

Susan M Rogers
May 28, 2008

Data Integrity Check for the Diabetes Prevention Project (DPP) 2008 Full Scale Release

As a partial check of the integrity of the 2008 Full Scale Release of the Diabetes Prevention Program (DPP) data archived in the NIDDK data repository, a series of tabulations was performed to verify that published results from the DPP study can be reproduced using the archived datasets. Several analyses were performed to duplicate results published by the DPP Research Group in the following publication:

Impact of Intensive Lifestyle and Metformin Therapy on Cardiovascular Disease Risk Factors in the Diabetes Prevention Program (2005) *Diabetes Care* 28:888-894

The full text of the selected articles can be found in Attachment 1. STATA (v10) code for our data integrity checks is included in Attachment 2.

Our replicated analyses produced similar results to the published tabulations for most of the measurements although mean changes in some metabolic measures over time varied slightly from the published results. Some discrepancies between the published and replicated analyses arose since only study clinic sites whose IRBs approved the distribution of their data to the NIDDK repository are included in the archived data; other discrepancies may be related to differences in sample sizes between the original data file and the repository data.

The 2008 Full-Scale Data Release: Baseline and Follow-up Data

The DPP Research Group reports results for 3,665 participants in the 2008 Full Scale Data Release who were randomly assigned to one of four treatment arms aimed at preventing type 2 diabetes in high-risk adults: metformin (N=1,027), troglitazone (N=584), intensive lifestyle (N=1,024), or placebo (N=1030). Eligibility criteria for the study included: age 25 years and older, a body mass index of 24 kg/m² or higher (22+ if Asian American), a fasting plasma glucose (FPG) level of 95-125 mg/dl (≤ 125 for American Indians), and a 2-hour plasma glucose concentration of 140-199 mg/dl (see 2008 Data Release Documentation for full details). Eligible participants were identified through a multi-step screening and recruitment process. Enrollment began in 1996 and participants were followed through 2001. Follow-up assessments (quarterly, semi-annual, annual, etc) included various physical measurements, medical history updates, questionnaire administration, medication adherence, and medical testing according to a standard protocol.

In de-identifying the data, all personal identifiers including all dates were removed from the DPP archive. Variables that could possibly identify a particular individual were grouped, e.g., race/ethnicity was recoded into 4 groups (Caucasian, African American, Hispanic and Other), age at baseline was recoded in 5-year groupings with truncation of those <40 and 65+, fasting glucose levels less than 100 at baseline appear as 99, and baseline BMI is provided in two alternative groupings (2 kg/m² with truncation of those ≤ 26 kg/m² and those ≥ 42 kg/m² and tertiles of <30, 30 to <35, and ≥ 35 kg/m²). Only research data are included in the released

dataset (screening and post-randomization visits, clinic visits, lifestyle visits, laboratory data). Non-research data and adverse event data are not included.

Data Forms

The NIDDK data repository includes 37 data files – 31 files of data collected on standard forms ('form' files) and 6 files of data not collected on forms ('nonform' files) that include laboratory data, nutrition, quality of well-being, CT-scan, a summary of event variables for diabetes, and a baseline file (treatment assignment, baseline age and BMI group, sex, and race/ethnicity).

Several variables are common to all datasets and can be used to link multiple files or to match specific visits across multiple forms. These include RELEASE_ID (unique subject ID), VISIT (baseline, screening, annual visits, interim unscheduled visit, etc), and DAYSRAND (number of days between a visit and randomization). For this verification exercise, data from 4 of the 37 files were accessed.

All form files are named with the prefix DPP_REL and either a S (screening form), F (follow-up form), TR (form for participants randomized to troglitazone), Q (questionnaires), L (lifestyle form), E (events form, e.g. pregnancy), or R (report, e.g. CHD risk status) suffix. One dataset exists for each DPP form and corresponding SAS dataset. Variables and measurement times for all non-form files, e.g., LAB, Nutrient Data, CT Scan Data, Quality of Well Being Data, Baseline Data, and Events Data, are provided in the Data Release Documentation (pages 16-26).

The Diabetes Prevention Program Research Group (2005) Impact of Intensive Lifestyle and Metformin Therapy on Cardiovascular Disease Risk Factors in the Diabetes Prevention Program. *Diabetes Care* 28(4):888-894.

The purpose of this manuscript (see Attachment 1) is to assess the prevalence of cardiovascular disease (CVD) and its risk factors among DPP participants (assigned to the intensive lifestyle, metformin, or placebo treatment) and to compare the effects of the intensive lifestyle intervention and metformin therapy with that of the placebo on CVD events. Of the 3,234 DPP participants randomized to either the lifestyle, metformin, or placebo treatment groups that are included in this manuscript, the NIDDK repository includes data on 3,081. A copy of the manuscript is included in Attachment 1. STATA 10 programming code to replicate Table 2 and Figure 2 in the published manuscript is shown in Attachment 2.

A set of tabulations was performed to replicate the results published in Table 2 of the manuscript. Table 2 illustrates the mean baseline blood pressure and changes in mean blood pressure by treatment group (regardless of pharmacologic therapy) over time, i.e., change from baseline to year 1, baseline to year 2, etc. Blood pressure measurements were recorded after being seated for 5 minutes and then repeated 30 seconds later after cuff deflation; the mean of these two measurements was used for analysis. Means were calculated by experimental condition and by year of study visit.

A comparison of the published and tabulated results is shown in **TABLE D**. The average blood pressure measurement at each time period (baseline and years 1 to 3) was calculated as the average of the 5-min and 30-seconds measurements. Mean blood pressure change at year 1 was calculated as the average difference in the mean blood pressure reading at year 1 compared to the

mean baseline measurement. Similar calculations were tabulated to assess change at years 2 and 3. Estimates from the data repository are reasonably close to the published estimates for the *baseline* systolic and diastolic blood pressure measurements and several similar trends are noted from the published and replicated results. In both the published and replicated analyses, the greatest declines in mean blood pressure readings occurred for those assigned to the intensive lifestyle treatment. Participants in the metformin group experienced similar declines in years 1 and 2, with substantially less change in year 3. Participants assigned to the placebo experienced a larger difference in year 1, with similar but smaller declines in mean blood pressure in years 2 and 3.

