User's Guide for Analysis of Usual Intakes

For use with versions 2.1 of the MIXTRAN, DISTRIB, and INDIVINT SAS[*] macros

Contents

[*] SAS Institute, Inc., Cary, NC, USA

1 Introduction

This user's guide describes the MIXTRAN, DISTRIB, and INDIVINT SAS macros that implement a statistical method for estimating usual dietary intakes of a single dietary component, whether consumed daily or episodically. This method can be used for a variety of applications, including:

- estimating the distribution of usual food or nutrient intake for a population or subpopulation;
- predicting individual food or nutrient intake for use in a disease model; and
- evaluating the effects of individual covariates on food or nutrient consumption.

The macros can perform a weighted analysis of survey data such as the National Health and Nutrition Examination Survey (NHANES), and this user's guide discusses several examples that illustrate the method using NHANES data. The SAS macros and example programs can be downloaded from the following website: http://appliedresearch.cancer.gov/diet/usualintakes/macros_single.html.
The files **mixtran_macro_v2.1.sas**, **distrib_macro_v2.1.sas**, and **indivint_macro_v2.1.sas** include the MIXTRAN, DISTRIB, and INDIVINT macros, respectively.

Please send your comments and questions to **usualintake@imsweb.com**. We will take this feedback into consideration when developing future versions of our software and other resources. We will not be able to send a response to each user that submits a comment or question.

1.1 Selecting the Appropriate Macro

The MIXTRAN macro can be used alone to evaluate the effects of individual covariates on food or nutrient consumption, and the MIXTRAN macro can be used to generate parameter estimates and linear predictor values used as input for the DISTRIB and INDIVINT macros. The MIXTRAN macro fits a nonlinear mixed effects model to repeat 24-hour recalls, using the SAS NLMIXED procedure. For an episodically consumed food or nutrient, the macro fits a two-part model that defines the distribution of reported intake on a given day as the probability of consumption multiplied by the conditional distribution of amount consumed on a consumption day. For a food or nutrient consumed every day, the macro fits a one-part (i.e. amount-only) model of the amount consumed.

The MIXTRAN and DISTRIB macros are used together to estimate the distribution of usual intake of a food or nutrient in a population, using reported intake from repeat 24-hour recalls and covariates of interest (Tooze et al., 2006). These macros can be used to fit the two-part model and estimate the distribution of usual intake for an episodically consumed food or nutrient, or these macros can be used to fit the amount-only model and estimate the distribution of usual intake for a food or nutrient consumed every day or nearly every day. The DISTRIB macro uses parameter estimates and linear predictor values from the MIXTRAN macro and uses a Monte Carlo method to estimate the distribution of usual intake.

The MIXTRAN and INDIVINT macros are used together to predict individual food or nutrient intake for use in a disease model. The INDIVINT macro uses parameter estimates and linear predictor values from the mixed effects model estimated by the MIXTRAN macro and predicts each individual's usual intake of an episodically consumed food or nutrient or a food or nutrient consumed every day (Kipnis et al., 2009). The INDIVINT macro performs adaptive Gaussian quadrature to predict individual usual intake and allows the user to provide a Box-Cox transformation parameter in order to predict individual usual intake on a transformed scale. As discussed by

Kipnis et al. (2009), the covariates in the health outcome model should also be included as covariates in the mixed effects model. Kipnis et al. (2009) also state, "Our methodology is suitable for analysis of a particular food and its relationship with a health outcome that involves no other dietary factors."

1.2 Standard Errors

The user must perform additional calculations, such as balanced repeated replication (BRR), in order to obtain standard errors and confidence intervals for the percentiles and mean from the distribution of usual intake. BRR calculations require that the user writes a SAS program that calls the macro(s) using the appropriate weight for each replication. MIXTRAN has several options (e.g. start_val1, start_val2, start_val3, and vcontrol) that are helpful for a user that needs to make repeated calls to MIXTRAN for BRR calculations.

1.3 Accessing the Macro Programs

The MIXTRAN, DISTRIB, and INDIVINT SAS macros are provided as separate programs. These programs can be saved in a SAS macro library, or called using the %include statement in a SAS program.

On the Linux/Unix platform the syntax is:
```
%include "/mypath /mixtran_macro_v2.1.sas";
%include "/mypath /distrib_macro_v2.1.sas";
%include "/mypath /indivint_macro_v2.1.sas";.
```

On the Windows platform the syntax is:
```
%include  'C:\mypath\mixtran_macro_v2.1.sas';
%include  'C:\mypath\distrib_macro_v2.1.sas';
%include  'C:\mypath\indivint _macro_v2.1.sas';.
```

Note that "mypath" is the path name to the directory in which the macros are stored.

2 The MIXTRAN Macro

The MIXTRAN macro allows three possible models to be fit using the SAS NLMIXED procedure. For foods or nutrients that are consumed episodically, the correlated or uncorrelated model can be specified. For those foods or nutrients that are typically consumed every day, the amount-only model can be specified. The MIXTRAN macro outputs data sets of the parameter estimates and of linear predictor values for use in subsequent steps, provides summary reports, and optionally can list the output from the SAS procedures.

2.1 Syntax for the MIXTRAN Macro

The MIXTRAN macro uses the following syntax:

%mixtran (data=, response=, modeltype=, subject=, repeat=, outlib=, covars_prob=, covars_amt=, foodtype=, seq=, lambda=, replicate_var=, weekend=, vargroup=, numvargroups=, vcontrol=, start_val1=, start_val2=, start_val3=, nloptions=, titles=, printlevel=, subgroup=);

The macro parameters in MIXTRAN are keyword parameters and are not positional parameters. These keyword parameters can be specified in any order so long as each value is preceded by the parameter name and the equals sign (e.g. data=mydata,). Separate each parameter specification with a comma. Parameters that are null can be omitted. The default value of a parameter is null unless otherwise noted in the documentation.

