
# Diabetes Prevention Program

# Data Release Documentation

# February 2008 Full Scale 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 (DPP)
# Full Scale Data Release
# Table of Contents

# 1. Introduction

## 1.1 General

The Diabetes Prevention Program (DPP) is a randomized clinical trial designed to investigate the efficacy of four treatment arms on the prevention of type 2 diabetes in high-risk adults. Detailed information about the DPP including DPP 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 complete public release of the DPP dataset, and is based on all data collected on or before July 31, 2001, after which the study group and participants were unmasked to study results. A brief description of the trial is given below.

## 1.2 Medical Visits

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.

## 1.3 Treatment Arms

At 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 point 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 discontinued after this point.

## 1.4 Diabetes Diagnosis and Subsequent Treatment

The complete definition of diabetes in the DPP 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 fasting hyperglycemia (defined by DPP to be fasting glucose $\geq$ 140 mg/dL, confirmed), coded medications were discontinued and the participant was sent to his or her local primary care provider for treatment; participation in the remainder of the DPP continued. Lifestyle participants continued with their intervention throughout the entire trial.

# 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 with IRB approval to distribute their data to the repository are included. Out of the 3819 original DPP participants, 3665 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. For example, data was sorted into small clearly-identifiable groups (sex*age) and collapsed if the sample size was small.
- Only research data is included in the released dataset, including data for all screening and post-randomization clinic visits, lifestyle visits, and laboratory data. Non-research data, including tracking forms, are not included. Adverse event and serious adverse event data were collected but are also not included in the data release. This data was not adjudicated and is not considered research data.
- All available data from each form and central unit database is included. 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 DPP_REL. One dataset exists for each DPP 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 DPP_REL 'directory for the SAS datasets on your host';
filename tranfile 'name of the transport file on your host';
proc cimport data=DPP_REL.data infile=tranfile;
run;
```

For example to import file DPP_REL.F01:

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

- The contents of variables in these datasets are provided.

## 2.3 De-identified Data

The DPP 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. In addition, variables that might identify a particular individual were collapsed into wider 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 modified fasting glucose or body mass index inclusion criteria that might identify specific participants, therefore fasting glucose values at baseline below 100 mg/dL were re-coded as 99 mg/dL and 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 all visits, including screening and run-in, baseline, quarterly, mid-year, annual, confirmation, primary endpoint, and interim visits through the Month 57 visit. Only 2 participants had a year 5 visit before the July 31, 2001 datalock; to protect participants' confidentiality, those observations have been removed. Section 4 describes in detail the data included.

The number of participants participating in each follow-up visit decreased over time owing to the 3-year period during which participants were randomized and the common July 31, 2001 termination date of the study. The table below shows the number of participants at each regularly-scheduled follow-up visit in the complete dataset and in this release dataset.

**Number of participants who completed in-clinic baseline, quarterly, mid-year and annual visits by treatment arm**
**For the original study sample and the database in NIDDK repository**

| | Lifestyle | | Metformin | | Placebo | | Troglitazone* | |
|---|---|---|---|---|---|---|---|---|
| VISIT | Study | Repository | Study | Repository | Study | Repository | Study | Repository |
| BAS | 1,079 | 1,024 | 1,073 | 1,027 | 1,082 | 1,030 | 585 | 584 |
| M03 | 1,043 | 990 | 1,027 | 981 | 1,031 | 979 | 531 | 530 |
| M06 | 1,046 | 993 | 1,030 | 984 | 1,026 | 976 | 556 | 556 |
| M09 | 1,002 | 950 | 999 | 955 | 1,017 | 966 | 365 | 365 |
| Y01 | 1,026 | 973 | 1,017 | 971 | 1,027 | 976 | 535 | 535 |
| M15 | 964 | 914 | 970 | 925 | 976 | 924 | 178 | 178 |
| M18 | 995 | 941 | 1,002 | 956 | 1,005 | 953 | 525 | 529 |
| M21 | 963 | 913 | 973 | 931 | 992 | 940 | 59 | 59 |
| Y02 | 1,000 | 947 | 1,006 | 960 | 1,015 | 963 | 525 | 524 |
| M27 | 924 | 876 | 941 | 900 | 957 | 907 | 0 | 0 |
| M30 | 885 | 844 | 876 | 838 | 898 | 858 | 521 | 520 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| M33 | 738 | 706 | 743 | 710 | 768 | 730 | 2 | 2 |
| Y03 | 638 | 604 | 625 | 601 | 657 | 623 | 525 | 524 |
| M39 | 503 | 476 | 510 | 489 | 514 | 485 | 3 | 3 |
| M42 | 434 | 411 | 445 | 427 | 438 | 418 | 415 | 414 |
| M45 | 347 | 330 | 335 | 322 | 333 | 318 | 4 | 4 |
| Y04 | 241 | 230 | 227 | 218 | 238 | 226 | 233 | 232 |
| M51 | 151 | 143 | 138 | 132 | 161 | 152 | 5 | 5 |
| M54 | 96 | 91 | 86 | 81 | 90 | 87 | 96 | 96 |
| M57 | 33 | 30 | 30 | 29 | 35 | 33 | 4 | 4 |

* Participants randomized to troglitazone discontinued quarterly visits after July 1998 when troglitazone was discontinued.