The reduction in blood pressure across all treatment groups over the three years is observed in the replicated results, although differences in the magnitude of the change are evident between the replicated and published results. The largest difference, change in mean systolic blood pressure at year 3 in the lifestyle group (-3.27 ± 0.5 in the published ms and -2.68 ± 0.6 in the replicated results), could be due to a slightly higher mean blood pressure at baseline tabulated in the replicated results, as well as to differences in year 3 measurements. Sample Ns were not included in the published manuscript so it is not possible to compare changes in the size of the cohort by treatment over time.

An alternate estimation of change in mean blood pressure over time is shown at the bottom of TABLE D. This estimate subtracts the mean blood pressure value at each year from the mean value at baseline (see *Difference*) for each systolic and diastolic measurement. Again, while the results are similar to the published figures in Table 2 of the manuscript, the published mean change estimates are slightly higher in the placebo group and lower in the metformin group than in the replicated results.

NIDDK repository variables used to replicate Table 2.

Table 2 Variable	NIDDK variable used in replication
Treatment assignment	DPP_REL.Basedata, assign
Blood pressure, baseline	DPP_REL.S03, sosbp1, sosbp2, sosbpa, sodbp1, sodbp2, sodbpa
Blood pressure, Years 1-3	DPP_REL.F02, apsbp1, apsbp2, apdbp1, apdbp2
Visit	DPP_REL.F02, visit [BAS, Y01, Y02, Y03]

Figure 2 in the published manuscript graphs mean triglyceride levels (Fig 2A), mean HDL cholesterol levels (Fig 2B), and the percent of participants with LDL phenotype B (Fig 2C) by treatment assignment and year. The replicated results are shown in **FIGURES 1A-1C**. The raw data are not provided for Figure 2 in the manuscript, so it is not possible to precisely match our repository estimates (means \pm SE by treatment are shown in **TABLE E**) to the published data. Consequently, *p*-values are not presented in the replicated results. We note, however, that the bar graphs presented in our replicated results appear similar to the published manuscript and that results derived from the repository data are consistent with text provided in the Results section of the *Diabetes Care* article (Results, paragraph 4, p890-891). Clearly, mean triglyceride levels are highest in each of the treatment groups at baseline with substantial declines in Years 1 to 3 in the lifestyle group. Similarly, mean HDL cholesterol levels appear higher overall in the lifestyle and

metformin groups with slight increases in Year 1 returning to near baseline levels in Years 2 and 3. Contrary to the published graph (Fig 2B) the mean HDL cholesterol levels in the placebo group in the replicated results decline to levels slightly below baseline in years 2 and 3, and mean values for the metformin group are higher in year 3 than in year 2. The mean percentage of participants with LDL phenotype B appears similar in the replicated and published results. Mean percentages are slightly higher in the placebo group with the lowest levels in years 1 to 3 in the lifestyle group.

NIDDK variables used in replication of Figure 2.

Figure 2 Variable	NIDDK variable used in replication
Treatment assignment	DPP_REL.Basedata, assign
Triglycerides	DPP_REL.LAB, trig [visit==BAS,Y01,Y02, Y03]
HDL cholesterol	DPP_REL.LAB, chdl [visit==BAS,Y01,Y02, Y03]
LDL phenotype B	DPP_REL.LAB, ldlz [visit==BAS,Y01,Y02, Y03]

TABLE D. Comparison of changes in blood pressure by treatment group as reported in *Diabetes Care* 28(4):890, 2005 with tabulations calculated from the 2008 DPP Full Scale Data Release in NIDDK repository

Table 2. Blood pressure by treatment group regardless of pharmacologic therapy

Original published table ^a

Blood pressure (mmHg)	Placebo		Metformin		Intensive lifestyle	
	Systolic	Diastolic	Systolic	Diastolic	Systolic	Diastolic
Baseline (means ± SD)	123.5 ± 14.4	78.0 ± 9.2	124.0 ± 14.9	78.2 ± 9.5	123.7 ± 14.8	78.6 ± 9.2
Change at year 1	-0.90 ± 0.4	-0.89 ± 0.2	-0.91 ± 0.4	-1.26 ± 0.2	-3.4 ± 0.4	-3.6 ± 0.2
Change at year 2	-0.52 ± 0.4	-1.07 ± 0.2	-0.94 ± 0.4	-1.06 ± 0.2	-3.4 ± 0.4	-3.33 ± 0.2
Change at year 3	-0.57 ± 0.5	-1.88 ± 0.3	-0.29 ± 0.5	-1.59 ± 0.3	-3.27 ± 0.5	-3.82 ± 0.3

Data are means ± **SE** unless otherwise noted.

Calculations from NIDDK repository ^b

Blood pressure ^c (mmHg)	Placebo		Metformin		Intensive lifestyle	
	Systolic	Diastolic	Systolic	Diastolic	Systolic	Diastolic
Baseline (means ± SD)	123.8 ± 14.4	78.2 ± 9.2	124.45 ± 14.9	78.5 ± 9.5	124.15 ± 14.8	78.8 ± 9.2
Change at year 1	-0.82 ± 0.4	-0.78 ± 0.3	-1.12 ± 0.4	-1.25 ± 0.3	-3.3 ± 0.4	-3.7 ± 0.3
Change at year 2	-0.42 ± 0.4	-0.98 ± 0.3	-1.15 ± 0.4	-1.12 ± 0.3	-3.3 ± 0.4	-3.4 ± 0.3
Change at year 3	-0.43 ± 0.6	-1.84 ± 0.4	-0.69 ± 0.6	-1.56 ± 0.4	-2.68 ± 0.6	-3.77 ± 0.4

^a From: The Diabetes Prevention Program Research Group. Impact of intensive lifestyle and metformin therapy on cardiovascular disease risk factors in the Diabetes Prevention Program, *Diabetes Care* 28(4):888-894, 2005. Tabulations based on 3,234 baseline observations.

^b Calculations derived from 2008 DPP Full Scale Data Release: DPP_REL.basedata, DPP_REL.F02, DPP_REL.S03. Tabulations based on 3,081 randomized observations (584 participants randomized to the troglitazone arm of the study in the NIDDK repository were excluded from these tabulations). An additional 153 participants (15 men and 138 women) that were included in the published manuscript were excluded from the archive dataset.

^c Blood pressure calculated as the average of 2 blood pressure measurements: seated arm blood pressure after sitting 5 minutes and after waiting 30 seconds. Change in blood pressure at years 1-3 calculated as mean difference in measurement between each year and baseline (see Attachment 2, STATA output, variables chng1s-chng3s, chng1d-chng3d).

