Diabetes Prevention Program

Data Release Documentation

November 2019 Supplemental 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 3

2. Release Information 4

2.1 General Information 4

2.2 Data Location 4

2.2.1 Structure of the SAS Data Files: 4

2.3 De-identified Data 4

2.4 Structure of the Datasets 5

3. File Descriptions 6

3.1 Variables Common to All Datasets 6

3.2 Results Datasets 6

3.2.1 Sample collection and storage 6
3.2.2 DPP_REL.Antibodies 6
3.2.3 DPP_REL.Biomarkers 6
3.2.4 DPP_REL.Hormone 7
3.2.5 DPP_REL.ProBNP 7
3.2.6 DPP_REL.VIT_D 7

4. References: 8

1. Introduction

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. Detailed information about the DPP including DPP protocols, intensive lifestyle manuals, references, publication list, and links to PubMed abstracts and manuscripts is available at https://dppos.bsc.gwu.edu/web/dppos/about.

This is the public release of results from supplemental measurements using blood samples that had been stored at the central Biochemistry Laboratory at the University of Washington. Only results from samples collected and measured on DPP samples are included – that is, no samples from the long-term outcomes study (DPPOS) are included. Further information about DPP can be found in the full release data documentation or the DPP website above. DPP and DPPOS publications, including those using these data, can be found here.

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 were included in the original data release in 2008. This release includes measurements from samples taken on only those participants.
- No measurements are included for participants assigned to the original troglitazone arm.
- All available results are included. Missing data was caused by a variety of reasons: the visit was not completed in its entirety; the sample was not available, the samples was accidentally not collected or measured; the sample volume was insufficient; the sample was unusable for other reasons.

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.Hormone:

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

- The contents of variables in these datasets are provided.

2.3 De-identified Data

The dataset was de-identified in the following manner. All personal identifiers were removed, including DPP participant ID and other personal identifiers (PHI), clinical center, and all dates.

2.4 Structure of the Datasets

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

3. File Descriptions

3.1 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:
  - BAS: Baseline (randomization) visit or prior to baseline (e.g., during screening)
  - M06: Month 6 visit, approximately 6 months after DPP randomization
  - Y01, Y02: Annual visits at DPP Years 1 and 2
- DAYSRAND: The number of days a particular visit occurred after randomization. Baseline and screening are assigned 0.

3.2 Results Datasets

SAS Proc Contents are available for each dataset below, and include the variable names and unit of measurement.

3.2.1 Sample collection and storage

All samples were collected during DPP participant visits, centrifuged and aliquoted by the clinical center, and sent frozen on dry ice to the DPP Central Biochemistry Laboratory at the Northwest Lipid Research Laboratory, University of Washington, Seattle, WA, USA. Upon arrival, samples were frozen at −70°C and stored until use by the relevant project.

3.2.2 DPP_REL.Antibodies

DPP data **Antibodies** includes results for Glutamic acid decarboxylase 65 (GAD65) autoantibodies and insulinoma-associated-2 autoantibodies (IA-2) measured in 2009 on a sample of DPP participants using stored, fasting, plasma EDTA samples from DPP baseline (visit=BAS). A total of 625 participant results are included.

Diabetes autoantibody measurements were performed at the DPP Central Biochemistry Laboratory. GAD65 and IA-2 autoantibodies were analyzed using the harmonized protocol [2]. A radio-labeled human recombinant GAD65 was produced (Promega Inc., Madison, WI, USA) using a pThGAD65 clone. The concentration of GAD autoantibodies was determined using the standard curve and was expressed in GAD65 harmonized DK units. A DK value of ≥33, equivalent to the 99th percentile in 500 Diabetes Autoantibodies Standardization Program control subjects, was considered positive, with 88% sensitivity and 94.9% specificity compared with Diabetes Autoantibodies Standardization Program 2010 workshop standards. The insulinoma-associated-2 autoantibody assay was identical to the GAD65 autoantibody assay except that 35S-labelled insulinoma-associated-2 was used as a tracer. The assay used a pSP64-PolyA-IA-2ic clone (amino acids 601–979) in the in vitro insulinoma-associated-2 IC (intracytoplasmic construct) transcription and translation system to produce 35S-IA2 trace. Using the standard curve a sample with DK value > 5 was positive (99th percentile in Diabetes Autoantibodies Standardization Program control subjects) for a 62% sensitivity and 100% specificity, similar across laboratories [1].

3.2.3 DPP_REL.Biomarkers

DPP data Biomarkers includes results for leptin, Interleukin 6 (IL-6), soluble E-selectin (sE-selectin), intercellular adhesion molecule 1 (sICAM-1), and monocyte chemotactic protein 1 (MCP-1). These biomarkers were measured in 2009-2010 on stored, fasting, plasma EDTA samples from DPP baseline (visit=BAS) and DPP Year 1 (visit=Y01).