# 3. Statistical Considerations

## 3.1 Definition of Diabetes

The primary endpoint for the DPP 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 June 24, 1997 through April 1, 2001:
  - 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 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.

## 3.4 Intent-to-treat

The DPP was 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 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 in a gray box at the top of each form, and additional instructions are often included in section C at the bottom of the 1st page or throughout the form as required. The DPP 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.

Over the course of DPP 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, with the exception of the variables assessing coded troglitazone compliance as described in the specific forms below.

### 4.1.2 Variable Names on Data Forms

- Variable names for each released variable are embedded in blue on the data form.
- All datasets are HIPAA compliant. Information that might identify a specific participant has been excluded from the release datasets, and is indicated in light gray on the forms. This includes the original DPP participant ID, screening ID, clinical center, date of birth, participant initials, 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.
- Text information written on forms is indicated by underscore lines, and was not data entered and therefore not included in the release datasets.

## 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 baseline (Screening Step 3) and Year 1.
- Quality of well being: A self-administered Quality of Well Being (QWB) Questionnaire was completed at baseline and annual visits, beginning in mid-1997. One record is included of analyzed QWB data for each participant visit where the QWB was administered.
- CT-scan: A CT scan of L2-L3 and L4-L5 was completed on a random subset of participants at

Baseline and Year 1. One record for each participant of valid CT scan data is included for all CT scans completed.

- An EVENTS file includes summary event variables for diabetes, fasting hyperglycemia (fasting glucose ≥140 mg/dL) and death, as well as times to events and censoring data. This file has one record for each participant.
- A BASELINE file with one record for each participant which includes treatment assignment, baseline age group, baseline BMI group, sex, and race-ethnicity.

## 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:
  - SCR: Screening visits (initial visit to determine preliminary eligibility).
  - RUN: Run-in visits which are visits taking place after the screening visit and prior to randomization. Note that much of the data collected at the Step 3 run-in visits form the baseline data for DPP.
  - BAS: Baseline (randomization) visit.
  - M01, M02, …, M54, M57: Regularly scheduled non-annual visits.
  - Y01, Y02, Y03, Y04, Y05: Annual visits.
  - INT: Interim (unscheduled) visits.
  - UNS: No particular visit was completed (similar to INT).
  - 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 before (negative numbers) or after (positive values) randomization.

## 4.4 Screening and Run-in Forms (S-forms)

All participants completed a screening and run-in period prior to randomization. The screening and run-in period took at least 3 visits prior to randomization, over a period of 4 to 13 weeks. This period was used to:

- Assess eligibility,
- Determine if the participant was able and willing to complete the study tasks, such as taking medication (placebo), completing logs, and attending visits, and
- Collect baseline data.

The screening visit consisted of an eligibility OGTT as well as measurement of eligibility body mass index and collection of basic demographic information.

After the initial eligibility was confirmed participants entered a 3-week run-in period during which time the

preponderance of baseline data were collected and participants' ability to complete the various tasks of the study was assessed. Participants were allowed a 2nd run-in if eligibility tasks weren't completed satisfactorily; additional data-collection forms were not completed during the 2nd run-in.

Some participants completed 2 or 3 sets of screening and run-in periods prior to becoming eligible for the DPP. Only the final (eligible) data are included in the data release.

#### 4.4.1 DPP_REL.S01: ELIGIBILITY CHECKLIST

DPP Form S01 was used throughout the entire screening and run-in period and finalized prior to randomization. This form was used to record and verify eligibility criteria. Variable VISIT = RUN for this form.

#### 4.4.2 DPP_REL.S03: SCREENING STEP 2 INVENTORY

DPP Form S03 was used to record information collected at the initial screening visit. Variable VISIT = SCR for this form.

Several variables on this form are not included in the data release because it might identify a particular individual. These include all measures of weight as well as specific Asian and Hispanic ethnicities. In addition, only "White" and "Black" are included for variable SOETHN; responses 3 through 7 were recoded to missing to protect participant's identities.

For concomitant medications listed on forms S03, S07, F01, F02 and F06, "route" variables were included as a way to match concomitant medication to a DPP-purchased database. These variables are not included in the data release.

#### 4.4.3 DPP_REL.S05: SCREENING STEP 3 INVENTORY – START

DPP Form S05 was used to record information collected at the start of run-in (Step 3-Start). Variable VISIT = RUN for this form. The majority of the DPP baseline information was collected at this visit.

