Diabetes Prevention Program

Data Release Documentation

October 2011
Bridge 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)
Bridge 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 bridge period was the time period following the main DPP trial beginning in August 2001 and ending with the start of the Diabetes Prevention Program Outcomes Study (DPPOS), 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 the DPP bridge period dataset, and is based on all data collected after the main DPP visits ended in July 2001 and prior to the DPPOS Phase 1 period which started in the fall of 2002. A brief description of the trial is given below. Full details of the DPP and DPPOS Phase 1 public data releases are available separately.

1.2 Medical Visits

1.2.1 DPP (1996 – July 2001)

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

1.2.2 DPP Bridge (including Washout and Lifestyle training period, August 2001 – Fall 2002)

No changes were made to participant visit schedules during the bridge period. Comprehensive annual assessments continued, including physical measurements, medical history updates, adverse event assessment, medication adherence and dispensing, questionnaires, and a 2-hour 75g oral glucose tolerance test (OGTT). The briefer mid-year visits also continued, and included a subset of physical measurements, adverse event assessment, medication adherence and dispensing, and a fasting glucose test. Lastly, the brief quarterly visits continued as well, and included only adverse event assessment and medication adherence and dispensing. No changes were made to study outcomes during this bridge period.

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 routine quarterly visits were discontinued after this point.

1.3.2 DPP Bridge period (August 2001 – Fall 2002)

During the first 5 months of the bridge period, all participants were invited to a group-wide study-debriefing where the main trial results were described. In addition, participants were invited to one-on-one sessions with clinical center staff where personal results were distributed, including masked intervention status and clinical and lab measurements. These visits were not documented as data and are not included in the release dataset.

Washout visits: Prior to the one-on-one meetings, nondiabetic participants randomized to one of the masked medication-intervention arms (metformin or placebo) were asked to discontinue their coded medications for 10-14 days before returning to the clinic for an OGTT to assess diabetes status off-medication. If glucose results were elevated, a repeat OGTT was performed within 6 weeks. This washout visit data is contained in a special dataset included with this data release. After these visits, participants were scheduled for their one-on-one unmasking sessions.

Following the washout visit(s) and unmasking, placebo was discontinued. Metformin was continued open-label in participants who had been randomized to metformin and who were willing to continue. Intensive Lifestyle Intervention participants continued individual lifestyle sessions as staffing allowed.

Lifestyle Training: From January through June 2002, all DPP participants were invited to attend group-implemented lifestyle sessions with content identical to the Intensive Lifestyle Intervention core curriculum. These sessions were known as HELP (Healthy Lifestyle Program). Attendance at HELP sessions is documented in this data release.

1.4 Diabetes Diagnosis and Subsequent Treatment

The complete definition of diabetes as defined 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 advanced diabetes (defined during DPP as a fasting plasma glucose ≥140 mg/dl), 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 DPP bridge continued. Former placebo, troglitazone and lifestyle participants continued without respect to fasting glucose levels.

1.5 Exclusions from released data

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

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 repository are included. Out of the 3819 DPP participants, 3655 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 DPP bridge-period clinic visits, individual lifestyle visits, group HELP 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, as adverse events were 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 DPPBR_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 DPPBR_REL 'directory for the SAS datasets on your host';
filename tranfile 'name of the transport file on your host';
proc cimport data=DPPBR_REL.data infile=tranfile;
run;
```

For example to import file DPPBR_REL.F01:

```
libname DPPBR_REL 'c:\mysasfiles';
filename tranfile 'c:\myxptfiles\F01.XPT';
proc cimport data=DPPBR_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 those participants with consent for the bridge data release. 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 all visits including mid-year, annual, diabetes confirmation, and interim visits during the bridge period from August 2001 and ending in the fall of 2002 (varying dates depending on IRB approval dates for DPPOS).

