

User's Guide for Analysis of Usual Intakes

May 2009

For use with versions 1.1 of the Mixtran, Distrib, and Indivint SAS^a macros

Ruth Parsons¹
Stella S. Munuo¹
Dennis W. Buckman¹
Janet A. Tooze²
Kevin W. Dodd³

^a SAS Institute, Inc., Cary, NC, USA

¹ Information Management Services, Inc., Rockville, MD, USA

² Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA

³ Biometry, Division of Cancer Prevention, National Cancer Institute, Rockville, MD, USA

Contents

	Page
1 Introduction	3
1.1 Selecting the Appropriate Macro	3
1.2 Standard Errors	4
2 Mixtran Macro	4
2.1 Including and Calling the Mixtran Macro in the SAS Program	4
2.2 Mixtran Macro Parameters	4
2.3 SAS Data Sets Saved by the Mixtran Macro	7
2.4 Variable Names for the Covariates in the Mixtran Macro	9
3 Distrib Macro	10
3.1 Including and Calling the Distrib Macro in the SAS Program	10
3.2 Distrib Macro Parameters	10
3.3 SAS Data Sets Saved by the Distrib Macro	12
4 Indivint Macro	13
4.1 Including and Calling the Indivint Macro in the SAS Program	13
4.2 Indivint Macro Parameters	13
5 Examples and Programming Notes	15
5.1 Examples	15
5.2 Programming Notes	18
6 References	19

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 foods and nutrients. 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: riskfactor.cancer.gov/diet/usualintakes/macros.html.

The files **mixtran_macro_v1.1.sas**, **distrib_macro_v1.1.sas**, and **indivint_macro_v1.1.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 is 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 intake for an episodically consumed food or nutrient, or these macros can be used to fit an amount-only model and estimate the distribution of 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. An extension to a multivariate case with several foods and nutrients ... is another area for future research."

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. We plan to provide, in the future, an additional macro that can perform BRR calculations.

Caution: Note that the standard errors and p-values output by the Mixtran macro are only valid for an unweighted analysis (i.e. analysis of a simple random sample). In a weighted analysis (e.g. analysis of NHANES data), calculation of these standard errors requires additional programming to implement a replication method such as BRR.

2 Mixtran Macro

2.1 Including and Calling the Mixtran Macro in the SAS Program

To use the Mixtran macro in a SAS program, insert the following **%include** statement into your program prior to calling the macro:

```
%include "/mypath /mixtran_macro_v1.1.sas" ;
```

where "/mypath/" is the path name to the directory in which the file **mixtran_macro_v1.1.sas** is stored. After this **%include** statement, the Mixtran macro is called using the syntax:

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

2.2 Mixtran Macro Parameters

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

data	Mandatory. 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	Mandatory. Specifies the 24-hour recall variable for the food or nutrient.						
foodtype	Mandatory. 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.						
subject	Mandatory. Specifies the variable that uniquely identifies each subject.						
repeat	Mandatory. 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.						
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.						
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.						
outlib	<p>Mandatory. 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:</p> <pre style="margin-left: 40px;">LIBNAME mylib "/myoutputpath/";</pre> <p>Then the syntax in the macro call is:</p> <p style="margin-left: 40px;">outlib=mylib.</p>						
modeltype	<p>Mandatory. 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:</p> <table border="0" style="margin-left: 40px;"> <tr> <td style="padding-right: 20px;">"corr"</td> <td>All 3 sections of the macro will be executed.</td> </tr> <tr> <td>"nocorr"</td> <td>The probability and amount sections of the macro will be executed, but the correlated section will not be executed.</td> </tr> <tr> <td>"amount"</td> <td>Only the amount section of the macro will be executed.</td> </tr> </table>	"corr"	All 3 sections of the macro will be executed.	"nocorr"	The probability and amount sections of the macro will be executed, but the correlated section will not be executed.	"amount"	Only the amount section of the macro will be executed.
"corr"	All 3 sections of the macro will be executed.						
"nocorr"	The probability and amount sections of the macro will be executed, but the correlated section will not be executed.						
"amount"	Only the amount section of the macro will be executed.						
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.
seq	<p>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) can 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.</p> <p>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.</p>
weekend	Optional. Specifies the weekend (Fri.-Sun.) indicator variable to account for a weekend effect. A value of 1 represents a Fri.-Sun. record, and a value of 0 represents a Mon.-Thurs. record. This variable can not also appear in the "covars_prob" or "covars_amt" parameters.
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 weekend variable can be used.
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.
subgroup	Optional. Specifies one categorical variable used for the calculation of a separate usual intake distribution for each subgroup. This variable can be created from a combination of other variables (e.g. age and sex), but all variables used to define the subgroup variable must also be among the covariates in the model. The subgroup variable is used in the Distrib macro; however, the subgroup variable must be in the call to the Mixtran macro because preparatory calculations are performed in the Mixtran macro.

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 NL MIXED (i.e. NL MIXED 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 NL MIXED (i.e. NL MIXED 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 NL MIXED (i.e. NL MIXED for correlated model).
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.
nloptions	Optional. Specifies a list of options to be added to the PROC NL MIXED statement for all calls to PROC NL MIXED, for example: nloptions=qpoints=1 gconv=1e-12 itdetails. 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 NL MIXED procedure. Printlevel=2 is the default value. Printlevel=3 prints summary reports and output from all of the statistical procedures.

2.3 SAS Data Sets Saved by the Mixtran Macro

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 NL MIXED 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 also 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.

2.4 Variable Names 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 amount and probability parts of the model, and these variables need to be differentiated during the 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 "agegrp2", it will become "A02_AGEGRP2",

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 intercept is always named "A01_INTERCEPT" or "P01_INTERCEPT" for the amount and probability intercepts, respectively.