Alternate calculation assessing difference in mean readings between years 1-3 and mean at baseline

Sample sizes and mean blood pressure values based on calculations from NIDDK repository

Mean Blood pressure: systolic (mmHg)					Mean Blood pressure: diastolic (mmHg)				
	N	Mean	S.E.	Difference		N	Mean	S.E.	Difference
<u>Placebo</u>					<u>Placebo</u>				
Baseline	(1030)	123.83	14.4		Baseline	(1030)	78.2	9.2	
Year 1	(972)	123.04	14.8	-0.79	Year 1	(972)	77.37	9.5	-0.83
Year 2	(962)	123.49	14.5	-0.34	Year 2	(962)	77.23	9	-0.97
Year 3	(622)	123.27	14.1	-0.56	Year 3	(622)	76.56	8.9	-1.64
<u>Metformin</u>					<u>Metformin</u>				
Baseline	(1027)	124.45	14.9		Baseline	(1027)	78.5	9.5	
Year 1	(971)	123.4	14.6	-1.05	Year 1	(971)	77.07	9.1	-1.43
Year 2	(960)	123.17	14.4	-1.28	Year 2	(960)	77.21	9.3	-1.29
Year 3	(599)	124.17	14.3	-0.28	Year 3	(599)	77.05	9.4	-1.45
<u>Intensive lifestyle</u>					<u>Intensive lifestyle</u>				
Baseline	(1024)	124.15	14.8		Baseline	(1024)	78.8	9.2	
Year 1	(972)	120.73	15.3	-3.42	Year 1	(972)	74.95	9.2	-3.85
Year 2	(946)	120.64	14.7	-3.51	Year 2	(946)	75.2	9.1	-3.6
Year 3	(602)	120.44	14.6	-3.71	Year 3	(602)	75.04	9.5	-3.76

Difference calculated as the difference between yearly mean (Years 1, 2, and 3) and baseline mean

Note: Sample Ns (noted in parentheses) are not provided in the published table.

TABLE E. Tabulation of mean lipid levels by treatment assignment and year as reported in Diabetes Care 28(4):891, 2005 with data from the 2008 DPP Full Scale Data Release in NIDDK repository

Tabulations for Published Figure 2. Mean lipid levels by treatment assignment and year (Data shown in Figures 1a to 1c)

2A. Mean triglyceride levels

<i>Treatment assignment</i>	<i>Baseline</i>	<i>Year 1</i>	<i>Year 2</i>	<i>Year 3</i>
<i>Placebo</i>	1.88 ± 1.04	1.78 ± 1.07	1.77 ± 1.04	1.71 ± 0.93
<i>Metformin</i>	1.78 ± 1.01	1.73 ± 1.05	1.70 ± 1.00	1.75 ± 1.03
<i>Lifestyle</i>	1.83 ± 1.09	1.54 ± 0.93	1.58 ± 0.95	1.56 ± 0.92

2B. Mean HDL cholesterol levels

<i>Treatment assignment</i>	<i>Baseline</i>	<i>Year 1</i>	<i>Year 2</i>	<i>Year 3</i>
<i>Placebo</i>	1.16 ± 0.30	1.16 ± 0.29	1.14 ± 0.29	1.14 ± 0.29
<i>Metformin</i>	1.19 ± 0.30	1.21 ± 0.31	1.19 ± 0.31	1.20 ± 0.32
<i>Lifestyle</i>	1.19 ± 0.32	1.23 ± 0.33	1.21 ± 0.33	1.19 ± 0.32

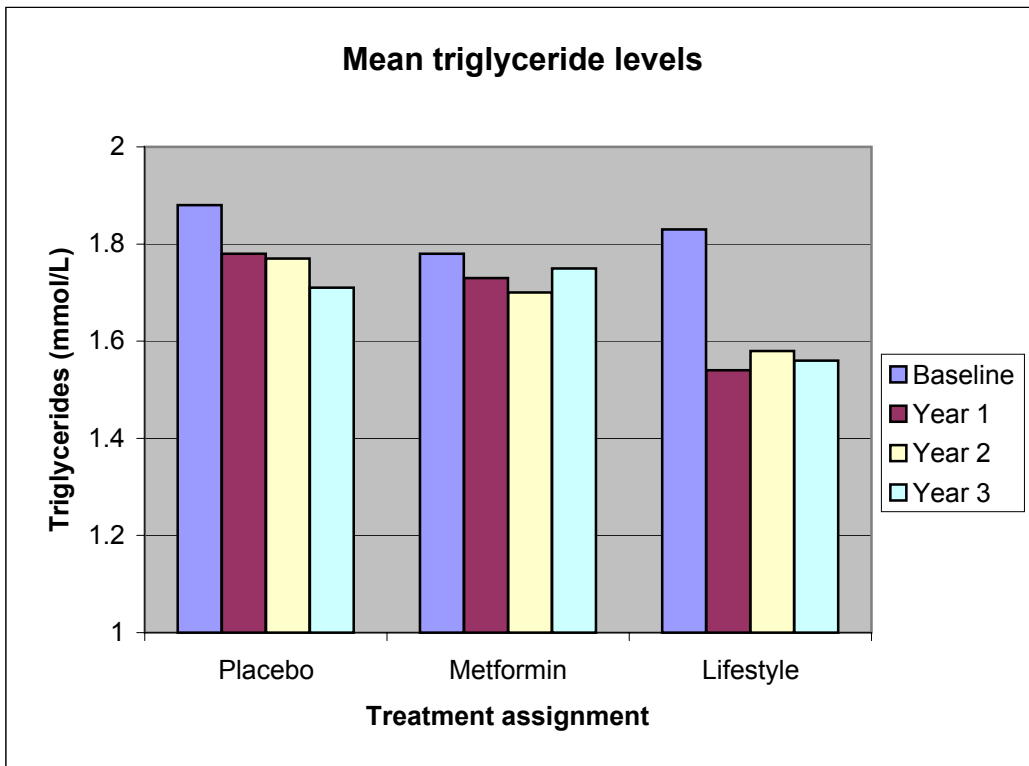
2C. Percent with LDL phenotype B

LDL phenotype B is defined as LDL particle size < .263 Rf

<i>Treatment assignment</i>	<i>Baseline</i>	<i>Year 1</i>	<i>Year 2</i>	<i>Year 3</i>
<i>Placebo</i>	34.3%	32.4%	33.8%	33.5%
<i>Metformin</i>	30.0%	30.3%	33.0%	28.4%
<i>Lifestyle</i>	29.5%	21.8%	25.1%	24.9%

Note: Calculated from 2008 DPP Full Scale Data Release, DPP_REL.Basedata, DPP_REL.LAB. The published article does not include the raw data values for Figure 2. Data are means ± SD unless otherwise noted.

