

## **Integrity Check for Look AHEAD (Action for Health in Diabetes): Baseline Data**

As a partial check of the integrity of the Look AHEAD Baseline dataset archived in the NIDDK data repository, a set of tabulations was performed to verify that published results can be reproduced using the archived dataset. Analyses were performed to duplicate selected results for the baseline data published by the Look AHEAD Research Group [1] in *Diabetes and Vascular Disease Research* in 2006 (Attachment 1). The results of this dataset integrity check (DSIC) are described below. The Stata code for our tabulations is included in Attachment 2.

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.

**Background.** The Look AHEAD study is a multicenter randomized clinical trial in overweight and obese individuals with type 2 diabetes to evaluate the long-term effects of an intensive weight loss intervention on major cardiovascular events. Eligibility requirements are diagnosis of type 2 diabetes (determined by self-report and verification) in individuals age 45–76 years, BMI >25 kg/m<sup>2</sup> (>27 kg/m<sup>2</sup> if currently taking insulin). A total of 5,145 participants from 16 clinical centers were randomized between 2001 and 2004. The published manuscript by the Look AHEAD Research Group provides the baseline characteristics of this randomized cohort.

Users of the Look AHEAD Baseline data should note that data on 239 participants from American Indian clinical sites were not included in the data transferred to the NIDDK repository.

**Archived Datasets.** The DCC submitted 32 ‘forms’ datasets which correspond with study data collection forms, 4 ‘extra datasets’ that provide information on baseline data and other common measures, diet and the DEXA scan substudy, and a file identifying the participant randomization assignment. For this DSIC, tabulations were completed using the ‘additional’ composite data file, *baseline\_combined.sas7bdat* and the participant randomization file, *treatment.sas7bdat*; SAS datasets were converted to Stata format using StatTransfer (Circle Systems).

**DSIC Analysis.** Our analyses focus on Tables 1-3 and Figure 5. Table 1 provides the overall distribution of baseline characteristics of the Look AHEAD participants by treatment group assignment. Table 2 shows the distribution of demographic and socio-economic characteristics of the Look AHEAD participants by gender. Clinical characteristics of the male Look AHEAD participants by racial/ethnic group are shown in

Table 3 and baseline medication use by education is shown in Figure 5. In the tables below, variable names, or *Var. Name*, refer to variables in the dataset, *baseline\_combined*.

Means, standard deviations, and frequencies calculated from archived data were compared to published results. All tabulations for this DSIC were conducted in Stata v12.

**DSIC Results.** Distributions of baseline demographic and clinical characteristics closely matched published statistics (Tables 1-3). Sample Ns did not match in all analyses due to the exclusion of participants from American Indian sites in the repository dataset. Small differences in percentages and means could be attributable to rounding or restriction of cases in the published analyses.

**Table 1. Participant Characteristics by Randomized Intervention Assignment: Published and DSIC Results**

[Published results extracted from Table 1 in LookAHEAD Research Group, *Diab Vasc Dis Res* (2006) 3(3)202-215]