An example of a call to the MIXTRAN macro is:

%mixtran (data=child,
          response=f_total,
          modeltype=corr,
          subject=seqn,
          repeat=day,
          foodtype=f_total_child,
          outlib=mylib,
          covars_prob=sex1 age2,
          covars_amt=sex1 age2,
          seq=day2,
          lambda=0.4,
          replicate_var=rndw1,
          weekend=weekend,
          vargroup=vgrp,
          numvargroups=2,
          vcontrol=1,
          start_val1=,
          start_val2= mylib._parmsf2_f_total_child,
          start_val3=,
          nloptions=qmax=61,
          titles=4,
          printlevel=2,
          subgroup=);

The following list provides an explanation of each of the parameters in the MIXTRAN macro.

data   **Required.** Specifies a SAS data set, sorted by individual, with one or more observations for each individual. The names "data" and "data0" are reserved for the MIXTRAN macro.

response   **Required.** Specifies the 24-hour recall variable for the food or nutrient.

modeltype   **Required.** Specifies one of three available model types: a two-part model with correlated random effects (CORR), a two-part model with independent random effects (NOCORR), and an amount-only model of the amount of the nutrient (or food) consumed (AMOUNT). The options for this parameter are:

           corr   correlated model. An initial two-part uncorrelated model is fit to obtain starting values when not provided by the user, and using the available starting values, a two-part correlated model is fit. This model is used for episodically consumed dietary components;

           nocorr uncorrelated model. A two-part uncorrelated model is fit. This model is used for episodically consumed dietary components;

           amount amount model. An amount-only model is fit. This model is used for dietary components consumed daily or nearly every day.

subject   **Required.** Specifies the variable that uniquely identifies each subject.

repeat   **Required.** Specifies the name of the variable that indexes repeated observations for each subject. This variable indexes which 24-hour recall appears on a given record for a subject, so the value on a record should be an integer value of 1 or more.

foodtype   **Required.** Specifies a character string used to name the parameter and predicted data sets output by the MIXTRAN macro. The value must be a string valid in a SAS data set name.
(For an explanation of the data sets output by the MIXTRAN macro, see the section SAS Data Sets Output by the MIXTRAN Macro.)

outlib   **Required.** Specifies a library name reference to the directory where the data sets output from the macro will be saved. This library name must be specified in the SAS program prior to calling the macro, for example:
           LIBNAME mylib "/myoutput/path";.
      Then the syntax in the macro call is:
           outlib=mylib.

covars_prob   **Optional.** Specifies a list of covariates for the first part of the model that models the probability of consumption. Covariates must be separated by spaces. Interactions must be in the order specified by the SAS GENMOD procedure. In an amount-only model, this parameter should be left blank. Do not include any variable mentioned in either the seq= or weekend= parameters as these will automatically be added to the list of covariates by the macro.

covars_amt   **Optional.** Specifies a list of covariates for the second part of the model that models the consumption-day amount. Covariates must be separated by spaces. Interactions must be in the order specified by the SAS GENMOD procedure. These covariates are used in all model types. Do not include any variable mentioned in either the seq= or weekend= parameters as these will automatically be added to the list of covariates by the macro.

seq   **Optional.** Specifies one or more sequence indicator variables to account for effects due to the sequence number of a subject's records. This list of variable(s) must not also appear in covars_prob or covars_amt. To illustrate, a data set with two records per person will have just one seq indicator variable, set to 0 for the person's first record and set to 1 for the second record. If the data set had 3 records per person, then two indicator variables could be used to identify a subject's second and third records, respectively.

                        Note: The DISTRIB macro estimates distributions of intake under the assumption that the values of all indicator variables specified in the "seq=" parameter have value zero. This process adjusts the usual intake for time-in-sample or sequence effects often noted in analysis of recall data, where reported intake on the first application of the recall is distributed differently from data collected on subsequent recalls. Alternative coding for the indicator variables representing sequence may be used, but the user must understand the consequences of using such coding given the behavior of the DISTRIB macro.

lambda   **Optional.** Specifies a value for the Box-Cox transformation parameter, lambda. If a value is not provided, lambda will be estimated as a model parameter.

replicate_var   **Optional.** Specifies a weight variable to be used in the "replicate" statement of the SAS NLMIXED procedure. The specified variable must be integer valued. The same variable will be used in all SAS procedures where a "freq" or "weight" statement is appropriate, and this value is used in the DISTRIB macro.

weekend   **Optional.** Specifies the weekend indicator variable to account for a weekend effect. A value of 1 represents a weekend (e.g. Fri.-Sun.) record, and a value of 0 represents a weekday record. Originally intended to differentiate between intake on weekdays or weekends, the parameter name of weekend has been retained, but advanced users could specify an alternative indicator variable. The specified weekend variable must not appear in the parameters covars_prob or covars_amt. If the weekend parameter is specified then the vargroup= and numvargroups= parameters may also be initiated.

vargroup   **Optional.** Specifies a variable that groups observations to allow the model to incorporate a separate residual variance parameter for each of these groups of observations. If the output from this macro is to be used in the DISTRIB macro, then only the variable specified in the weekend parameter can be used. *This parameter can only be specified if a variable has been specified for the weekend parameter*.
                        If supplied this is the name of a variable (for example vgrp), with values of 1 or 2, which groups observations and allows the model to incorporate a separate residual variance parameter for each of these groups of observations.

In this case if the weekend variable value is 0, the vargroup variable value must be 2, and if the weekend variable value is 1 then the vargroup variable value must be 1.
If the vargroup parameter is specified then the numvargroups parameter MUST be specified.

numvargroups   **Optional.** Specifies the number of groups defined by the "vargroup" variable.
If the output from this macro is to be used in the DISTRIB macro and a weekend variable is the "vargroup" variable, then the number of groups is 2. *This parameter can only be specified if variables have been specified for the weekend parameter **and** the vargroup parameter.* In such a case this parameter MUST be specified.

Note: The next four parameters - relating to version control and starting values for the various SAS NLMIXED procedures - apply almost exclusively to repetitive executions of the macro that are subsequent to an initial execution. We have adopted the terminology of a **base** run to denote a first or initial execution of the macro. A **rerun** refers to subsequent executions of the macro with all the parameters the same excepting replicate_var (the weight variable). In brief, reruns are indicated by a value in the vcontrol parameter. Reruns require starting values: uncorrelated models (nocorr) require data in start_val1 and start_val2; amount models require data in start_val2; correlated models (corr) require data in start_val3.

vcontrol   **Optional.** (Use only when starting values from a previous execution of this macro are also used.) Specifies a 1 to 6 character name to differentiate output data sets for runs using the same food. See the parameters start_val1, start_val2, and start_val3. The default is null.

start_val1   **Optional.** (Use only when vcontrol and parameter estimates (i.e. _parmsf1_"foodtype") from a previous execution of this macro are used.) Specifies the starting values data set for the 1st PROC NLMIXED (i.e. NLMIXED for probability model).

start_val2   **Optional.** (Use only when vcontrol and parameter estimates (i.e. _parmsf2_"foodtype") from a previous execution of this macro are used.) Specifies the starting values data set for the 2nd PROC NLMIXED (i.e. NLMIXED for amount model).

start_val3   **Optional.** (Use only when vcontrol and parameter estimates (i.e. _parmsf3_"foodtype") from a previous execution of this macro are used.) Specifies the starting values data set for the 3rd PROC NLMIXED (i.e. NLMIXED for correlated model).

nloptions   **Optional.** Specifies a list of options to be added to the PROC NLMIXED statement for all calls to PROC NLMIXED, for example: nloptions=qpoints=1 gconv=1e-12**.** See SAS documentation for further details regarding these options.

titles   **Optional**. Specifies the number of title lines (0-4) to be reserved for the user's titles. The remaining title lines are used by the macro. The default value is 0.

printlevel   **Optional.**   Specifies 1, 2, or 3 to control the amount of information printed in the list file. Printlevel=1 prints only the summary reports. Printlevel=2 prints summary reports and output from the calls to the NLMIXED procedure. Printlevel=2 is the default value. Printlevel=3 prints summary reports and output from all of the statistical procedures.

subgroup   **Optional**. The subgroup parameter has been kept in the MIXTRAN macro solely for compatibility with previous versions. The parameter will likely disappear from MIXTRAN in future versions. The subgroup parameter is now available in the DISTRIB macro instead. Please refer to the subgroup parameter instructions for the DISTRIB macro.

