MEMORANDUM

TO:        CARDIA Investigators and First Authors

FROM:      Lucia Juarez, Director
Data Analysis and Verification Unit
CARDIA Coordinating Center
lucia@uab.edu

RE:       Manuscript Verification Submission Instructions

As part of the CARDIA manuscript approval process, the Coordinating Center (CC) reviews
and verifies the findings presented in each manuscript to ensure consistency and accuracy in
CARDIA publications.

In January 2012, the CARDIA Steering Committee (SC) endorsed a recommendation from
the Publications and Presentations Committee (P&P) to refine the data set verification process in
an effort to better serve the Study and scientific community. The aim of the Study is to expedite
the verification time, resulting in a more rapid submission of manuscripts to journals.

The new process is summarized below:

1. The CC will no longer conduct the full verification of manuscripts.
2. The CC will continue to verify the final analytic data set for all manuscripts.
3. The verification of the final analytic data set may be conducted at any time, but must be
   completed before the manuscript will be released (approved by P&P) for submission to a
   journal. If any changes are made to the analytic data set between the time it was verified and
   the submission of the manuscript to the P&P Committee. The analytic data set will need to be
   verified again.
4. The above process will be utilized for one year, after which it will be revisited.

The Study requests that this memo be reviewed carefully, even if you have submitted a
manuscript for verification in the past. If the person who performs the statistical analyses is not
you, the first author, please share this memo with that person as it contains information that may
influence the analysis process as well as the preparation of the verification materials.

If there are questions regarding the verification process or this memorandum, please contact
Lucia Juarez (lucia@uab.edu) at the CC at (205) 934-0786.
1. Verification Process Overview

After the CC has received a complete set of materials from you (see detail below), the current version of the distributed data will be used to confirm that the data set used for analyses was created correctly. Elements which are evaluated include the merging of the component data sets, exclusions of participants from the analyses, definitions of new variables, and proper handling of missing values.

Upon completion of the verification of the analytic data set as part of the final manuscript, you will be notified in writing of any discrepancies between your data set and the verified data set. A copy will be sent to the chief reviewer of your manuscript. The verification of the analytic data set can be submitted at any stage of your analysis. However, once an analytic data set has been verified, we will expect this to be used throughout the manuscript’s analysis. If the analytic data set changes between the time of verification and the conclusion of your analyses, the authors will have to resubmit the modified analytic data set for verification. CARDIA policies state that manuscripts should not be submitted for publication until their analytic data sets are verified and the resultant corrections applied.

2. Instructions for Data Set Verification

A. Data Set Verification

Many of the problems encountered during analytic data set verifications involve merging issues, creation of filters for exclusion criteria, creation of new variables and not using the latest version of CARDIA data. The Study encourages you to submit your analytic data set for verification prior to the analysis.

The following verification materials should be submitted as attachments, via e-mail, to Lucia Juarez at lucia@uab.edu. Use the following descriptive e-mail subject line: Data Set Verification request for manuscript # xxx.

1. Programs and text files
   a. All programs used for data set creation, thoroughly documented. This includes programs that make exclusions, create new variables and merge multiple data sets;
   b. A text file that lists and describes all files in the attachment; and

2. A description of exclusions and created (derived) variables. If this document does not explicitly state “How Missing” and “Don’t Know” responses are handled for variables, also include a description of those decisions.

3. First author: Include all contact information:
   Name, e-mail address, telephone number and mailing address

4. Programmer: Include name and e-mail address

A checklist of these items is attached to this memo to facilitate your submission. We recommend early verification of the data set to eliminate the necessity to rerun analyses.
While most analyses of CARDIA data will be done in SAS and result in programs to be used in SAS, other analysis packages may be used. Instructions regarding non-SAS analyses can be found on page 3. Other related files may be submitted in WORD and EXCEL, or other such packages.

B. Data Set Verification resubmission

If your data set changed during the writing of your manuscript, resubmit the data set and follow the instructions outlined above for the Data Set Verification.