Variable	Var. Name	Look AHEAD Research Group		DSIC	
		DSE	Lifestyle	DSE	Lifestyle
<b>Number</b>		<b>2575</b>	<b>2570</b>	<b>2455</b>	<b>2451</b>
<b>Age (yrs)</b>	Age	58.85±6.86	58.55±6.77	59.1±6.8	58.7±6.7
<b>Sex (%F)</b>	Female	1534 (59.6%)	1524 (59.3%)	1438(58.6%)	1436 (58.6%)
<b>Race/ethnicity</b>	Racevar				
<b>AA/Black</b>		405 (15.8%)	398 (15.5%)	404(16.5%)	400 (16.3%)
<b>Amer Ind/Alas</b>		<b>129 (5.0%)</b>	<b>131 (5.1%)</b>		
<b>Asian/PI</b>		21 (0.8%)	29 (1.1%)		
<b>White</b>		1628 (63.4%)	1618 (63.1%)	1631 (66.4%)	1621 (66.1%)
<b>Hispanic</b>		337 (13.1%)	338 (13.2%)	338 (13.8%)	338 (13.8%)
<b>Other</b>		49 (1.9%)	49 (1.9%)	82(3.3%)	92(3.8%)
<b>Weight (kg)</b>	Baselinewgt_kg	100.86 ±18.83	100.54 ±19.65	101.2±18.8	100.9±19.6
<b>Height (cm)</b>	Eshgt_mean	167.27±9.86	167.22±9.59	167.6±9.8	167.5±9.5
<b>BMI (kg/m<sup>2</sup>)</b>	Bmi	36.0±5.76	35.89±6.01	36±5.7	35.9±6
<b>Waist circ (cm)</b>	Waistcm_mean	114.06±13.55	113.8±14.35	114.1±13.6	113.8±14.4
<b>Ankle/brac Index</b>	Minabi	1.07±0.26	1.07±0.26	1.1±1	1.12±0.1
<b>HbA1c</b>	Hba1cpct	7.31±1.20	7.25±1.14	7.3±1.2	7.2±1.1
<b>FPG (mg/dL)</b>	Glucosemgdl	154.04±46.5	152.19±44.71	154±46.0	152.4±45
<b>Plasma creat</b>	Screatmgdl	0.82±0.2	0.82±0.2	0.8±0.2	0.8±0.2
<b>Tot chol (mg/dL)</b>	Cholmgdl	190.65±37.04	191.33±38.16	191.0±37	191.5±38
<b>LDL chol (mg/dL)</b>	Ldlchlmgdl	112.23±32.3	112.36±32.21	112.7±32.2	112.5±32.3
<b>HDL chol (mg/dL)</b>	Hdlchlmgdl	43.47±11.81	43.38±11.83	43.5±11.9	43.4±11.8
<b>Triglyc (mg/dL)</b>	Trigmhdl	179.94±117.92	181.96±117.69	179.2±115.7	182.3±115.2
<b>Systol BP (mmHg)</b>	Bssbp_mean	129.45±17.09	128.19±17.26	129.8±17.1	128.4±17.2
<b>Diast BP (mmHg)</b>	Bsdbp_mean	70.4±9.72	69.93±9.55	70.5±9.6	70±9.5
<b>Dur diabetes(yr)</b>	Diab_dur	6.82±6.42	6.77±6.66	6.7±6.3	6.7±6.5
<b>Fitness</b>					
<b>Max MET value</b>	Maxexmets	7.18±1.99	7.21±1.05	7.2±2	7.2±1.9
<b>MaxMET (80%)</b>	Metsbase1	5.10±1.54	5.17±1.53	Missing	Missing
<b>Metab Syndr</b>	Metabolic_syndrome	2415 (93.8%)	2387 (92.9%)	2307 93.97%	2285 93.23%
<b>Fam hx diabetes</b>	Diabetes_fam	1723 (66.9%)	1622 (63.1%)	1628 (66.3%)	1532 (62.5%)
<b>History CVD</b>	Cvdhis	353 (13.7%)	373 (14.5%)	335 (13.6%)	356 (14.5%)
<b>Hypertension</b>	Hypertension	2069 (80.3%)	2063 (80.3%)	2053 83.63%	2061 84.09%

Note: DSE, Diabetes Support and Education; Lifestyle, Lifestyle Intervention

**Table 2. Demographic and Socioeconomic Characteristics by Sex: Published and DSIC Results**  
*[Published results extracted from Table 2 in Look AHEAD Research Group, Diab Vasc Dis Res (2006) 3(3)202-215]*