2.2 SAS Data Sets Output by the MIXTRAN Macro

This section describes the various data sets output by the macro, and gives some explanation of the variable names in these data sets.

The MIXTRAN macro creates a number of SAS data sets that are saved to disk for later use. Each data set is saved to the directory named by the reference in the "outlib" parameter as described in the previous section. After each execution of an NLMIXED procedure, the parameter estimates from that procedure are saved as a data set. This data set can be used as starting values input for later executions of the MIXTRAN macro. In addition, data sets of the predicted values and the parameter estimates for both the correlated and the uncorrelated models are saved for possible use as input for the DISTRIB or INDIVINT macros.
The data sets saved, for use as input in future re-runs of the MIXTRAN macro, are the tables of parameter estimates output by the SAS NLMIXED procedure and an expression to calculate the predicted values. The data sets are named using the following conventions:
        outlib._dsn_foodtype

described as follows.

        outlib   is the library name given in the MIXTRAN parameter outlib.

        _dsn   is one of the following names.

                _parmsf1   This data set captures the parameter estimates output by the SAS NLMIXED procedure for the probability model in a base run. It can be used as input for the starting values for this model in a re-run. In a re-run this data set is referenced in the MIXTRAN parameter "start_val1".

                _parmsf2   This data set captures the parameter estimates output by the SAS NLMIXED procedure for the amount model in a base run. It can be used as input for the starting values for this model in a re-run. In a re-run this data set is referenced in the MIXTRAN parameter "start_val2".

                _parmsf3   This data set captures the parameter estimates output by the SAS NLMIXED procedure for the correlated model in a base run. It can be used as input for the

starting values for this model in a re-run. In a re-run this data set is referenced in the MIXTRAN parameter "start_val3".

etas   This data set contains character strings that are interpreted by the MIXTRAN macro to calculate the predicted values. This data set is only output by a base run and only used in a re-run, and it is automatically utilized if the "vcontrol" parameter is in use.

_foodtype   is obtained from the "foodtype" parameter value specified in the call to the MIXTRAN macro.

Several data sets are saved for use as input for the DISTRIB or INDIVINT macros. The DISTRIB and INDIVINT macros use the parameter estimates and predicted values calculated by the MIXTRAN macro. The parameter estimates and predicted values are saved from the amount-only model, the uncorrelated model, and the correlated model for the "amount", "nocorr", and "corr" model types, respectively. Also, for the "corr" model type, the parameter estimates and predicted values are saved from the uncorrelated model which is fit in order to calculate starting values for the correlated model. The data sets are named using the following conventions:

outlib._param_unc_foodtype_vcontrol and
outlib._pred_unc_foodtype_vcontrol

described as follows.

outlib   is the library name given in the MIXTRAN parameter "outlib".

_param   is used to indicate a parameter data set. This data set consists of one record and always includes "_param" in the data set name.

_pred   is used to indicate a data set of predicted values. This data set consists of one record per person and always includes "_pred" in the data set name.

_unc   is used to indicate the parameter and predicted data sets from the uncorrelated model and the amount-only model. The MIXTRAN macro will include the string "_unc" in these data set names. The MIXTRAN macro will not include the string "_unc" for the output data sets from the correlated model.

_foodtype   is obtained from the "foodtype" parameter value specified in the call to the MIXTRAN macro.

_vcontrol   is obtained from the "vcontrol" parameter value specified in the call to the MIXTRAN macro. Only re-runs will have this value appended to the data set name.

For example, if we fit a correlated model using a response variable named "fish", the output data sets would be:

mylib._parmsf1_fish
mylib._parmsf2_fish
mylib._parmsf3_fish
mylib.etas_fish

mylib._param_unc_fish
mylib._pred_unc_fish
mylib._param_fish
mylib._pred_fish.

2.3 Description of Variables in the Predicted Data Set Output by the MIXTRAN macro

The variables in the data set containing the predicted values depend in part on the parameters used in the call to the MIXTRAN macro.

Amount Models

Without the weekend parameter

Variable	Explanation
Subject	The Subject ID variable name from the subject parameter.
Replicate	The weight variable named in the replicate_var parameter, if any.
Subgroup	The name of the variable in the subgroup parameter, if any.
X2b2	The estimate of the predicted value from the consumption-day amount calculations. This is an intermediate value used in subsequent programs such as the DISTRIB macro.

With the weekend parameter

Variable	Explanation
Subject	The Subject ID variable name from the subject parameter.
Replicate	The weight variable named in the replicate_var parameter, if any.
Subgroup	The name of the variable in the subgroup parameter, if any.
X2b2_0	The estimate of the predicted value from the consumption-day amount calculations for value 0 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.
X2b2_1	The estimate of the predicted value from the consumption-day amount calculations for value 1 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.

Uncorrelated Models

Without the weekend parameter

Variable	Explanation
Subject	The Subject ID variable name from the subject parameter.
Replicate	The weight variable named in the replicate_var parameter, if any.
Subgroup	The name of the variable in the subgroup parameter, if any.
X1b1	The estimate of the predicted value from the probability of consumption calculations. This is an intermediate value used in subsequent programs such as the DISTRIB macro.
X2b2	The estimate of the predicted value from the consumption-day amount calculations. This is an intermediate value used in subsequent programs such as the DISTRIB macro.

With the weekend parameter

Variable	Explanation
Subject	The Subject ID variable name from the subject parameter.
Replicate	The weight variable named in the replicate_var parameter, if any.
Subgroup	The name of the variable in the subgroup parameter, if any.
X1b1_0	The estimate of the predicted value from the probability of consumption calculations for value 0 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.
X1b1_1	The estimate of the predicted value from the probability of consumption calculations for value 1 in the weekend variable.
X2b2_0	The estimate of the predicted value from the consumption-day amount calculations for value 0 in the weekend variable.
X2b2_1	The estimate of the predicted value from the consumption-day amount calculations for value 1 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.