FIGURE 1A. Tabulation of mean triglyceride levels by treatment assignment and year from the 2008 DPP Full Scale Data Release in NIDDK repository

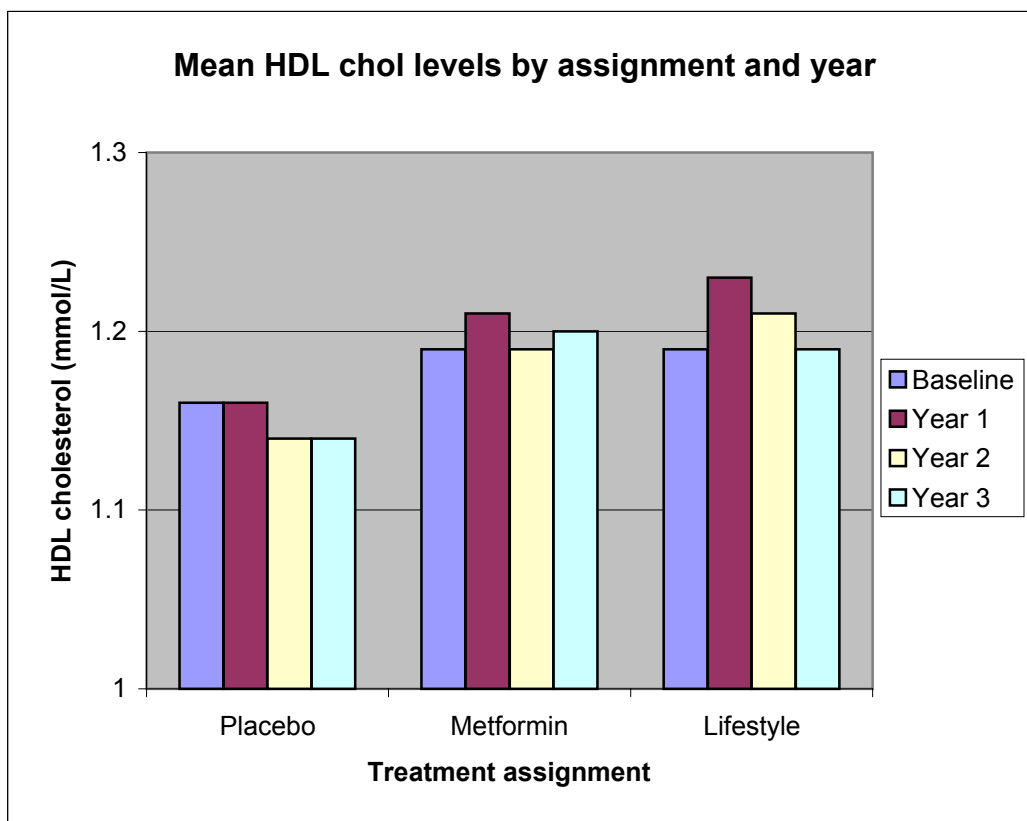


Note: Calculated from 2008 DPP Full Scale Data Release, DPP_REL.Basedata, DPP_REL.LAB. Original table published in: Diabetes Prevention Program Research Group, Impact of Intensive lifestyle and Metformin Therapy on Cardiovascular Disease Risk Factors in the Diabetes Prevention Program. *Diabetes Care* 28:888-894, 2005 (Figure 2)

DATA (means ± SD presented in TABLE E)

	Baseline	Y1	Y2	Y3
Placebo	1.88	1.78	1.77	1.71
Metformin	1.78	1.73	1.70	1.75
Lifestyle	1.83	1.54	1.58	1.56

FIGURE 1B. Tabulation of mean HDL cholesterol levels by treatment assignment and year calculated from the 2008 DPP Full Scale Data Release in NIDDK repository

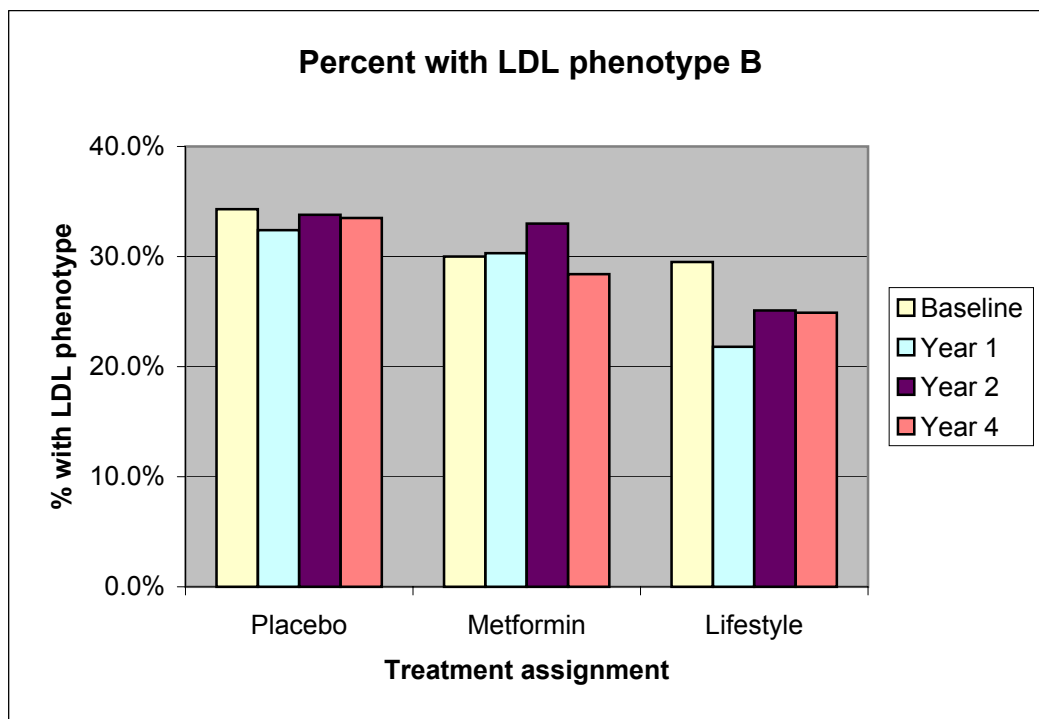


Note: Calculated from 2008 DPP Full Scale Data Release, DPP_REL.Basedata, DPP_REL.LAB. Original table published in: Diabetes Prevention Program Research Group, Impact of Intensive lifestyle and Metformin Therapy on Cardiovascular Disease Risk Factors in the Diabetes Prevention Program. *Diabetes Care* 28:888-894, 2005 (Figure 2)