Variable	Var. Name	Look AHEAD Research Group		DSIC	
		Male 2087	Female 3058	Male 2032	Female 2874
<b>Age (yrs)</b>	Age_t2				
<b>45-55</b>		527 (25.3%)	1093 (35.8%)	497 (24.5%)	986 (34.3%)
<b>56-65</b>		1116 (53.5%)	1534 (50.2%)	1098 (54%)	1477 (51.4%)
<b>66-75</b>		444 (21.3%)	430 (14.1%)	437 (21.5%)	411 (14.3%)
<b>Race/ethnicity</b>	Racevar				
<b>AA</b>		190 (9.1%)	614 (20.1%)	189(9.3%)	615(21.4%)
<b>Amer Ind/Alas</b>		52 (2.6%)	204 (6.7%)		
<b>Asian/PI</b>		16 (0.8%)	34 (1.1%)		
<b>White</b>		1580(76%)	1664 (54.5%)	1584(78%)	1668(58%)
<b>Hispanic</b>		196 (9.4%)	480 (15.7%)	196(9.6%)	480(16.7%)
<b>Other</b>		42 (2%)	57 (1.9%)	63(3.1%)	111(3.9%)
<b>Employ Stat</b>	Emp_status				
<b>Full/PT</b>		1395 (87%)	1837 (65.4%)	1363 (67.1%)	1730 (60.2%)
<b>Homemaker</b>		137 (7.7%)	772 (27.5%)	130 (6.4%)	727 (25.3%)
<b>Student</b>		4 (0.2%)	16 (0.6%)	3 (0.2%)	15 (0.5%)
<b>Not employed</b>		196 (11.3%)	184 (6.6%)	190 (9.4%)	170 (5.9%)
<b>Education (yrs)</b>	Educ_yrs				
<b>&lt;13</b>		261 (12.7)	763 (25.6%)	243 (12%)	687 (23.9%)
<b>13-16</b>		695 (33.9%)	1216 (40.8%)	669 (32.9%)	1133 (39.4%)
<b>&gt;16</b>		1094 (53.4%)	999 (33.5%)	1089 (53.6%)	985 (34.3%)
<b>Alcohol past yr</b>	Aldrksyr	1470 (70.4%)	1599 (52.3%)	1460 (71.9%)	1574 (54.8%)
<b>Family Income</b>	Inc_t2				
<b>&lt;\$20,000</b>		137 (7.2%)	452 (16.6%)	122 (6%)	384 (13.4%)
<b>\$20-\$40,000</b>		259 (13.6%)	725 (26.6%)	250 (12.3%)	680 (23.7%)
<b>\$40-\$60,000</b>		346 (18.1%)	607 (23.3%)	333 (16.4%)	576 (20%)
<b>\$60-\$80,000</b>		339 (17.8%)	412 (15.1%)	329 (16.2%)	404 (14.1%)
<b>&gt;\$80,000</b>		827 (43.3%)	532 (19.5%)	823 (40.5%)	523 (18.2%)
<b>Smoking status</b>	Tqsmk100				
<b>Never</b>	Tqsmkevr	781 (37.6%)	1794 (58.8%)	761 (37.5%)	1667 (58%)
<b>Past</b>		1201 (57.8%)	1126 (36.9%)	1176 (57.9%)	1093 (38%)
<b>Present</b>		96 (4.6%)	132 (4.3%)	90 (4.4%)	118 (4.1%)
<b>Marital status</b>	Sdmarstat				
<b>Never Married</b>		118 (5.7%)	268 (8.8%)	112 (5.5%)	249 (8.7%)
<b>Married</b>		1731 (83.1%)	1729 (56.6%)	1698 (83.6%)	1637 (57%)
<b>Widowed</b>		36 (1.7%)	323 (10.6%)	33 (1.6%)	294 (10.2%)
<b>Divor/sep</b>		199 (9.5%)	735 (24.1%)	199 (9.5%)	693 (24.1%)

*[DSIC results tabulated from baseline\_combined.dta using Stata v12]*

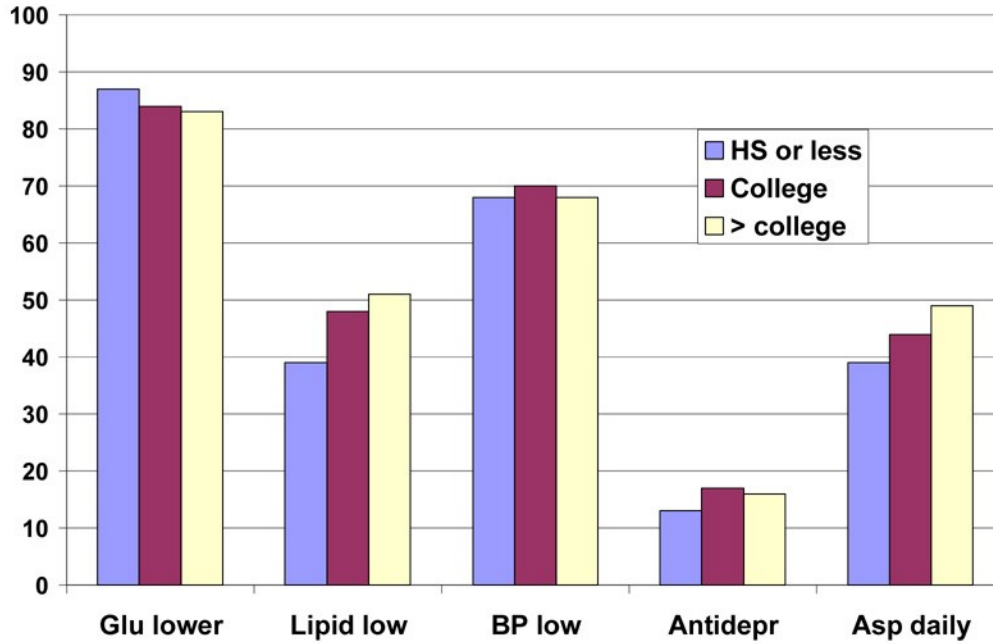
**Table 3. Clinical Characteristics in Male Subjects by Race/Ethnicity: Published and DSIC Results**  
*[Published results extracted from Table 3 in Look AHEAD Research Group, Diab Vasc Dis Res (2006) 3(3)202-215]*  
**[Note: DSIC results tabulated from baseline\_combined.dta using Stata v12; tabulations limited to 'All Subjects', Caucasian, and Hispanic; Ns (%) or Mean (±SD)]**