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

Number of participants who completed in-clinic quarterly, mid-year and annual visits based on forms F01, F02 and TR1
by treatment arm
For the original study sample and the Bridge database in NIDDK repository

Bridge Period August 2001 – Fall 2002

	Lifestyle		Metformin		Placebo		Troglitazone*	
VISIT	**Study**	**Repository**	**Study**	**Repository**	**Study**	**Repository**	**Study**	**Repository**
Y02	1	1	0	0	0	0	0	0
M27	15	11	13	9	5	4	0	0
M30	99	88	94	86	95	84	0	0
M33	200	185	184	174	180	170	0	0
Y03	334	316	342	324	318	302	2	2
M39	385	361	353	331	360	339	1	1
M42	435	410	423	398	450	421	102	102
M45	396	372	419	395	414	388	7	7
Y04	402	375	421	395	418	392	275	274
M51	357	334	341	323	327	304	40	40
M54	339	315	346	327	351	329	314	312
M57	325	301	330	312	308	288	47	47
Y05	276	255	247	235	244	227	244	243
M63	170	158	148	141	168	156	50	50
M66	108	97	103	98	103	92	101	101
M69	60	54	62	58	54	51	14	14
Y06	9	5	7	6	11	9	9	9
M75	1	1	1	1	1	1	0	0

* Participants randomized to troglitazone discontinued routine 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 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 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

During the main DPP trial, 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, with careful consideration of the impact of the group-lifestyle intervention offered during this bridge period.

3.4 Intent-to-treat

Throughout DPP and the bridge period, 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 and the bridge period 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 and version 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 and codes were added, and other variables were removed. Only the final PDF version of each form is distributed with this data release, with the exception of forms that were changed mid-way through the bridge period when placebo was discontinued, metformin was distributed open-label and the lifestyle HELP sessions were available to all participants. These forms are described in details below. All data collected are included in the released 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.
- 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.
- Washout data: One record for each participant for each visit where laboratory measurements were completed for the washout visits. No specific form was completed at these visits. Variables include glucose levels and OGTT times.
- 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.

- An EVENTS file includes summary event variables for diabetes, fasting hyperglycemia (fasting glucose ≥140 mg/dL), 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 BASELINE data released with the DPP Full Scale Release dataset but includes only those participants with consent for the bridge data 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:
  - M01, M02, …, M54, M57: Regularly scheduled non-annual visits.
  - Y01, Y02, Y03, Y04, Y05, Y06: 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).
  - WOV: washout visits (washout data only)
  - WCV: washout confirmation visits (washout data only)
- DAYSRAND: The number of days a particular visit occurred after randomization.

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

4.4.1 DPPBR_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: This form was changed as of January 1, 2002 therefore there are 2 versions of the form in the release dataset. Variable VERS identifies version 2 vs. version 3.

4.4.2 DPPBR_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. This form was changed as of January 1, 2002, therefore there are 2 versions of the form in the release dataset. Variable VERS identifies version 2 vs. version 3.

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

4.4.3 DPPBR_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 during the bridge period is documented in Section C 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 DPPBR_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.

4.4.5 DPPBR_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".

4.4.6 DPPBR_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 limited data were collected at home visits.

4.5 Forms for Participants Randomized to Troglitazone (TR-forms)

4.5.1 DPPBR_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 through early 2002. Variable VISIT is used to identify the visit completed. Beginning in 2002, troglitazone participant visits were captured on forms F01 and F02.

4.6 Questionnaires (Q-forms)

4.6.1 DPPBR_REL.Q01: BECK QUESTIONNAIRES

DPP Form Q01 includes both the Beck Depression Inventory and the Beck Anxiety Inventory. Form Q01 was self-administered at 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 21 BDI and BAI questions is 3, therefore the highest possible total for the whole BDI or BAI is 63 and the lowest possible score is 0.

4.6.2 DPPBR_REL.Q02: HEALTH SURVEY QUESTIONNAIRE

DPP Form Q02 is the MOS SF-36 questionnaire. Form Q02 was self-administered at 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.

4.6.3 DPPBR_REL.Q03: MODIFIABLE ACTIVITY QUESTIONNAIRE

DPP Form Q03 is the Modifiable Activity Questionnaire. Form Q03 was interviewer-administered at 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.

