Diabetes Prevention Program Outcomes Study

Data Release Documentation

October 2011

DPPOS Phase 1 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 1 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 Outcome 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 http://www.bsc.gwu.edu/dpp. This report describes the public release of Phase 1 of the DPPOS dataset, and is based on all DPPOS data collected after the final DPP Bridge visits in 2002 and prior to the Phase 1 datalock in August, 2008. A brief description of the trial is given below. Full details of the DPP and DPP-bridge releases are available separately.

1.2 Medical Visits

1.2.1 DPP including Washout and Bridge period (1996 – Fall 2002)

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. 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 Phase 1 of DPPOS (Fall 2002 – August 2008)

Comprehensive annual assessments continued in DPPOS 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.

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 Phase 1 of DPPOS (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 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.

1.4 Diabetes Diagnosis and Subsequent Treatment

The complete definition of diabetes in the DPP and DPPOS is given in section 3.1. After a participant was confirmed to have diabetes, the intervention was continued and reinforced. However, once a participant was diagnosed with advanced 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, troglitazone and lifestyle participants continued with DPPOS without respect to $HbA1_c$ levels.

1.5 Exclusions from released data

Data that are part of the DPPOS Phase 2 primary (microvascular disease) or major secondary outcomes (cardiovascular events and death) are not included in this data release. Therefore, data related to microvascular disease, macrovascular disease and deaths are not included.

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.

- Only clinics and participants with IRB approval and informed consent to distribute their data to the NIDDK repository are included. Out of the 3250 participants who consented to DPPOS Phase 1, 3049 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 at 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 OS1_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 given below.

```
libname OS1_REL “directory for the SAS datasets on your host”;
filename tranfile “name of the transport file on your host”;
proc cimport data=OS1_REL.data infile=tranfile;
run;
```

For example to import file OS1_REL.F01:

```
libname OS1_REL “c:\mysasfiles”;
filename tranfile “c:\myxptfiles\F01.XPT”;
proc cimport data=OS1_REL.F01 infile=tranfile;
run;
```

- The contents of variables in these datasets are provided.

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.

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 indicating 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 1 through 6 visits including mid-year, annual, diabetes confirmation, and interim visits beginning in fall 2002 (varying dates depending on IRB approval dates for DPPOS) through August 2008.

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 2002 – August 2008								
	Lifestyle		Metformin		Placebo		Troglitazone	
VISIT	Study	Repository	Study	Repository	Study	Repository	Study	Repository
01A	859	800	886	831	886	821	471	464
01M	812	760	833	785	830	775	449	442
02A	828	770	857	801	874	806	454	447
02M	801	742	840	785	844	779	447	440
03A	818	758	836	776	845	779	442	436
03M	806	747	823	767	826	762	442	435
04A	813	756	826	769	831	764	433	426
04M	798	739	816	759	808	741	435	428
05A	825	766	844	788	848	780	440	433
05M	775	719	794	740	799	734	416	409
06A	743	693	749	707	762	706	402	397
06M	738	695	735	689	756	708	397	392

3. Statistical Considerations

3.1 Definition of 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 as diabetic. 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.

3.2 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 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. Endpoints after DPP formally ended in July 2001 include diabetes diagnosed during the washout and DPPOS Phase 1, with careful consideration of the impact of the group-lifestyle intervention offered during this bridge period.

3.4 Intent-to-treat

The DPP and DPPOS were analyzed as an intent-to-treat trial; that is, the treatment groups were compared without regard to compliance to medication or lifestyle during the trial.

3.5 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, with SAS Proc GLM.

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. In general, "specify" 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 Years 1, 2 and 5.
- Quality of well being: A self-administered Quality of Well Being (QWB) Questionnaire was completed at annual visits during DPPOS Phase 1. 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 diabetics and a sample of non-diabetics during DPPOS Year 1.
- 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 1 release.

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:
  - 01M, 02M, 03M… 06M: Regularly scheduled DPPOS mid-year visits.
  - 01A, 02A, 03A… 06A: 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
01M, 01A, 01Y	September 2002 – October 2003
02M, 02A	July 2003 – October 2004
03M, 03A	July 2004 – October 2005
04M, 04A	July 2005 – October 2006
05M, 05A	July 2006 – October 2007
06M, 06A	July 2007 – August 2008

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

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

4.4.1 OS1_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.

4.4.2 OS1_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.

4.4.3 OS1_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 below 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 OS1_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 OS1_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 1, 116 such visits took place and limited data were collected at home visits.

4.4.6 OS1_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.

4.5 Questionnaires (Q-forms)

4.5.1 OS1_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 1 annual 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 sixty-three and the lowest possible score is zero.