		Look AHEAD Research Group			DSIC		
Variable	Var. Name	All	Caucasian	Hispanic	All	Caucasian	Hispanic
<b>Number</b>		<b>2081</b>	<b>1581</b>	<b>196</b>	<b>2032</b>	<b>1584</b>	<b>196</b>
Age (yrs)	Age_t2						
<b>45-55</b>		524 (25.2)	371 (23.5)	58 (29.6)	497(24.5)	369(23.3)	59(30.1)
<b>56-65</b>		1115(53.6)	859 (54.3)	109 (55.6)	1098(54)	862(54.4)	108(55.1)
<b>66-75</b>		442(21.2)	351 (22.2)	29 (14.8)	437(21.5)	353(22.3)	29(14.8)
BMI(kg/m2)	Bmigrp						
<b>25-27</b>		46 (2.2)	32 (2)	3 (1.5)	43 (2.1)	32(2)	3(1.5)
<b>27-30</b>		294 (14.1)	216 (13.7)	29 (14.8)	285 (14)	215(13.6)	29(14.8)
<b>30-35</b>		833 (40)	623 (39.4)	92 (46.9)	813(40)	630(39.8)	92(46.9)
<b>35-40</b>		535 (25.7)	417 (26.4)	41 (20.9)	523(25.7)	415(26.2)	40(20.4)
<b>≥40</b>		372 (17.9)	292 (18.5)	31 (15.8)	368(18.1)	292(18.4)	32(16.3)
<b>BP (mmHg)</b>							
<b>Systolic</b>	Bssbp_mean	128.5 ±16.6	128.7±16.7	128.37±15.51	128.6±16.6	128.7±16.8	128.4±15.5
<b>Diastolic</b>	Bsdbp_mean	73.21±9.17	72.65±9.02	74.25±9.09	73.2±9.2	72.7±9.0	74.3±9.1
<b>Antihyp Medication</b>	Any_Htndrug	1450 (69.7)	1121(70.9)	127 (64.8)	1515 (74.6)	1181(74.6)	140(71.4)
FPG(mg/dL)	Glucosemgdl	156.07±46.1	157.7±45.65	156.7±46.13	156.3±46.3	157.5±45.8	158.0±47.3
<b>HbA1c</b>	Hba1cpct						
<b>&lt;6</b>		170 (8.4)	129(8.4)	19 (9.8)	171(8.4)	133(8.4)	20(10.2)
<b>6.0-6.5</b>		344 (17)	265(17.3)	30 (15.5)	431(21.2)	340(21.5)	35(17.9)
<b>6.5&lt;-7.0</b>		410 (20.3)	337(22)	25 (13)	411(20.2)	349(22)	26(13.3)
<b>7.0&lt;-8.0</b>		619 (30.6))	468(30.5)	64 (33.2)	591(29.1)	451(28.5)	61(31.2)
<b>8.0&lt;-9.0</b>		317 (15.7)	233(15.2)	33 (17.1)	269(13.2)	205(12.9)	30(15.3)
<b>9.0+</b>		161 (8)	103(6.7)	22 (11.4)	159(7.8)	106(6.7)	24(12.2)
<b>Plasma Lipids(mg/dL)</b>							
<b>Total chol</b>	Cholmgdl	182.2±36.2	182.02±36.23	185.9 ±36.58	182.4±36.1	181.6±36.0	186±36.4
<b>HDL</b>	Hdlchlmgdl	38.05±9.09	37.64±8.79	37.09±8.57	38±9.1	37.7±8.8	36.8±8.5
<b>LDL</b>	Ldlchlmgdl	107.07±30.61	106.12±30.12	109.37±33.45	106.9±30.5	106±30	109.4±33.3
<b>Triglyc</b>	Trigmgdl	192.01±129.75	198.2±129.64	203.92±138.28	192±127	196.8±124.7	205.7±137.9
<b>Lipid-lower medication</b>	Any_Lipidrug	1148 (53.2)	935 (59.1)	79 (40.3)	1217(59.9)	999(63.1)	92(46.9)
<b>CVD Hx</b>	cvdhis	442 (21.2)	362 (22.5)	29 (14.8)	430(21.2)	360(22.7)	29(14.8)