4.6.4 DPPBR_REL.Q04: LOW-LEVEL PHYSICAL ACTIVITY RECALL

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

4.6.5 DPPBR_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 mid-year visits throughout DPP.

4.6.6 DPPBR_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.

4.6.7 DPPBR_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.

4.6.8 DPPBR_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.

4.7 Intensive Lifestyle Forms (L-forms)

DPP forms: As described in detail in the lifestyle materials available on the DPP website (http://www.bsc.gwu.edu/dpp), during DPP intensive lifestyle participants completed a 16-session core curriculum followed by a lifestyle maintenance curriculum. Each DPP intensive lifestyle participant had a designated 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.

DPP bridge forms: As during DPP, DPP Form L04 records each participant who came to an activity session and DPP form L05 records each participant who attended optional group sessions. However, beginning in January 2002, form L04 and L05 were modified to record group-implemented lifestyle sessions for all participants. Form L05 was used to record the group-implemented core curriculum sessions as well as additional optional group sessions. DPP Form L06 was used for visits beginning in January 2002 only for participants randomized to intensive lifestyle. This form is a modified version of form L03 and was used during the bridge period to record information collected at in-person lifestyle visits. A series of codes are required for these forms – coding for type of session can be found in the file "Lifestyle Coding for L03, L04 and L05.pdf".

4.7.1 DPPBR_REL.L03: LIFESTYLE CONTACT– IN PERSON

DPP Form L03 was used throughout DPP to record information collected at in-person lifestyle visits. This form was for visits prior to January 2002 and includes only participants randomized to intensive lifestyle.

4.7.2 DPPBR_REL.L04: LIFESTYLE PHYSICAL ACTIVITY LOG

Up to 30 participants could have been entered on one activity log. Version 1 of this form was for visits prior to January 2002 and includes only participants randomized to intensive lifestyle. Version 2 of this form was implemented in January 2002 and includes participants from all treatment arms during the HELP training. Variable VERS identifies version 1 vs. version 2.

4.7.3 DPPBR_REL.L05: LIFESTYLE GROUP SESSION LOG

Up to 30 participants could have been entered on one group session log. Variable VERS identifies version 1 vs. version 2.

- Version 1 of this form was for visits prior to January 2002 and includes only participants randomized to intensive lifestyle. Codes for Version 1 are included in the file "Lifestyle Coding for L03, L04 and L05.pdf".
- Version 2 of this form was implemented in January 2002 and includes participants from all treatment arms during the group HELP training. Coding for type of session can be found in the file "Lifestyle Coding for L03, L04 and L05.pdf" and codes for the group-implemented core curriculum sessions exclusively used for Version 2 are shown below.

Variable	Brief description	Type	Coding	Details
GRTYPE	Type of session	Numeric	801 = Core session 1 802 = Core session 2 … 816 = Core session 16	Values correspond to the 16-session lifestyle core curriculum during the group-implemented HELP training.

4.7.4 DPPBR_REL.L06: BRIDGE PERIOD – INTENSIVE LIFESTYLE CONTACT – IN PERSON

DPP Form L06 was used for visits beginning in January 2002 and includes only participants randomized to intensive lifestyle. This form is a modified version of form L03 and was used during the bridge period to record information collected at in-person lifestyle visits.

4.8 Event Forms (E-forms)

4.8.1 DPPBR_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).

4.8.2 DPPBR_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).

4.9 Report Forms (R-forms)

4.9.1 DPPBR_REL.R04: CHD Risk Status Report

DPP 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.

4.10 Central Unit Datasets

4.10.1 DPPBR_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.

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.

