

# **Dataset Integrity Check for the Diabetes Prevention Program (DPP) Bridge Period Data Files**



**Prepared by**

**RTI International**  
3040 Cornwallis Road  
Research Triangle Park, NC 27709-2194  
**September, 2012**

## Revision History

| Version | Author/Title             | Date            | Comments |
|---------|--------------------------|-----------------|----------|
| 1.0     | Norma Pugh, Statistician | September, 2012 | Original |

# Table of Contents

|   |                           |   |
|---|---------------------------|---|
| 1 | Standard Disclaimer ..... | 1 |
| 2 | Study Background .....    | 1 |
| 3 | Archived Datasets.....    | 2 |
| 4 | Statistical Methods ..... | 2 |
| 5 | Conclusions .....         | 2 |
| 6 | References .....          | 2 |

## 1 Standard Disclaimer

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected on a first (or second) exercise in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, *unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff*. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

## 2 Study Background

The Diabetes Prevention Program (DPP) was a major multicenter clinical research study aimed at discovering whether modest weight loss through dietary changes and increased physical activity or treatment with the oral diabetes drug metformin (Glucophage) could prevent or delay the onset of type 2 diabetes in study participants. At the beginning of the DPP, participants were all overweight and had prediabetes [1].

The DPP found that participants who lost a modest amount of weight through dietary changes and increased physical activity sharply reduced their chances of developing diabetes. Taking metformin also reduced risk, although less dramatically. The DPP resolved its research questions earlier than projected and, following the recommendation of an external monitoring board, the study was halted a year early. The researchers published their findings in the February 7, 2002, issue of the New England Journal of Medicine [1], [2].

Following unblinding of the Diabetes Prevention Program (DPP) results, a 16-session lifestyle intervention program (DPP Bridge) was offered to all study participants, including those who had initially

been randomized to lifestyle treatment. Venditti et al compare the effects of the lifestyle program between participants who had previous exposure and those who had not [3].

### **3 Archived Datasets**

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the DPP Bridge data folder in the “DPP\_Bridge\_Official\_Archive\DPP\_BRIDGE\_Data\Form\_based” and “DPP\_Bridge\_Official\_Archive\DPP\_BRIDGE\_Data\non\_form\_based” sub-folders of the Official Archive. 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 the DPP Bridge datasets [4].

### **4 Statistical Methods**

Analyses could not be performed to duplicate results published for these data by Venditti et al [3] in the International Journal of Obesity in October 2008. The DCC created archived datasets separately by study phase (DPP, DPP Bridge and DPPOS Phase 1). While a dataset merge could theoretically have been performed, this was not possible due to the removal of some required information during dataset de-identification prior to deposit of the datasets in the Repository. It was thus not possible for us to carry out our standard Dataset Integrity Check (DSIC) procedure.

### **5 Conclusions**

Based on our Dataset Integrity Checks of previous datasets from the DPP studies, we have every reason to believe that the DPP Bridge data files distributed by the NIDDK Repository are true copies of the study data. Requestors should proceed, however, with the understanding that these data files have not been subject to our normal Dataset Integrity Check procedure.

### **6 References**

1. National Diabetes Information Clearinghouse (NDIC) website,  
<http://diabetes.niddk.nih.gov/dm/pubs/preventionprogram/index.aspx>
2. Diabetes Prevention Program Research Group, Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *New England Journal of Medicine*, 2002 Feb 7; 346(6):393-403.

3. E.M. Venditti, et al. (2008) First versus repeat treatment with a lifestyle intervention program: attendance and weight loss outcomes. *International Journal of Obesity*; 2008 Oct; 32(10):1537-44.
  
4. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Data Coordinating Center (DCC) (2012) DPP Bridge Documentation, stored in the “DPP\_Bridge\_Official\_Archive\DPP\_BRIDGE\_Documentation” folder.