Correlated Models

Without the weekend parameter

Variable	Explanation
Subject	The Subject ID variable name from the subject parameter.
Replicate	The weight variable named in the replicate_var parameter, if any.
Subgroup	The name of the variable in the subgroup parameter, if any.
X1b1	The predicted value from the probability of consumption calculations. This is an intermediate value used in subsequent programs such as the DISTRIB macro.
X2b2	The predicted value from the consumption-day amount calculations. This is an intermediate value used in subsequent programs such as the DISTRIB macro.

With the weekend parameter

Variable	Explanation
Subject	The Subject ID variable name from the subject parameter.
Replicate	The weight variable named in the replicate_var parameter, if any.
Subgroup	The name of the variable in the subgroup parameter, if any.
X1b1_0	The estimate of the predicted value from the probability of consumption calculations for value 0 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.
X1b1_1	The estimate of the predicted value from the probability of consumption calculations for value 1 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.
X2b2_0	The estimate of the predicted value from the consumption-day amount calculations for value 0 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.
X2b2_1	The estimate of the predicted value from the consumption-day amount calculations for value 1 in the weekend variable. This is an intermediate value used in subsequent programs such as the DISTRIB macro.

2.4 Variable Name Changes for the Covariates in the MIXTRAN Macro

The variable names for the covariates used in the MIXTRAN macro are amended during the execution of the macro. This modification is performed in part because it is possible to use the same variables for both the probability of consumption and the consumption-day amount parts of the model, and these variables need to be differentiated during macro processing. The new variable names also allow the macro to maintain the order of the variables as originally entered by the user. The covariates named in the "covars_amt" parameter will be modified, so the variable name is prefixed by the letter "A", a sequence number, and an underscore, so the format is "Ann_". The first named covariate will be prefixed by the string "A02_", the second variable in the list will be prefixed by the string "A03_" and so on. For example, if the first variable is "agegrp1", it will become "A02_AGEGRP1", and if the second variable is "agegrp3", it will become "A03_AGEGRP3". Similarly, the covariates named in the parameter "covars_prob" for the probability part of the model will be prefixed using the format "Pnn_". The output from the SAS procedures and the saved data sets will reflect the new variable names. The intercepts are always named "P01_INTERCEPT" or "A01_INTERCEPT" for the probability and amount intercepts, respectively.

3 The DISTRIB Macro

The DISTRIB macro can be called subsequent to the MIXTRAN macro in the same program, or in a later program. It can only be called after the MIXTRAN macro has been executed successfully.

The DISTRIB macro uses results from the MIXTRAN macro and estimates the distribution of usual intake for episodically consumed foods or nutrients, or foods or nutrients consumed every day. The data can then be used to calculate percentiles and cut points, and optionally the percent meeting or failing to meet the recommended daily intake for a population.

The DISTRIB macro performs two main functions.

The first function reads the data sets of parameter estimates and predicted values output by the MIXTRAN macro, and uses Monte Carlo simulation of the random effect(s) to estimate the distribution of usual intake. This data can be saved for future analyses.

The second function uses these estimates to calculate the percentiles and other descriptive values of the distribution of usual intake. The addition of one categorical subgroup variable is accommodated, so that statistics can be calculated for each subgroup level and for the overall data set. The variable(s) in the subgroup must have been included in the model statement of the MIXTRAN macro, and the group levels should be the same. The percent of subjects meeting or failing to meet recommended daily intake values can also be calculated. If a subgroup is used, the statistics will be calculated for each level of the subgroup and overall.

To accomplish this and allow flexibility, the DISTRIB macro contains two sub-macros which can be called separately or together – MC and PC.

The sub-macro MC uses Monte Carlo simulation of the random effect(s) to estimate the usual intake values. The output data set can be saved to disk or retained temporarily as a SAS work data set to save storage.

> Note: The Taylor Linearization method to back-transform the amount of consumption to the original scale has been replaced with a 9-point approximation method.

The sub-macro PC reads in the usual intake values calculated in the sub-macro MC, normalizes the weights, calculates the percentiles of usual intake, cut points if requested, and optionally compares the amount of intake to a recommended amount of intake. A single subgroup variable can be accommodated. The resulting data set can be saved for future use.

3.1 Syntax for the DISTRIB Macro

The DISTRIB macro uses the following syntax:

%DISTRIB (call_type=, modeltype=, outlib=, subject=, mcsimda=, pred=, param=, seed=, nsim_mc=, food=, byvar=, cutpoints=, ncutpnt=, wkend_prop=, add_da=, subgroup=, recamt=, recamt_co=, recamt_hi=, titles=);

The macro parameters in DISTRIB are keyword parameters and are not positional parameters. These keyword parameters can be specified in any order so long as each value is preceded by the parameter name and the equals

sign and followed by a comma (e.g. "modeltype=amount,"). Parameters that are null or accept the default can be omitted from the call.

An example of a call to the DISTRIB macro follows:

%distrib(call_type=full,
      modeltype=corr,
      outlib=mylib,
      subject=ID,
      mcsimda= mylib._mcsim_f_total,
      pred= mylib._pred_ f_total,
      param= mylib._param_ f_total,
      seed=76253,
      nsim_mc=100,
      food=f_total,
      byvar=strata,
      cutpoints=0.3 0.7 1.2, ncutpnt=3,
      add_da=mylib.nhanes.subset,
      wkend_prop=2/7,
      subgroup=age_sex_grp,
      recamt= min_rec, recamt_co=r, recamt_hi=max_rec,
      titles=4);

The following list provides an explanation of each of the parameters in the DISTRIB macro.

calltype   **Required.** Specifies which sections of the macro should be invoked. The default is FULL.
                           Full   invokes both sub-macros (MC and PC).
                           MC   invokes only the MC sub-macro which calculates the estimated usual intake of foods or nutrients using Monte Carlo simulations.
                           PC   invokes only the PC sub- macro that reads in the estimated usual intake of foods or nutrients (created in the sub-macro MC) and calculates the percentiles. The PC sub-macro offers options for cut points and also a comparison of the usual intake with a recommended amount of intake. This sub-macro should not be invoked until the usual intake has already been calculated in the MC sub-macro.

modeltype   **Required.** Specifies the model that was used by the MIXTRAN macro to prepare the data for the DISTRIB macro. The value must be the same as the model declared for the MIXTRAN macro. The possible values are:
                           corr   fit two-part correlated random effects model (the default),
                           nocorr   fit two-part model with independent random effects,
                           amount fit amount-only model.
        The declaration of the modeltype affects which parameter and predicted data sets are used as input for the DISTRIB macro. If the modeltype in the MIXTRAN macro was "corr", and the correlated model was fit successfully, then the correlated data sets can be used as input for the DISTRIB macro. If the modeltype in the MIXTRAN macro was "nocorr" or "amount", then only the data sets with "_unc" in the name can be used as input for the DISTRIB macro.

