Diabetes Prevention Program Outcomes Study

Data Release Documentation

February 2016
DPPOS Phase 2 Data Release

Prepared by the DPP Coordinating Center

The George Washington University Biostatistics Center
6110 Executive Boulevard, Suite 750
Rockville, MD 20852

Telephone: (301) 881-9260
Fax: (301) 881-8752

Diabetes Prevention Program Outcomes Study (DPPOS)

Phase 2 Data Release

Table of Contents

1. Introduction

1.1 General

The Diabetes Prevention Program (DPP) was a randomized clinical trial designed to investigate the efficacy of four treatment arms on the prevention of type 2 diabetes in high-risk adults. The Diabetes Prevention Program Outcomes Study (DPPOS) is the long-term followup of the original DPP cohort. Detailed information about the DPP and DPPOS including protocols, intensive lifestyle manuals, references, publication list, and links to MEDLINE abstracts and manuscripts is available at https://dppos.bsc.gwu.edu/. This report describes the public release of Phase 2 of the DPPOS dataset, and is based on all DPPOS data collected after the DPPOS Phase 1 datalock in 2008 and prior to the Phase 2 datalock in January 2014. A brief description of the trial is given below.

1.2 Descriptions DPP and DPPOS

Separate datasets are available for each of the DPP and DPPOS periods described below. Further details of each phase are available from the NIDDK repository corresponding to each data release.

1.2.1 DPP masked intervention period (1996 – July 31, 2001)

Randomization into the DPP began in July 1996 and continued for nearly 3 years through May 1999. Participants were seen at quarterly visits after randomization until the study was terminated by the Data and Safety Monitoring Board. Comprehensive baseline and annual assessments included physical measurements, medical history updates, adverse event assessment, medication adherence and dispensing, questionnaires, and a 2-hour 75g oral glucose tolerance test (OGTT). Mid-year visits were briefer and included a subset of physical measurements, adverse event assessment, medication adherence and dispensing, and a fasting glucose test. Quarterly visits were very brief and included only adverse event assessment and medication adherence and dispensing. OGTTs were discontinued after a confirmed diagnosis of diabetes.

1.2.2 DPP Washout and Bridge period (August 1, 2001 – Fall 2002)

Beginning in August 2001, all participants were unmasked to the study results. Participants assigned to metformin active or placebo participated in a 2-week washout period after which they were unblinded to their masked medications (see DPP Bridge documentation). Between January and July 2002, all participants, including those randomized to lifestyle, were offered the full 16-session lifestyle program in group format. Annual, mid-year and quarterly visits continued as during DPP.

1.2.3 DPPOS Phase 1 (Fall 2002 – August 2008)

DPPOS Phase 1 began in the fall of 2002, with variable start times depending on each clinic's IRB approval. Comprehensive annual assessments continued in DPPOS Phase 1, and included physical measurements, medical history updates, adverse event assessment, medication adherence and dispensing, questionnaires, and a 2-hour 75g oral glucose tolerance test (OGTT). Mid-year visits were briefer and included a subset of physical measurements, some questionnaires, adverse event assessment, medication adherence and dispensing, and a fasting glucose test. OGTTs were discontinued after a confirmed diagnosis of diabetes. Fundus photos were taken on a subset of participants during DPPOS year 1 and on all willing participants in DPPOS year 5.

1.2.4 DPPOS Phase 2 (Fall 2008 – October 2013)

No substantial changes were made between DPPOS Phase 1 and Phase 2. Measurements, medical history updates, adverse event assessment, medication adherence and dispensing, questionnaires, and a 2-hour 75g oral glucose tolerance test (OGTT) continued. Brief mid-year visits included a subset of physical measurements, some questionnaires, adverse event assessment, medication adherence and dispensing, and a fasting glucose test. Some questionnaires were moved from annual to mid-year during DPPOS Phase 2 to reduce the length of the annual visit. OGTTs were discontinued after a confirmed

diagnosis of diabetes. For the first time, cognitive and physical function were assessed in DPPOS Years 8 and 10. Fundus photos were taken on all willing participants in DPPOS year 11.

1.3 Treatment Arms

1.3.1 DPP (1996 – July 2001)

At DPP randomization, participants were randomly assigned to one of four treatment groups: metformin, troglitazone, lifestyle or double-placebo. Participants assigned to one of the medication groups (metformin, troglitazone or placebo) were masked to which medication they were taking, and were given one of three medication regimes: active metformin and troglitazone placebo, active troglitazone and metformin placebo, or double placebo. Participants were given their coded medication at the randomization visit and at all quarterly visits thereafter. The troglitazone arm of the study was discontinued in mid-1998 due to medication toxicity, after which participants assigned to troglitazone were followed off-medication on a modified protocol. Placebo-troglitazone was discontinued in participants assigned to the metformin and placebo arms, while maintaining the masked intervention among those participants. Troglitazone participants continued with mid-year and annual visits, but quarterly visits were not required after this point.

1.3.2 DPP Washout and Bridge period (August 2001 – Fall 2002)

During the DPP washout and bridge period (see DPP Bridge documentation), placebo was discontinued, metformin was continued open-label in participants who had been randomized to metformin and who were willing to continue, and individual lifestyle sessions continued as staffing allowed. Between January and July 2002, all participants, including those randomized to lifestyle, were offered the full 16-session lifestyle program in group format.

1.3.3 DPPOS Phase 1 (Fall 2002 – August 2008)

During Phase 1 of DPPOS, the metformin and lifestyle participants were kept on their study interventions to the extent possible. For participants randomized to Lifestyle, individual lifestyle sessions were discontinued, and instead group-implemented boost lifestyle sessions were held semi-annually. Metformin was continued open-label in participants who had been randomized to metformin and who were willing to continue, In addition, all participants were invited to quarterly Healthy Lifestyle Program (HELP) classes.

1.3.4 DPPOS Phase 2 (Fall 2008 – October 2013)

During Phase 2, as during Phase 1, the metformin and lifestyle participants were kept on their study interventions to the extent possible. For participants randomized to Lifestyle, group-implemented boost lifestyle sessions continued to be held semi-annually. Metformin was continued open-label in participants who had been randomized to metformin and who were eligible and willing to continue. In addition, all participants continued to be invited to quarterly Healthy Lifestyle Program (HELP) classes. Former troglitazone participants were no longer followed during DPPOS Phase 2 and therefore no data is available or included in this release for troglitazone participants.

1.4 Diabetes Diagnosis and Subsequent Treatment

The complete definition of diabetes in the DPP and DPPOS is given in section 3.1 and is unchanged from previous releases. After a participant was confirmed to have diabetes, the assigned intervention was continued and reinforced. However, once a participant was diagnosed with worsening diabetes (defined during DPPOS to be an $HbA1_c \geq 7.0\%$), study metformin was discontinued and the participant was sent to his or her local primary care provider for treatment; participation in the remainder of the DPPOS continued. Former placebo and lifestyle participants continued with DPPOS Phase 2 without respect to $HbA1_c$ levels.

1.5 Exclusions from released data

Data that are part of the DPPOS Phase 2 primary microvascular disease outcomes are included in this data release. Not included in this data release are major secondary outcomes including cancer, cardiovascular events and death which were underpowered during DPPOS Phase 2 and are designated as primary outcomes during DPPOS Phase 3.