Several variables on this form are not contained in the data release because it might identify a particular individual. These include all measures of weight as well as specific dates of events. In addition, sagittal diameter was found to have been measured incorrectly on many participants; these measurements have been removed from all databases.

#### 4.4.4 DPP_REL.S06: SCREENING STEP 3 INVENTORY – END

DPP Form S06 was used to record information collected at the end of the run-in (Step 3-End). Variable VISIT = RUN for this form.

#### 4.4.5 DPP_REL.S07: SCREENING STEP 4 INVENTORY – RANDOMIZATION

DPP Form S07 was used to record information collected at the randomization visit. At the randomization visit, participants signed the final study informed consent, were given their randomization code, and began the intervention. The data collected at the randomization visit are recorded on form S07. Variable VISIT = BAS for this form.

There are several participants who had longer than 3 months between the initial screening and randomization, or shorter than 19-days of run-in. These and other minor randomization exceptions were approved by the DPP Screening and Eligibility Committee prior to randomization. No major randomization criteria were exempted.

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

#### 4.5.1 DPP_REL.F01: STANDARD FOLLOW-UP VISIT INVENTORY

DPP Form F01 was used throughout DPP to record information collected at quarterly and mid-year visits (NOT annual visits). Variable VISIT is used to identify the visit completed.

Note: One participant erroneously had an annual visit completed on an F01 form (coded as M36).

#### 4.5.2 DPP_REL.F02: MAJOR FOLLOW-UP VISIT INVENTORY

DPP Form F02 was used throughout DPP to record information collected at annual visits after baseline. Variable VISIT is used to identify the visit completed. Several variables on this form are not included in the data release because they might identify a particular individual.

Note: During the course of DPP, it was discovered that sagittal diameter was being measured incorrectly on a subset of participants. This variable was discovered to be uncorrectable and therefore is not included in the data release.

#### 4.5.3 DPP_REL.F03: INTERIM FOLLOW-UP VISIT INVENTORY

DPP Form F03 was used throughout DPP to record information collected at interim visits (e.g. not quarterly, mid-year or annual visits). The reason for interim visits is documented in section C and includes reasons such as coded medication management, blood pressure or other concomitant disease and concomitant medication management, etc. Although scheduled visits were conducted on a quarterly basis and recorded on forms F01 and F02, scheduled interim visits were completed at Month 1 throughout DPP for all medication arm participants for the purpose of titrating coded metformin to full dose. During the period while troglitazone was used, visits conducted at Month 1 also collected safety laboratory measurements (liver function). In addition, due to changes in the labeling for troglitazone, Form F03 was used to document liver function tests conducted monthly through Month 7 on all medication arm participants during the period in which troglitazone was used. Interim visits do not have a standard VISIT recorded, therefore VISIT = INT for all F03 forms.

#### 4.5.4 DPP_REL.F04: MISSED FOLLOW-UP VISIT REPORT

DPP Form F04 was used throughout DPP to record information about a quarterly, mid-year or annual visit that was missed and therefore no data are available. Form F04 was not used to collect missed visit information on a missed interim visit. Variable VISIT is used to identify the missed scheduled visit completed.

#### 4.5.5 DPP_REL.F05: MEDICATION ADHERENCE INTERVIEW

DPP Form F05 was used beginning in 1997 to record information collected at Month 1 and all scheduled visits regarding medication adherence. Variables VISIT and MAVSTWK are used to identify the visit completed. Coding for this form can be found in the file “F05codes.pdf”.

#### 4.5.6 DPP_REL.F06: HOME VISIT INVENTORY

DPP Form F06 was used beginning in late-1999 throughout DPP to record information about a mid-year or annual visit that was completed outside the clinic (at home). Only 5 such visits took place and limited data were collected at home visits.

### 4.6 Forms for Participants Randomized to Troglitazone (TR-forms)

#### 4.6.1 DPP_REL.TR1: PARTICIPANTS RANDOMIZED TO TROGLITAZONE FOLLOW-UP VISIT INVENTORY

DPP Form TR1 was used for participants randomized to troglitazone after the troglitazone arm of the protocol was stopped. Form TR1 was used to record information collected at mid-year and annual visits beginning in mid-1998. Variable VISIT is used to identify the visit completed.

#### 4.6.2 DPP_REL.TR2: PARTICIPANTS RANDOMIZED TO TROGLITAZONE GROUP SESSION LOG

DPP form TR2 records each participant who attended optional group sessions offered to former troglitazone participants after troglitazone was discontinued. Up to 30 participants could have been entered on one group session log. A series of codes are required for this form – coding can be found in

the file “Troglitazone lifestyle MoO.pdf”.

