Diabetes Prevention Program Outcomes Study

Data Release Documentation

April 2024
DPPOS Phase 3 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 3 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 follow-up of the original DPP cohort. Detailed information about the DPP and DPPOS, including protocols, intensive lifestyle intervention manuals, references, publication list, and links to abstracts and manuscripts is available at https://dppos.bsc.gwu.edu/. This report describes the public release of Phase 3 of the DPPOS dataset, and is based on all DPPOS data collected after the DPPOS Phase 2 data lock in 2013 and prior to the Phase 3 data lock in February 2020. A brief description of the study is given below.

1.2 Descriptions DPP and DPPOS

Separate datasets are available for each of the DPP and DPPOS phases described below. Further details and materials from 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. Major DPP eligibility inclusion criteria included age ≥25 years, fasting glucose 95-125 mg/dl, two-hour post-load glucose 140-199 mg/dl, body mass index (BMI) ≥24 kg/m^2 (≥22 kg/m^2 among Asian-American participants), and exclusion criteria included taking medications known to alter glucose tolerance, illnesses that could reduce life expectancy, or the ability to participate in the trial.

During DPP, participants were seen at quarterly visits after randomization until the study was terminated at the recommendation of the Data and Safety Monitoring Board in 2001. 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. The primary outcome of DPP was time to first confirmed diabetes as further described in Section 3.3.

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

Beginning in August 2001, all participants were unmasked to the DPP study results. Participants assigned to metformin 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 the lifestyle intervention, 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 (DPPOS Years 1-6, Fall 2002 – August 2008)

DPPOS Phase 1 began in the fall of 2002, with variable start times depending on each clinic's IRB approval. Participants randomized to the Metformin intervention continued to take open-label metformin until such time as they had developed diabetes with an HbA1c ≥ 7%. Participants randomized to intensive lifestyle were offered a twice-yearly boost lifestyle curriculum (BOOST). All participants were offered a quarterly Healthy Lifestyle Program (HELP). 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 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. The primary outcome

of DPPOS Phase 1 was continued time to first confirmed diabetes as further described in Section 3.3.

1.2.4 DPPOS Phase 2 (DPPOS Years 6-11, Fall 2008 – October 2013)

No substantial changes were made between DPPOS Phase 1 and Phase 2, including the interventions, which continued as during DPPOS Phase 1. Measurements, medical history updates, adverse event assessment, medication adherence and dispensing, questionnaires, and 2-hour 75g oral glucose tolerance tests (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. For the first time, cognitive and physical function were assessed in DPPOS Years 8 and 10. Coronary Artery Calcification (CAC) scans were obtained on all willing participants in DPPOS Year 10, and fundus photos were taken in DPPOS Year 11. The primary outcome of DPPOS Phase 2 was incident microvascular disease before or at the DPPOS Year 11 visit as further described in Section 3.4 and in the DPPOS Phase 2 data documentation.

1.2.5 DPPOS Phase 3 (DPPOS Years 12-18, mid 2013 – February 2020)

Beginning during DPPOS Phase 3, most midyear visits were conducted by phone starting with the DPPOS Year 13 visits, with no physical or laboratory measurements performed. The BOOST lifestyle intervention and the quarterly HELP sessions were discontinued. Metformin continued to be provided to the original Metformin participants as during DPPOS Phases 1 and 2. Measurements, medical history updates, adverse event assessment, medication adherence and dispensing, questionnaires, cognitive and physical function testing, and a 2-hour 75g OGTT continued during annual visits. More detailed neuropathy testing began in DPPOS Year 17. Fundus photos and Optical Coherence Tomography (OCT) scans were obtained once in DPPOS Year 16. DEXA scans were obtained in DPPOS Years 12 and 17. The brief mid-year visits, usually conducted by phone, included some questionnaires, adverse event assessment, and medication adherence. The two primary outcomes for DPPOS Phase 3 were time to first incident cancer and time to first incident CVD as further described in Section 3.5.

All data in this DPPOS Phase 3 release include visits or events which occurred prior to widespread impacts of the COVID-19 pandemic, including the DPPOS clinic closures that began in March 2020.

1.3 Treatment Arms

1.3.1 DPP (1996 – July 2001)

At DPP randomization, participants were randomly assigned to one of four treatment groups: intensive lifestyle, metformin (850 mg twice per day as tolerated), troglitazone (400 mg once daily), or double-placebo.

The lifestyle intervention was individually administered during 16 in-person sessions over 6 months, with continued individual and group sessions throughout DPP. The lifestyle curriculum was designed to produce ≥7% body weight loss and moderately intensive physical activity of ≥150 minutes per week.

Participants assigned to one of the medication interventions (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 the lifestyle intervention, were offered the full 16-session lifestyle program in group format.

1.3.3 DPPOS Phase 1 (Fall 2002 – August 2008)

During DPPOS Phase 1, the metformin and lifestyle participants were kept on their study interventions to the extent possible. For participants randomized to the lifestyle intervention, 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, and maintained an HbA1c < 7%. In addition, all participants were invited to quarterly HELP classes.

1.3.4 DPPOS Phase 2 (Fall 2008 – October 2013)

During DPPOS Phase 2, as during Phase 1, the original metformin and lifestyle participants were kept on their study interventions to the extent possible. For participants randomized to intensive 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 HELP sessions. 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.

1.3.5 DPPOS Phase 3 (mid 2013 – February 2020)

During DPPOS Phase 3, metformin was continued open-label in participants who had been randomized to metformin and who were eligible and willing to continue. For participants randomized to intensive lifestyle, group-implemented BOOST lifestyle sessions were discontinued in 2014. In addition, the quarterly HELP classes previously offered to all participants were discontinued. No HELP or BOOST session data are included in the DPPOS Phase 3 data release.

1.4 Diabetes Diagnosis and Subsequent Treatment

The complete definition of diabetes in the DPP and DPPOS is provided in section 3.3 and is unchanged from previous study periods with the exception of the mid-year fasting glucose diagnosis, which was discontinued in DPPOS Phase 3. After a participant was confirmed to have diabetes, the metformin intervention was continued and reinforced. However, once a participant was diagnosed with worsening diabetes (defined during DPPOS to be an HbA1c $\geq$ 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 3 without respect to HbA1c levels.