C. Analyses Using Programs Other than SAS

While CARDIA data are distributed for use with SAS, other software packages are sometimes used to perform part or all of the data preparation before analysis. When this is the case, program code may not be sufficient to allow us to duplicate your analytic data set. Please include the following:

1. All documentation to describe how the data set was created (particularly important when the software is interactive). Please include information about “How Missing” and “Don’t Know” responses were handled.

2. A text file containing a list of all CARDIA variables used in the analyses, using the actual SAS variable names from the distributed data (e.g., A10SMOKE).

3. A text file containing names and formulae for all created variables, including formats, reformats, and recodes. Again, please use the actual SAS variable names from the distributed data when noting these details.

4. A text file containing exclusions in term of the SAS variable names from the distributed data (e.g., A22DURTN<500).

5. All programs used for analyses. If the package does not generate programs for analyses, text files containing code-like statements that correspond to the analyses done and using CARDIA variable names.

6. For calculations done in spreadsheets, please provide us with the formula used.

D. Data from Ancillary Studies or Non-CARDIA Studies

Manuscripts using data from ancillary studies should be verified. If the data for an ancillary study are not on file at the CC, you should provide those data along with the other verification materials. If your manuscript involves data from another study, please contact the CC so that we may determine which portion of the analytic data set should be verified.
3. **Other Useful Information**

Authors are encouraged to call or e-mail the CC at any time to discuss issues or questions regarding analytic data set creation. While the documentation distributed with the data is intended to be complete, questions often arise, particularly regarding some of the more complicated data sets such as pregnancies or medications. In the Version 8.2 data documentation update, we will provide a list of common mistakes made in CARDIA analyses. It is important to confirm that the programs are operational prior to submitting the verification. Often programs have been created from larger programs and will not operate on their own.

Verifications are assigned to various CC staff. It requires approximately three weeks to complete the verification after all necessary materials are received. This time frame is extended if further information is required from you. The CC verification staff have graduate training in statistics. Thus, suggestions and/or comments may be included in the completed verification that we deem may be useful to carry out the analysis of your manuscript.

4. **Attached Materials**

Attached to this memo are additional materials which you may find useful. First, the required materials for verifications are summarized in a checklist. We request that you use the checklist when preparing the verification packet, sign and date it once the packet is complete. Submit the verification files, as attachments, via e-mail to: Lucia Juarez at lucia@uab.edu. Use the following descriptive e-mail subject line: Data Set Verification request for manuscript # xxx.

Second, we have included sample SAS programs complete with the type of documentation we would like to have in programs you submit. Also, included are example annotated manuscript pages illustrating the type of annotation we prefer.
CARDIA DATA SET VERIFICATION CHECKLIST

☐ 1. All data set creation programs merging distributed CARDIA data, thoroughly documented.

☐ 2. All programs that make exclusions, thoroughly documented.

☐ 3. All programs that create variables, thoroughly documented.

☐ 4. A text file that lists and describes all files enclosed.

☐ 5. E-mail message with text attachments containing Items 1-4, labeled with first author and CARDIA manuscript number.

☐ 6. Confirm that submitted programs are operational.

☐ 7. A description of exclusions and created (derived) variables. If this document does not explicitly state “How Missing” and “Don’t Know” responses are handled for variables, also include a description of those decisions.

☐ 8. First author: Include all contact information:
   Name, e-mail address, telephone number and mailing address

☐ 9. Programmer: Include name and e-mail address

___________________________  _________________________
Signature of First Author      Date Signed
**SAMPLE SAS PROGRAMS**

*---------------------------------------------------------------------*
** PROGRAM: P999 - A.SAS                                               *
** Purpose: This program builds the data set P999.sas7bdat for         *
** paper #999 by merging blood chemistry, smoking                      *
** dietary data at Year 0 and 7                                       *
**                                                               *
** Apply exclusion criteria                                           *
** DATE OF LAST RUN: 8/27/01, 02/04/03 last MODIFIED                   *
*---------------------------------------------------------------------*

libname home 'O:\YKIM\Verif\P999';