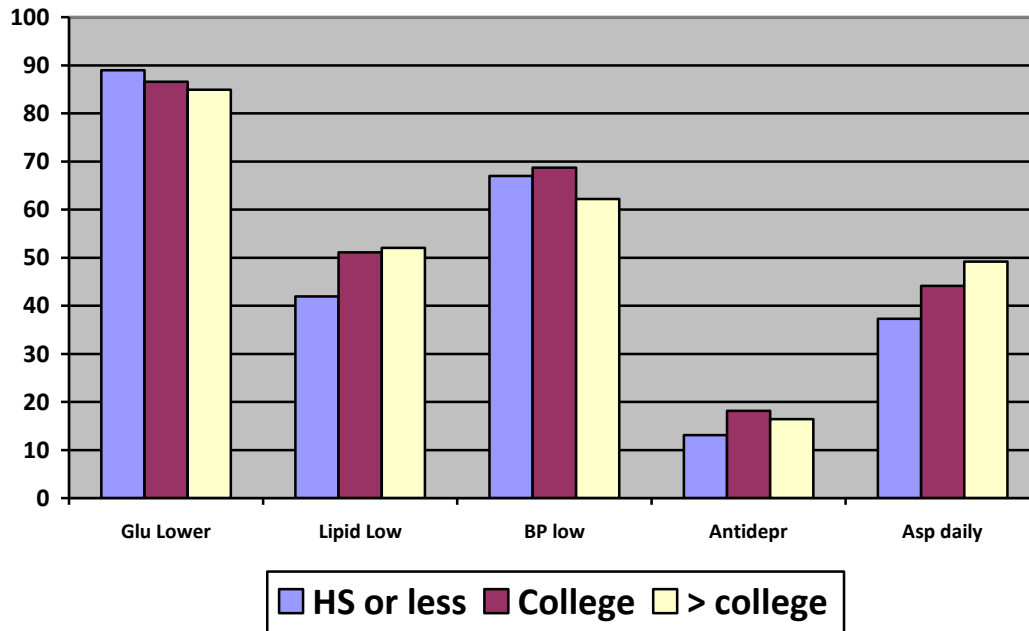
Figure 5 in the publication tabulates baseline medication use (glucose lowering, lipid lowering, lowering of blood pressure, antidepressants, and daily aspirin use) by education group in the Look AHEAD baseline study. Estimates of medication use are expressed as percentages of each education group. Medication variables used for this analysis include *diabdrug*, *lipidrug*, *mihighbp*, *antidep*, and *aspirin*. DSIC tabulations were similar to published results.

**Figure 5. Medications by Education: Published and DSIC Results**  
[Published results extracted from Figure 5 in *Diab Vasc Dis Res* (2006) 3(3)202-215]

### Medications by education



### Medications by Education: DSIC Results



Susan M Rogers  
(revised) April 2014

### References

[1] The Look AHEAD Research Group. Baseline characteristics of the randomized cohort from the Look AHEAD (Action for Health in Diabetes) Research Study (2006) *Diab Vasc Dis Res* 3(3) 202-215, doi: 10.3132/dvdr.2006.031

### Attachments

[1] Article text: The Look AHEAD Research Group. Baseline characteristics of the randomized cohort from the Look AHEAD (Action for Health in Diabetes) Research Study (2006) *Diab Vasc Dis Res* 3(3) 202-215, doi: 10.3132/dvdr.2006.031

[2] Stata code to generate DSIC analyses and SAS code (provided by the DCC) to generate race/ethnicity variable, racevar.