In addition, non-research data, including tracking forms, are not included. Serious adverse event data were collected but are also not included in the data release as this data was not adjudicated and is not considered research data.

2. Release Information

2.1 General Information

- No participant identifying information is included.
- A randomly generated 9-character RELEASE_ID uniquely identifies each participant.
- Clinic and other location identifiers have been removed.
- No dates are included; all time points are given as days from randomization.

- The visit schedule in Chapter of the DPPOS Protocol shows the forms and assessments completed at each visit. All data collected during DPPOS Phase 2 years 07 through 11 are included in this release except where indicated.
- Only clinics and participants with IRB approval and informed consent to distribute their data to the NIDDK repository are included. Out of the 2776 participants (excluding those originally randomized to troglitazone) who consented to DPPOS, 2607 participants are included in this release dataset.
- In accordance with HIPAA regulations and to protect the identification of DPP participants, the data has been modified to ensure that no participant is identifiable.
- Only research data is included in the released dataset, including data for all DPPOS clinic visits, lifestyle visits, and laboratory data. Non-research data, including tracking forms, are not included. Serious adverse event data were collected but are also not included in the data release as this data was not adjudicated and is not considered research data.
- All available data from each form and central unit database is included to the extent possible. Missing data was caused by a variety of reasons: the visit was not completed in its entirety; the variable was accidentally not collected or measured; the variable was completed incorrectly; the visit was missed, etc.

2.2 Data Location

Data are released from the DPP Coordinating Center at the George Washington University Biostatistics Center to the Data Repository of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health.

2.2.1 Structure of the SAS Data Files

- Multiple SAS datasets are available in transport files, under the library OS2_REL. One dataset exists for each DPPOS form or dataset.
- The files are included as SAS datasets within transport files with the same name as the embedded form or dataset name and the extension XPT. The SAS code to import each dataset is provided in the appendix.
- The contents of variables in these datasets are provided.

2.2.2 Merging DPP and DPPOS SAS Data Files

The Appendix provides examples of merging the various DPP and DPPOS datasets together, as well as sample analyses.

2.3 De-identified Data

The DPPOS dataset was de-identified in the following manner. All personal identifiers were removed, including participant ID and other personal identifiers (initials, date of birth, etc), clinical center, and all dates. Medications taken by 10 or fewer participants have been removed to protect confidentiality.

Baseline data such as age at randomization, race/ethnicity and body mass index (BMI) were de-identified in the DPPOS data release and are available in the DEMOGRAPHIC dataset. Data in this file is ***identical*** to the BASEDATA data included in the DPP Full Scale data release but includes only participants who participated in DPPOS. For those data, variables that might identify a particular individual were collapsed into wide groupings. For example, race/ethnic groups were coded as Caucasian, African American, Hispanic (anyone who indicated Yes to Hispanic origin), and All Other. Age at baseline was collapsed into 5-year age groups, with truncation of those <40 and those ≥65. Some ethnic groups had body mass index inclusion criteria that might identify specific participants, therefore baseline body mass index (BMI) is given in the following two alternative groupings:

1. collapsed into 2 kg/m^2 groupings; participants with a BMI $\leq$ 26 kg/m^2 were combined, as were those with a BMI $\geq$ 42 kg/m^2
2. collapsed into approximate tertiles of <30 kg/m^2 , ≥30 to <35 kg/m^2 and $\geq$ 35 kg/m^2

2.4 Structure of the Datasets

One record exists in each file for each participant for each visit at which that particular form was completed or data was collected. Variable RELEASE_ID is used to identify a particular participant and variable VISIT to identify which visit was completed.

This dataset includes data collected at DPPOS years 6 through 11 visits including mid-year, annual, diabetes confirmation, and interim visits beginning in fall through October 2013. Data included herein are not included in the DPPOS Phase 1 data release.

The number of participants participating in each follow-up visit is shown in the table below for each regularly-scheduled in-person follow-up visit in the complete dataset and in this release dataset.

Number of participants who completed in-clinic mid-year and annual visits based on forms F01 and F02 by original DPP randomized treatment arm For the original study sample and the DPPOS Phase 1 database in NIDDK repository DPPOS Phase 1 Fall 2008 – October 2013						
	Lifestyle		Metformin		Placebo	
VISIT	Study	Repository	Study	Repository	Study	Repository
06A	41	33	51	38	47	37
06M	30	19	31	22	29	14
07A	749	702	766	724	777	732
07M	747	693	753	707	771	728
08A	758	709	783	740	766	728
08M	749	700	775	730	755	713
09A	740	690	778	731	762	715
09M	732	684	768	720	758	710
10A	728	676	757	709	763	713
10M	726	674	751	706	751	703
11A	719	665	750	703	754	705
11M	697	644	723	677	729	678

3. Statistical Considerations

3.1 Analysis according to original DPP randomization

The DPP was analyzed by intent-to-treat and continued throughout DPPOS according to their treatment groups assigned at DPP randomization. Thus, in general analyses were without regard to adherence to medication or lifestyle during the trial.

3.2 Repeated Measures

Much of the data in DPP and DPPOS were collected at several time points over the years of follow-up. To account for the repeated and variable measurements over time, the average mean change from baseline, as well as comparisons of the changes from baseline among the three treatment groups were computed using SAS Proc MIXED, adjusted for the baseline value of the covariate where appropriate. Changes from baseline to a specific visit were computed and compared across treatment groups using analysis of covariance, adjusted for the baseline value, using SAS Proc GLM.

3.3 DPP and DPPOS Phase 1 Primary Outcome: Diabetes

The primary endpoint for the DPP and DPPOS Phase 1 was time to diabetes as defined by the protocol at the time of the visit:

- Visits through June 23, 1997:
    - fasting glucose >=140 mg/dL, or

- 2-hour post challenge glucose >=200 mg/dL

- Visits on or after June 24, 1997:
  - fasting glucose >= 126 mg/dL, or
  - 2-hour post challenge glucose >=200 mg/dL

An OGTT was completed at annual visits, with only fasting glucose measured at mid-year visits. If a participant had elevated glucose levels at either an annual visit (either fasting or 2-hour glucose) or a mid-year visit (fasting glucose only), diabetes was confirmed at a subsequent visit, usually within 6 weeks, in order for the participant to be diagnosed with diabetes. Confirmation visits included the same glucose measurements as the visits where confirmation was triggered. That is, the confirmation visit following a trigger at an annual visit included an OGTT, whereas a confirmation visit following a trigger at a mid-year visit included a fasting glucose only. Confirmation at an annual visit was based on ***either*** the fasting or the 2-hour glucose level without regard to which glucose value (fasting, 2-hour, both) was elevated at the main (trigger) annual visit.

Many participants had elevated glucose levels at a visit but these levels were not confirmed at the subsequent visit. Visits of this sort were not used to define diabetes.

3.3.1 Time to Diabetes

For the DPP and DPPOS data analyses, the time to diabetes was computed using interval censoring with each interval lasting 6 months, e.g. 3 months before and after the target visit date for semi-annual or annual visits. The diagnosis of diabetes is the time interval during which diabetes was first diagnosed.