3 Distrib Macro

3.1 Including and Calling the Distrib Macro in the SAS Program

To use the Distrib macro in a SAS program, insert the following **%include** statement into your program prior to calling the macro:

```
%include "/mypath /distrib_macro_v1.1.sas" ;
```

where "/mypath/" is the path name to the directory in which the file **distrib_macro_v1.1.sas** is stored. After this **%include** statement, the Distrib macro is called using the syntax:

```
%Distrib (seed=, nsim_mc=, modeltype=, pred=, param=,
          outlib=, cutpoints=, ncutpnt=, byvar=, subgroup=,
          subject=, titles=, food=);.
```

3.2 Distrib Macro Parameters

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

seed	Mandatory. Specifies the seed for the random number generator used for the Monte Carlo simulation of the random effects u1 and u2.
nsim_mc	Mandatory. Specifies the number of repetitions to be used in the Monte Carlo simulation. For each subject, one record will be output for each repetition.
modeltype	<p>Mandatory. 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:</p> <p>"corr" or null fit two-part correlated random effects model, "nocorr" fit two-part model with independent random effects, "amount" fit amount-only model.</p> <p>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 labeled with "_unc" can be used as input for the Distrib macro.</p>

- pred** **Mandatory.** 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** **Mandatory.** 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.
- outlib** **Mandatory.** 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.
- cutpoints** **Optional.** Specifies one or more cutpoints for which to calculate the proportion of the population below the cutpoint. Each cutpoint value must be separated by a single space. If no cutpoints 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.
- byvar** **Optional.** Specifies a list of by-variables that are in the parameter and predicted data sets, indicating that the Mixtran model was fit separately for each by group. That is, a separate call to the Mixtran macro was made for each by group value. For example if males and females were each passed through the Mixtran macro separately, the parameter and predicted data sets for each could be concatenated, merging the appropriate parameter estimates to the predicted values, and passed to the Distrib macro together. The Distrib macro will produce distributions for the entire population, not distributions within

each bygroup. Use the "subgroup" parameter to obtain distributions for subpopulations. The "byvar" and "subgroup" parameters can be used together.

- 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 also be in the call to the Mixtran macro because preparatory calculations are performed in the macro.
- subject** **Optional.** (Required when "weekend" is used in Mixtran). Specifies the variable that uniquely identifies each subject.
- 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.
- food** **Optional.** Specifies a name for the analysis, used to label the output data set.

3.3 SAS Data Sets Saved by the Distrib Macro

The Distrib macro outputs one SAS data set that contains descriptive statistics for usual intake. These statistics include the mean and percentiles from the estimated distribution of usual intake and optionally the cutpoint probabilities. This data set consists of one record summarizing the results from all subjects in the data set and one additional record for each level of the subgroup variable. The data set is named using the following conventions:

outlib.descript_food_freq_var

described as follows.

- outlib** is the library name specified in the Distrib parameter "outlib".
- descript** is the name used to distinguish the distribution data set and is always "descript".
- _food** is obtained from the "food" parameter value specified in the call to the Distrib macro.
- _freq_var** is obtained from the "replicate_var" parameter value specified in the call to the Mixtran macro. If no replicate variable was specified, this value will be a null string. Note that the specified value is saved in the parameter data sets and is therefore available to the Distrib macro.

For example, if the library reference is "mylib", the food is "add_sug", and the weight variable is "rndw1", the output data set would be:

mylib.descript_add_sug_rndw1.

4 Indivint Macro

4.1 Including and Calling the Indivint Macro in the SAS Program

To use the Indivint macro in a SAS program, insert the following **%include** statement into your program prior to calling the macro:

```
%include "/mypath /indivint_macro_v1.1.sas" ;
```

where "/mypath/" is the path name to the directory in which the file **indivint_macro_v1.1.sas** is stored. After this **%include** statement, the Indivint macro is called using the 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=); .
```

4.2 Indivint Macro Parameters

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 obtained by considering all of the available consumption-day amounts and using the smallest value. 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. The macro does not allow the Box-Cox parameter to be zero.
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: riskfactor.cancer.gov/diet/usualintakes/macros.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 need; 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 "subgroup" parameter is assigned the value "agegroup" and 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" and "mylib._pred_unc_add_sug".

The Distrib macro calculates the percentiles, cutpoint probabilities, and mean intake, by subgroup level and for all levels combined, and writes out a data set named "mylib.descript_add_sug_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. The same subgroup variable that was named in the call to the Mixtran macro 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 White Potatoes 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 white potatoes 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 white potato cup equivalents. The covariates are sequence, weekend, age group, race, and sex. The sequence covariate and the weekend covariate are entered into the macro using the "seq" and "weekend" parameters, respectively. The "subgroup" parameter is assigned the value "agegroup" and 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_v_potato" and "mylib._pred_v_potato".

The Distrib macro calculates the percentiles, cutpoint probabilities, and mean intake, by subgroup level and for all levels combined, and writes out a data set named "mylib.descript_v_potato_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. The same subgroup variable that was named in the call to Mixtran is used in Distrib. 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_child", "add_sug_male", and "add_sug_female". The Mixtran macro created the data files: "mylib._param_unc_add_sug_child"; "mylib._pred_unc_add_sug_child"; "mylib._param_unc_add_sug_male"; "mylib._pred_unc_add_sug_male"; "mylib._param_unc_add_sug_female"; and "mylib._pred_unc_add_sug_female".

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, cutpoint 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 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 cutpoint 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 cutpoint 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 to 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 cutoffs (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 BUG**
 - 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.

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; DOI: 10.1111/j.1541-0420.2009.01223.x.

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; 106:1575-1587.