# **ATTACHMENT 1**

## **Full Text of Article**

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 Code for DSIC Analyses

```
log using lookahead_tables
```

```
use "\Look Ahead\stata\baseline_combined.dta", clear
sort p_id
merge m:1 p_id using "\Look Ahead\stata\treatment.dta"
save "\Look Ahead\stata\dsic_file1.dta"
```

```
/* Table 1. Participant characteristics by randomized intervention assignment */
sort randarm
```

```
by randarm : tab1 female racevar metabolic_syndrome diabetes_fam cvdhis hypertension
by randarm : summ age baselinewgt_kg eshgt_mean bmi waistcm_mean minabi hba1cpct glucosemgdl
screatmgdl cholmgdl ldlchlmgdl hdlchlmgdl trigmgdl ///
      bssbp_mean bsdbp_mean diab_dur maxexmets metsbase1
```

```
/* Table 2. Demographic and SE characteristics by sex. */
codebook female
```

```
/*Table 3. clinical characteristics in male subjects by race/ethnicity */
codebook racevar
sort racevar
```

```
recode hba1cpct (4/5.9=1 lessthan6) (6/6.5=2 6-6.5) (6.51/7=3 6.5-7) (7.1/8=4 7-8) (8.1/9=5 8-9)
(9.1/14.5=6 >9), gen(hba1c_r)
*egen hba1c_cat = cut(hbalcpct), at (6, 6.5, 7, 8, 9, 15)
```

```
tab1 hba1c_r hba1cpct
by racevar: tab1 age_t2 bmigrp hba1c_r any_htndrug any_lipidrug cvdhis if racevar!="Missing" &
female=="No"
summ bssbp_mean bsdbp_mean glucosemgdl cholmgdl hdlchlmgdl ldlchlmgdl trigmgdl if female=="No"
summ bssbp_mean bsdbp_mean glucosemgdl cholmgdl hdlchlmgdl ldlchlmgdl trigmgdl if
racevar=="White" & female=="No"
*summ bssbp_mean bsdbp_mean glucosemgdl cholmgdl hdlchlmgdl ldlchlmgdl trigmgdl if
racevar=="African American / Black (not Hispanic)" & female=="No"
summ bssbp_mean bsdbp_mean glucosemgdl cholmgdl hdlchlmgdl ldlchlmgdl trigmgdl if
racevar=="Hispanic" & female=="No"
*summ bssbp_mean bsdbp_mean glucosemgdl cholmgdl hdlchlmgdl ldlchlmgdl trigmgdl if
racevar=="American Indian / Native American / Ala" & female=="No"
*summ bssbp_mean bsdbp_mean glucosemgdl cholmgdl hdlchlmgdl ldlchlmgdl trigmgdl if
racevar=="Asian/Pacific Islander" & female=="No"
*summ bssbp_mean bsdbp_mean glucosemgdl cholmgdl hdlchlmgdl ldlchlmgdl trigmgdl if
racevar=="Other/Mixed" & female=="No"
```



```
/* Table 7. Baseline medications by education group
codebook educ_yrs
replace educ_yrs="" if educ_yrs=="Missing"
tab educ_yrs, missing

by educ_yrs, sort: tab1 migluclower insulin tzd sulfonylurea biguanides statins fibrates niacin htndrug
ace arb ccblock bb diuretics ///
    antidep aspirin if educ_yrs!="" */

/* Figure 5. Medications by education */
encode aspirin, gen(aspdaily)
codebook aspdaily
recode aspdaily (2=2) (3 4=3)
mvdecode aspdaily, mv(1=.)
tab aspdaily
replace educ_yrs="" if educ_yrs=="Missing"
tab educ_yrs, missing

    /*convert string to numeric, identify missing values*/
    codebook diabdrug lipidrug mihighbp
    encode diabdrug, gen(gluclow)
    encode lipidrug, gen(lipid)
    encode mihighbp, gen(bplow)
    mvdecode gluclow lipid, mv(1=.)
    mvdecode bplow, mv(1 2=.)
    tab1 gluclow lipid bplow

by educ_yrs, sort: tab1 gluclow lipid bplow antidep aspdaily
```

## **SAS Code for constructing race/ethnicity variable, provided by DCC**

### **Race/Ethnicity Data**

Provide the “race/ethnicity data” to RTI.