DATA (means ± SD presented in TABLE E)

	Baseline	Y1	Y2	Y3
Placebo	1.16	1.16	1.14	1.14
Metformin	1.19	1.21	1.19	1.2
Lifestyle	1.19	1.23	1.21	1.19

FIGURE 1C. Tabulation of the percent of participants with LDL phenotype B by treatment assignment and year calculated from the 2008 DPP Full Scale Data Release in NIDDK repository



Note: Calculated from 2008 DPP Full Scale Data Release, DPP_REL.Basedata, DPP_REL.LAB. Original table published in: Diabetes Prevention Program Research Group, Impact of Intensive lifestyle and Metformin Therapy on Cardiovascular Disease Risk Factors in the Diabetes Prevention Program. *Diabetes Care* 28:888-894, 2005 (Figure 2)

DATA (means ± SD presented in TABLE E)

	Baseline	Y1	Y2	Y3
Placebo	34.3%	32.4%	33.8%	33.5%
Metformin	30.0%	30.3%	33.0%	28.4%
Lifestyle	29.5%	21.8%	25.1%	24.9%

ATTACHMENT 1

The full text of the article referenced will be provided to approved data requestors along with the archived data.

The Diabetes Prevention Program Research Group (2005) Impact of Intensive Lifestyle and Metformin Therapy on Cardiovascular Disease Risk Factors in the Diabetes Prevention Program. *Diabetes Care* 28(4):888-894.

NOTE. Single copies of articles published in scientific journals are included with this documentation. These articles are copyrighted, and the repository has purchased ONE reprint from their publisher to include with this documentation. If additional copies are made of these copyrighted articles, users are advised that payment is due to the copyright holder (typically the publisher of the scientific journal).

ATTACHMENT 2

**STATA/SE10 Code for Tabulations of Treatment Assignment on
Cardiovascular Disease Risk Factors from the DPP Dataset in the
NIDDK Repository**
[*Diabetes Care* 28: 888-894, 2005; Table 2, Figure 2]

```

-----
log: C:\DPP\analyses\DC2005\Attach3.log
log type: text
opened on: 15 May 2008, 09:48:56
***DPP_F02.2 for blood pressure readings at years 1-3
** take average of BP Readings 1&2 for each yearly visit
/* get file mydata\f02bp.dta list release_id visit apsbp1 apsbp2 in 1/4
SAVE FILE c:\DPP\Analyses\DC2005\F02BP_123.dta */

tab visit
tab apsbp1 if visit=="Y01"
gen sbp1_1=apsbp1 if visit=="Y01"
label var sbp1_1 "systolic BP1 Y01"
tab sbp1_1
tab apsbp2 if visit=="Y01"
gen sbp2_1=apsbp2 if visit=="Y01"
label var sbp2_1 "systolic BP2 Y01"
tab sbp2_1
gen avsbp_1=(sbp1_1+sbp2_1)/2
label var avsbp_1 "av systolic BP Y01"

tab apdbp1 if visit=="Y01"
gen dbp1_1=apdbp1 if visit=="Y01"
label var dbp1_1 "diastolic BP1 Y01"
tab dbp1_1
tab apdbp2 if visit=="Y01"
gen dbp2_1=apdbp2 if visit=="Y01"
label var dbp2_1 "diastolic BP2 Y01"
tab dbp2_1
gen avdbp_1=(dbp1_1+dbp2_1)/2
label var avdbp_1 "av diastolic BP Y01"

summarize avsbp_1
summarize avdbp_1
list release_id visit sbp1_1 sbp2_1 dbp1_1 dbp2_1 avsbp_1 avdbp_1 in 1/10

keep if visit=="Y01"
save c:\DPP\Analyses\DC2005\F02BP_1.dta

**get mydata\f02bp.dta
*Y02 visit, BP readings
tab apsbp1 if visit=="Y02"
gen sbp1_2=apsbp1 if visit=="Y02"
label var sbp1_2 "systolic BP1 Y02"
tab sbp1_2
tab apsbp2 if visit=="Y02"
gen sbp2_2=apsbp2 if visit=="Y02"
label var sbp2_2 "systolic BP2 Y02"
tab sbp2_2
gen avsbp_2=(sbp1_2+sbp2_2)/2
label var avsbp_2 "av systolic BP Y02"

tab apdbp1 if visit=="Y02"
gen dbp1_2=apdbp1 if visit=="Y02"
label var dbp1_2 "diastolic BP1 Y02"
tab dbp1_2
tab apdbp2 if visit=="Y02"
gen dbp2_2=apdbp2 if visit=="Y02"
label var dbp2_2 "diastolic BP2 Y02"
tab dbp2_2
gen avdbp_2=(dbp1_2+dbp2_2)/2
label var avdbp_2 "av diastolic BP Y02"

```

```

summarize avsbp_2
summarize avdbp_2
list release_id visit sbp1_2 sbp2_2 dbp1_2 dbp2_2 avsbp_2 avdbp_2 in 1/10
keep if visit=="Y02"
list release_id visit sbp1_2 sbp2_2 dbp1_2 dbp2_2 avsbp_2 avdbp_2 in 1/10
save c:\DPP\Analyses\DC2005\F02BP_2.dta

```

```

*get file mydata\f02bp.dta
***Y03 visit, BP readings
tab apsbp1 if visit=="Y03"
gen sbp1_3=apbsp1 if visit=="Y03"
label var sbp1_3 "systolic BP1 Y03"
tab sbp1_3
tab apsbp2 if visit=="Y03"
gen sbp2_3=apbsp2 if visit=="Y03"
label var sbp2_3 "systolic BP2 Y03"
tab sbp2_3
gen avsbp_3=(sbp1_3+sbp2_3)/2
label var avsbp_3 "av systolic BP Y03"