Variable (concentration for lab measurements)	**Variable name**	**Measurement times**
Serum AST (U/L)	SGOT	M30, Y03, M42, Y04, M54, Y06, M66 – medication arm participants only.
Serum ALT (U/L)	SGPT	M30, Y03, M42, Y04, M54, Y06, M66 – medication arm participants only.
Serum creatinine (mg/dL)	CREA	M30, Y03, M42, Y04, M54, Y06, M66 – medication arm participants only.
HbA1c (%)	HBA1	Y02, M30, Y03, M42, Y04, M54, Y05, M66, Y06, CON, POV
Total cholesterol (mg/dL)	CHOL	Y02, Y03, Y04, Y05, Y06
Triglycerides (mg/dL)	TRIG	Y02, Y03, Y04, Y05, Y06
HDL (mg/dL)	CHDL	Y02, Y03, Y04, Y05, Y06
LDL (mg/dL)	CLDL	Y02, Y03, Y04, Y05, Y06
VLDL (mg/dL)	VLDL	Y02, Y03, Y04, Y05, Y06
LDL-B subfraction (mg/dL)	LDLB	Y02, Y03, Y04, Y05, Y06
LDL-C subfraction (mg/dL)	LDLC	Y02, Y03, Y04, Y05, Y06
LDL particle size (mg/dL)	LDLZ	Y02, Y03, Y04, Y05, Y06
OGTT measurements +		
Time started drinking glucola +	DRNK0M	Y02, Y03, Y04, Y05, Y06, CON, POV
30-minute blood draw time +	DRNK30M	Y02, Y03, Y04, Y05, Y06, CON, POV
2-hour blood draw time +	DRNK2H	Y02, Y03, Y04, Y05, Y06, CON, POV
Fasting Plasma Glucose (mg/dL)	G000	Y02, M30, Y03, M42, Y04, M54, Y05, M66, Y06, CON, POV
30 Minute Plasma Glucose (mg/dL)	G030	Y02, Y03, Y04, Y05, Y06, CON, POV
2 Hour Plasma Glucose (mg/dL)	G120	Y02, Y03, Y04, Y05, Y06, CON, POV
Fasting Insulin (uU/mL)	I000	Y02, Y03, Y04, Y05, Y06, CON, POV

30 Minute Insulin (uU/mL)	I030	Y02, Y03, Y04, Y05, Y06, CON, POV
Fasting Proinsulin (pM)	PIN	Y02, Y03, Y04, Y05, Y06, CON, POV

+ Blood draw times can be found on form P07.

4.10.2 DPPBR_REL.WASHOUT: Washout Laboratory Data

DPP data WASHOUT includes the glucose results from all washout visits on nondiabetic metformin and placebo participants. The laboratory results outlined in the table below were measured from August 2001 through early 2002.

OGTT measurements ++		
Time started drinking glucola ++	DRNK0M	WOV, WCV
30-minute blood draw time ++	DRNK30M	WOV, WCV
2-hour blood draw time ++	DRNK2H	WOV, WCV
Fasting Plasma Glucose (mg/dL)	G000	WOV, WCV
30 Minute Plasma Glucose (mg/dL)	G030	WOV, WCV
2 Hour Plasma Glucose (mg/dL)	G120	WOV, WCV

4.10.3 DPPBR_REL.QWB: Quality of Well Being Data

DPP data QWB includes annual (Y02, Y03 and Y04) 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 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.

4.11 Created Datasets

4.11.1 DPPBR_REL.DEMOGRAPHIC: Demographic Data

DPP data DEMOGRAPHIC includes one record for each participant in the released database. Data in this file is identical to the BASELINE data included in the DPP Full Scale data release, but includes only those participants with consent for the bridge data release, and contains 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.

4.11.2 DPPBR_REL.EVENTS: Events Data

DPP data EVENTS includes one record for each participant. This file is updated from the full data release, and includes also diabetes diagnosed during the washout visit. The following variables are available:

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. Computed based on fasting and/or 2-hour glucose values from the central laboratory.
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 11 = Month 66 12 = Year 6	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.
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.
FASTHYPF	Indicator of fasting hyperglycemia	Numeric	0 = No 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 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 11 = Month 66 12 = Year 6	True time interval at which fasting hyperglycemia was diagnosed and NOT necessarily the VISIT that was conducted – OR – 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 – OR – 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 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 time in study through last visit of any type (quarterly, mid-year, annual or interim) as of the final DPP visit in fall 2002.