Please see the section on [SAS Data Sets Saved by the MIXTRAN Macro](#) for an explanation of the data set naming conventions. The parameter and predicted data set names will be declared with the [pred= and param=](#) parameters explained below.

outlib   **Required.** Specifies the library reference to the parameter and predicted data sets saved by the MIXTRAN macro and the directory to which the data set of distributions will be written by the DISTRIB macro. Since the input data sets for the DISTRIB macro were output by the MIXTRAN macro, the value of "outlib" should be identical to the value of the "outlib" parameter in the MIXTRAN macro. An explanation of the naming conventions for the data set output by the DISTRIB macro follows the explanation of the parameters.

subject   **Optional.** Specifies the variable that uniquely identifies each subject. (Required if the parameter "weekend" is used in MIXTRAN or any of the parameters "recamt", recamt_co, or recamt_hi are specified in the current call to DISTRIB).

mcsimda   **Optional.** Specifies the name of the data set containing the estimated intake derived from the monte carlo simulations. To save the data set include a libname and valid SAS data set name. This data set is created in the sub-macro MC and used to calculate the distribution of intake in the sub-macro PC.

pred   **Required.** Specifies the name of the data set containing predicted values for each subject. The input data to be used depends on the "modeltype" used in the MIXTRAN macro. If the "modeltype" in the MIXTRAN macro was "corr", and the correlated model was fit successfully, then the correlated data set can be used as input for the DISTRIB macro. If the "modeltype" in the MIXTRAN macro was "nocorr" or "amount", then only the data set labeled with "_unc" can be used as input for the DISTRIB macro. The full name of the data set is also determined by the values of the parameters "outlib", "foodtype", and "vcontrol" that were specified in the call to the MIXTRAN macro.

param   **Required.** Specifies the name of the data set containing the parameter estimates. The input data to be used depends on the "modeltype" used in the MIXTRAN macro. If the "modeltype" in the MIXTRAN macro was "corr", and the correlated model was fit successfully, then the correlated data set can be used as input for the DISTRIB macro. If the "modeltype" in the MIXTRAN macro was "nocorr" or "amount", then only the data set labeled with "_unc" can be used as input for the DISTRIB macro. The full name of the data set is also determined by the values of the parameters "outlib", "foodtype", and "vcontrol" that were specified in the call to the MIXTRAN macro.

seed   **Required if the calltype=FULL or MC.** Specifies the seed for the random number generator used for the Monte Carlo simulation of the random effects $u1$ and $u2$.

nsim_mc   **Required if the calltype=FULL or MC.** Specifies the number of repetitions to be used in the Monte Carlo simulation. One record will be output for each repetition of each subject.

food   Optional. Specifies a name for the analysis, used to label the output data set. (For an explanation of the data sets output by the DISTRIB macro, see the section [SAS Data Set Output by the DISTRIB Macro](#).)

byvar   Optional. If the MIXTRAN model was fit separately for one or more by-groups the user can list the name of the variable(s) identifying the by-groups here. Separate each by variable by a space. The estimates used in the calculation of usual intake differ based on the by-group, however the DISTRIB macro will produce final estimates for the population (not distribution within each by-group). This parameter can be invoked if the parameter calltype is FULL or MC.

Note: the next two parameters – cutpoints and ncutpnt - should be used as a pair. Either both or neither should be called. They set up the cut points for analysis in the distribution of the estimates of usual intake. They are an option in the sub-macro PC.

cutpoints   Optional. Specifies one or more cutoff points for which to calculate the proportion of the population below the cutoff points. Each cutoff point value must be separated by a single space. If no cutoff points are supplied, then the DISTRIB macro will only calculate the mean and percentiles of intake.

ncutpnt   Optional. Specifies the number of cutoff points in the "cutpoints" list. If cutoff points are specified, then "ncutpnt" must also be specified. This parameter is only available in the FULL or PC calltype, and should only be used with the cutpoints parameter.

wkend_prop   Optional. The proportional weight for value 1 of a binary variable if this variable was used in the weekend parameter of MIXTRAN. Enter a value between 0 and 1 (not inclusive). Either a fraction or decimal number can be used. The default is 3/7 if the weekend parameter was used in MIXTRAN. This parameter is only available if the calltype is FULL or MC.

add_da   Optional. Specifies the name of the data set containing variables needed to process the optional parameters subgroup, recamt, recamt_co or recamt_hi. The data set must also include the ID variable declared in the subject parameter.

subgroup   Optional. Specifies a single categorical variable used for the calculation of a separate usual intake distribution for each subgroup. The distribution of usual intake will also be calculated for the overall data set (i.e. all subjects). The subgroup variable must be included in the data set named by the add_da parameter. The subgroup is available if the call_type is FULL or PC.

recamt   Optional. Specifies the variable name that contains the cut-off value of the recommended amount of intake. This variable must be included in the data set named by the add_da parameter. If the comparison is to be within a range then this variable will indicate the lower cut-off level. A comparison operator must be supplied in the parameter recamt_co. This parameter is available if the calltype is FULL or PC.

recamt_co	**Optional.** Specifies the comparison operator between individual intake and the recommended cut-off amount. The available comparisons are:

        LT – less than the value in the recamt variable;

        LE – less than or equal to the value in the recamt variable;

        GT – greater than the value in the recamt variable;

        GE – greater than or equal to the value in the recamt variable;

        R – A range between two cutoff levels. The lower cutoff is found in the recamt variable and the upper cutoff in the recamt_hi variable.

recamt_hi	**Optional.** Specifies the variable name that contains the upper cut-off value of the recommended amount of intake if a range was specified in the reacamt_co parameter. This variable must be included in the data set named by the add_da parameter. This parameter is available if the calltype is FULL or PC.
titles	**Optional.** Specifies the number of title lines (0-4) to be reserved for the user's titles. The remaining title lines are used by the macro. The default value is 0.

3.2 SAS Data Sets Output by the DISTRIB Macro

The DISTRIB macro outputs two data sets. The intake estimates calculated by the MC sub-macro (referred to as the "mcsim" data), and the percentiles and other descriptive data calculated in the PC sub-macro (referred to as the "descript" data).

The name of the mcsim data set is decided by the user in the mcsimda parameter of the DISTRIB macro. Depending on the number of replications and the size of the input date file, this data set can be quite large.