```

```

tab apdbp1 if visit=="Y03"
gen dbp1_3=apdbp1 if visit=="Y03"
label var dbp1_3 "diastolic BP1 Y03"
tab dbp1_3
tab apdbp2 if visit=="Y03"
gen dbp2_3=apdbp2 if visit=="Y03"
label var dbp2_3 "diastolic BP2 Y03"
tab dbp2_3
gen avdbp_3=(dbp1_3+dbp2_3)/2
label var avdbp_3 "av diastolic BP Y03"

```

```

summarize avsbp_3
summarize avdbp_3
list release_id visit sbp1_3 sbp2_3 dbp1_3 dbp2_3 avsbp_3 avdbp_3 in 1/10
keep if visit=="Y03"
list release_id visit sbp1_3 sbp2_3 dbp1_3 dbp2_3 avsbp_3 avdbp_3 in 1/10
save c:\DPP\Analyses\DC2005\F02BP_3.dta

```

```

***sort by ID and merge Y01, Y02, and Y03 files
list release_id avsbp_1 avsbp_2 avsbp_3
keep release_id sbp1_1 sbp2_1 dbp1_1 dbp2_1 sbp1_2 sbp2_2 dbp1_2 dbp2_2 sbp1_3
sbp2_3 dbp1_3 dbp2_3 avsbp_1 avdbp_1 avsbp_2 avdbp_2 avsbp_3 avdbp_3
save c:\DPP\Analyses\DC2005\F02BP_123.dta

```

```

/* merge BP measurements at Y1-Y3 with baseline data & measurements
select variables from basedata and S03 and save S03BP.dta
get file my files\baseS1S3 */

```

```

*random variable excludes Rs assigned to Troglitazone
gen random=.
replace random=1 if assign=="Lifestyle"
replace random=2 if assign=="Metformin"
replace random=3 if assign=="Placebo"
label define random 1"Lifestyle" 2"Metformin" 3"Placebo"
label val random random
keep if random >=1
list release_id visit random sosbp1 sosbp2 sodbp1 sodbp2 sosbpa sodbpa in 1/4
keep release_id visit random sosbp1 sosbp2 sodbp1 sodbp2 sosbpa sodbpa
save c:\DPP\Analyses\DC2005\S03BP.dta

```

```
*merge with F02BP_123.dta see BP files.do
save c:\DPP\Analyses\DC2005\BPfinal.dta
```

```
. use "C:\DPP\analyses\DC2005\BPfinal.dta", clear
```

```
. **TABLE 2. BP by treatment group regardless of therapy
. tab random
```

random	Freq.	Percent	Cum.
Lifestyle	1,024	33.24	33.24
Metformin	1,027	33.33	66.57
Placebo	1,030	33.43	100.00
Total	3,081	100.00	

```
. list sosbp1 sosbp2 sosbpa sbp1_1 sbp2_1 avsbp_1 sodbp1 sodbp2 sodbpa in 1/5
```

	sosbp1	sosbp2	sosbpa	sbp1_1	sbp2_1	avsbp_1	sodbp1	sodbp2	sodbpa
1.	106	100	103	96	98	97	66	70	68
2.	98	100	99	108	114	111	78	74	76
3.	150	148	149	118	116	117	80	82	81
4.	124	124	124	114	114	114	88	92	90
5.	122	124	123	134	136	135	84	86	85

```
.
. sort random
. by random: summarize sosbpa avsbp_1 avsbp_2 avsbp_3
```

```
-> random = Lifestyle
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sosbpa	1024	124.1543	14.7619	84	175
avsbp_1	972	120.7284	15.29833	87	191
avsbp_2	946	120.6438	14.72228	85	209
avsbp_3	602	120.4435	14.63804	80	181

```
-> random = Metformin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sosbpa	1027	124.4508	14.90353	89	179
avsbp_1	971	123.4032	14.60181	82.5	180
avsbp_2	960	123.1656	14.39469	90	180
avsbp_3	599	124.1728	14.32412	89.5	186

```
-> random = Placebo
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sosbpa	1030	123.832	14.39981	80	176
avsbp_1	972	123.0386	14.85265	89	187

```

    avsbp_2 |      962    123.4927    14.54642      90      185
    avsbp_3 |      622    123.2749    14.09426      91      178

```

```
. by random: summarize sodbpa avdbp_1 avdbp_2 avdbp_3
```

```
-----
-
-> random = Lifestyle
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sodbpa	1024	78.82715	9.159625	48	105
avdbp_1	972	74.95165	9.236298	50	111
avdbp_2	946	75.22146	9.109902	50	107
avdbp_3	602	75.04485	9.525147	44	105

```
-----
-
-> random = Metformin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sodbpa	1027	78.45959	9.510796	47	105
avdbp_1	971	77.06746	9.132125	43	105
avdbp_2	960	77.20885	9.352018	49	109
avdbp_3	598	77.05268	9.442242	41	114

```
-----
-
-> random = Placebo
```

Variable	Obs	Mean	Std. Dev.	Min	Max
sodbpa	1030	78.17864	9.203938	25	104
avdbp_1	972	77.37397	9.472047	49	110
avdbp_2	962	77.23233	9.046724	41	120
avdbp_3	622	76.55707	8.858089	50	111

```
. *mean change from baseline to year one
. gen chngls=.
(3356 missing values generated)
```

```
. replace chngls=avsbp_1-sosbpa
(2915 real changes made)
```

```
. mean chngls, over (random)
```

```
Mean estimation          Number of obs    =    2915
```

```

Lifestyle: random = Lifestyle
Metformin: random = Metformin
Placebo: random = Placebo

```

Over	Mean	Std. Err.	[95% Conf. Interval]	
chngls				
Lifestyle	-3.288066	.4342065	-4.139449	-2.436683
Metformin	-1.11689	.4468502	-1.993064	-.2407156
Placebo	-.8153292	.4176811	-1.634309	.0036509


```
. gen chng1d=.
(3356 missing values generated)
. replace chng1d=avdbp_1-sodbpa
(2915 real changes made)
. mean chng1d, over (random)
```

Mean estimation Number of obs = 2915

```
Lifestyle: random = Lifestyle
Metformin: random = Metformin
Placebo: random = Placebo
```

	Over	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
chng1d					
Lifestyle		-3.740741	.2900138	-4.309394	-3.172088
Metformin		-1.254892	.2941041	-1.831565	-.6782188
Placebo		-.7813786	.2860282	-1.342217	-.2205406