Previously aliquoted samples were thawed and measured at the DPP Central Biochemistry Laboratory. IL-6 and MCP-1 were measured by ELISA assays from R&D Systems (Minneapolis MN), sE-selectin and sICAM-1 were measured by R&D ELISA assays, and leptin was measured by radioimmunoassay (Millipore Inc.). The within-run and between-run coefficients of variation (CV) for all biomarker assays ranged between 2.10-7.40% and 2.60-9.25% respectively.

3.2.4 DPP_REL.Hormone

DPP data Hormone includes results from sex hormones and sex hormone binding globulin (SHBG) measured in 2011-2013 using non-fasting heparinized plasma samples collected during DPP screening (visit=BAS) and at DPP Year 1 (visit=Y01). Sex steroids measured include estrone, estrone-sulfate, estradiol, estradiol-sulfate, testosterone, dehydroepiandrosterone (DHEA), dehydroepiandrosterone-sulfate (DHEA-S), dihydrotestosterone (DHT), along with SHBG.

Samples were sent frozen on dry ice to Endoceutics (Quebec City, Canada) for measurement of SHBG and sex steroids. SHBG was measured using an ELISA (Bioline) with interassay coefficients of variation of 7.8 and 5.0 at 18.2 and 63.1 nmol/l, respectively. Sex steroids were measured using gas chromatography/mass spectrometry [2]. The lower limits of detection (LLD) for DHEA, DHT, testosterone, estradiol, estrone and estrone-sulfate were 100, 2, 10, 0.2, 0.8, and 10 pg/mL, respectively and 20 ng/mL for DHEA-S. The lower limits of quantification (LLOQ) for DHEA, DHT, testosterone, estradiol, estrone and estrone-sulfate were 500, 10, 50, 1, 4, and 50 pg/mL, respectively and 100 ng/mL for DHEA-S. Inter-assay variation (coefficient of variation, %CV) was 10.2, 10.7, 7.0, and 12.5 for DHEA, testosterone, estradiol and estrone respectively at the LLOQ level. Values were extrapolated below the lower limit of quantification using Mass Hunter.

3.2.5 DPP_REL.ProBNP

DPP data PROBNP includes a single result for circulating N-terminal pro B-type natriuretic peptide (NT-proBNP), measured in 2011 on stored, fasting, plasma EDTA samples from DPP baseline (visit=BAS) and DPP Year 2 (visit=Y02).

Previously aliquoted samples were shipped frozen on dry ice to the Biomarker Research Laboratory/TIMI Clinical Trials Laboratory at the Brigham and Women's Hospital. Samples were thawed, spun and centrifuged for 5 min at 1,500 rpm. NT-proBNP concentrations were measured using a cobas e601 immunoanalyser and proBNP II assay (both Roche Diagnostics, Indianapolis, IN, USA).

3.2.6 DPP_REL.VIT_D

DPP data Vit_D includes results for vitamin D2, vitamin D3 and total vitamin D on participants randomized to placebo and intensive lifestyle only. These were measured in 2009 on stored, fasting, plasma EDTA samples from DPP Baseline (visit=BAS), Month 6 (visit=M06), and Years 1 and 2 (visit=Y01, Y02).

Plasma 25OHD concentrations were measured in samples stored at −70°C since collection. Samples were analyzed in the fewest batches possible, and all samples from a single participant analyzed in the same run. Stability of vitamin D metabolites during transport and long-term freezing has been documented previously [3, 4]. Plasma 25OHD concentration was measured at the Metabolic Laboratory at Tufts Medical Center by liquid chromatography, tandem mass spectrometry (LC/MS/MS) (Waters Acquity UPLC with TQD triple quadrupole mass spectrometer), certified through the National Institute of Standards and Technology (NIST) vitamin D quality assurance program [5]. In the most recent testing, correlation with the NIST external standard for total 25OHD concentration was 0.994.

4. References:

1) Bonifacio E, Yu L, Williams AK, Eisenbarth GS, Bingley PJ, Marcovina SM et al. Harmonization of Glutamic Acid Decarboxylase and Islet Antigen-2 Autoantibody Assays for National Institute of Diabetes and Digestive and Kidney Diseases Consortia. J Clin Endocrinol Metab 2010; 95: 3360–3367.

2) Ke Y, Bertin J, Gonthier R, Simard JN, Labrie F. A sensitive, simple and robust LC-MS/MS method for the simultaneous quantification of seven androgen- and estrogen-related steroids in postmenopausal serum. The Journal of steroid biochemistry and molecular biology. 2014;144 Pt B:523-34

3) Lissner D, Mason RS, Posen S. Stability of vitamin D metabolites in human blood serum and plasma. Clin Chem. 1981;27(5):773–4.

4) Hankinson SE, et al. Effect of transport conditions on the stability of biochemical markers in blood. Clin Chem. 1989;35(12):2313–6

5) National Institute of Standards and Technology, Standard Reference Materials. 11/3/2010]; Available from: http://www.nist.gov/srm/