The data output from the Monte Carlo simulations in the MC sub-macro

Variable	Explanation
Subject	The name of the Subject ID variable from the subject parameter.
Mcsim_wt	The original weight divided by the number of simulations.
Numsims	The number of simulations.
Mc_t	The simulated usual intake. For a weekend run, mc_t is the weighted average of the simulated weekday (value 0) usual intake and simulated weekend (value 1) usual intake.
Mc_a	Consumption day amount. An intermediate variable used in the calculation of mc_t.
Mc_p	Probability to consume. An intermediate variable used in the calculation of mc_t.
Mc_p_0	Probability to consume for a weekday (value 0) in a weekend run. An intermediate variable used in the calculation of mc_t.

Variable	Explanation
Mc_p_1	Probability to consume for a weekend day (value 1) in a weekend run. An intermediate variable used in the calculation of mc_t.

The data set output by the PC sub-macro is named using the following conventions:

outlib.descript_food_weight

where:

outlib   is the library name (libname) given in the parameter outlib, and thus depends on user input. If null, the data set is not saved to disk.

descript   This is a literal string and so will not change.

food   The value of the parameter food from the call to DISTRIB. It serves to name the data set, and need not be a food or nutrient.

weight   The name of the variable used in the replicate_var parameter of the previous MIXTRAN run, taken automatically from the parameter file input to DISTRIB. If unweighted this is a null string.

For example a data set where outlib = "mylib", the food parameter="potato" and the weight = "rndw1" the name of the data set will be mylib.descript_potato_rndw1.

The data output from the Percentile Calculations in the PC sub-macro

Variable Type	Explanation
Subgroup	The name of the variable in the subgroup parameter, if any.
Mean_mc_t	The mean value of the simulated usual intake (the variable mc _t) estimated in the Monte Carlo simulations.
Percent_rec_amt	The percent testing true to the recommended amount test (e.g. the percent below the recommend amount of intake.) The variable exists only if the recommended amount parameters were invoked.
numsubjects	The number of subjects.
Tpercentile0-tpercentile100	The percentiles.
Cutprob1-cutprobN	The probability values estimated using the cutoff points, if any. If 10 cutoff points were included then there will be 10 cutprob variables.

4 The INDIVINT Macro

The INDIVINT macro calculates predicted values for regression calibration using methods from Kipnis et al. (Biometrics, 2009) and using results from an amount-only model or a two-part model fit using the MIXTRAN macro. The INDIVINT macro performs adaptive Gaussian quadrature to predict usual intake for each individual, and the macro allows the user to provide a Box-Cox transformation parameter in order to calculate the predicted values on a transformed scale. The results from this macro are intended for use in a subsequent regression model as discussed by Kipnis et al. (Biometrics, 2009).

4.1 Syntax for the INDIVINT Macro

The INDIVINT macro uses the following syntax:

%indivint(model12=, subj1recdata=, recid=, r24vars=, min_amt=, var_u1=, var_u2=, cov_u1u2=, var_e=,
        lambda=, xbeta1=, xbeta2=, boxcox_t_lamt=, lamt=, dencalc=, denopt=, u1nlmix=, u2nlmix=,
        titles=, notesprt=); .

The macro parameters in INDIVINT are keyword parameters and are not positional parameters. These keyword parameters can be specified in any order so long as each value is preceded by the parameter name and the equals sign (e.g. subj1recdata=mydata,). Separate each parameter call with a comma. Parameters that are null can be omitted. The default value of a parameter is null unless otherwise noted in the documentation.

An example of a call to the INDIVINT macro is:

%indivint (model12=2,
        subj1recdata=paramsubj1rec,
        recid=subjectid,
        r24vars=R1 R2,
        min_amt=min_amt,
        var_u1=p_var_u1,
        var_u2=a_var_u2,
        cov_u1u2=cov_u1u2,
        var_e=a_var_e,
        lambda=a_lambda,
        xbeta1=x1b1,
        xbeta2=x2b2,
        boxcox_t_lamt=y,
        lamt=lamt,
        dencalc=y,
        denopt=y,
        u1nlmix=,
        u2nlmix=,
        titles=2,
        notesprt=y);

The following list provides an explanation of each of the parameters in the INDIVINT macro.

model12   Specifies the type of model that was fit prior to calling this macro. A value of 1 indicates that an amount-only model was fit, and a value of 2 indicates that a two-part model was fit where part 1 is the probability part of the model, and part 2 is the amount part of the model.

subj1recdata   Specifies a data set with 1 record for each subject and with required variables discussed by Kipnis et al. (2009) and described in this section. The input data set includes parameter estimates and predicted values obtained by fitting a nonlinear mixed model.

recid   Specifies an identification (ID) variable that uniquely identifies each subject's record.

r24vars   Specifies the 24-hour recall variables with values that are either non-negative or a SAS missing value if the 24-hour recall is not available. Variables must be space delimited as illustrated in the following example: "r24vars=r24hr1 r24hr2". Note for Advanced Users: If all 24-hour recall values are missing for each subject, then the denominator integration should not be performed, so the "dencalc" macro parameter should be specified as "dencalc=n".

min_amt   Specifies a variable that provides the minimum intake amount. This value may be selected as the smallest value among the observed consumption-day amounts. Note that the specified variable provides the same value for each individual. This value will be divided in half and used in the calculations for the numerator integration.

var_u1   Specifies a variable that provides the variance estimate for u1, the random effect from the probability part of the model. If a variable is specified, then the macro will use its value as a diagonal entry of the covariance matrix which is either a 1x1 matrix or a 2x2 matrix depending on the number of random effects that are in the model.

var_u2   Specifies a variable that provides the variance estimate for u2, the random effect from the amount part of the model or from an amount-only model. If a variable is specified, then the macro will use its value as a diagonal entry of the covariance matrix which is either a 1x1 matrix or a 2x2 matrix depending on the number of random effects that are in the model.

cov_u1u2   Specifies a variable that provides the estimate of the covariance of (u1, u2) from the two-part model. If the two-part model was an uncorrelated model, then the specified variable should have a value of zero for every individual's record.

var_e   Specifies a variable that provides the variance estimate for e, the within-person error term from the amount part of the model or from an amount-only model.

lambda   Specifies a variable that provides the estimate of the Box-Cox parameter, lambda, from the amount part of the model or from an amount-only model.

xbeta1   Specifies a variable that provides the linear predictor values calculated using the covariates and estimates of the fixed effects parameters from the probability part of the model.

xbeta2   Specifies a variable that provides the linear predictor values calculated using the covariates and estimates of the fixed effects parameters from the amount part of the model or from an amount-only model.

boxcox_t_lamt   If "boxcox_t_lamt=y" or "boxcox_t_lamt=Y" then individual usual intake will be predicted on a transformed scale where the Box-Cox transformation is used with the Box-Cox parameter value provided by the "lamt" macro parameter. The default value for "boxcox_t_lamt" is "n".