```
.
. gen chng2s=.
(3356 missing values generated)
. replace chng2s=avsbp_2-sosbpa
(2868 real changes made)
. mean chng2s, over (random)
```

Mean estimation Number of obs = 2868

```
Lifestyle: random = Lifestyle
Metformin: random = Metformin
Placebo: random = Placebo
```

	Over	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
chng2s					
Lifestyle		-3.27167	.4318737	-4.118484	-2.424856
Metformin		-1.152083	.4461665	-2.026923	-.2772437
Placebo		-.4189189	.4374185	-1.276605	.4387676

```
. gen chng2d=.
(3356 missing values generated)
. replace chng2d=avdbp_2-sodbpa
(2868 real changes made)
. mean chng2d, over (random)
```

Mean estimation Number of obs = 2868

```
Lifestyle: random = Lifestyle
Metformin: random = Metformin
Placebo: random = Placebo
```

	Over	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
chng2d					
Lifestyle		-3.433932	.2934095	-4.009247	-2.858617
Metformin		-1.120313	.2872632	-1.683576	-.5570492
Placebo		-.9786902	.2955802	-1.558261	-.399119

```
. gen chng3s=.
(3356 missing values generated)
. replace chng3s=avsbp_3-sosbpa
(1823 real changes made)
. mean chng3s, over (random)
```

Mean estimation Number of obs = 1823

Lifestyle: random = Lifestyle
Metformin: random = Metformin
Placebo: random = Placebo

Table with columns: Over, Mean, Std. Err., [95% Conf. Interval]. Rows: chng3s, Lifestyle, Metformin, Placebo.

```
. gen chng3d=.
(3356 missing values generated)
. replace chng3d=avdbp_3-sodbpa
(1822 real changes made)
. mean chng3d, over (random)
```

Mean estimation Number of obs = 1822

Lifestyle: random = Lifestyle
Metformin: random = Metformin
Placebo: random = Placebo

Table with columns: Over, Mean, Std. Err., [95% Conf. Interval]. Rows: chng3d, Lifestyle, Metformin, Placebo.

```
.
end of do-file
```

```
. use "C:\DPP\Data\Non-Form Data\base-lab.dta", clear
```

```
. do "C:\DOCUME~1\smr\LOCALS~1\Temp\STD06000000.tmp"
```

```
. *****Figure 2
. * use base-lab.data
```

```
. /* random variable excludes Rs assigned to Troglitazone
> gen random=.
> replace random=1 if assign=="Lifestyle"
> replace random=2 if assign=="Metformin"
> replace random=3 if assign=="Placebo"
> label define random 1"Lifestyle" 2"Metformin" 3"Placebo"
> label val random random */
```

```
. sort random
```

```

.
. /*Figure 2A. Mean triglyceride levels by treatment
> gen trigR=trig/89 */
. summarize trigR if visit=="BAS" & random >=1

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	3660	1.837515	1.089796	.3483146	14.1573

```

. by random:summarize trigR if visit=="BAS" & random >=1

```

```

-----
--> random = Lifestyle

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	1022	1.831252	1.092557	.3483146	8.94382

```

-----
--> random = Metformin

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	1027	1.780281	1.014228	.4044944	10.33708

```

-----
--> random = Placebo

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	1027	1.878844	1.0406	.4157303	9.247191

```

-----
--> random = .

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	584	1.876443	1.279494	.3595506	14.1573

```

.
. summarize trigR if visit=="Y01" & random >=1

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	3438	1.670173	1.004233	.2921348	12.68539

```

. by random: summarize trigR if visit=="Y01" & random >=1

```

```

-----
--> random = Lifestyle

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	969	1.541622	.9295092	.3146068	9.393258

```

-----
--> random = Metformin

```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	966	1.725591	1.048554	.3258427	12.68539

--> random = Placebo

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	974	1.782791	1.067591	.3483146	9.449438

--> random = .

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	529	1.597099	.8979138	.2921348	8.089888

. summarize trigR if visit=="Y02" & random >=1

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	3381	1.673503	.9862476	.2808989	11.48315

. by random: summarize trigR if visit=="Y02" & random >=1

--> random = Lifestyle

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	946	1.578165	.9507271	.3483146	7.550562

--> random = Metformin

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	956	1.699567	1.004376	.2808989	11.48315

--> random = Placebo

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	960	1.766468	1.043485	.3595506	9.898876

-> random = .

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	519	1.627308	.8866862	.3258427	7.16854

. summarize trigR if visit=="Y03" & random >=1

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	2343	1.658035	.9389407	.3483146	10.70786

. by random: summarize trigR if visit=="Y03" & random >=1

-

```
-> random = Lifestyle
```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	601	1.56105	.9174685	.3820225	8.089888

```
-> random = Metformin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	598	1.748281	1.034187	.4157303	10.1236

```
-> random = Placebo
```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	621	1.710778	.9348008	.3483146	10.70786

```
-> random = .
```

Variable	Obs	Mean	Std. Dev.	Min	Max
trigR	523	1.603669	.8368647	.3595506	5.651685

```
. /* Figure 2B. Mean HDL cholesterol by treatment  
> gen chdlR=chdl/38.67 */  
. summarize chdlR if visit=="BAS" & random >=1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	3660	1.179668	.3054864	.4913369	2.715283

```
. by random:summarize chdlR if visit=="BAS" & random >=1
```

```
-> random = Lifestyle
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	1022	1.193171	.3238833	.4913369	2.663563

```
-> random = Metformin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	1027	1.19061	.2968285	.5171968	2.663563

```
--> random = Placebo
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	1027	1.156894	.2955464	.5689165	2.715283

--> random = .

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	584	1.176844	.3029259	.5430566	2.456685

. summarize chdlR if visit=="Y01" & random >=1

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	3438	1.198067	.3114319	.4137574	2.922162

. by random:summarize chdlR if visit=="Y01" & random >=1

--> random = Lifestyle

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	969	1.226861	.3284897	.5171968	2.922162

--> random = Metformin

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	966	1.208747	.3104418	.4654771	2.715283

--> random = Placebo

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	974	1.156737	.2928049	.5430566	2.353245

--> random = .

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	529	1.20192	.3074066	.4137574	2.249806

. summarize chdlR if visit=="Y02" & random >=1

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	3381	1.183916	.3094908	.4913369	3.051461

. by random:summarize chdlR if visit=="Y02" & random >=1

--> random = Lifestyle

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	946	1.215358	.3269114	.5171968	2.792863