1.5 Data Included in DPPOS Phase 3 release

Data included in this release begin with DPPOS Year 12 visits through DPPOS Year 18 visits truncated on February 23, 2020. This includes the DPPOS Phase 3 primary cancer and cardiovascular disease outcomes, along with deaths, which were underpowered during prior phases of DPPOS.

Non-research data, including tracking forms, are not included. Serious adverse event data were collected but are also not included in the data release, except for cancer and CVD, as SAE data were generally not adjudicated and are 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 12 of the DPPOS Protocol shows the forms and assessments completed at each visit. Data collected during DPPOS Phase 3 Years 12 through 18, prior to mid-February 2020, 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 2514 participants (excluding those originally randomized to troglitazone) who consented to DPPOS-2 or DPPOS-3, 2263 participants are included in this release dataset.
- Cancer, CVD and mortality events includes all DPP participants with consent to release data to the NIDDK Repository.
- 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, as further described in Section 2.3.
- 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; the variable was removed due to small sample size; 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, under the library OS3_REL. One dataset exists for each DPPOS form or dataset.
- 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 and responses by 10 or fewer participants overall or within various demographic categories have been removed to protect confidentiality.

2.4 Structure of the Datasets

Except where indicated, one record exists in each file for each participant for each visit at which that particular form or assessment 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 during DPPOS Years 12 through 18 completed by February 23, 2020, including mid-year, annual, diabetes confirmation, and interim visits beginning in mid-2013 until February 2020. Data included were not included in the DPPOS Phase 2 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, F02 and F06 by original DPP randomized treatment arm For the original study sample and the DPPOS Phase 3 database in NIDDK repository

DPPOS Phase 3 Mid 2013 – February 2020

	Lifestyle		Metformin		Placebo	
VISIT	**Study**	**Repository**	**Study**	**Repository**	**Study**	**Repository**
12A	700	624	728	654	740	671
12M	696	618	712	642	725	656
13A	714	643	739	667	752	690
13M	718	644	739	670	758	695
14A	704	639	723	659	736	681
14M	704	641	727	663	738	683
15A	690	627	705	644	717	664
15M	688	624	706	645	709	656
16A	665	603	696	632	691	637
16M	677	614	690	627	701	649
17A	658	594	674	610	679	628
17M	660	597	667	604	687	634
18A*	274	250	285	268	317	306
18M*	286	268	291	268	288	271

* 18A and 18M visit numbers are low because this release only includes partial year 18 visits

3. Statistical Considerations

3.1 Analysis according to original DPP randomization

The DPP was analyzed by intent-to-treat, which 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 or after the randomized trial.

3.2 Repeated Measures

Data in DPP and DPPOS were collected at multiple time points over the years of follow-up. To account for the repeated 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, 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.

3.3 DPP and DPPOS Phase 1 Primary Outcome: Diabetes

The primary endpoint for the DPP and DPPOS Phase 1, and a secondary outcome during DPPOS Phase 2 and Phase 3, 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

Throughout DPP and DPPOS, an OGTT was completed at annual visits, with a fasting glucose measured at mid-year visits through 2013. 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, through 2013), diabetes was confirmed at a subsequent visit, usually within 6 weeks, in order for the participant to be diagnosed with diabetes. Confirmation visits were either standalone visits (VISIT=CON) or at the following in person visit, and generally included the same glucose measurements as the visit where confirmation was triggered. That is, the confirmation visit following a trigger at an annual visit included an OGTT, whereas a standalone confirmation visit (e.g., VISIT=CON) 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 DPP and DPPOS data analyses, 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, often at 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 through DPPOS Phases 1, 2 and 3, with careful consideration of the impact of the group-lifestyle intervention offered during the bridge period and continued through DPPOS Phase 2 for all participants.

3.4 DPPOS Phase 2 Primary Outcome: Microvascular Disease

Data for the DPPOS Phase 2 primary outcome were provided in the DPPOS Phase 2 data release. A brief summary is included here.

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).

3.4.1 Analytic Considerations for DPPOS Phase 2

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.

3.5 DPPOS Phase 3 Primary Outcomes: Cancer and Cardiovascular Disease

There were two primary outcomes for DPPOS Phase 3:

- The first occurrence of any cancer, excluding non-melanoma skin cancer, was the primary outcome during the first 5-years of DPPOS-3. Obesity-related cancer was a major secondary outcome.
- During a planned second 5-year period of DPPOS Phase 3, the primary outcome was intended as the first occurrence of a Major Atherosclerotic Cardiovascular Event (MACE: fatal of non-fatal heart attack or stroke). A pre-specified interim analysis after the first 5 years of DPPOS Phase 3 found futility for this outcome.

3.5.1 Cancer

During DPPOS Phase 3, the first occurrence of any cancer, excluding non-melanoma skin cancer, was the primary outcome, and obesity-related cancer was a major secondary outcome. Incident cancer cases were ascertained at each annual follow-up visit and/or were reported as a serious adverse event if there was hospitalization or death. In 2012, a retrospective collection of medical records associated with all previously reported cancer events was undertaken, and records were collected prospectively from 2012 onward, including those reported on a death certificate. Detailed clinical records, including pathology reports, were sought from treating physicians or healthcare institutions. A National Cancer Institute (NCI) staff oncologist, blinded to participant treatment group and metformin use, conducted adjudication according to the Surveillance, Epidemiology, and End Results (SEER) Program guidelines. Only primary incident cases that occurred after randomization were included in the analysis. The dataset provided includes incident cases of any first cancer. If a participant experienced cancer at additional anatomical sites, only their first occurrence contributed to the composite “Cancer” event, while their first occurrence

of each specific anatomical-type contributed to the individual anatomical-type (e.g., “Breast cancer”) or subgroup (e.g. “Obesity cancer”).Obesity-related cancer were based on CDC definitions and include: meningioma, multiple myeloma, or cancer of the esophagus, kidneys, uterus, ovaries, thyroid, breast (postmenopausal), liver, gallbladder, upper stomach, pancreas, and colon/rectum. Specific cancer diagnoses for anatomical sites of origin are provided for sites with sufficient numbers of events.