4.5.2 OS1_REL.Q02: HEALTH SURVEY QUESTIONNAIRE

DPPOS Form Q02 is the MOS SF-36 questionnaire. Form Q02 was self-administered at DPPOS Phase 1 annual 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.

4.5.3 OS1_REL.Q03: MODIFIABLE ACTIVITY QUESTIONNAIRE

DPPOS Form Q03 is the Modifiable Activity Questionnaire. Form Q03 was interviewer-administered at DPPOS Phase 1 annual 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.

4.5.4 OS1_REL.Q13: URINARY INCONTINENCE QUESTIONNAIRE

DPPOS Form Q13 was used to record participants' issues related to urinary incontinence during the past year. This form was self-administered at DPPOS Phase 1 annual visits.

4.5.5 OS1_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 4. VISIT = 04A or 04M for this form. This is similar to form Q12 used during DPP.

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 OS1_REL.L07: LIFESTYLE SESSION LOG

DPPOS form DPP 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 OS1_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 (below) by the "Date of Positive Pregnancy Test" variable (transformed to days from randomization).

4.7.2 OS1_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).

4.8 Report Forms (R-forms)

4.8.1 OS1_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. This form was completed at all visits where LDL was measured.

4.9 Central Unit Datasets

4.9.1 OS1_REL.LAB: Laboratory Data

DPPOS data LAB includes the 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.

Records which include OGTT data also include the blood draw times. 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.

Variable (concentration for lab measurements)	Variable name	Measurement times
Serum creatinine (mg/dL)	CREA	01A, 02A, 03A, 04A, 05A, 06A
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	01A, 02A, 03A, 04A, 05A, 06A, INT
Triglycerides (mg/dL)	TRIG	01A, 02A, 03A, 04A, 05A, 06A, INT
HDL (mg/dL)	CHDL	01A, 02A, 03A, 04A, 05A, 06A, INT
LDL (mg/dL)	CLDL	01A, 02A, 03A, 04A, 05A, 06A, INT
VLDL (mg/dL)	VLDL	01A, 02A, 03A, 04A, 05A, 06A, INT
LDL-B subfraction (mg/dL)	LDLB	01A, 05A
LDL-C subfraction (mg/dL)	LDLC	01A, 05A
LDL particle size (mg/dL)	LDLZ	01A, 05A
CRP (mg/dL)	CRP	01A, 05A
Fibrinogen (mg/dL)	FIBR	01A, 05A

OGTT measurements +		
Time started drinking glucola +	DRNK0M – SAS TIME5. Format – seconds since midnight	01A, 02A, 03A, 04A, 05A, 06A, INT, CON, POV
30-minute blood draw time +	DRNK30M – SAS TIME5. Format – seconds since midnight	01A, 02A, 03A, 04A, 05A, 06A, INT, CON, POV
2-hour blood draw time +	DRNK2H – SAS TIME5. Format – seconds since midnight	01A, 02A, 03A, 04A, 05A, 06A, 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	01A, 02A, 03A, 04A, 05A, 06A, 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	01A, 02A, 03A, 04A, 05A, 06A, 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	01A, 02A, 03A, 04A, 05A, 06A, 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	01A, 02A, 03A, 04A, 05A, 06A, INT, CON, POV. Measured at CON, POV or INT to capture insulin as close as possible to the diagnosis of diabetes.

+ Blood draw times can be found on form P07.

4.9.2 OS1_REL.NCC: Nutrient Data

DPPOS data NCC includes DPPOS Years 1, 2 and 5 data based on an interviewer-administered semi-quantitative food frequency questionnaire. The original questionnaire is not available for release. Data released includes the summary information outlined below which was coded by the Nutrition Coding Center at the University of South Carolina. Only coded nutrient variables are included in the released NCC dataset.

Most questionnaires were completed during the annual visit but some were completed at mid-year visits. VISIT=BRI were NCC questionnaires that were completed at the end of the BRIDGE between DPP and DPPOS but prior to formal enrollment in DPPOS – these may be merged with the DPPOS year 1 (01A and 01M) data.