-

```
-> random = Metformin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	956	1.1915	.3056945	.4913369	3.051461

```
-> random = Placebo
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	960	1.144917	.2940055	.4913369	2.663563

```
--> random = .
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	519	1.184769	.3049428	.5689165	2.327386

```
. summarize chdlR if visit=="Y03" & random <=1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	601	1.194286	.3237026	.5430566	2.456685

```
. by random:summarize chdlR if visit=="Y03" & random >=1
```

```
--> random = Lifestyle
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	601	1.194286	.3237026	.5430566	2.456685

```
--> random = Metformin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	598	1.199153	.3187892	.5430566	3.051461

```
--> random = Placebo
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	621	1.144371	.2909613	.5430566	2.223946

```
--> random = .
```

Variable	Obs	Mean	Std. Dev.	Min	Max
chdlR	523	1.186141	.3053153	.5430566	3.077321

```
. /* Percent with LDL phenotype B by treatment  
> LDC phenotype B defined as LDL particle size <.263Rf */
```

```
. gen LDLpB=ldlz
(14488 missing values generated)
```

```
. replace LDLpB=1 if ldlz <.263
(5075 real changes made)
```

```
. replace LDLpB=0 if ldlz >=.263
(26740 real changes made)
```

```
. tab LDLpB
```

LDLpB	Freq.	Percent	Cum.
0	26,740	84.05	84.05
1	5,075	15.95	100.00
Total	31,815	100.00	

```
. tab ldlz if visit=="BAS" & random >=1
```

ldl particle size	Freq.	Percent	Cum.
.145	2	0.05	0.05
.158	11	0.30	0.36
.171	2	0.05	0.41
.184	32	0.87	1.28
.197	7	0.19	1.48
.211	220	6.01	7.49
.224	44	1.20	8.69
.237	749	20.47	29.16
.25	72	1.97	31.13
.263	1,062	29.02	60.15
.276	88	2.41	62.56
.289	1,047	28.61	91.17
.303	51	1.39	92.57
.316	247	6.75	99.32
.329	5	0.14	99.45
.342	17	0.46	99.92
.368	2	0.05	99.97
.474	1	0.03	100.00
Total	3,659	100.00	

```
. tab LDLpB if visit=="BAS" & random <=1
```

LDLpB	Freq.	Percent	Cum.
0	722	70.51	70.51
1	302	29.49	100.00
Total	1,024	100.00	

```
. by random: tab LDLpB if visit=="BAS" & random >=1
```

```
-----
-> random = Lifestyle
```

LDLpB	Freq.	Percent	Cum.
-------	-------	---------	------

	Freq.	Percent	Cum.
0	722	70.51	70.51
1	302	29.49	100.00
Total	1,024	100.00	

-> random = Metformin

LDLpB	Freq.	Percent	Cum.
0	719	70.01	70.01
1	308	29.99	100.00
Total	1,027	100.00	

--> random = Placebo

LDLpB	Freq.	Percent	Cum.
0	677	65.73	65.73
1	353	34.27	100.00
Total	1,030	100.00	

--> random = .

LDLpB	Freq.	Percent	Cum.
0	408	69.86	69.86
1	176	30.14	100.00
Total	584	100.00	

. tab LDLpB if visit=="Y01" & random >=1

LDLpB	Freq.	Percent	Cum.
0	2,498	72.41	72.41
1	952	27.59	100.00
Total	3,450	100.00	

. by random: tab LDLpB if visit=="Y01" & random >=1

--> random = Lifestyle

LDLpB	Freq.	Percent	Cum.
0	759	78.17	78.17
1	212	21.83	100.00
Total	971	100.00	

--> random = Metformin

LDLpB	Freq.	Percent	Cum.
-------	-------	---------	------

	Freq.	Percent	Cum.
0	676	69.69	69.69
1	294	30.31	100.00
Total	970	100.00	

--> random = Placebo

LDLpB	Freq.	Percent	Cum.
0	660	67.62	67.62
1	316	32.38	100.00
Total	976	100.00	

--> random = .

LDLpB	Freq.	Percent	Cum.
0	403	75.61	75.61
1	130	24.39	100.00
Total	533	100.00	

. tab LDLpB if visit=="Y02" & random >=1

LDLpB	Freq.	Percent	Cum.
0	2,340	69.11	69.11
1	1,046	30.89	100.00
Total	3,386	100.00	

. by random: tab LDLpB if visit=="Y02" & random >=1

--> random = Lifestyle

LDLpB	Freq.	Percent	Cum.
0	709	74.95	74.95
1	237	25.05	100.00
Total	946	100.00	

--> random = Metformin

LDLpB	Freq.	Percent	Cum.
0	641	66.98	66.98
1	316	33.02	100.00
Total	957	100.00	

--> random = Placebo

LDLpB	Freq.	Percent	Cum.
-------	-------	---------	------

	Freq.	Percent	Cum.
0	637	66.22	66.22
1	325	33.78	100.00
Total	962	100.00	

--> random = .

LDLpB	Freq.	Percent	Cum.
0	353	67.75	67.75
1	168	32.25	100.00
Total	521	100.00	

. tab LDLpB if visit=="Y03" & random >=1

LDLpB	Freq.	Percent	Cum.
0	1,678	71.56	71.56
1	667	28.44	100.00
Total	2,345	100.00	

. by random: tab LDLpB if visit=="Y03" & random >=1

--> random = Lifestyle

LDLpB	Freq.	Percent	Cum.
0	452	75.08	75.08
1	150	24.92	100.00
Total	602	100.00	

--> random = Metformin

LDLpB	Freq.	Percent	Cum.
0	429	71.62	71.62
1	170	28.38	100.00
Total	599	100.00	

--> random = Placebo

LDLpB	Freq.	Percent	Cum.
0	413	66.51	66.51
1	208	33.49	100.00
Total	621	100.00	

--> random = .

LDLpB	Freq.	Percent	Cum.
-------	-------	---------	------

0	384	73.42	73.42
1	139	26.58	100.00

Total	523	100.00	

log close

log: C:\DPP\analyses\DC2005\Attach3.log

log type: text

closed on: 15 May 2008, 09:54:31