3.5.2 Cardiovascular Disease

The CVD outcome was defined as the first occurrence of a Major Atherosclerotic Cardiovascular Event (MACE: non-fatal heart attack, non-fatal stroke, or CVD death). Each individual component of MACE, along with Extended MACE (encompassing any cardiovascular event: non-fatal MI, non-fatal stroke, coronary or peripheral revascularization, hospitalized congestive heart failure, CHD by angiography, silent MI, or fatal CVD) were important secondary outcomes. CVD outcomes were adjudicated by a team of DPPOS investigators blinded to treatment arm, who had access to detailed medical record data. The DPPOS CVD and mortality adjudication procedures manual is available as supplemental documents. The dataset includes the combined MACE and Extended MACE outcomes as well as the individual components with sufficient numbers of events.

3.5.3 Analytic Considerations for DPPOS Phase 3

The primary outcome analysis for DPPOS Phase 3 compared the original metformin intervention group with the original placebo group with respect to both the cancer and cardiovascular outcomes. Comparison of the original lifestyle intervention group with both the metformin and placebo groups were secondary analyses.

The primary analyses used the “intention-to-treat” paradigm including all participants and data as of the data lock. For the primary analysis, cumulative incidence rates accounted for competing risk due to non-cancer or CVD deaths using Fine-Gray’s estimates in Cox proportional hazards models to assess treatment and covariate effects that may mediate or moderate treatment effects. The primary analysis used the first incident cancer or MACE event based on adjudication.

3.6 DPPOS Phase 3 Microvascular Assessments

For DPPOS Phase 3, the main composite diabetes-related microangiopathic outcome was refined from DPPOS Phase 2, as time to first occurrence of nephropathy or retinopathy.

3.6.1 Nephropathy in DPPOS Phase 3

Nephropathy was adapted from KDIGO for categories of moderately increased albuminuria or moderately to severely decreased eGFR and defined as development of any of the following: (1) confirmed elevated albuminuria (≥30 mg/g creatinine) based on annual spot urine; (2) confirmed eGFR by CKD-epi <45 ml/min; (3) kidney transplant; or (4) dialysis for end-stage renal disease.

3.6.2 Retinopathy in DPPOS Phase 3

Fundus photos and Optical Coherence Tomography (OCT) scans were obtained once in DPPOS Year 16. Digital photographs were taken locally by trained photographers using FPRC 7-standard field color fundus photography on the left eye.

Spectral domain OCT is a noninvasive imaging technology that uses principles of light interferometry and Fourier analysis to generate a voxel (3-dimensional) rendering of the retinal architecture. Macular edema by OCT was measured a secondary outcome.

Photographs were read by the DPPOS Fundus Photo Reading Center at the University of Wisconsin.

The DPPOS Phase 3 definition of retinopathy was defined as mild diabetic retinopathy (ETDRS≥35) or adjudicated treatment for retinopathy. Note that this is a change from the DPPOS Phase 2 definition of microaneurysms (ETDRS≥20). The ETDRS grade was modified in recognition of the expected increase in time and level of hyperglycemia that will result in more diabetes-specific lesions and to minimize noise from aging-related retinal findings.

3.6.3 Neuropathy in DPPOS Phase 3

Neuropathy was defined in DPPOS Phase 3 beginning in DPPOS Year 17 as distal symmetric polyneuropathy (DSPN) when new measures of vibration and sharp (pinprick) sensation were added. This definition of DSPN includes:

- Symptoms from the Michigan Neuropathy Screening Instrument (Form Q15)
- Signs of neuropathy as recommended in the 2017 ADA Position statement are detected using quantitative measurements collected on Form P14 of:
    - Vibration
    - Light touch
    - Sharp (pinprick) sensation
- The principal DSPN measure in DPPOS Phase 3 was defined as the presence of DSPN symptoms (score ≥4 on the MNSI questionnaire) or any signs of bilateral DSPN. The signs of bilateral DSPN require abnormal finding on both toes for pinprick testing (score ≥5), for vibration testing (score≥5), or for light touch sensing (<8 detected from 10 monofilament applications on each toe).

4. File Descriptions

4.1 Data Forms

4.1.1 General

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

Data collection forms are available as PDFs 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 the 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 text-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 but are generally excluded so as not to inadvertently identify a participant.

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. Deleted variables are not included.

4.1.2 Variable Names on Data Forms

- Variable names for each released variable are embedded on the data form. Variable names for non-released variables have been grayed out or removed.
- All datasets are HIPAA compliant. Information that might directly identify a specific participant has been excluded from the release datasets. 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 generally excluded so as not to inadvertently identify a participant.

4.2 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 across all DPP and DPPOS data releases.
- 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:
    - 12A, 13A, 14A, 15A, 16A, 17A, 18A: Regularly scheduled DPPOS annual visits.
    - 12M, 13M, 14M, 15M, 16M, 17M, 18M: Regularly scheduled DPPOS mid-year visits (mostly completed by phone beginning in DPPOS Year 13).
    - INT, UNS: 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 if necessary after glucose confirmation to obtain measures that were not collected at the visit where the participant's glucose was first elevated (e.g., trigger visit).

- DAYSRAND: The number of days a particular visit or event 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. DPPOS-3 visits occurred at the following time ranges:

DPPOS Visits	Calendar year
12M, 12A	July 2013 – October 2014
13M, 13A	July 2014 – October 2015
14M, 14A	July 2015 – October 2016
15M, 15A	July 2016 – October 2017
16M, 16A	July 2017 – October 2018
17M, 17A	July 2018 – October 2019
18M, 18A	July 2019 – February 2020. DPPOS Year 18 visits are truncated mid-way through the study year, and all were prior to the COVID-19 pandemic.

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.

4.3 Created Datasets (Non-Form Data)

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

- A DEMOGRAPHIC file with one record for each participant which includes the original DPP treatment assignment, baseline age, baseline BMI, sex, and race-ethnicity. This file includes only participants in the DPPOS Phase 3 release.

- A DIABETES file includes variables for diabetes as well as censoring data. This file includes one record for every DPPOS Phase 3 participant.

- DSPN: Records of DSPN (distal symmetric polyneuropathy) for each participant for DPPOS Years 17 and 18. This data represents the summary measures included in the DSPN outcome for DPPOS Phase 3 as described in Section 3.6.3.