On occasion, participants came to clinic visits well outside their targeted visit window. Participants who missed an annual visit but came to the clinic much later in the year, maybe for a mid-year visit, took part in the full annual visit that he or she missed (including the OGTT) instead of the mid-year visit. In such cases, all measurements are included with the annual visit data, as noted on the case report form. However, if the participant was diagnosed with diabetes at that out-of-window visit, the actual date of diagnosis was used; therefore, the interval for the diagnosis of diabetes is the window in which the glucose measurements were actually taken (e.g. the mid-year visit). The remaining mid-year visit data are missing in such cases.

3.3.2 Life Table Analysis

The three treatment arms of the DPP were compared using life table analysis with the log rank test, and proportional hazards models with the “ties=discrete” option in SAS Proc PHREG (see Appendix Section 5.4). Endpoints after DPP formally ended in July 2001 include diabetes diagnosed during the washout and DPPOS Phases 1 and 2, with careful consideration of the impact of the group-lifestyle intervention offered during this bridge period and continued through DPPOS Phase 2 for all participants.

3.4 DPPOS Phase 2 Primary Outcome: Microvascular Disease

The primary outcome for DPPOS Phase 2 was the presence of one or more of the following at the DPPOS year 11 visit, or if the participant is deceased or lost to follow-up before Year 11, the presence of one or more of the following as of his/her last assessment:

a. Nephropathy: micro- or macro-albuminuria (≥30 mg/gram creatinine, confirmed), or renal dysfunction (end-stage renal disease, dialysis or renal transplant) or GFR < 45 ml per min based on serum creatinine, using the CKD-EPI equation or another validated algorithm; the qualifying criteria confirmed)
b. Retinopathy: retinopathy by fundus photography (ETDRS grade of 20 or greater) or adjudicated history of laser or other treatment for retinopathy or
c. Neuropathy: reduction or absence of light touch sensation to monofilament (Semmes-Weinstein 10 gram) in either foot (< 8 of 10 applications detected).

If a participant was taking antihypertensive drugs at the last assessment and does not meet the ACR or eGFR criteria at that time, he or she was considered to have reached the nephropathy outcome if the nephropathy criteria were met at 2 consecutive past visits. This substituted for the occurrence of nephropathy at DPPOS year 11.

The microvascular dataset includes each individual component separately as well as the combined outcome. The variables were adjudicated by a small DPPOS team, which had access to additional data not available for release due to possible de-identification (e.g., dates, uncommon medical details, or small samples sizes).

3.4.1 Analytic Considerations

The primary outcome analysis for DPPOS Phase 2 compared the three intervention groups with respect to the DPPOS year 11 assessment of the components of the microvascular outcome. The global test was used to estimate average prevalence and account for correlations among the 3 components using general estimating equations (GEE) among DPPOS enrolled participants. The global test gives the component outcomes equal weight, and is interpreted as testing for a consistent difference between groups across the component outcomes (see Appendix Section 5.5).

4. File Descriptions

4.1 Data Forms

4.1.1 General

Multiple data collection forms were completed for each participant at every clinical visit. This release includes research data for each data form completed at every visit.

Each form is available as a PDF for use in approved data-release analyses only – ***no form is to be used for primary data collection without specific permission from the Diabetes Prevention Program Research Group or the original source***. Instructions for completing each form are included at the top of each form, and additional instructions are included throughout the form as required. The DPPOS form number can be found at the top-right and the form name at the top-center of all forms.

Data-entry included responses in both the check-boxes and the data-boxes on the data collection forms. Specify-style questions and other questions with responses written on ***underscore lines*** were not data entered; this information is unavailable for analysis and was available only for use by the clinical centers. Specify-style questions that are ***within boxes*** were data entered and are included in this release.

Over the course of DPPOS many forms were changed – new variables were added, new codes were added, and variables were removed. Only the final PDF version of each form is distributed with this data release, although all data collected are included in the data files. Variables that were added will have missing data prior to the addition of the variable and are noted under each specific form below. Deleted variables are not included.

4.1.2 Variable Names on Data Forms

- Variable names for each released variable are embedded in blue on the data form. Variable names for non-released variables are in faded gray.
- All datasets are HIPAA compliant. Information that might identify a specific participant has been excluded from the release datasets, and is indicated in faded gray on the forms. This includes the original DPP participant ID, screening ID, clinical center, and all dates.
- Coding and formats for all variables can be found on the original data form except where described below.
- The numerical value entered for check-box style categorical variables is noted inside the check-boxes with the exception of Yes/No variables which have been re-coded to 1=YES and 0=NO.

- Text information written on forms that is indicated by underscore lines was not data entered and therefore not included in the release datasets. Text information entered in boxes is included.

4.2 Datasets for Non-Form Data

Data not collected on forms but for which datasets are included in this release are as follows:

- Laboratory data: One record for each participant for each visit where laboratory measurements were completed.
- Nutrition: One record of analyzed nutrition data for each participant at DPPOS Year 11.
- Quality of well being: A self-administered Quality of Well Being (QWB) Questionnaire was completed at annual visits during DPPOS Phase 2. One record is included of analyzed QWB data for each participant visit where the QWB was administered.
- Fundus (eye) photos: Two records (one for each eye) of analyzed fundus photo data for all participants during DPPOS Years 5 and 11. This is one of the few instances where data for visits during DPPOS Phase 1 is included.
- An EVENTS file includes summary event variables for diabetes as well as times to events and censoring data. This file has one record for each participant.
- A DEMOGRAPHIC file with one record for each participant which includes treatment assignment, baseline age group, baseline BMI group, sex, and race-ethnicity. This file is identical to the data released with the DPP Full Scale Release dataset but includes only participants in the DPPOS Phase 2 release.
- A MICROVASCULAR file includes summary event variables for each microvascular based on the DPPOS protocol. This file has one record for each participant.
- A CAC file with one record of Coronary Artery Calcification (CAC) CT scans for each participant during DPPOS Year 10.

4.3 Variables Common to All Datasets

Several variables are used to identify a specific participant, visit and time on all datasets. These include:

- RELEASE_ID: This is a randomly generated ID used to link a participant to all other records, and is unique to each participant.
- VISIT: This identifies the visit and is used along with RELEASE_ID to match a participant's visit across the multiple forms completed for that visit. VISIT is coded as follows:
    - 06M, 07M, 08M… 11M: Regularly scheduled DPPOS mid-year visits.
    - 06A, 07A, 08A… 11A: Regularly scheduled DPPOS annual visits.
    - INT: Interim (unscheduled) visits.
    - CON: Confirmation visits to confirm or not-confirm diabetes status; usually completed within 6 weeks of the trigger visit.
    - POV: Primary outcome visits completed after glucose confirmation. Note: Data collected at primary outcome visits included all data that were not collected at the visit where the participant's glucose was first elevated (trigger visit).
- DAYSRAND: The number of days a particular visit occurred after (positive values) randomization.
- **IMPORTANT NOTE:** Visit coding changed from DPP to DPPOS. During DPP visits were coded based on the time from randomization as M03, M06, M09, Y01, M15, etc. During DPPOS however, visits were coded as Annual (corresponding to the approximate month and day of randomization) or Mid-year during each calendar year of DPPOS allowing for a 2-month window around each visit. Thus visits occurred at the following time ranges:

DPPOS Visits	Calendar year
06M, 06A	July 2007 – October 2008
07M, 07A	July 2008 – October 2009
08M, 08A	July 2009 – October 2010
09M, 09A	July 2010 – October 2011
10M, 10A	July 2011 – October 2012
11M, 11A	July 2012 – October 2013

Therefore to order visits as time from randomization, the variable DAYSRAND needs to be used in conjunction with VISIT as shown in Appendix Section 5.2.