SAS code for identifying ethnic groups

- 1) 3 variables involving race and ethnicity are formed. Hispanic is just yes/no to Hispanic ethnicity. Race\_only is race independent of ethnicity. Racevar combines race and ethnicity.

```
/* Define ethnicity variable */
if race then do; /* if they have the year 1 race form */
    if hisp = 1 then hispanic = 1;
    else hispanic = 0;
```

```
end;
if not race then do; /* if they do not have the year 1 race form */
  if pshisp = 1 or psethnic = 5 then hispanic = 1;
  else hispanic = 0;
end;

/* Define race variable that doesn't consider ethnicity */
if race then do; /* if they have the year 1 race form */
  if race1 = 1 and (race2 ne 1 and race3 ne 1 and race4 ne 1 and race5 ne 1
and race6 ne 1) then race_only = 1; /* Black */
  else if race2 = 1 and (race1 ne 1 and race3 ne 1 and race4 ne 1 and
race5 ne 1 and race6 ne 1) then race_only = 2; /* Amer Ind*/
  else if (race3 = 1 or race6 = 1) and (race1 ne 1 and race2 ne 1 and
race4 ne 1 and race5 ne 1) then race_only = 3; /* Asian */
  else if race4 = 1 and (race1 ne 1 and race2 ne 1 and race3 ne 1 and
race5 ne 1 and race6 ne 1) then race_only = 4; /* White */
  else if (race1 = 1 or race2 = 1 or race3 = 1 or race4 = 1 or race5 = 1
or race6 = 1) then race_only = 10; /* Mixed */
end;
/* Next check for prescreen for new race variables */
if race_only = . then do;
  if psrace1 = 1 and (psrace2 ne 1 and psrace3 ne 1 and psrace4 ne 1 and
psrace5 ne 1) then race_only = 1; /* Black */
  else if psrace2 = 1 and (psrace1 ne 1 and psrace3 ne 1 and psrace4 ne 1
and psrace5 ne 1) then race_only = 2; /* Amer Ind*/
  else if psrace3 = 1 and (psrace1 ne 1 and psrace2 ne 1 and psrace4 ne 1
and psrace5 ne 1) then race_only = 3; /* Asian */
  else if psrace4 = 1 and (psrace1 ne 1 and psrace2 ne 1 and psrace3 ne 1
and psrace5 ne 1) then race_only = 4; /* White */
  else if (psrace1 = 1 or psrace2 = 1 or psrace3 = 1 or psrace4 = 1 or
psrace5 = 1) then race_only = 10; /* Mixed */
end;
/* Next check for prescreen old race variable psethnic */
if race_only = . and psethnic ne . then do;
  race_only = psethnic;
  if race_only not in (1,2,3,4) then race_only = 10;
end;
if race_only = . then race_only = 10;

/* Define Race Variable */
/* First check for Year 1 Race/Ethnicity Form--this is the most accurate if
available */
if race then do; /* if they have the year 1 race form */
  race_source = 'Year 1 Race/Ethnicity Form';
  if hisp = 1 then racevar = 9; /* Hispanic */
  else if race1 = 1 and (race2 ne 1 and race3 ne 1 and race4 ne 1 and
race5 ne 1 and race6 ne 1) then racevar = 1; /* Black */
  else if race2 = 1 and (race1 ne 1 and race3 ne 1 and race4 ne 1 and
race5 ne 1 and race6 ne 1) then racevar = 2; /* Amer Ind*/
  else if (race3 = 1 or race6 = 1) and (race1 ne 1 and race2 ne 1 and
race4 ne 1 and race5 ne 1) then racevar = 3; /* Asian */
  else if race4 = 1 and (race1 ne 1 and race2 ne 1 and race3 ne 1 and
race5 ne 1 and race6 ne 1) then racevar = 4; /* White */
  else if (race1 = 1 or race2 = 1 or race3 = 1 or race4 = 1 or race5 = 1
or race6 = 1) then racevar = 10; /* Mixed */
end;
```