- A CVD_MORT_CANC file includes summary event variables for each adjudicated cardiovascular and cancer event, as well as mortality. Unlike other files in this data release, this file includes all participants randomized to DPP who have consent to release data to the NIDDK repository. Included in this file are:

  - CVD variables include a summary event variables for MACE and Extended MACE for all participants who had an adjudicated, post-randomization CVD event, as well time to first specific events with sufficient sample size (including: non-fatal MI, non-fatal stroke, CVD death, coronary revascularization, hospitalized congestive heart failure).

  - Cancer variables include a summary event variables for the first incident cancer and the first incident obesity-related cancer for all participants who had an adjudicated, post-

randomization cancer, as well as specific cancer types with a sufficient number of events.

- Mortality includes all deaths as well as the adjudicated cause of death for all participants who passed away during DPP or DPPOS.
- Each event has its own event/censoring time variable relevant to that event. For participants who did not have the event, censoring time is the last participant contact.

4.4 Datasets for Central Unit Data

Data provided through DPPOS central reading centers and laboratories are included in this release are as follows:

- LAB: Laboratory data as centrally measured by the DPPOS Central Laboratory at the University of Washington. One record for each participant for each visit where laboratory measurements were completed.
- ULAB: Urine laboratory data as centrally measured by the DPPOS Central Laboratory at the University of Washington. One record for each participant for each visit **throughout all of DPPOS (Phase 1, 2 and 3)** where urine laboratory measurements were completed. This is the first time urine creatinine and albumin results have been provided to the NIDDK repository.
- FUNDUS: Fundus (eye) photo data for assessment of retinopathy. One record of centrally analyzed fundus photo data by the DPPOS Central Fundus Photo Reading Center at the University of Wisconsin. Fundus photos were obtained for all participants during DPPOS Year 16.
- OCT: Optical Coherence Tomography data. One record of centrally analyzed OCT data by the DPPOS Central Fundus Photo Reading Center at the University of Wisconsin. OCT was obtained for all participants during DPPOS Year 16.
- QWB: Quality of Well Being data. Records of centrally analyzed by the DPPOS Central QWB Reading Center at the University of California, San Diego. QWB data was collected for all participants during DPPOS Year 16.
- DXA: Records of centrally analyzed by the DPPOS Central DEXA Reading Center at the University of California, San Francisco. Whole body DEXA scans were obtained for all participants during DPPOS Years 12 and 17.

Further details of each of these datasets can be found below.

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

4.5.1 OS3_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. Mid-year visits were completed in person through DPPOS Year 12, and generally by phone beginning in DPPOS Year 13. There was a major F01 form change starting with DPPOS Year 13 at which time many variables were removed.

4.5.2 OS3_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. Annual visits were generally completed in person throughout DPP and DPPOS. There was a major F02 form change midway through DPPOS Year 13; many variables are new or modified after that point. Neuropathy was collected on Form F02 through DPPOS Year 16 after which it was moved to Form P14 as described below.

4.5.3 OS3_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.

4.5.4 OS3_REL.F06: HOME VISIT INVENTORY

DPPOS Form F06 was used to record information about an annual visit that was completed outside the clinic (e.g. at home, nursing home, etc.). Limited data were collected at home visits. Variable VISIT is used to identify the visit completed. There was a major F06 form change midway through DPPOS Year 13; many variables are new or modified after that point. Neuropathy was collected on Form F06 through DPPOS Year 16 after which it was moved to Form P14 as described below.

4.5.5 OS3_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 for discontinuing metformin was reported in section B, additional F07 forms were not required. For participants not taking metformin temporarily and eligible to restart, form F07 was completed at every visit where 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.

4.5.6 OS3_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. All questions regarding pregnancy have been removed due to small sample sizes.

4.6 Questionnaires (Q-forms)

4.6.1 OS3_REL.Q01: BECK QUESTIONNAIRES

DPPOS Form Q01 includes the Beck Depression Inventory (BDI) and the Beck Anxiety Inventory (BAI). Form Q01 was self-administered at the DPPOS Phase 3 12M and 15A visits. Variable VISIT is used to identify the visit completed. The BAI was discontinued after the 12M visit, and two additional non-Beck questions were added to the Q01.

To score the BDI, add up the score for each of the 21 BDI questions (exclude BDI question 19b) and obtain the total. The highest score on each of the 21 BDI questions is 3, therefore the highest possible total for the whole BDI is 63 and the lowest possible score is 0.

To score the BAI, add up the score for each of the 21 items on the BAI and obtain the total. The highest score on each of the 21 BAI questions is 3, therefore the highest possible total for the whole BAI is 63 and the lowest possible score is 0.

4.6.2 OS3_REL.Q02: HEALTH SURVEY QUESTIONNAIRE

DPPOS Form Q02 is the MOS SF-36 questionnaire. Form Q02 was self-administered at the DPPOS Phase 3 12M and 15A visits. 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.

4.6.3 OS3_REL.Q03: MODIFIABLE ACTIVITY QUESTIONNAIRE

DPPOS Form Q03 is the Modifiable Activity Questionnaire. Form Q03 was interviewer-administered at the DPPOS Phase 3 12M and 18A 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.

4.6.4 OS3_REL.Q13: URINARY INCONTINENCE QUESTIONNAIRE

DPPOS Form Q13 was self-administered at the DPPOS Phase 3 12A and 15A visits. The Q13 records participants' issues related to urinary incontinence during the past year. Variable VISIT is used to identify the visit completed.

4.6.5 OS3_REL.Q15: NEUROPATHY (MNSI) QUESTIONNAIRE

DPPOS Form Q15 is the Neuropathy (MNSI) questionnaire. Form Q15 was self-administered at DPPOS Phase 3 annual visits: 12A, 13A, 17A, and 18A. Variable VISIT is used to identify the visit completed.

4.6.6 OS3_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 Phase 3 Years 15A and 18A. Variable VISIT is used to identify the visit completed. The measures of cognitive function administered include the Modified Mini Mental Status Examination (3MSE), the Spanish-English Verbal Learning Test (SEVLT), and the Digit Symbol Substitution Test of the Wechsler Adult Intelligence Scale-Revised (WAIS-R). These measures tap cognitive domains including global mental status, verbal learning and memory, and psychomotor speed, respectively.