4.4 Follow-up Visit Inventory Forms (F-forms)

4.4.1 OS2_REL.F01: STANDARD FOLLOW-UP VISIT INVENTORY

DPPOS Form F01 was used to record information collected at mid-year visits (NOT annual visits). Variable VISIT is used to identify the visit completed. Note that after Year 7, medications were collected on a supplemental form but have been re-merged into the corresponding F01. As a result, medication variable names change across visit years and prior releases.

4.4.2 OS2_REL.F02: MAJOR FOLLOW-UP VISIT INVENTORY

DPPOS Form F02 was used to record information collected at annual visits. Variable VISIT is used to identify the visit completed. Note that after Year 7, medications were collected on a supplemental form but have been re-merged into the corresponding F02. As a result, medication variable names change across visit years and prior releases.

4.4.3 OS2_REL.F03: INTERIM FOLLOW-UP VISIT INVENTORY

DPPOS Form F03 was used to record information collected at interim visits (e.g. not mid-year or annual visits). The reason for interim visits is documented and includes reasons such as coded medication management, blood pressure or other concomitant disease and concomitant medication management, etc. Interim visits do not have a standard VISIT recorded, therefore VISIT = INT for all F03 forms.

4.4.4 OS2_REL.F04: MISSED FOLLOW-UP VISIT REPORT

DPPOS Form F04 was used to record information about a mid-year or annual visit that was missed and therefore no data are available. Variable VISIT is used to identify the missed scheduled visit.

4.4.5 OS2_REL.F06: HOME VISIT INVENTORY

DPPOS Form F06 was used to record information about a mid-year or annual visit that was completed outside the clinic (e.g. at home, nursing home, etc). During DPPOS Phase 2, 168 such visits took place and limited data were collected at home visits. Variable VISIT is used to identify the visit completed. Note that after Year 7, medications were collected on a supplemental form but have been re-merged into the corresponding F06. As a result, medication variable names change across visit years and prior releases.

4.4.6 OS2_REL.F07: METFORMIN DISCONTINUATION FORM

DPPOS Form F07 was used to record information about metformin participants who were not taking study metformin. If a permanent condition was reported in section B, additional F07 forms were not required. For participants off metformin temporarily and eligible to restart, form F07 was completed every

time study metformin was not dispensed. PNP (Participant Not Present) was marked for visit if form F07 was completed without the participant's presence. Variable VISIT is used to identify the visit completed.

4.4.7 OS2_REL.F08: METFORMIN SAFETY AND ADHERENCE FORM

DPPOS Form F08 was used to record information about metformin participants' adherence to study-provided metformin, the barriers to medication adherence, adherence strategies used, and to record safety pregnancy testing and CBC. NOTE: This data was collected on forms F01, F02 and F06 during prior phases of DPP and DPPOS. Variable VISIT is used to identify the visit completed. This form was only completed for participants who took at least some study-provided metformin since the prior study visit.

4.5 Questionnaires (Q-forms)

4.5.1 OS2_REL.Q01: BECK QUESTIONNAIRES

DPPOS Form Q01 includes both the Beck Depression Inventory and the Beck Anxiety Inventory. Form Q01 was self-administered at DPPOS Phase 2 mid-year visits. Part II is the Beck Depression Inventory and Part III is the Beck Anxiety Inventory. Variable VISIT is used to identify the visit completed.

To score the BDI or BAI, add up the score for each of the questions (exclude BDI question 19b) and obtain the total. The highest score on each of the twenty-one BDI and BAI questions is three, therefore the highest possible total for the whole BDI or BAI is 63 and the lowest possible score is zero.

4.5.2 OS2_REL.Q02: HEALTH SURVEY QUESTIONNAIRE

DPPOS Form Q02 is the MOS SF-36 questionnaire. Form Q02 was self-administered at DPPOS Phase 2 mid-year visits 08M, 10M and 11M. Variable VISIT is used to identify the visit completed. The scoring algorithm for this questionnaire is available at http://www.rand.org/health/surveys_tools/mos/mos_core_36item_scoring.html.

4.5.3 OS2_REL.Q03: MODIFIABLE ACTIVITY QUESTIONNAIRE

DPPOS Form Q03 is the Modifiable Activity Questionnaire. Form Q03 was interviewer-administered at DPPOS Phase 2 mid-year visits. Variable VISIT is used to identify the visit completed. To score the MAQ, each activity is weighted by its estimated relative intensity or MET value and added across the year.

4.5.4 OS2_REL.Q13: URINARY INCONTINENCE QUESTIONNAIRE

DPPOS Form Q13 was used to record participants' issues related to urinary incontinence during the past year. Variable VISIT is used to identify the visit completed. This form was self-administered at DPPOS Phase 2 annual visits.

4.5.5 OS2_REL.Q15: NEUROPATHY QUESTIONNAIRE

DPPOS Form Q15 is the Neuropathy (MNSI) questionnaire. Form Q15 was self-administered at DPPOS Phase 1 and Phase 2 annual visits. Variable VISIT is used to identify the visit completed. IMPORTANT NOTE: Because neuropathy data was not released with DPPOS Phase 1, both DPPOS Phase 1 and Phase 2 data are included in this release.

4.5.6 OS2_REL.Q16: ECONOMIC EVALUATION QUESTIONNAIRE

DPPOS Form Q16 was used to record participants' costs and time related to food purchase and preparation, exercise behavior and equipment, and medical care during DPPOS participation. This form was self-administered one time by each participant during DPPOS Year 10. Variable VISIT = 10A or 10M for this form. This is similar to form Q12 used during DPP.

4.5.7 OS2_REL.Q17: COGNITIVE ASSESSMENTS QUESTIONNAIRE

DPPOS Form Q17 was used to record participants' cognitive function testing results. Cognitive function testing was completed during annual visits in DPPOS Years 8 and 10. Variable VISIT is used to identify the visit completed. The measures of cognitive function administered in the DPPOS include the six-item screener, the Spanish English Verbal Learning Test, the Word Fluency Test of the Multilingual Aphasia Examination (letters F or P {Spanish speakers} and Animal Category), and the Digit Symbol Substitution Test of the Wechsler Adult Intelligence Scale-Revised (WAIS-R). These measures tap five cognitive domains: global mental status, verbal learning and memory, word fluency, and psychomotor speed, respectively.

4.5.8 OS2_REL.Q19: BEHAVIORAL, DIET, AND ACTIVITY QUESTIONNAIRE

DPPOS Form Q19 is a Behavioral, Diet, and Activity Questionnaire completed at all mid-year visits during DPPOS Phase 2. Variable VISIT is used to identify the visit completed.

4.6 Lifestyle Forms (L-forms)

During DPPOS all participants were invited to quarterly Healthy Lifestyle Program (HELP) sessions, and participants randomized to the original lifestyle group were invited to semi-annual boost classes.