*---------------------------------------------------------------*
Item 1: exclusion of study subjects-> final sample size: 3146          *
*---------------------------------------------------------------*

data home.P999; merge y0.a4ref (keep=id exameage sex race center)   *
y0.a4f10 (keep=id a10smoke)                                          *
y0.a5f06a (keep=id a06calc)                                          *
y0.a4cham (keep=id a13_ggt)                                          *
y7.dif10 (keep=id d10smoke)                                          *
y7.direfa (keep=id ex4_age x4status)                                 *
y7.d2f06 (keep=id d06calc)                                          *
home.foodnew (in=ref); by id;

if ref;
* exclusion criteria                                                *
  1) attend year 7 exam                                              *
  2) no data on yr 0 GGT                                            *
  3) no data on yr 0 diet or yr 7 diet                              *
  4) impossible total calorie                                       *
  5) No diabetes on yr 0 or yr 7;                                   *

if x4status in ('E','H','1','2','3','4');
if a13_ggt=. then delete;
if a06calc=. or d06calc=. then delete;

if sex=1 and (a06calc<800 or a06calc>8000) then delete;
if sex=1 and (d06calc<800 or d06calc>8000) then delete;
if sex=2 and (a06calc<600 or a06calc>6000) then delete;
if sex=2 and (d06calc<600 or d06calc>6000) then delete;

if diab=1 or diab7=1 then delete;
run;

*---------------------------------------------------------------*
Item 3: TEXT in RESULTS                                             *
*---------------------------------------------------------------*

title 'No. of subjects by race and sex';
proc freq data=home.P999;
tables race*sex; run;

title 'mean age and education';
proc means data=home.P999;
var exameage a03ed; run;
title 'food groups and year 10 GGT';
proc glm data=heme_F999;
model lne_ggt=drink
   calo07 race sex center examage a20bmi a18total a10smoke;
run;

proc glm data=home_F999;
model lne_ggt=mamm07q
   calo07 race sex center examage a20bmi a18total a10smoke;
run;
Purpose: The purpose of this study was to examine dietary correlates of serum gamma glutamyltransferase (CGT) level, motivated by the observation of a strong dose response relationship of serum CGT level with incident diabetes.

Methods: Study subjects were 3146 black and white men and women aged 18-30 years in 1985-1986. Diet was measured at years 0 and 7 using an interviewer-administered quantitative food frequency questionnaire. Food items were classified into 14 food groups; alcohol, meat, poultry, fish, fresh or frozen vegetables, fried or breaded/battered or canned vegetables, fruit, fruit juice, refined grain, whole grain, dairy, legumes, nuts, and coffee.

Results: After adjusting for non-dietary factors and other food groups, serum CGT level was positively associated with alcohol consumption and meat intake. Geometric means of year 10 CGT across categories of alcohol consumption (0g, 1-9g/d, 10-19g/d, 20-29g/d, ≥30g/d) were 17.7, 18.8, 20.4, 21.8 and 24.8 U/L (P for trend <0.01); corresponding means across quintiles of meat intake were 19.2, 20.2, 20.5, 21.8, and 21.2 (P for trend <0.01). However, serum CGT level was inversely associated with various plant foods, especially fruit, and fish intake. Among possible meat constituents, dietary heme iron, but not saturated fat, was associated with serum CGT level. Beneficial dietary constituents typical of plant foods, including beta-carotene, vitamin C, and fiber showed an inverse association. In contrast, vitamin A, vitamin C, and folate from supplements were positively associated with serum CGT level.

Conclusion: Serum CGT level increased in a dose-response manner as alcohol and meat consumption increased and fruit consumption decreased. Heme iron contained in meats and beneficial dietary micronutrients contained in fruits may influence CGT metabolism. However, micronutrients taken as supplements did not show a consistent relation with CGT level.