4.6.7 OS3_REL.Q21: ACTIVITIES OF DAILY LIVING/INSTRUMENTAL ACTIVITIES OF DAILY LIVING

DPPOS Form Q21 was used to record self-reported Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL). The Q21 was interviewer-administered at the DPPOS Phase 3 annual visits 13A, 15A, and 17A.

4.6.8 OS3_REL.Q22: MALE SEXUAL FUNCTION

DPPOS Form Q22 was used to record self-reported sexual function for men. The Q22 was self-administered once at the DPPOS Phase 3 15A visits.

4.6.9 OS3_REL.Q23: FEMALE SEXUAL FUNCTION

DPPOS Form Q23 was used to record self-reported sexual function for women. The Q23 was self-administered once at the DPPOS Phase 3 15A visits.

4.6.10 OS3_REL.Q24: RESPIRATORY QUESTIONNAIRE

DPPOS Form Q24 is the Respiratory Questionnaire, which was self-administered at the DPPOS Phase 3 15A and 18A visits.

4.6.11 OS3_REL.Q25: DEMENTIA SCREENING (AD8)

DPPOS Form Q25 is the Dementia Screening Questionnaire (AD8). This form was interview-administered beginning in DPPOS Year 16 to a consented proxy (preferred) or to participant (if a proxy was not available or consented).

At DPPOS Year 16, the Q25 was administered to participants who were inactive, were active but missed their Year 16 annual visit, or were active and completed their Year 16 annual visit by phone, or did not complete the Year 15 cognitive function testing (for any reason).

In DPPOS Year 17, this form was interviewed-administered to a consented proxy (preferred) for all participants who missed their in-person Annual Visit (regardless of reason).

Starting in DPPOS Year 18, this form was interview administered for all participants at the Annual Visit.

4.6.12 OS3_REL.Q26: ECONOMIC EVALUATION QUESTIONNAIRE

DPPOS Form Q26 was used to record participants' costs and time related to exercise behavior, health

insurance, medical care, living situation and updated income. The Q26 was self- or interview-administered by participants during DPPOS Years 16 and 18. Section C was added for DPPOS Year 18. The Q26 replaces Form Q12 used during DPP and Q16 used during DPPOS Phases 1 and 2.

4.7 Event Forms (E-forms)

4.7.1 OS_REL.E09: EYE PROCEDURES

DPPOS Form E09 was used to document any eye procedures or surgery reported by participants. The dates on this form have been transformed into days since randomization as indicated on the PDF version of the form. Completion of this form was triggered via the F01, F02, F03 or F06 follow-up forms. Variables that have very few or no responses were removed.

4.7.2 OS_REL.E11: GASTRIC REDUCTION SURGERY

DPPOS Form E11 was used to document any gastric reduction surgery reported by a participant. Reversals of prior gastric reduction surgery are also reported. 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. Completion of this form was triggered via the F01, F02, F03 or F06 follow-up forms.

4.7.3 OS_REL.E14: INFECTION

DPPOS Form E14 was used to document any hospitalized infection reported by participants which could be reported at any visit. Completion of this form is triggered via the F01, F02, F03 or F06 follow-up forms. Variables that have very few or no responses were removed.

4.7.4 OS_REL.E15: FRACTURES

DPPOS Form E15 was used to document any ***hospitalized*** fracture reported by participants. Fractures could be reported at any visit. Completion of this form is triggered via the F01, F02, F03 or F06 follow-up forms. Most specific body part fracture locations had very few or no responses, therefore only hip fracture is specifically indicated, although all available fracture records with time to fracture, degree of trauma, and treatment are included.

4.7.5 OS_REL.E16: JOINT REPLACEMENTS

DPPOS Form E16 was used to document any joint replacement surgery reported by participants. Completion of this form is triggered via the F01, F02, F03 or F06 follow-up forms.

4.8 Procedures (P-forms)

4.8.1 OS3_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. The timing of the fasting, 30-minute and 2- hour blood draws are included with the LAB data. The P07 is ***not*** released as a separate dataset.

4.8.2 OS3_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 15 and 18. Variable VISIT is used to identify the visit completed.

4.8.3 OS3_REL.P12: SIX MINUTE WALK TEST

DPPOS Form P12 records results from the six-minute walk test. Course length varied by clinic, 10 or 20m, and was not recorded on the data collection form. It was added as a variable at end of the dataset. The six-minute walk test was completed during annual visits in DPPOS Years 15 and 18. Variable VISIT is used to identify the visit completed. A few variables with very sparse data were removed.

4.8.4 OS3_REL.P14 : NEUROPATHY ASSESSMENTS

DPPOS Form P14 records results from the neuropathy assessments in DPPOS Years 17 and 18. In addition to the 10g monofilament collected at prior visits, beginning in DPPOS Year 17, two additional bilateral neuropathy assessments were added: A pinprick sensation test, and a vibratory sensation test as described in Section 3.6.3. Variable VISIT is used to identify the visit completed. During DPPOS Years 12-16, neuropathy assessment data was collected on Form F02.

4.9 Report Forms (R-forms)

4.9.1 OS3_REL.R16: ENROLLMENT REPORT

DPPOS Form R16 was used to document enrollment in each phase of DPPOS. During DPPOS Phase 3, questions were added to update socioeconomic status, race/ethnicity (not included), and family health history. Variable VISIT is used to identify the visit completed. This form was completed once during DPPOS-3.

4.9.2 OS3_REL.R25: FALLS REPORT

DPPOS Form R25 was used to document details anytime a participant reported a fall. Variable VISIT is used to identify the visit completed. Completion of this form was triggered via the F01, F02, F03 or F06 follow-up forms.

4.10 Created Datasets

4.10.1 OS3_REL.DEMOGRAPHIC: Demographic Data

DPPOS data DEMOGRAPHIC includes one record corresponding to DPP baseline 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 3 of DPPOS, 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”	Participant ID for repository
AGE	Age at randomization (years)	Numeric		Computed based on date of randomization and birth date, from DPP screening form S07.
ASSIGN	DPP Treatment assignment	Character	Lifestyle Metformin Placebo (Troglitazone)	Randomized treatment assignment. Not available on any data form.