4.6.1 OS2_REL.L07: LIFESTYLE SESSION LOG

DPPOS form L07 records each participant who attended group HELP or boost sessions, along with session information. Up to 30 participants could have been entered on one lifestyle session log. The codes required for this form can be found in the file "Lifestyle Coding for L07.PDF".

4.7 Event Forms (E-forms)

4.7.1 OS2_REL.E04: PREGNANCY CONFIRMATION REPORT

DPPOS Form E04 was used to document a confirmed pregnancy. The dates on this form have been transformed into days since randomization as indicated on the PDF version of the form. This form is filled out for every confirmed pregnancy, and is matched to the E05 by the "Date of Positive Pregnancy Test" variable (transformed to days from randomization).

4.7.2 OS2_REL.E05: PREGNANCY OUTCOME REPORT

DPPOS Form E05 was used to document pregnancy outcomes. The dates on this form have been transformed into days since randomization as indicated on the PDF version of the form. This form is filled out for every confirmed pregnancy and is matched to the E04 (above) by the "Date of Positive Pregnancy Test" variable (transformed to days from randomization).

4.7.3 OS_REL.E11: GASTRIC REDUCTION SURGERY EVENT

DPPOS Form E11 was used to document any gastric reduction surgery reported by a participant. The dates on this form have been transformed into days since randomization as indicated on the PDF version of the form. Possible types of gastric reduction surgery include gastric banding, gastric bypass and other types of gastric surgery intended to treat obesity. Reversals of prior gastric reduction surgery are also reported. Completion of this form is triggered via the F01, F02, F03 or F06 follow-up forms. Variable VISIT is used to identify the visit the form was completed.

4.8 Procedures (P-forms)

4.8.1 OS2_REL.P07: BLOOD DRAW PROCEDURES

DPPOS Form P07 Records blood draw times for all oral glucose tolerance tests (OGTTs). This information includes the time the participant started drinking the glucola, the time of the 30-minute blood draw, and the time of the 2-hour blood draw. This information has been appended to the LAB data and is

not released as a separate P07 dataset.

4.8.2 OS2_REL.P09: PHYSICAL FUNCTION ASSESSMENTS

DPPOS Form P09 Records results from the physical function assessments. Evaluation of physical performance for each participant was based on the short-physical performance battery (SPPB) including the grip strength test, balance test, gait speed test, and chair stand test. Physical function testing was completed during annual visits in DPPOS Years 8 and 10. Variable VISIT is used to identify the visit completed.

4.9 Report Forms (R-forms)

4.9.1 OS2_REL.R04: CHD Risk Status Report

DPPOS Form R04 was used to identify coronary heart disease risk factors to determine LDL goals as defined by NCEP guidelines for adults. Variable VISIT is used to identify the visit completed. This form was completed at all visits where LDL was measured.

4.10 Central Unit Datasets

4.10.1 OS2_REL.LAB: Laboratory Data

DPPOS data LAB includes laboratory results from all regularly scheduled visits. The laboratory results outlined in the table below were measured at the given measurement times. Only regularly scheduled laboratory data are included. Variable VISIT is used to identify the visit completed.

Records which include OGTT results include corresponding blood draw times as collected on Form F07. Form F07 collected the time the participant started drinking the glucola, the time of the 30-minute blood draw, and the time of the 2-hour blood draw.

Variable (concentration for lab measurements)	Variable name	Measurement times
Serum creatinine (mg/dL)	CREA	06A, 07A, 08A, 09A, 10A,11A
Urine creatinine (mg/dL)	UCRE	06A, 07A, 08A, 09A, 10A,11A
HbA1c (%)	HBA1	Any visit – measured annually per protocol and semi-annually after diabetes diagnosis. When diabetes diagnosed at a semi-annual visit, also measured at CON, POV or INT to capture HbA1c as close as possible to the diagnosis of diabetes.
Total cholesterol (mg/dL)	CHOL	06A, 07A, 08A, 09A, 10A,11A, INT
Triglycerides (mg/dL)	TRIG	06A, 07A, 08A, 09A, 10A,11A, INT
HDL (mg/dL)	CHDL	06A, 07A, 08A, 09A, 10A,11A, INT
LDL (mg/dL)	CLDL	06A, 07A, 08A, 09A, 10A,11A, INT
VLDL (mg/dL)	VLDL	06A, 07A, 08A, 09A, 10A,11A, INT
LDL-B subfraction (mg/dL)	LDLB	06A, 07A, 08A, 09A, 10A,11A, INT
LDL-C subfraction (mg/dL)	LDLC	06A, 07A, 08A, 09A, 10A,11A, INT
LDL particle size (mg/dL)	LDLZ	06A, 07A, 08A, 09A, 10A,11A, INT
CRP (mg/dL)	CRP	11A

Fibrinogen (mg/dL)	FIBR	11A

OGTT measurements +		
Time started drinking glucola +	DRNK0M – SAS TIME5. Format – seconds since midnight	06A, 07A, 08A, 09A, 10A, 11A, INT, CON, POV
30-minute blood draw time +	DRNK30M – SAS TIME5. Format – seconds since midnight	06A, 07A, 08A, 09A, 10A, 11A, INT, CON, POV
2-hour blood draw time +	DRNK2H – SAS TIME5. Format – seconds since midnight	06A, 07A, 08A, 09A, 10A, 11A, INT, CON, POV
Fasting Plasma Glucose (mg/dL)	G000	Any visit – measured semi-annually per protocol and at diabetes confirmation (CON). Also measured at POV or INT during OGTT when diabetes was diagnosed by a fasting glucose only.
30 Minute Plasma Glucose (mg/dL)	G030	06A, 07A, 08A, 09A, 10A, 11A, INT, CON, POV. Measured at CON, POV or INT to capture OGTT as close as possible to the diagnosis of diabetes.
2 Hour Plasma Glucose (mg/dL)	G120	06A, 07A, 08A, 09A, 10A, 11A, INT, CON, POV. Measured at CON, POV or INT to capture OGTT as close as possible to the diagnosis of diabetes.
Fasting Insulin (uU/mL)	I000	06A, 07A, 08A, 09A, 10A, 11A, INT, CON, POV. Measured at CON, POV or INT to capture insulin as close as possible to the diagnosis of diabetes.
30 Minute Insulin (uU/mL)	I030	06A, 07A, 08A, 09A, 10A, 11A, INT, CON, POV. Measured at CON, POV or INT to capture insulin as close as possible to the diagnosis of diabetes.

+ Blood draw times were collected on form P07 and included on all records where an OGTT was completed.

4.10.2 OS2_REL.NCC: Nutrient Data

DPPOS data NCC includes DPPOS Year 11 data based on an interviewer-administered semi-quantitative food frequency questionnaire. The original questionnaire is not available for release. Data released includes summary information as coded by the Nutrition Coding Center at the University of South Carolina. The listing below shows categories and coding for categorical variables; all other coding can be found in the SAS Contents. Please note that some NCC dataset variable names have changed from prior releases.