Susan M Rogers  
(revised) April 2014

```
/* Next check for prescreen for new race variables.  Checkboxes where they
can select more than one choice */
if racevar = . then do;
  race_source = 'Prescreen Form (not original)';
  if pshisp = 1 then racevar = 9; /* Hispanic */
  else if psmrace1 = 1 and (psrace2 ne 1 and psmrace3 ne 1 and psmrace4 ne 1
and psmrace5 ne 1) then racevar = 1; /* Black */
  else if psmrace2 = 1 and (psrace1 ne 1 and psmrace3 ne 1 and psmrace4 ne 1
and psmrace5 ne 1) then racevar = 2; /* Amer Ind*/
  else if psmrace3 = 1 and (psrace1 ne 1 and psmrace2 ne 1 and psmrace4 ne 1
and psmrace5 ne 1) then racevar = 3; /* Asian */
  else if psmrace4 = 1 and (psrace1 ne 1 and psmrace2 ne 1 and psmrace3 ne 1
and psmrace5 ne 1) then racevar = 4; /* White */
  else if (psrace1 = 1 or psmrace2 = 1 or psmrace3 = 1 or psmrace4 = 1 or
psrace5 = 1) then racevar = 10; /* Mixed */
end;
/* Next check for prescreen old race variable psethnic -- this is the
original form that was a dropdown and they could only select one--this is a
last resort*/
if racevar = . and psethnic ne . then do;
  racevar = psethnic;
  if racevar = 5 then racevar = 9;
  race_source = 'Prescreen (Original)';
end;

/*****
*****
***Program: /prj/niddk/ims_analysis/Look_AHEAD/prog_initial_analysis/Look_AHEAD_dsic.sas;
***Programmer: Michael Spriggs
***Date Created: 03/13/2014
*** Must be run with the option -V93
*****
*****/;

title1 "%sysfunc(getoption(sysin))";
title2 " ";

options nofmterr;

libname sasdataa
'/prj/niddk/ims_analysis/Look_AHEAD/private_orig_data/LookAHEAD/Baseline/Analysis Data Sets';

proc format;
  value HbA1cf
    0-<6      ="0-<6      "
    6.0-6.5="6.0-6.5"
    6.5-<7.0="6.5-<7.0"
    7.0-<8.0="7.0-<8.0"
    8.0-<9.0="8.0-<9.0"
    9.0-high="9.0+"
  ;

data baseline_combined  ; set sasdataa.baseline_combined;

proc freq data=baseline_combined;
  tables (female racevar metabolic_syndrome DIABETES_FAMILY cvdhis hypertension)*treatmentarm
  /missing nopercnt norow;
  title3 'Integrity Check for Look AHEAD (Action for Health in Diabetes):  Baseline Data Table
1';

proc means data=baseline_combined mean STDDEV;
  var baselinewgt_kg
  Eshgt_mean
  Bmi
  Waistcm_mean
```

Susan M Rogers  
(revised) April 2014

```
Minabi
HbA1cpct
Glucosemgdl
Screatmgdl
Cholmgdl
Ldlchlmgdl
Hdlchlmgdl
Trigmgdl
Bssbp_mean
Bsdbp_mean
Diab_dur
Maxexmets
;
class treatmentarm;
title3 'Integrity Check for Look AHEAD (Action for Health in Diabetes): Baseline Data Table
1';

proc freq data=baseline_combined;
tables (Age_t2
Racevar
Emp_status
Educ_yrs
Aldrksyr
Inc_t2
Tqsmk100
TQSMKCUR
Sdmarstat
)*female
/missing nopercnt norow;
title3 'Integrity Check for Look AHEAD (Action for Health in Diabetes): Baseline Data Table
2';

proc freq data=baseline_combined;
tables Age_t2
Bmigrp
Any_Htndrug
HbA1cpct
Any_Lipidrug
cvdhis
/missing;
where female='No';
format HbA1cpct HbA1cf.;
title3 'Integrity Check for Look AHEAD (Action for Health in Diabetes): Baseline Data Table
3';

proc freq data=baseline_combined;
tables (Age_t2
Bmigrp
Any_Htndrug
HbA1cpct
Any_Lipidrug
cvdhis
)*racevar
/missing nopercnt norow;
where female='No';
format HbA1cpct HbA1cf.;
title3 'Integrity Check for Look AHEAD (Action for Health in Diabetes): Baseline Data Table
3';

proc means data=baseline_combined mean STDDEV;
var Bssbp_mean
Bsdbp_mean
Glucosemgdl
Cholmgdl
Hdlchlmgdl
Ldlchlmgdl
Trigmgdl
;
where female='No';
title3 'Integrity Check for Look AHEAD (Action for Health in Diabetes): Baseline Data Table
3';
```

Susan M Rogers  
(revised) April 2014

```
proc means data=baseline_combined mean STDDEV;  
  var Bssbp_mean  
      Bsdbp_mean  
      Glucosemgdl  
      Cholmgdl  
      Hdlchlmgdl  
      Ldlchlmgdl  
      Trigmgdl  
  ;  
  class racevar;  
  where female='No';  
  title3 'Integrity Check for Look AHEAD (Action for Health in Diabetes): Baseline Data Table  
3';
```