Variable	Brief description	Type	Coding	Details
BMI	BMI (kg/m^2)	Numeric		Body mass index at DPP baseline. Computed based on height and weight as measured on DPP 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	Summarized, self-reported race/ethnicity based on the 1990 census questionnaire during screening on DPP Form S03.
SEX	Sex	Numeric	1 = Male 2 = Female	Collected during screening on DPP Form S03.

4.10.2 OS3_REL.DIABETES: Diabetes Data

DPPOS data DIABETES includes one record for each participant. This file is updated from the DPP Phase 2 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”	Participant ID for repository
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 no diabetes by final visit. This should be used as the censoring time for diabetes.
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 in participants without diabetes 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.

Variable	Brief description	Type	Coding	Details
TOTALTIM	Years in study	Numeric		Total years in study through last visit of any type (quarterly, mid-year, annual or interim) as of February 2020.

4.10.3 OS3_REL.DSPN: DSPN Data

DPPOS data DSPN (distal symmetric polyneuropathy) includes a record for each participant for DPPOS Years 17 and 18. This data represents the summary measures included in the DSPN outcome for DPPOS Phase 3 as described in Section 3.6.3. Raw data comprising these summary variables is included in form Q15 (Neuropathy Questionnaires - MSNI) and P14 (Neuropathy Procedure Log). 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”	Participant ID for repository
DAYSRAND	Days since randomization	Numeric		Days from DPP randomization
VISIT	Visit completed	Character	17A; 18A	
neu_signs	Indicator of Signs of DSPN	Numeric	0 = No 1 = Yes	DSPN Signs from any of pinprick, vibration or monofilament
neu_symptoms	Indicator of Symptoms of DSPN	Numeric	0 = No 1 = Yes	DSPN Symptoms from the MNSI (Form Q15)
neu_signsymp	DSPN Signs or Symptoms	Numeric	0 = No 1 = Yes	Summary DSPN variable of signs or symptoms

4.10.4 OS3_REL.CVD_MORT_CANCER: CVD, Cancer and Mortality Data throughout DPP and DPPOS

DPPOS data CVD_MORT_CANCER includes all adjudicated CVD, Cancer and mortality events for all DPP participants with approval to share data with the NIDDK Repository.

Specific MACE, Extended MACE, and Cancer events are included only if the sample size was sufficiently large. Participants may have had more than one type of event on the same day, and may have had a second event at a later time. All first events of each type are included. The censoring time for each outcome in participants who did not develop the event during DPP or DPPOS is the days from randomization to last participant contact included in the DAYS variables.

This file includes the following variables:

Variable	Type	Coding	Details
RELEASE_ID	Char		Participant ID for repository
MACE	Num	0 = No 1 = Yes	Any MACE: Non-fatal MI, non-fatal stroke, fatal CVD
DAYSRAND_MACE	Num		Days from randomization to first date of any MACE or censoring
MI	Num	0 = No 1 = Yes	Non-fatal MI
DAYSRAND_MI	Num		Days from randomization to first date of non-fatal

Variable	Type	Coding	Details
			MI or censoring
STROKE	Num	0 = No 1 = Yes	Non-fatal stroke
DAYSRAND_STROKE	Num		Days from randomization to first date of non-fatal stroke censoring
CVDDTH	Num	0 = No 1 = Yes	CVD Death
DAYSRAND_CVDDTH	Num		Days from randomization to date CVD Death or censoring
EXTMACE	Num	0 = No 1 = Yes	Any Extended MACE: Non-fatal MI, non-fatal stroke, coronary or peripheral revascularization, hospitalized congestive heart failure, CHD by angiography, silent MI, or fatal CVD
DAYSRAND_EXTMACE	Num		Days from randomization to first date extended MACE or censoring
CHF	Num	0 = No 1 = Yes	Hospitalized Congestive Heart Failure
DAYSRAND_CHF	Num		Days from randomization to first date of Hospitalized Congestive Heart Failure or censoring
COR_REVASC	Num	0 = No 1 = Yes	Coronary Revascularization: CABG, PTCA, or Coronary stent, laser, atherectomy
DAYSRAND_ COR_REVASC	Num		Days from randomization to first date Coronary Revascularization
CANCER	Num	0 = No 1 = Yes	Any cancer
DAYSRAND_CANCER	Num		Days from randomization to first date any cancer/censor
OBCANCER	Num	0 = No 1 = Yes	Obesity-related cancer (meningioma, multiple myeloma, esophagus, kidney, uterine, ovarian, thyroid, breast (postmenopausal), liver, gallbladder, upper stomach, pancreas, colon, and rectum)
DAYSRAND_OBCANC	Num		Days from randomization to first date obesity-related cancer/censor
BREASTCANCER	Num	0 = No 1 = Yes	Breast cancer
DAYSRAND_BREAST	Num		Days from randomization to first date breast cancer/censor
PROSTATECANCER	Num	0 = No 1 = Yes	Prostate cancer
DAYSRAND_PROSTATE	Num		Days from randomization to first date prostate cancer/censor
HEMOCANCER	Num	0 = No 1 = Yes	Hematological cancer
DAYSRAND_HEMO	Num		Days from randomization to first date

Variable	Type	Coding	Details
			Hematological cancer/censor
LUNGCANCER	Num	0 = No 1 = Yes	Lung cancer
DAYSRAND_LUNG	Num		Days from randomization to first date lung cancer/censor
COLORECTALCANCER	Num	0 = No 1 = Yes	Colon/rectum cancer
DAYSRAND_COLOREC	Num		Days from randomization to first date colon/rectum cancer/censor
UTERINECANCER	Num	0 = No 1 = Yes	Uterine cancer
DAYSRAND_UTERINE	Num		Days from randomization to first date uterus cancer/censor
MORTALITY	Num	0 = No 1 = Yes	Any Mortality
DAYSRAND_MORT	Num		Days from randomization to date any Mortality/censor
DESC_MORT	Char		Adjudicated mortality category

4.11 Central Unit Datasets

4.11.1 OS3_REL.LAB: Laboratory Data

DPPOS data LAB includes centrally-measured 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. Laboratory data were measured by the DPPOS Central Laboratory at the University of Washington.

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. OGTTs were discontinued after a confirmed diagnosis of diabetes.