Variable	Description	Units	Coding
NCCCHKSKIN	"How often do you eat the skin on chicken?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCFG1	Bread, Cereal, Rice & Pasta (High Fiber/Low Fat)	Servings per day	
NCCFG2	Bread, Cereal, Rice & Pasta (Low Fiber/High Fat)	Servings per day	
NCCFG3	Bread, Cereal, Rice & Pasta (Low Fiber/Low Fat)	Servings per day	
NCCFG4	Vegetable (Tomato)	Servings per day	
NCCFG5	Vegetable (Dark Green/Deep Yellow)	Servings per day	
NCCFG6	Vegetable (Cruciferous)	Servings per day	
NCCFG7	Vegetable (Other)	Servings per day	
NCCFG8	Fruit & Fruit Juice (Citrus)	Servings per day	
NCCFG9	Fruit & Fruit Juice (Other)	Servings per day	
NCCFG10	Dairy (High Fat)	Servings per day	
NCCFG11	Dairy (Low Fat – Including up to 2% Milk)	Servings per day	
NCCFG12	Fish (High Fat)	Servings per day	
NCCFG13	Fish (Low Fat)	Servings per day	
NCCFG14	Fish (High Omega 3 Fatty Acids)	Servings per day	
NCCFG15	Dried Beans	Servings per day	
NCCFG16	Eggs	Servings per day	
NCCFG17	Meat (High Fat)	Servings per day	
NCCFG18	Meat (Low Fat)	Servings per day	
NCCFG19	Poultry (High Fat)	Servings per day	
NCCFG20	Poultry (Low Fat)	Servings per day	
NCCFG21	Sweets & Desserts	Servings per day	
NCCFG22	Fats & Oils	Servings per day	
NCCFG23	Soy Products	Servings per day	
NCCFG24	Nuts & Seeds	Servings per day	

Variable	Description	Units	Coding
NCCFG25	Coffee & Tea	Servings per day	
NCCFG26	Meal Replacements (Instant Breakfast / Slimfast)	Servings per day	
NCCFG27	Alcohol	Servings per day	
NCCFRSTMEAL	"How soon after you wake up do you have your first meal of the day?"	N/A	. = Missing
NCCHERBS	"Have you taken Herbs/Bot Supp During past Month?"		0= Yes, regularly 1= Yes, not regularly 2= No
NCCHR_MINS	Unit of measure for NCCFRSTMEAL	N/A	1=Hours 2=Minutes . = Missing
NCCLARGMEAL	"Which meal is usually your largest meal?"	N/A	1= 1st 2= 2nd 3= 3rd 4= 4th 5= 5th 6= 6th 7= 7th 8= 8th 9= 9th .=Missing
NCCLFATBACN	"If you eat bacon or sausage, how often do you eat low-fat bacon or sausage?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCLFATBEEF	"If you eat ground beef, how often do you use lean or extra lean ground beef?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCLFATCAKE	"If you eat cookies or cake, how often do you eat low-fat cookies or cakes?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCLFATCHPS	"If you eat snacks such as chips or popcorn, how often do you eat low-fat chips, etc?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCLFATLMTS	"If you eat hot dogs, bologna or other lunch meats, how often do you eat low-fat lunch meats?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCLFATYGRT	"If you eat yogurt, how often do you eat low-fat yogurt?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCLFCHZ	"If you eat cheese (cottage cheese, cheddar cheese, cream cheese, American), how often do you eat low-fat cheese?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing

Variable	Description	Units	Coding
NCCMEATFAT	"How often do you eat the fat on meat?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCMLPERDAY	"How many meals per day do you usually eat?"	N/A	0 – 7 . = Missing
NCCMOREDRNK	"How often do you drink 7+ drinks w/l 24 Hrs?"		If YES to SEVENALC, then: 1 = Once a week or more 2 = No answer 3 = Less than once a month 4 = 3 times per month
NCCOFTENFAT	"How often is fat or oil used in cooking the foods you eat?	N/A	1=Never/<1 Per Month 2=1 Per Month 3=2-3 Per Month 4=1 Per Week 5=2 Per Week 6=3-4 Per Week 7=5-6 Per Week 8=1 Per Day 9=2+ Per Day . = Missing
NCCPFG1	Bread, Cereal, Rice & Pasta	Servings per day	
NCCPFG2	Vegetable	Servings per day	
NCCPFG3	Fruit	Servings per day	
NCCPFG4	Milk, Yogurt & Cheese	Servings per day	
NCCPFG5	Meat, Poultry, Fish, Dry Beans, Eggs & Nuts	Servings per day	
NCCPFG6	Fats, Oils & Sweets	Servings per day	
NCCSALT	"How often do you add salt to your food at the table?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
NCCSNACKDAY	"How many snacks do you usually have per day? (This does not include diet beverages, coffee, tea or water)"	N/A	0 – 99 . = Missing
NCCVITAMINS	"During the past month have you taken any vitamins or minerals?"	N/A	0= Yes, regularly 1= Yes, not regularly 2= No

4.10.3 OS2_REL.QWB: Quality of Well Being Data

DPPOS data QWB includes the annual data based on a self-administered quality of well being questionnaire (QWB-SA). This survey inquired of health problems that had occurred in the 3 days prior to the questionnaire, not including the day the questionnaire was administered. Data released include the summary variable TOTALQWB as coded by the Quality of Well Being Center at the University of California, San Diego. Questions from the original survey are not available in the dataset with the exception of Question 9 A, B and C.

4.10.4 OS2_REL.FUNDUS: Fundus (eye) Photo Data

DPPOS data FUNDUS includes results from fundus photos during Years 5 and 11 of DPPOS (with VISIT coded as 05Y or 11Y). Note that DPPOS Year 5 fundus photos were collected during DPPOS Phase 1 but not released at the time as the results were used for the DPPOS Phase 2 primary outcome. Digital or film (Year 5 only) photographs were taken locally by trained photographers using FPRC 7-standard field color fundus photography. Photographs were read by the DPP Fundus Photo Reading Center at the University of Wisconsin.

4.10.5 OS2_REL.CAC: Coronary Artery Calcification Data

DPPOS data CAC includes results from Coronary Artery Calcification (CAC) CT scans obtained in Year 10 of DPPOS. Scans were read by the DPP CAC Photo Reading Center at the Los Angeles Biomedical Research Institute. DPPOS Participants underwent the following CT procedures:

1.) Coronary calcium scan: this was a prospectively gated, non-contrast CT scan of the heart. Most cephalad slice of the scan included the pulmonary artery bifurcation. The most caudad slice went beyond the apex of the heart.

2.) Single slice abdomen scan: this was a non contrast CT scan of the abdomen, between the L4-L5 interspace. Scan included entire girth of the abdomen in the field of view.

4.11 Created Datasets

4.11.1 OS2_REL.DEMOGRAPHIC: Demographic Data

DPPOS data DEMOGRAPHIC includes one record for each participant in the released database. Data in this file is ***identical*** to the BASEDATA data included in the DPP Full Scale data release but includes only participants who participated in Phase 2 of DPPOS, and includes the following variables:

Variable	Brief description	Type	Coding	Details
RELEASE_ID	DPP ID for public release datasets	Character	9-digit character number beginning with “100”	Randomly assigned.
AGEGROUP	Age group at randomization (years)	Numeric	1 = less than 40 2 = 40-44 3 = 45-49 4 = 50-54 5 = 55-59 6 = 60-64 7 = 65 and older	Computed based on date of randomization and birth date, from screening form S07.
ASSIGN	Treatment assignment	Character	Lifestyle Metformin Placebo Troglitazone	Randomized treatment assignment. Not available on any data form.