lamt   Specifies a variable that provides the Box-Cox parameter value when "boxcox_t_lamt=y" or "boxcox_t_lamt=Y". The macro does not allow the Box-Cox parameter to be zero.

dencalc   By default, "dencalc=y" so the denominator integration is performed. Note for Advanced Users: If all 24-hour recall variables are missing for each subject, then the denominator integration should not be performed, so the "dencalc" option should be specified as "dencalc=n".

denopt   By default, "denopt=y" so the denominator optimization is performed as part of the denominator integration calculations. Note for Advanced Users: In some situations the denominator optimization is redundant because the empirical Bayes estimates of u1 and u2 are available from the model fitting software; therefore, in these situations, setting "denopt=n" or "denopt=N" allows the macro to skip this optimization step and use the variables provided by the "u1nlmix" and "u2nlmix" macro parameters.

u1nlmix   Specifies a variable for an Advanced Users option. For details, see the description for the "denopt" macro parameter.

u2nlmix   Specifies a variable for an Advanced Users option. For details, see the description for the "denopt" macro parameter.

titles   Specifies the number of title lines to be reserved for the user's titles. One additional title line is used by the macro. The default value is 0.

notesprt   If "notesprt=n" or "notesprt=N" then notes are not printed to the SAS log. The default value for "notesprt" is "y".

5 Examples and Programming Notes

5.1 Examples

The examples discussed in this section can be downloaded from the following website: http://appliedresearch.cancer.gov/diet/usualintakes/macros_single.html. This website includes 4 SAS example programs and output files that illustrate the use of the MIXTRAN, DISTRIB, and INDIVINT macros. Examples 1, 2, and 3 illustrate the use of the MIXTRAN and DISTRIB macros to estimate the distribution of usual intake. Example 4 illustrates the use of the MIXTRAN and INDIVINT macros to evaluate the relationship between fish intake and blood mercury levels.

Example 1. Estimation of the Distribution of Usual Intake of Added Sugar for Age Subgroups Using an Amount-Only Model

Example 1 demonstrates the use of the MIXTRAN and DISTRIB macros for a food consumed nearly every day. The data set includes males with an age of 9 years or older surveyed using a complex sampling design. The primary goal is to estimate the distribution of usual intake of added sugar for age subgroups, adjusting for race in the model, and for sequence and weekend effects. Because added sugar is consumed almost every day by most members of the sample, only the second part of the model is needed; the model type is "amount" and the food of interest is added sugar. The covariates are sequence, weekend, age group, and race. The sequence covariate and the weekend covariate are entered into the macro using the "seq" and "weekend" parameters, respectively. The "replicate_var" parameter is assigned "rndw1" which is the name of the weight variable. An option is added to the NLMIXED procedure call. The parameter and predicted data sets output by MIXTRAN for use in the DISTRIB macro are named "mylib._param_unc_add_sug_male1" and "mylib._pred_unc_add_sug_male1". (The suffix 'male1' is used to indicate the sub-population and replicate variable used.)

The DISTRIB macro calculates the percentiles, cutoff point probabilities, and mean intake, by subgroup level and for all levels combined, and creates a Monte Carlo data set of the calculated intake, named "_mcsim_add_sug_male_rndw1" and a data set of the percentiles and cutoff point probabilities named "mylib.descript_add_sug_male_rndw1". The name of the weight variable used in the MIXTRAN macro is passed to the DISTRIB macro through a saved data set and is used in naming the output data file. A subgroup variable is used in the DISTRIB macro. Distributions are calculated for each subgroup level and for all levels combined. If the subgroup variable is a character variable, then "_overall" is used as a label for the results from all levels combined. In this example, the variable "agegroup" is not a character variable, so the overall group is assigned an "agegroup" value of -255.

Example 2. Estimation of the Distribution of Usual Intake of Total Fruit for Age Subgroups of Children Using a Two-Part Model with Correlated Random Effects

Example 2 demonstrates the use of the MIXTRAN and DISTRIB macros for a food consumed episodically. The data set includes children with an age of 1 to 8 years surveyed using a complex sampling design. The primary goal is to estimate the distribution of usual intake of total fruit for age subgroups, adjusting for race and sex in the model, and for sequence and weekend effects. The model type is "corr" (i.e. correlated), and the food of interest is total fruit cup equivalents. The covariates are sequence, weekend, age group, race, and sex. The sequence covariate

23

and the weekend covariate are entered into the macro using the "seq" and "weekend" parameters, respectively. The "replicate_var" parameter is assigned "rndw1" which is the name of the weight variable. An option is added to the NLMIXED procedure call. The parameter and predicted data sets output by MIXTRAN for use in the DISTRIB macro are named "mylib._param_f_total_child1" and "mylib._pred_ f_total_child1".

The DISTRIB macro calculates the percentiles, cutoff point probabilities, and mean intake, by subgroup level and for all levels combined, and creates a Monte Carlo data set of the calculated intake, named "_mcsim_f_total_child_rndw1" and a data set of the percentiles and cutoff point probabilities named "mylib.descript_f_total_child_rndw1". The name of the weight variable used in the MIXTRAN macro is passed to the DISTRIB macro through a saved data set and is used in naming the "descript" output data file. The "subgroup" parameter is assigned the value "agegroup". Distributions are calculated for each subgroup level and for all levels combined. If the subgroup variable is a character variable, then "_overall" is used as a label for the results from all levels combined. In this example, the variable "agegroup" is not a character variable, so the overall group is assigned an "agegroup" value of -255.

Example 3. Illustration of Programming Techniques that Minimize Effort Required to Estimate Distributions of Usual Intake

Example 3 illustrates two ways of minimizing the time and effort required to estimate distributions of usual intake. The first section minimizes the number of calls to the DISTRIB macro for strata, and the second section demonstrates a way to re-use the data output by the MIXTRAN macro in follow-up calls to the DISTRIB macro. This approach requires only a little SAS coding outside of the macros.

Three strata of the data (i.e. children age 1 to 8 years, males age 9 years and older, and females age 9 years and older) are each run through the MIXTRAN macro separately to get independent estimates. The parameters are nearly the same as those used in Example 1. Notice that the subgroup variable is carefully coded so there are no overlapping values between strata. This fact will be important later. The "foodtype" parameter in the call to the MIXTRAN macro is used to create output data files with distinct names. In this example the values of "foodtype" are "add_sug_child1", "add_sug_male1", and "add_sug_female1". The MIXTRAN macro created the data files: "mylib._param_unc_add_sug_child1"; "mylib._pred_unc_add_sug_child1"; "mylib._param_unc_add_sug_male1"; "mylib._pred_unc_add_sug_male1"; "mylib._param_unc_add_sug_female1"; and "mylib._pred_unc_add_sug_female1".