This file includes the following variables in addition to RELEASE_ID, DAYSRAND and VISIT:

Variable (concentration for lab measurements)	Variable name	Measurement times
Fasting Plasma Glucose (mg/dL)	G000	Any visit – measured annually and at 12M 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	12A, 13A, 14A, 15A, 16A, 17A, 18A, 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	12A, 13A, 14A, 15A, 16A, 17A, 18A, INT, CON, POV. Measured at CON, POV or INT to capture OGTT as close as possible to the diagnosis of diabetes.

Variable (concentration for lab measurements)	Variable name	Measurement times
Fasting Insulin (uU/mL)	I000	12A, 13A, 14A, 15A, 16A, 17A, 18A, 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	12A, INT, CON, POV. Measured at CON, POV or INT to capture insulin as close as possible to the diagnosis of diabetes. Regular measures of 30-minute insulin were discontinued after 12A.
HbA1c (%)	HBA1	Any visit – measured annually per protocol. When diabetes diagnosed at a semi-annual visit (prior to 13A), also measured at CON, POV or INT to capture HbA1c as close as possible to the diagnosis of diabetes.
Total cholesterol (mg/dL)	CHOL	12A, 14A, 16A,18A, INT*
Triglycerides (mg/dL)	TRIG	12A, 14A, 16A,18A, INT
HDL (mg/dL)	CHDL	12A, 14A, 16A,18A, INT
LDL (mg/dL)	CLDL	12A, 14A, 16A,18A, INT
VLDL (mg/dL)	VLDL	12A, 14A, 16A,18A, INT
Serum creatinine (mg/dL)	CREA	12A, 13A, 14A, 15A, 16A,17A,18A
Vitamin B12 (pg/mL)	VB12	14A, 16A

*1 participant had lipids measured at 12M and one at a CON visit when the annual visit lipids were not collected.

OGTT measurements +		
Time started drinking glucola	DRNK0M – SAS TIME5. Format – seconds since midnight	12A, 13A, 14A, 15A, 16A,17A,18A, INT, CON, POV
30-minute blood draw time	DRNK30M – SAS TIME5. Format – seconds since midnight	12A, 13A, 14A, 15A, 16A,17A,18A, INT, CON, POV
2-hour blood draw time	DRNK2H – SAS TIME5. Format – seconds since midnight	12A, 13A, 14A, 15A, 16A,17A,18A, INT, CON, POV

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

4.11.2 OS3_REL.ULAB: Urine Laboratory Data

DPPOS data ULAB includes laboratory results from **all DPPOS (Phases 1, 2 and 3) annual visits with urine albumin and creatinine laboratory measurements**. Variable VISIT is used to identify the visit completed. Laboratory data were measured by the DPPOS Central Laboratory at the University of Washington.

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).
DAYSRAND	Days since randomization	Numeric		Days from DPP randomization
VISIT	Visit completed	Character		01A – 18A, CON, UNS, INT, POV
UCRE	Urine creatinine	Numeric	mg/dL	
UALB	Urine albumin	Numeric	mg/dL	

4.11.3 OS3_REL.FUNDUS: Fundus Photos

DPPOS data FUNDUS includes microaneurysms, macular edema and diabetic retinopathy severity results from fundus photos of the left eye collected during Year 16 of DPPOS Phase 3 (see section 3.6.2). Fundus photos were read and scored by the DPPOS Fundus Photo Reading Center at the University of Wisconsin.

Variable	Type	Coding	Details
RELEASE_ID	Char	9-digit character number beginning with “100”	Participant id for repository
DAYSRAND	Num		Days since randomization
VISIT	Char	16A	Outcome visit
FPCOLOR_PQ	Char		Fundus Reflex Confidence Score
FPCSRSN	Char		Confidence Score Reason
FPFOCALPC	Char		Focal and/or Grid photocoagulation
FPSCATERPC	Char		Scatter (Panretinal) PC
FPMACOUNT	Char		Number of microaneurysms
FPHEMRG	Char		Hemorrhages within Grid
FPHEGRID	Char		Hard exudate within grid, Field 2
FPCSME	Char		Clinically significant Macular Edema
FPAMD	Char		Early AMD Presence
FPOCULABNO	Char		Confounding ocular abnormality
FPABNORM	Char		Comments related to Fundus Reflex Grading (including ocular abnormalities)
FPDRSEVETY	Char		Diabetic Retinopathy (DR) severity level for the eye

4.11.4 OS3_REL.OCT: Optical Coherence Tomography

DPPOS data OCT includes results from OCT of the left eye collected during Year 16 of DPPOS Phase 3 (see section 3.6.2). OCTs were read and scored by the DPPOS Fundus Photo Reading Center at the University of Wisconsin.

Variable	Type	Coding	Details
RELEASE_ID	Char	9-digit character number beginning with “100”	Participant id for repository
DAYSRAND	Num		Days since randomization
VISIT	Char	16A	Outcome visit
OCTOCT_PQ	Char		OCT Confidence Score
OCTCSRSN	Char		Confidence Score Reason
OCTCTRLOC	Char		The center of the OCT volume scan based on the grid placement
OCTCTRLBL	Char		Total Center Layers Label
OCTCTRPTTHCK	Char		Total Center Point Thickness
OCTTOTTHK_C	Char		Total Thickness Sector _C
OCTTOTTHK_II	Char		Total Thickness Sector _II
OCTTOTTHK_IN	Char		Total Thickness Sector _IN
OCTTOTTHK_IS	Char		Total Thickness Sector _IS
OCTTOTTHK_IT	Char		Total Thickness Sector _IT
OCTTOTRLBLTY	Char		Total Thickness Inner Subfield Reliable
OCTRETLYRS	Char		Inner Retinal Layers Label
OCTIRPTTHCK	Char		Inner Retinal Center Point Thickness
OCTINRTHK_C	Char		Inner Thickness Sector C
OCTINRTHK_II	Char		Inner Thickness Sector II
OCTINRTHK_IN	Char		Inner Thickness Sector IN
OCTINRTHK_IS	Char		Inner Thickness Sector IS
OCTINRTHK_IT	Char		Inner Thickness Sector IT
OCTINRRLBLTY	Char		Inner Retinal Subfield Reliable
OCTPRPTTHC	Char		Photoreceptor Center Point Thickness
OCTPRTHK_C	Char		Photreceptor Thickness Sector _C
OCTPRTHK_II	Char		Photreceptor Thickness Sector _II
OCTPRTHK_IN	Char		Photreceptor Thickness Sector _IN
OCTPRTHK_IS	Char		Photreceptor Thickness Sector _IS
OCTPRTHK_IT	Char		Photreceptor Thickness Sector _IT
OCTPRRLBLTY	Char		Photoreceptor Thickness Reliable
OCTSSRDPRES	Char		SSRD Presence
OCTSSRDHIC	Char		SSRD Height in Center
OCTCYSSPC	Char		Cystoid Spaces
OCTPVD	Char		PVD: Presence and severity of posterior vitreous detachment
OCTERM	Char		ERM: Presence and location of Epiretinal