Variable	Brief description	Type	Coding	Details
BMI_CAT	BMI categorized (kg/m^2)	Numeric	BMI categorized into the following groups: 1: <26 kg/m^2 2: ≥26 to <28 kg/m^2 3: ≥28 to <30 kg/m^2 4: ≥30 to <32 kg/m^2 5: ≥32 to <34 kg/m^2 6: ≥34 to <36 kg/m^2 7: ≥36 to <38 kg/m^2 8: ≥38 to <40 kg/m^2 9: ≥40 to <42 kg/m^2 10: ≥ 42 kg/m^2	Body mass index. Computed based on height and weight as measured during screening on screening form S03. Average of the 2 (or 3) measured heights and average of the 2 (or 3) measured weights were used. Used for eligibility.
BMIGROUP	BMI group (kg/m^2)	Numeric	BMI collapsed into the following groups: 1: <30 kg/m^2 2: ≥30 to <35 kg/m^2 3: ≥35 kg/m^2	Body mass index. Computed based on height and weight as measured during screening on screening form S03. Average of the 2 (or 3) measured heights and average of the 2 (or 3) measured weights were used. Used for eligibility.
RACE_ETH	Race/ethnicity	Numeric	1 = Caucasian 2 = African American 3 = Hispanic, of any race 4 = All other	Self-reported race/ethnicity based on the 1990 census questionnaire during screening on Form S03.
SEX	Sex	Numeric	1 = Male 2 = Female	Collected during screening on form S03.

4.11.2 OS2_REL.EVENTS: Events Data

DPPOS data EVENTS includes one record for each participant. This file is updated from the DPP Phase 1 data, and includes the following variables:

Variable	Brief description	Type	Coding	Details
RELEASE_ID	Participant ID for repository	Character	9-digit character number beginning with "100"	Randomly assigned ID (NOT DPP ID).
DIABF	Indicator of diabetes	Numeric	0 = No 1 = Yes	Indicator of ever diagnosed with diabetes during DPP or DPPOS. Computed based on fasting and/or 2-hour glucose values from the central laboratory.
DIABT	Years to first diabetes	Numeric		Number of years from randomization to visit where diabetes was diagnosed – OR – Number of years from randomization to final visit where glucose was measured if not diabetic by final visit.

Variable	Brief description	Type	Coding	Details
DIABV	Interval for diabetes	Numeric	1 = Month 6 2 = Year 1 3 = Month 18 4 = Year 2 5 = Month 30 6 = Year 3 7 = Month 42 8 = Year 4 9 = Month 54 10 = Year 5 Etc.	True time interval at which diabetes was diagnosed (NOT necessarily the VISIT that was conducted) – OR – The final visit where glucose was measured if not diabetic by final visit. Note: Intervals are defined as 3-months before and 3-month after the target visit date except for interval 1 which began at randomization.
RANDPER	Randomization period	Numeric	1 = July –September 1996 2 = October – December 1996 3 = January – March 1997 4 = April – June 1997 5 = July –September 1997 6 = October – December 1997 7 = January – March 1998 8 = April – June 1998 9 = July –September 1998 10 = October – December 1998 11 = January – March 1999 12 = April – May 1999	Along with TOTALTIM, the randomization period can be used to assess participant's completion of the trial.
TOTALTIM	Years in study	Numeric		Total years in study through last visit of any type (quarterly, mid-year, annual or interim) as of October 2013.

4.11.3 OS2_REL.MICROVASCULAR: Microvascular Events Data

DPPOS data MICROVASCULAR includes one record for each participant. This data represents the primary outcome for Phase 2 of DPPOS. Missing observations are those participants that did not have the necessary visit and/or data to determine the microvascular event status. This file includes the following variables:

Variable	Brief description	Type	Coding	Details
RELEASE_ID	Participant ID for repository	Character	9-digit character number beginning with "100"	Randomly assigned ID (NOT DPP ID).
EVTRET	Indicator of retinopathy	Numeric	0 = No 1 = Yes	Retinopathy at the DPPOS Year 11 visit or if missing at the DPPOS Year 5 or year 1 visit.
EVTNEU	Indicator of neuropathy	Numeric	0 = No 1 = Yes	Neuropathy at the DPPOS Year 11 visit or if missing at a previous visit.

Variable	Brief description	Type	Coding	Details
EVTNEP	Indicator of nephropathy	Numeric	0 = No 1 = Yes	Nephropathy at the DPPOS Year 11 visit or if missing or taking blood pressure medication, at a prior visit, confirmed.
EVTMICRO	Indicator of any microvascular disease	Numeric	0 = No 1 = Yes	Any of the 3 microvascular disease components at the DPPOS Year 11 visit, as described for each of the 3 components.

5. Appendix: SAS programs

5.1 Sample SAS program to import datasets

```
libname OS2_REL "directory for the SAS datasets on your host";
filename tranfile "name of the transport file on your host";
proc cimport data=OS2_REL.data infile=tranfile;
run;
```

/* For example to import file OS2_REL.F01: /

```
libname OS2_REL "c:\mysasfiles";
filename tranfile "c:\myxptfiles\F01.XPT";
proc cimport data=OS2_REL.F01 infile=tranfile;
run;
```

5.2 Sample SAS program to create datasets

/* DPP data */

```
libname DPPREL 'Directory where you stored DPP data';
%macro impfile(FN);
filename tranfile "Directory of DPP xpt files\&FN..XPT";
proc cimport data=DPPREL.&FN infile=tranfile;
run;
%mend;
        %impfile(S03);  * Baseline info;
        %impfile(F01);
        %impfile(F02);
        %impfile(F06);
        %impfile(BASEDATA);
        %impfile(EVENTS);

proc sort data=DPPREL.S03; by release_id; run;
proc sort data=DPPREL.basedata; by release_id; run;
```