## 4.7 Questionnaires (Q-forms)

### 4.7.1 DPP_REL.Q01: BECK QUESTIONNAIRES

DPP Form Q01 includes both the Beck Depression Inventory and the Beck Anxiety Inventory. Form Q01 was self-administered at the Step 3-Start visit and at post-randomization 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.

### 4.7.2 DPP_REL.Q02: HEALTH SURVEY QUESTIONNAIRE

DPP Form Q02 is the MOS SF-36 questionnaire. Form Q01 was self-administered at the Step 3-Start visit and at post-randomization 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.

### 4.7.3 DPP_REL.Q03: MODIFIABLE ACTIVITY QUESTIONNAIRE

DPP Form Q03 is the Modifiable Activity Questionnaire. Form Q03 was interviewer-administered at the Step 3-Start visit and at post-randomization 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.

### 4.7.4 DPP_REL.Q04: LOW LEVEL PHYSICAL ACTIVITY RECALL

DPP Form Q04 is the Low-level Physical Activity Questionnaire. Form Q04 was interviewer-administered at the Step 3-Start visit and at post-randomization annual visits. Variable VISIT is used to identify the visit completed.

### 4.7.5 DPP_REL.Q05: NHANES III PHYSICAL ACTIVITY SCALE

DPP Form Q05 is the NHANES Physical Activity Scale. Form Q05 was interviewer-administered at the Step 3-Start visit only, and was used to compare the DPP participants to the national sample of NHANES participants. The scoring algorithm for this questionnaire is available on the NHANES website. Variable VISIT = RUN is used for this form.

### 4.7.6 DPP_REL.Q06 RETENTION AND TREATMENT MONITORING MEASURES

DPP Form Q06 included 3 questionnaires designed to assess DPP retention and treatment monitoring: Life Events, Social Provisions Scale, and Family Assessment. Form Q06 was self-administered at the Step 3-End visit and at mid-year visits throughout DPP.

### 4.7.7 DPP_REL.Q08 INTERVAL HISTORY QUESTIONNAIRE

DPP Form Q08 was completed by study staff at all annual visits after baseline, and was used to record updated medical information.

### 4.7.8 DPP_REL.Q09: DPP-SPECIFIC SUPPORT MEASURE – BASELINE VISIT

DPP Form Q09 was used to assess participants’ anticipated social support of their DPP participation, and was self-administered prior to randomization at the Step 3-End visit.

### 4.7.9 DPP_REL.Q10: DPP-SPECIFIC SUPPORT MEASURE – FOLLOW-UP VISITS

DPP Form Q10 was used to assess participants’ observed social support of their DPP participation, and

was self-administered at each annual visit after randomization.

#### 4.7.10 DPP_REL.Q12: ECONOMIC EVALUATION QUESTIONNAIRE

DPP Form Q12 was used to record participants' costs and time related to food purchase and preparation, exercise behavior and equipment, and medical care during DPP participation. This form was self-administered one time by each participant during mid- to late-2000. VISIT = UNS is used for this form.

#### 4.7.11 DPP_REL.Q13: URINARY INCONTINENCE QUESTIONNAIRE

DPP Form Q13 was used to record participants' issues related to urinary incontinence during the past year. This form was self-administered one time by each participant during mid-2001.

### 4.8 Intensive Lifestyle Forms (L-forms)

As described in detail in the lifestyle materials available on the DPP website (http://www.bsc.gwu.edu/dpp), intensive lifestyle participants completed a 16-session core curriculum followed by a lifestyle maintenance curriculum. Each participant had a designed lifestyle coach who completed form L03 after each in-person lifestyle visit. In addition, Lifestyle participants were offered a minimum of two lifestyle activity sessions each week. DPP Form L04 records each participant who came to an activity session. DPP form L05 records each participant who attended optional group sessions. A series of codes are required for these forms – coding can be found in the file "Lifestyle Coding for L03, L04 and L05.pdf".

#### 4.8.1 DPP_REL.L03: LIFESTYLE CONTACT – IN PERSON

DPP Form L03 was used throughout DPP to record information collected at in-person lifestyle visits.

#### 4.8.2 DPP_REL.L04: LIFESTYLE PHYSICAL ACTIVITY LOG

Up to 30 participants could have been entered on one activity log.

#### 4.8.3 DPP_REL.L05: LIFESTYLE GROUP SESSION LOG

Up to 30 participants could have been entered on one group session log.

### 4.9 Event Forms (E-forms)

#### 4.9.1 DPP_REL.E04: PREGNANCY CONFIRMATION REPORT

DPP 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).

#### 4.9.2 DPP_REL.E05: PREGNANCY OUTCOME REPORT

DPP 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).

### 4.10 Report Forms (R-forms)

#### 4.10.1 DPP_REL.R04: CHD Risk Status Report

DPP Form R04 was used to identify major risk factors for LDL goals as defined by NCEP ATP II in 1993. This form was completed at all visits where LDL was measured.