Variable	Type	Coding	Details
			membrane
OCTRTD	Char		RTD: Presence and location of Retinal Traction and Distortion
OCTMACHOLE	Char		Macular Hole: Presence and type of macular hole
OCTAMD	Char		AMD Presence
OCTPRLBL	Char		Photoreceptor Layers Label

4.11.5 OS3_REL.QWB: QUALITY OF WELL BEING

DPPOS data QWB includes the self-administered quality of well-being questionnaire (QWB-SA) collected during DPPOS Phase 3 Years 12M (during the midyear visit) and 16A (during the annual visit) on form Q11. This survey inquired about health problems that had occurred in the 3 days prior to the questionnaire, not including the day the questionnaire was administered. Data released include Questions 9 A, B and C and the summary variable TOTALQWB as coded by the Quality of Well Being Center at the University of California, San Diego. The remaining responses from the original survey are not available in the dataset.

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).
DAYSRAND	Days since randomization	Numeric		Days from DPP randomization
VISIT	Visit completed	Character	12M, 16A	
Q9A	Current health	Numeric	1: Excellent 2: Very Good 3: Good 4: Fair 5: Poor	Q9A: Would you say that your health is:
Q9B	Heath compared to one year ago	Numeric	1: Much better now than a year ago 2: Somewhat better now than one year ago 3: About the same as a year ago 4: Somewhat worse than a year ago 5: Much worse than a year ago	9b. Compared to a year ago, how would you rate your health in general now?
Q9C	State of health over the last 3 days	Numeric	Scaled as 0-100 by scale	Q9c: Think about a scale of 0 to 100, with zero being the least desirable state of health that you could imagine and 100 being perfect health. What number from 0 to 100 would you give to the state of your health, on average, over the last 3 days?
TOTALQWB		Numeric		Total QWB score

4.11.6 OS3_REL.DXA: DEXA Scans

DPPOS data DXA includes results from centrally analyzed DEXA data for all participants during DPPOS Years 12 and 17. The DEXA data are extensive and include measures of fat mass, lean mass, bone mass, and bone density for several body areas and whole body. DEXA Scans were centrally read by the DPPOS DEXA Reading Center at the University of San Francisco. The list of DEXA variables is extensive and can be found in the Proc Contents.

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).
DAYSRAND	Days since randomization	Numeric		Days from DPP randomization
VISIT	Visit completed	Character	12M, 17A	

5. Appendix: Sample SAS programs

5.1 Sample SAS program to create datasets

```
/* DPP data */
libname DPPREL 'Directory where you stored DPP data';

/* Push together all DPP data from baseline through followup */
data dpp_follow; set DPPREL.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';

/* 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';

/* 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';

/* Push together DPPOS Phase 2 data */
data dppos_follow2; set DPPOSR2.f01 DPPOSR2.f02 DPPOSR2.f06;
run;
proc sort; by release_id; run;

/* DPPOS Phase 3 data */
```

```
libname DPPOSR3 'Directory where you stored DPPOS Phase 3 data';

/* Push together DPPOS Phase 3 data */
data dppos_follow3; set DPPOSR3.f01 DPPOSR3.f02 DPPOSR3.f06;
run;
proc sort; by release_id; run;

/* Merge all visit data together from DPP+Bridge+DPPOS1+DPPOS2+DPPOS03 */
Data follow;  set dpp_follow(in=indpp) dppbr_follow(in=inbridge) dppos_follow1(in=indppos1)
dppos_follow2(in=indppos2) dppos_follow3(in=indppos3);
        if indpp then DPP=1;
        if inbridge then Bridge=1;
        if indppos1 then DPPOS1=1;
        if indppos2 then DPPOS2=1;
        if indppos3 then DPPOS3=1;

*** Compute weight at each followup visit;
        if QPWGHT1>. then WEIGHT = MEAN(QPWGHT1,QPWGHT2,QPWGHT3);
        else if APWGHT1>. then WEIGHT=MEAN(APWGHT1,APWGHT2,APWGHT3);
        else if KGWGHT1>. then WEIGHT=MEAN(KGWGHT1, KGWGHT2, KGWGHT3);
        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 - 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 data=follow; by release_id semi; run;

/* Combine diabetes events datasets */
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;
proc sort data=DPPOSR3.diabetes by release_id; run; *Note revised filename for DPPOS-3;
```

```
/* Combine diabetes events datasets from all 3 time periods */
data events;  set
        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)
        DPPOSR3.diabetes(in=indppos3 keep=release_id diabf diabt diabv totaltim);
        length last_event $6.;

        *** Keep the record from the latest the participant was in the study;
        if indppos3
                or (indppos2 and ~indppos3)
                or (indppos1 and ~indppos2 and ~indppos3)
                or (inbridge and ~indppos1 and ~indppos2 and ~indppos3)
                or (indpp and ~inbridge and ~indppos1 and ~indppos2 and ~indppos3);

        *** Label the visit type;
        if indppos3 then last_event='DPPOS3';
        else 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. Drop Troglitazone pts */
proc sort data=DPPREL.basedata; by release_id; run;
data events_demo; merge events DPPREL.basedata (where=(assign ne 'Troglitazone')); by release_id;
run;
```

5.2 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<=40; *Truncate at 20 years – data goes out a little further;
        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, 2 and 3';
run;
```

5.3 Sample SAS program for Cox Proportional Hazards Model

```
/* Sample Cox Proportional Hazards Model for time to diabetes during the full study:
DPP+Bridge+DPPOS1+DPPOS2+DPPOS3 */
PROC PHREG DATA=events_demo;
 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+DPPOS3';
RUN;
```