes;
       var diab_dur;
       class diabf;
       types () diabf;
       title3 'Table 1, DPPOS End, Diabetes duration, Total cohort';

proc means data = dppos_end_diabetes;
       var diab_dur;
       title3 'Table 1, DPPOS End, Diabetes duration, Subjects who developed diabetes';

proc means data = dppos_end_diabetes;
       var fpg;
       class diabf;
       types () diabf;
```

```
        title3 'Table 1, DPPOS End, FPG';

proc means data = dppos_end_diabetes;
        var hba1;
        class diabf;
        types () diabf;
        title3 'Table 1, DPPOS End, HbA1c, Total Cohort';

proc means data = dppos_end_diabetes;
        var hba1;
        where diabf = 1;
        class diabf;
        types () diabf;
        title3 'Table 1, DPPOS End, HbA1c, Subjects who developed diabetes';
```