## 4.11 Central Unit Datasets

### 4.11.1 DPP_REL.LAB: Laboratory Data

DPP data LAB includes the laboratory results from all regularly scheduled visits. Troglitazone participants had fewer measurements after troglitazone was discontinued in 1998. The laboratory results outlined in the table below were measured at the given measurement times. Additional measurements at other times were completed upon clinic request, usually for safety concerns. Only regularly scheduled laboratory data are included. Baseline measurements collected at screening (all glucose measures) were combined with measurements collected at baseline into the BAS record.

Most post-randomization 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. Blood draw time data are available on form S03 for the screening (eligibility) OGTT.

| **Variable (concentration for lab measurements)** | **Variable name** | **Measurement times** |
|---|---|---|
| Serum Sodium (µmol/L) | NA | BAS |
| Serum Potassium (µmol/L) | K | BAS |
| Serum Bicarbonate (µmol/L) | HCO3 | BAS |
| Serum AST (U/L) | SGOT | BAS – all participants<br>M03, M06, Y01, M18, Y02, M30, Y03, M42, Y04, M54, and M01, M02, M04, M05, M07 and M09 for a brief period of time – medication arm participants only. |
| Serum ALT (U/L) | SGPT | BAS – all participants<br>M03, M06, Y01, M18, Y02, M30, Y03, M42, Y04, M54, and M01, M02, M04, M05, M07 and M09 for a brief period of time – medication arm participants only. |
| Serum creatinine (mg/dL) | CREA | BAS – all participants<br>M06, Y01, M18, Y02, M30, Y03, M42, Y04, M54 – medication arm participants only. |
| HbA1c (%) | HBA1 | BAS, M06, Y01, Y02, Y03, Y04 |
| Fasting Proinsulin (pM) | PIN | BAS, Y01, Y02, Y03, Y04 |
| Total cholesterol (mg/dL) + | CHOL | BAS, M06, Y01, Y02, Y03, Y04 |
| Triglycerides (mg/dL) + | TRIG | BAS, M06, Y01, Y02, Y03, Y04. |
| HDL (mg/dL) + | CHDL | BAS, M06, Y01, Y02, Y03, Y04 |
| LDL (mg/dL) + | CLDL | BAS, M06, Y01, Y02, Y03, Y04 |
| VLDL (mg/dL) | VLDL | BAS, M06, Y01, Y02, Y03, Y04 |
| LDL-B subfraction (mg/dL) | LDLB | BAS, M06, Y01, Y02, Y03, Y04 |
| LDL-C subfraction (mg/dL) | LDLC | BAS, M06, Y01, Y02, Y03, Y04 |

| | | |
|---|---|---|
| LDL particle size (mg/dL) | LDLZ | BAS, M06, Y01, Y02, Y03, Y04 |
| TPA (ng/mL) * | TPA | BAS, Y01 |
| CRP (mg/dL) * | CRP | BAS, M06, Y01 |
| Fibrinogen (mg/dL) * | FIBR | BAS, Y01 |
| Adiponectin (µg/mL) ** | ADIPON | BAS, Y01 |
| Urine Albumin (mg/dL) *** | UALB | BAS, CON, POV |
| Urine Creatinine (mg/dL) *** | UCRE | BAS, CON, POV |

| **OGTT measurements ++** | | |
|---|---|---|
| Time started drinking glucola ++ | DRNK0M | Y01, Y02, Y03, Y04, CON, POV |
| 30-minute blood draw time ++ | DRNK30M | Y01, Y02, Y03, Y04, CON, POV |
| 2-hour blood draw time ++ | DRNK2H | Y01, Y02, Y03, Y04, CON, POV |
| Fasting Plasma Glucose (mg/dL) | G000 | BAS, M06, Y01, M18, Y02, M30, Y03, M42, Y04, M54, CON |
| 30 Minute Plasma Glucose (mg/dL) | G030 | BAS, Y01, Y02, Y03, Y04, CON |
| 2 Hour Plasma Glucose (mg/dL) | G120 | BAS, Y01, Y02, Y03, Y04, CON |
| Fasting Insulin (uU/mL) | I000 | BAS, Y01, Y02, Y03, Y04 |
| 30 Minute Insulin (uU/mL) | I030 | BAS, Y01, Y02, Y03, Y04 |

* Due to changes over time in collection of Fibrinogen, TPA and CRP, occasional other visits have one or more of these measures.
** Adiponectin (total circulating) was measured after the end of DPP on stored samples.
*** Urine albumin and creatinine were reported following diabetes confirmation at the CON visit or the POV visit, and for some participants towards the end of the study (various visits).
+ Some participants had additional safety lipid measurement at M03.
++ Blood draw times can be found on form S03 for screening visits.