Variable	Description	Units	Coding
ADDSALT	"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
ALCBEER	Daily intake of alcohol from beer	G	
ALCLIQU	Daily intake of alcohol from liquor	G	
ALCWINE	Daily intake of alcohol from wine	G	
ALC_DAY	Daily intake of alcohol from beer, wine & liquor	G	
DT_12_0	Daily intake of Lauric Acid (12:0) from diet	G	
DT_14_0	Daily intake of Myristic Acid (14:0) from diet	G	
DT_16_0	Daily intake of Palmitic Acid (16:0) from diet	G	
DT_18_0	Daily intake of Stearic Acid (18:0) from diet	G	
DT_18_3	Daily intake of Linolenic Acid (18:3) from diet	G	
DT_20_5	Daily intake of Eicosapentaenoic Acid (20:5) from diet	G	
DT_22_6	Daily intake of Docosahexaenoic Acid (22:6) from diet	G	
DT_ACAR	Daily intake of Alpha-Carotene from diet	MCG	
DT_ANZN	Daily intake of Zinc from animal sources from diet	MG	
DT_A_IU	Daily intake of Vitamin A (IU) from diet	I.U.	
DT_A_RE	Daily intake of Vitamin A (RE) from diet	R.E.	
DT_B1	Daily intake of Thiamin from diet	MG	
DT_B6	Daily intake of Vitamin B6 from diet	MG	
DT_BCAR	Daily intake of Beta-Carotene from diet	MCG	
DT_CALC	Daily intake of Calcium from diet	MG	
DT_CARB	Daily intake of Carbohydrate from diet	G	

Variable	Description	Units	Coding
DT_CHOL	Daily intake of Cholesterol from diet	MG	
DT_CRYP	Daily intake of Cryptoxanthin from diet	MCG	
DT_DFIB	Daily intake of Dietary Fiber from diet	G	
DT_FAT	Daily intake of Fat from diet	G	
DT_FE	Daily intake of Iron from diet	MG	
DT_FOL	Daily intake of Folate from diet,	MCG	*****CAUTION SEE DOCUMENTATION BELOW*****
DT_FRUC	Daily intake of Fructose from diet	G	
DT_GALAC	Daily intake of Galactose from diet	G	
DT_GLUC	Daily intake of Glucose from diet	G	
DT_KCAL	Daily intake of Calories from diet	Calories	
DT_LAC	Daily intake of Lactose from diet	G	
DT_LIN	Daily intake of Linoleic Acid from diet	G	
DT_LUT	Daily intake of Lutein from diet	MCG	
DT_LYC	Daily intake of Lycopene from diet	MCG	
DT_MG	Daily intake of Magnesium from diet	MC	
DT_NA	Daily intake of Sodium from diet	MC	
DT_NIAC	Daily intake of Niacin from diet	MC	
DT_OLEC	Daily intake of Oleic Acid from diet	G	
DT_PFA	Daily intake of Total Polyunsaturated Fat (n6 & n3)	G	
DT_PHOS	Daily intake of Phosphorus from diet	MG	
DT_POTA	Daily intake of Potassium from diet	MG	
DT_PROA	Daily intake of Provitamin A Carotenoids from diet	MCG	
DT_PROT	Daily intake of Protein from diet	G	
DT_RET	Daily intake of Retinol from diet	MCG	
DT_RIBO	Daily intake of Riboflavin from diet	MG	
DT_SFAT	Daily intake of Saturated Fat from diet	G	
DT_STAR	Daily intake of Starch from diet	G	
DT_SUCR	Daily intake of Sucrose from diet	G	
DT_TR_FA	Daily intake of Total Trans Fatty Acids from diet	G	
DT_VITC	Daily intake of Vitamin C from diet	MG	
DT_VITE	Daily intake of Vitamin E from diet	a-TE	
DT_ZINC	Daily intake of Zinc from diet	MG	
FATOIL	“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
FATMEAT	“How often do you eat the fat on meat?”	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
FG1	Bread, Cereal, Rice & Pasta (High Fiber/Low Fat)	Servings per day	

Variable	Description	Units	Coding
FG2	Bread, Cereal, Rice & Pasta (Low Fiber/High Fat)	Servings per day	
FG3	Bread, Cereal, Rice & Pasta (Low Fiber/Low Fat)	Servings per day	
FG4	Vegetable (Tomato)	Servings per day	
FG5	Vegetable (Dark Green/Deep Yellow)	Servings per day	
FG6	Vegetable (Cruciferous)	Servings per day	
FG7	Vegetable (Other)	Servings per day	
FG8	Fruit & Fruit Juice (Citrus)	Servings per day	
FG9	Fruit & Fruit Juice (Other)	Servings per day	
FG10	Dairy (High Fat)	Servings per day	
FG11	Dairy (Low Fat – Including up to 2% Milk)	Servings per day	
FG12	Fish (High Fat)	Servings per day	
FG13	Fish (Low Fat)	Servings per day	
FG14	Fish (High Omega 3 Fatty Acids)	Servings per day	
FG15	Dried Beans	Servings per day	
FG16	Eggs	Servings per day	
FG17	Meat (High Fat)	Servings per day	
FG18	Meat (Low Fat)	Servings per day	
FG19	Poultry (High Fat)	Servings per day	
FG20	Poultry (Low Fat)	Servings per day	
FG21	Sweets & Desserts	Servings per day	
FG22	Fats & Oils	Servings per day	
FG23	Soy Products	Servings per day	
FG24	Nuts & Seeds	Servings per day	
FG25	Coffee & Tea	Servings per day	
FG26	Meal Replacements (Instant Breakfast / Slimfast)	Servings per day	
FG27	Alcohol	Servings per day	