In the SAS program, after the successful execution of all three calls to the MIXTRAN macro but prior to calling the DISTRIB macro, the parameter files for the three strata are concatenated, and the predicted data files are also concatenated. The variable "stra" is coded in both the concatenated parameter file and the concatenated predicted file to designate the appropriate stratum for each record.

The DISTRIB macro is then called, in a manner very similar to the call in Example 1, but this time the "byvar" parameter is invoked for the variable "stra" (i.e. "byvar=stra"). The parameter and predicted data sets will be merged by the variable "stra" thus ensuring that the appropriate parameter estimates are attached to each record in the predicted data set.

The DISTRIB macro will calculate the count, percentiles, cutoff point probabilities, and mean for each level of the subgroup, and for all subjects combined, and save the data in one descriptive file. In this case the name would be "mylib.descript_add_sug_all_rndw1". The DISTRIB macro also creates the data set of estimated intakes named "mylib._mcsim_add_sug_rndw1".

The next section of Example 3 demonstrates an instance of avoiding unnecessary calls to the MIXTRAN macro, by using the saved parameter and predicted data sets in new calls to the DISTRIB macro.

In the SAS program the data sets saved by MIXTRAN for the male stratum are subset to males age 19 years and older. The variable "stra" is created and assigned a value of 2 in both the parameter and predicted data sets for this subset.

The DISTRIB macro is then called. The subgroup option is omitted because only the combined distribution for all males age 19 years and older is required. However, the "byvar" parameter must be invoked with the variable "stra" which is the stratification variable. The reason for this requirement is that the weights and subject counts have to be recalculated for the subset of the population. The value of the "food" parameter is changed, so the previous descriptive data set is not overwritten. In the example, the data set output by this execution of DISTRIB is called "mylib.descript_add_sug_m19_rndw1". The data contains the counts, mean, percentiles, and cutoff point probabilities for intake of added sugar by males age 19 years and older. Note, that the descriptive data set will not include a subgroup variable. If the descriptive data set is later concatenated with the first descriptive data set produced using subgroup as above, the data from this file will need to be assigned a distinct value for the subgroup variable.

Similarly the data saved by the MIXTRAN macro for the female stratum is subset to females age 19 years and older. The variable "stra" is created and assigned a value of 3.

The DISTRIB macro is called, with no subgroup parameter, and the byvar set to "byvar=stra". The food parameter is changed, and another data set of descriptive statistics is saved. It will contain the count, mean, percentiles, and cutoff point probabilities for intake of added sugar by females age 19 years and older.

Example 4. Modeling the Relationship between Fish Intake and Blood Mercury in Males Age 12 Years and Older Using NHANES Data

Example 4 illustrates the use of the MIXTRAN and INDIVINT macros to assess the relationship between fish intake and blood mercury levels among males age 12 years and older using NHANES data. To estimate the difference in log mercury between males with an average of 0.1 oz and 1.0 oz of fish per day, the calculations involve fitting a nonlinear mixed model using the MIXTRAN macro with weights. After using the results of this fit as input for the INDIVINT macro, the INDIVINT macro results are used as the covariate in a linear regression model for log mercury.

5.2 Programming Notes

- Because of the complexity of the model, the MIXTRAN macro can require considerable computing time. When the model type is **amount** (i.e. "**modeltype=amount**"), we have seen that the macro generally requires less than 10 minutes of computing time for NHANES data. When the model type is **uncorrelated** (i.e. "**modeltype=nocorr**") or **correlated** (i.e. "**modeltype=corr**"), however, the required computing time can be considerably longer; in analyzing NHANES data, we have seen correlated models require from as little as 20 minutes to 2 hours or longer.

- Check the log file for lines beginning with "##". These lines act as informal documentation of the parameters used in the MIXTRAN and DISTRIB macros.

- Base runs are defined as being a call to the MIXTRAN macro with no user supplied starting values for the three calls to the SAS NLMIXED procedure; therefore, start_val1, start_val2, and start_val3 will not be assigned a value. To obtain variance estimates using BRR, the same models are run numerous times with different weights. Instead of letting the MIXTRAN macro calculate the starting values for each call to the SAS NLMIXED procedure, the parameter estimates output by the SAS NLMIXED procedure from the first run (i.e. base run) can be used as starting values.

- It is possible to re-run the DISTRIB macro without re-running the MIXTRAN macro once the MIXTRAN macro has successfully executed. The data sets needed for input to the DISTRIB macro have been saved and are available for use at any time. For example, if the user decides to try different cutoff points (i.e. cutpoints) in the DISTRIB macro, it would not be necessary to re-run the MIXTRAN macro. However, please be absolutely sure to use the proper data sets as input to the DISTRIB macro. Example 3, discussed in the previous section, uses the DISTRIB macro without re-running the MIXTRAN macro.

- At the end of the DISTRIB macro execution, the data sets in the work library are deleted. To keep the data set _mcsim1 to use for further analysis in the program, look for the phrase "to keep _mcsim1 for further analysis" in the DISTRIB macro code. This search will find the correct line, which can then be commented out. Since _mcsim1 contains all the simulated records, it might use a lot of storage space.

- **WARNING: SAS VERSION 9 BUGS**

  - In SAS version 9 there is a bug in the SAS NLMIXED procedure.
    If the effect names in the model become too long and too numerous, the procedure will never complete. It has not been possible to discover a definitive limit to the size of the model. This problem has not been encountered in SAS version 8.

  - According to SAS technical support as of 5/2/2007
    "…there is no definitive answer to the size of the model (in terms of pure symbol or character storage) that PROC NLMIXED can handle. NLMIXED has to store derivatives of the model behind the scenes and those derivatives can double or even quadruple the size of the "symbol space" needed to process the model. It is best to stick with shorter effect names when creating a model in NLMIXED, at least until SAS 9.2 comes out. Tech Support does not have an official release date for that version of SAS yet."
    We note that SAS 9.2 has been released; however, we still provide this information to give users a more complete understanding of macro performance in various versions of SAS.

  - Note that the INDIVINT macro requires SAS IML. The SAS Institute has reported an error that can occur when running SAS IML in SAS 9.2 TS1MO - the error relates to variables with missing values. SAS has provided a Hot Fix for this issue. Also, the problem is fixed in SAS 9.2 TS2M2, and this error is not encountered in SAS 9.1.3.

6 References

Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF, Tooze JA, Carroll RJ, Freedman LS. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics* 2009 Dec;65(4):1003-10.

Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ, Kipnis V. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc* 2006 Oct;106(10):1575-87.

Tooze JA, Kipnis V, Buckman DW, Carroll RJ, Freedman LS, Guenther PM, Krebs-Smith SM, Subar AF, Dodd KW. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. *Stat Med* 2010 Nov 30;29(27):2857-68.