#### 4.11.2 DPP_REL.NCC: Nutrient Data

DPP data NCC includes the baseline (RUN) and Year 1 (Y01) 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.

| **Variable** | **Description** | **Corres-ponding with Supp. or Dietary Variable** | **Units** | **Coding** |
|---|---|---|---|---|
| ADDSALT | "How often do you add salt to your food at the table?" | | N/A | 1=Seldom/Never or N/A<br>2=Sometimes<br>3=Often/Always<br>. = 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 | WS_14_0 | G | |
| DT_16_0 | Daily intake of Palmitic Acid (16:0) from diet | WS_16_0 | G | |
| DT_18_0 | Daily intake of Stearic Acid (18:0) from diet | WS_18_0 | 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 | WS_20_5 | G | |
| DT_22_6 | Daily intake of Docosahexaenoic Acid (22:6) from diet | WS_22_6 | 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 | WS_A_IU | I.U. | |
| DT_A_RE | Daily intake of Vitamin A (RE) from diet | WS_A_RE | R.E. | |
| DT_B1 | Daily intake of Thiamin from diet | WS_B1 | MG | |
| DT_B6 | Daily intake of Vitamin B6 from diet | WS_B6 | MG | |
| DT_BCAR | Daily intake of Beta-Carotene from diet | WS_BCAR | MCG | |
| DT_CALC | Daily intake of Calcium from diet | WS_CALC | MG | |
| DT_CARB | Daily intake of Carbohydrate from diet | | G | |
| 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 | WS_FAT | G | |
| DT_FE | Daily intake of Iron from diet | WS_FE | MG | |
| DT_FOL | Daily intake of Folate from diet, | WS_FOL | 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 | WS_KCAL | 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 | WS_MG | MC | |
| DT_NA | Daily intake of Sodium from diet | | MC | |
| DT_NIAC | Daily intake of Niacin from diet | WS_NIAC | MC | |
| DT_OLEC | Daily intake of Oleic Acid from diet | WS_OLEC | G | |
| DT_PFA | Daily intake of Total Polyunsaturated Fat (n6 & n3) | WS_PFA | G | |

| | | | | |
|---|---|---|---|---|
| DT_PHOS | Daily intake of Phosphorus from diet | WS_PHOS | MG | |
| DT_POTA | Daily intake of Potassium from diet | WS_POTA | MG | |
| DT_PROA | Daily intake of Provitamin A Carotenoids from diet | WS_PROA | MCG | |
| DT_PROT | Daily intake of Protein from diet | | G | |
| DT_RET | Daily intake of Retinol from diet | WS_RET | MCG | |
| DT_RIBO | Daily intake of Riboflavin from diet | WS_RIBO | 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 | WS_VITC | MG | |
| DT_VITE | Daily intake of Vitamin E from diet | WS_VITE | a-TE | |
| DT_ZINC | Daily intake of Zinc from diet | WS_ZINC | MG | |
| FATMEAT | "How often do you eat the fat on meat?" | | N/A | 1=Seldom/Never or N/A<br>2=Sometimes<br>3=Often/Always<br>. = Missing |
| FATOIL | "How often is fat or oil used in cooking the foods you eat? | | N/A | 1=Never/<1 Per Month<br>2=1 Per Month<br>3=2-3 Per Month<br>4=1 Per Week<br>5=2 Per Week<br>6=3-4 Per Week<br>7=5-6 Per Week<br>8=1 Per Day<br>9=2+ Per Day<br>. = Missing |
| FG1 | Bread, Cereal, Rice & Pasta (High Fiber/Low Fat) | | Servings per day | |
| 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 | |
| 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<br>2=Minutes<br>. = Missing |
| HOWOFTEN | “About how often is it that you have had 7 or more alcoholic beverages within a 24 hour period?” | | N/A | 1=Once/week or more<br>2=No answer<br>3=< once/month<br>4=3 times per month<br>. = Missing |
| IRON | Take Iron Supplement? | | N/A | 0=No<br>1=Yes |
| LARGMEAL | “Which meal is usually your largest meal?” | | N/A | 1 – 9<br>.=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<br>2=Sometimes<br>3=Often/Always<br>. = 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<br>2=Sometimes<br>3=Often/Always<br>. = 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<br>2=Sometimes<br>3=Often/Always<br>. = 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<br>2=Sometimes<br>3=Often/Always<br>. = 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<br>2=Sometimes<br>3=Often/Always<br>. = 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<br>2=Sometimes<br>3=Often/Always<br>. = Missing |
| LFYOGURT | "If you eat yogurt, how often do you eat low-fat yogurt?" | | N/A | 1=Seldom/Never or N/A<br>2=Sometimes<br>3=Often/Always<br>. = Missing |