Variable	Description	Units	Coding
FMEALTM1	"How soon after you wake up do you have your first meal of the day?"	N/A	. = Missing
FMEALTM2	Unit of measure for FMEALTM1	N/A	1=Hours 2=Minutes . = Missing
HERBS	"Have you taken Herbs/Bot Supp During past Month?"		0= Yes, regularly 1= Yes, not regularly 2= No
LARGMEAL	"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
LEANMEAT	"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
LFBACON	"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
LFCAKE	"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
LFCHEESE	"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
LFCHIPS	"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
LFLUNCH	"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
LFYOGURT	"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
MEALSDAY	"How many meals per day do you usually eat?"	N/A	0 – 7 . = Missing
MOREDRNK	"How often do you drink 7+ drinks w/I 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

Variable	Description	Units	Coding
PERCCARB	Percent of Calories from Carbohydrate	N/A	
PERCFAT	Percent of Calories from Fat	N/A	
PERCLIN	Percent of Calories from Linoleic	N/A	
PERCOLEC	Percent of Calories from Oleic	N/A	
PERCPFAT	Percent of Calories from Polyunsaturated Fat (n6 &n3)	N/A	
PERCPROT	Percent of Calories from Protein	N/A	
PERCSFAT	Percent of Calories from Saturated Fat	N/A	
PFG1	Bread, Cereal, Rice & Pasta	Servings per day	
PFG2	Vegetable	Servings per day	
PFG3	Fruit	Servings per day	
PFG4	Milk, Yogurt & Cheese	Servings per day	
PFG5	Meat, Poultry, Fish, Dry Beans, Eggs & Nuts	Servings per day	
PFG6	Fats, Oils & Sweets	Servings per day	
SERVBEER	Servings of Beer Per Day	Servings per day	0 – 6
SERVLIQU	Servings of Liquor Per Day	Servings per day	0 – 6
SERVWINE	Servings of Wine Per Day	Servings per day	0 – 6
SEVENALC	"During the last year, have you ever had 7 or more alcoholic beverages within a 24 hour period (including mixed drinks, shots, beer and/or wine)"	N/A	1=No 2=Yes . or M = Missing
SKINCHIC	"How often do you eat the skin on chicken?"	N/A	1=Seldom/Never or N/A 2=Sometimes 3=Often/Always . = Missing
SNACKS	"How many snacks do you usually have per day? (This does not include diet beverages, coffee, tea or water)"	N/A	0 – 99 . = Missing
VITAMINS	"During the past month have you taken any vitamins or minerals?"	N/A	0= Yes, regularly 1= Yes, not regularly 2= No

NOTE ON FOLATE: On January 1, 1998, the US Department of Health and Human Services required that all enriched cereal grains be fortified with folate at 1.4 mg/kg of grain. These changes affect the estimates of folate intake in epidemiologic studies relying on nutrient databases which were impossible to update due to changing nutrient content of food in the months preceding January 1, 1998. The period of data collection for the DPP baseline and 1-year follow-up covered the time of rapid change in the marketplace. The decision was made to include the variable reflecting folate intake in the DPP baseline datasets, warning investigators of the potential biases in using this variable, but allowing them to pursue adjustment, or correction, procedures. Investigators should be reminded that the actual adjusted value for folate intake from food cannot be reconstructed and be further made aware that it will not be possible to use change in folate intake in analyses.

4.9.3 OS1_REL.QWB: Quality of Well Being Data

DPPOS data QWB includes the annual data based on a self-administered quality of well being questionnaire. 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.

4.9.4 OS1_REL.FUNDUS: Fundus (eye) Photo Data

DPPOS data FUNDUS includes results from fundus photos taken on all diabetics and a sample of nondiabetics during Year 1 of DPPOS (with VISIT coded as 01Y). Photographs were taken locally by trained photographers using FPRC 7-standard field color fundus photography using 35mm film (7Std-f). Photographs were sent to and read by the DPP Fundus Photo Reading Center at the University of Wisconsin.

4.10 Created Datasets

4.10.1 OS1_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 1 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.
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.

4.10.2 OS1_REL.EVENTS: Events Data

DPPOS data EVENTS includes one record for each participant. This file is updated from the DPP Bridge 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.
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 the datalock in August 2008.