/* Use S03 data to get baseline height and merge with BASEDATA to estimate weight at baseline */

```
        data S03; merge DPPREL.S03 DPPREL.basedata(keep=release_id BMI_CAT); by release_id;
```

```
   if SOHGHT1>. then HEIGHT = MEAN(SOHGHT1,SOHGHT2,SOHGHT3);

   *** Estimate weight at baseline from BMI categories;
select (bmi_cat);
      when (1) BMI=25;
      when (2) BMI=27;
      when (3) BMI=29;
      when (4) BMI=31;
      when (5) BMI=33;
      when (6) BMI=35;
      when (7) BMI=37;
      when (8) BMI=39;
      when (9) BMI=41;
      when (10) BMI=45;
                  otherwise;
                  end;

 weight=(height/100)*(height/100)*bmi;
visit='BAS';
drop bmicat;
run;

/* Push together all DPP data from baseline through followup */
data dpp_follow; set S03 DPPREL.f01 DPPREL.f02 DPPREL.f06;
run;
proc sort; by release_id; run;

/* DPP-bridge data */
libname DPPBRREL 'Directory where you stored Bridge data';
%macro impfile(FN);
filename tranfile "Directory of Bridge xpt files\&FN..XPT";
proc cimport data=DPPBRREL.&FN infile=tranfile;
run;
%mend;
          %impfile(F01);
          %impfile(F02);
          %impfile(EVENTS);
```

```
/* Push together DPP-bridge data */
data dppbr_follow; set DPPBRREL.f01 DPPBRREL.f02;
run;
proc sort; by release_id; run;

/* DPPOS Phase 1 data */
libname DPPOSR1 'Directory where you stored DPPOS Phase 1 data';
%macro impfile(FN);
filename tranfile "Directory of DPPOS1 xpt files\&FN..XPT";
proc cimport data=DPPOSR1.&FN infile=tranfile;
run;
%mend;
        %impfile(F01);
        %impfile(F02);
        %impfile(F06);
        %impfile(EVENTS);

/* Push together DPPOS Phase 1 data */
data dppos_follow; set DPPOSR1.f01 DPPOSR1.f02 DPPOSR1.f06;
run;
proc sort; by release_id; run;

/* DPPOS Phase 2 data */
libname DPPOSR2 'Directory where you stored DPPOS Phase 2 data';
%macro impfile(FN);
filename tranfile "Directory of DPPOS2 xpt files\&FN..XPT";
proc cimport data=DPPOSR2.&FN infile=tranfile;
run;
%mend;
        %impfile(F01);
        %impfile(F02);
        %impfile(F06);
        %impfile(EVENTS);

/* Push together DPPOS Phase 2 data */
data dppos_follow2; set DPPOSR1.f01 DPPOSR2.f02 DPPOSR2.f06;
```

```
run;
proc sort; by release_id; run;

/* Merge all visit data together from DPP+Bridge+DPPOS1+DPPOS2 */
Data follow;  set dpp_follow(in=indpp) dppbr_follow(in=inbridge) dppos_follow1(in=indppos1)
dppos_follow2(in=indppos2);
        if indpp then DPP=1;
        if inbridge then Bridge=1;
        if indppos1 then DPPOS1=1;
        if indppos2 then DPPOS2=1;

*** Compute weight at each followup visit;
        if QPWGHT1>. then WEIGHT = MEAN(QPWGHT1,QPWGHT2,QPWGHT3);
        else if APWGHT1>. then WEIGHT=MEAN(APWGHT1,APWGHT2,APWGHT3);
        label weight = "Current weight (kg)";

**** Define regularly-scheduled quarterly and semi-annual visits for DPP and DPPOS;
  *** DPP and Bridge visits - assign based on VISIT as listed on form;
        select (visit);
      when ('SCR') QUARTER=0;
      when ('BAS') QUARTER=0;
      when ('M03') QUARTER=1;
      when ('M06') QUARTER=2;
      when ('M09') QUARTER=3;
      when ('Y01') QUARTER=4;
      when ('M15') QUARTER=5;
      when ('M18') QUARTER=6;
      when ('M21') QUARTER=7;
      when ('Y02') QUARTER=8;
      when ('M27') QUARTER=9;
      when ('M30') QUARTER=10;
      when ('M33') QUARTER=11;
      when ('Y03') QUARTER=12;
      when ('M39') QUARTER=13;
      when ('M42') QUARTER=14;
      when ('M45') QUARTER=15;
      when ('Y04') QUARTER=16;
```

```
      when ('M51') QUARTER=17;
      when ('M54') QUARTER=18;
      when ('M57') QUARTER=19;
      when ('Y05') QUARTER=20;
      when ('M63') QUARTER=21;
      when ('M66') QUARTER=22;
      when ('M69') QUARTER=23;
      when ('Y06') QUARTER=24;
       otherwise;
    end;

  if mod(quarter,2)=0 then semi=quarter/2;

  *** DPPOS Visits - must assign semi-annual visits based on days since randomization;
  if substr(visit,3,1) in ('A','M') then semi = floor(daysrand/182.625);

    label quarter = "Quarter of visit - DPP"
          semi    = "Semi-annual visit";

        keep release_id quarter semi visit weight dpp bridge dppos;
        run;

proc sort; by release_id; run;

data all_data;
        merge follow events_demo ; by release_id;
run;

proc sort data=all_data; by release_id semi; run;

/* Combine complete events dataset */
proc sort data=DPPREL.events; by release_id; run;
proc sort data=DPPBRREL.events; by release_id; run;
proc sort data=DPPOSR1.events; by release_id; run;
proc sort data=DPPOSR2.events; by release_id; run;

/* Combine events datasets from all 3 time periods */
```

```
data events;
        merge DPPREL.events(in=indpp keep=release_id diabf diabt diabv totaltim)
        DPPBRREL.events(in=inbridge keep=release_id diabf diabt diabv totaltim)
        DPPOSR1.events(in=indppos1 drop=randper);
        DPPOSR2.events(in=indppos2 drop=randper);
          by release_id;
        length last_event $6.;

        *** Keep the record from the latest the participant was in the study;
        if indppos2
                or (indppos1 and ~indppos2)
                or (inbridge and ~indppos1 and ~indppos2)
                or (indpp and ~inbridge and ~indppos1 and ~indppos2);

        *** Label the visit type;
        if indppos2 then last_event='DPPOS2';
        else if indppos1 then last_event='DPPOS1';
        else if inbridge then last_event='BRIDGE';
        else if indpp then last_event='DPP';
run;
proc sort data=events; by release_id; run;

/* Combine complete events dataset with baseline demographics */
proc sort data=DPPREL.basedata; by release_id; run;
data events_demo; merge events DPPREL.basedata; by release_id; run;
```

5.3 Sample SAS program for variables over time by treatment group

```
/* Sample table of variables over time by treatment group */
proc tabulate data=all_data;
        where .<semi<=30 and assign ne 'Troglitazone';
        class semi assign;
        var weight;
        tables semi='Semi-annual visit',assign=' '*weight=' '*(n='N'*f=4.0 mean='Mean'*f=5.1
std='Standard Deviation'*f=5.1);
        title 'Diabetes Prevention Program';
        title2 'Weight at each semi-annual visit by treatment arm';
```

```
        title3 'DPP + Bridge + DPPOS Phases 1 and 2';
run;
```

5.4 Sample SAS program for Cox Proportional Hazards Model

```
/* Sample Cox Proportional Hazards Model for the full DPP+Bridge+DPPOS1+DPPOS2 */
PROC PHREG DATA=events_demo; where assign ne 'Troglitazone';
 class assign;
 MODEL diabv*diabf(0)=assign
     /TIES=discrete ALPHA=0.05 RL;
 title 'Diabetes Prevention Program';
 title2 'Time to diabetes';
 title3 'DPP+Bridge+DPPOS1+DPPOS2';
RUN;
```

5.5 Sample SAS program for the Microvascular Comparison using GEE

```
proc sql;
 create table microgee as
 select *
 from os2_rel.demographic a, os2_rel.microvascular b
 where a.release_id=b.release_id;
 run;

 data microgee2;
  set microgee;
   length triolab $20;
  triolab='Nephropathy';
  trio = max(evtnep);
  output;
  triolab='Retinopathy';
  trio = max(evtret);
  output;
  triolab='Neuropathy';
  trio = evtneu;
  output;
```

```
run;

** Global test: All participants combined;
proc genmod data=MIcrogee2 descending;
 where trio not in (0.5,.) ;
class release_id assign triolab ;
model trio= assign  triolab triolab*assign  / dist=binomial link=log type3;
repeated subject=release_id / type=un corrw covb within=triolab;
lsmeans triolab*assign assign / pdiff oddsratio cl stepdown exp ;
ods output lsmeans=lsmrr diffs=diffrr;
run;
```