| MEALSDAY | "How many meals per day do you usually eat?" | | N/A | 0 – 7<br>. = Missing |
| OTHERVIT | Take Other Supplement (Including herbs and extracts)? | | N/A | 0=No<br>1=Yes |
| 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 | |
| SELENIUM | Take Selenium Supplement? | | N/A | 0=No<br>1=Yes |
| SERVBEER | Beer | | Servings per day | 0 – 6 |
| SERVLIQU | Liquor | | Servings per day | 0 – 6 |
| SERVWINE | Wine | | 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<br>2=Yes<br>. = Missing |
| SKINCHIC | "How often do you eat the skin on chicken?" | | N/A | 1=Seldom/Never or N/A<br>2=Sometimes<br>3=Often/Always<br>. = Missing |
| SNACKS | "How many snacks do you usually have per day? (This does not include diet beverages, coffee, tea or water)" | | N/A | 00 – 12<br>. = Missing |
| VITAMIN | "During the past month have you taken any vitamins or minerals?" | | N/A | 1=No<br>2=Yes, fairly regularly<br>3=Yes, but not regularly<br>. = Missing |
| WS_14_0 | Daily intake of Myristic Acid (14:0) from diet and supplements | DT_14_0 | G | |
| WS_16_0 | Daily intake of Palmitic Acid (16:0) from diet and supplements | DT_16_0 | G | |
| WS_18_0 | Daily intake of Stearic Acid (18:0) from diet and supplements | DT_18_0 | G | |
| WS_20_5 | Daily intake of Eicosapentaenoic Acid from diet and supplements | DT_20_5 | G | |
| WS_22_6 | Daily intake of Docosahexaenoic Acid from diet and supplements | DT_22_6 | G | |
| WS_A_IU | Daily intake of Vitamin A (IU) from diet and supplements | DT_A_IU | I.U. | |
| WS_A_RE | Daily intake of Vitamin A (RE) from diet and supplements | DT_A_RE | R.E. | |
| WS_B1 | Daily intake of Thiamin from diet and supplements | DT_B1 | MG | |
| WS_B6 | Daily intake of Vitamin B6 from diet and supplements | DT_B6 | MG | |
| WS_BCAR | Daily intake of Beta-Carotene from diet and supplements | DT_BCAR | MCG | |
| WS_CALC | Daily intake of Calcium from diet and supplements | DT_CALC | MG | |
| WS_CHOL | Daily intake of Cholesterol from diet and supplements | DT_CHOL | MG | |
| WS_FAT | Daily intake of Fat from diet and supplements | DT_FAT | G | |
| WS_FE | Daily intake of Iron from diet and supplements | DT_FE | MG | |
| WS_FOL | Date intake of Folate from diet and supplements, | DT_FOL | MCG | *****CAUTION SEE DOCUMENTATION BELOW***** |
| WS_KCAL | Daily intake of Calories from diet and supplements | DT_KCAL | Calories | |
| WS_MG | Daily intake of Magnesium from diet and supplements | DT_MG | MG | |
| WS_NIAC | Daily intake of Niacin from diet and supplements | DT_NIAC | MG | |
| WS_OLEC | Daily intake of Oleic Acid from diet and supplements | DT_OLEC | G | |

| | | | | |
|---|---|---|---|---|
| WS_PFA | Daily intake of Total Polyunsaturated Fat (n6 & n3) from diet and supplements | DT_PFA | G | |
| WS_PHOS | Daily intake of Phosphorus from diet and supplements | DT_PHOS | MG | |
| WS_POTA | Daily intake of Potassium from diet and supplements | DT_POTA | MG | |
| WS_PROA | Daily intake of Provitamin A Carotenoids from diet and supplements | DT_PROA | MCG | |
| WS_RET | Daily intake of Retinol from diet and supplements | DT_RET | MCG | |
| WS_RIBO | Daily intake of Riboflavin from diet and supplements | DT_RIBO | MG | |
| WS_VITC | Daily intake of Vitamin C from diet and supplements | DT_VITC | MG | |
| WS_VITE | Daily intake of Vitamin E from diet and supplements | DT_VITE | a-TE | |
| WS_ZINC | Daily intake of Zinc from diet and supplements | DT_ZINC | MG | |
| YEAST | Take Yeast Supplement? | | N/A | 0=No<br>1=Yes |
| ZINC | Take Zinc Supplement? | | N/A | 0=No<br>1=Yes |

**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.

### 4.11.3 DPP_REL.CT: CT Scan Data

DPP data CT includes the baseline (RUN) and Year 1 (Y01) data based on a lumbar spine CT scan. CT scans were performed on a subset of participants beginning in mid-1997 prior to baseline and at year 1. Attempts were made to obtain year 1 CT scans on all participants with baseline CT scans, but were not obtained in participants without a baseline CT. Visceral and subcutaneous measurements of adipose tissue are available at L2/L3 and L4/L5.

### 4.11.4 DPP_REL.QWB: Quality of Well Being Data

DPP data QWB includes the baseline (RUN) and annual (Y01, Y02, Y03 and Y04) data based on a self-administered quality of well being questionnaire. The questionnaire was implemented beginning in mid-1997 so data is only available beginning at that time. 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 information as coded by the Quality of Well Being Center at the University of California, San Diego. Questions on the original survey are not available in the dataset with the exception of Question 9 A, B and C.

## 4.12 Created Datasets

### 4.12.1 DPP_REL.BASEDATA: Baseline Data

DPP data BASEDATA includes one record for each participant in the released database. This file 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<br>2 = 40-44<br>3 = 45-49<br>4 = 50-54<br>5 = 55-59<br>6 = 60-64<br>7 = 65 and older | Computed based on date of randomization and birth date, from screening form S07. |
| ASSIGN | Treatment assignment | Character | Lifestyle<br>Metformin<br>Placebo<br>Troglitazone | Randomized treatment assignment. Not available on any data form. |
| BMICAT | BMI categorized (kg/m$^2$) | Numeric | BMI categorized into the following groups:<br>1: <26 kg/m$^2$<br>2: ≥26 to <28 kg/m$^2$<br>3: ≥28 to <30 kg/m$^2$<br>4: ≥30 to <32 kg/m$^2$<br>5: ≥32 to <34 kg/m$^2$<br>6: ≥34 to <36 kg/m$^2$<br>7: ≥36 to <38 kg/m$^2$<br>8: ≥38 to <40 kg/m$^2$<br>9: ≥40 to <42 kg/m$^2$<br>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:<br>1: <30 kg/m$^2$<br>2: ≥30 to <35 kg/m$^2$<br>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<br>2 = African American<br>3 = Hispanic, of any race<br>4 = All other | Self-reported race/ethnicity based on the 1990 census questionnaire during screening on Form S03. |
| SEX | Sex | Numeric | 1 = Male<br>2 = Female | Collected during screening on form S03. |

### 4.12.2 DPP_REL.EVENTS: Events Data

DPP data EVENTS includes one record for each participant. 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). |
| DEATH | Indicator of death | Numeric | 0 = No<br>1 = Yes | |
| DEATHDAYS | Number of days from randomization through date of death. | Numeric | | Missing if participant alive at end of DPP |
| DIABF | Indicator of diabetes | Numeric | 0 = No<br>1 = Yes | Indicator of ever diagnosed with diabetes during DPP. Computed based on fasting and/or 2-hour glucose values from the central laboratory. |
| DIABV | Interval for diabetes | Numeric | 1 = Month 6<br>2 = Year 1<br>3 = Month 18<br>4 = Year 2<br>5 = Month 30<br>6 = Year 3<br>7 = Month 42<br>8 = Year 4<br>9 = Month 54<br>10 = Year 5 | True time interval at which diabetes was diagnosed (NOT necessarily the VISIT that was conducted)<br>– OR –<br>The final visit where glucose was measured if not diabetic by final visit.<br>Note: Intervals are defined as 3-months before and 3-month after the target visit date except for interval 1 which began at randomization. |
| DIABT | Years to first diabetes | Numeric | | Number of years from randomization to visit where diabetes was diagnosed<br>– OR –<br>Number of years from randomization to final visit where glucose was measured if not diabetic by final visit. |
| FASTHYPF | Indicator of fasting hyperglycemia | Numeric | 0 = No<br>1 = Yes | Indicator of ever diagnosed with fasting hyperglycemia (fasting glucose ≥ 140 mg/dL) during DPP. Computed based on fasting values from the central laboratory. |

| Variable | Brief description | Type | Coding | Details |
|---|---|---|---|---|
| FASTHYPV | Interval for fasting hyperglycemia | Numeric | 1 = Month 6<br>2 = Year 1<br>3 = Month 18<br>4 = Year 2<br>5 = Month 30<br>6 = Year 3<br>7 = Month 42<br>8 = Year 4<br>9 = Month 54<br>10 = Year 5 | True time interval at which fasting hyperglycemia was diagnosed and NOT necessarily the VISIT that was conducted<br>– OR –<br>The final visit where glucose was measured if not hyperglycemic by final visit. |
| FASTHYPT | Years to first fasting hyperglycemia | Numeric | | Number of years from randomization to visit where fasting hyperglycemia was diagnosed<br>– OR –<br>Number of years from randomization to final visit where glucose was measured if not hyperglycemic by final visit. |
| RANDPER | Randomization period | Numeric | 1 = July –September 1996<br>2 = October – December 1996<br>3 = January – March 1997<br>4 = April – June 1997<br>5 = July –September 1997<br>6 = October – December 1997<br>7 = January – March 1998<br>8 = April – June 1998<br>9 = July –September 1998<br>10 = October – December 1998<br>11 = January – March 1999<br>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 time in study through last visit of any type (quarterly, mid-year, annual or interim) as of the datalock on 7/31/2001. |