

**Paper 147-29****A SAS® Macro for Generating Randomization Lists in Clinical Trials Using Permuted Blocks Randomization**

Mazen Abdellatif, M.S., Hines VA CSPCC, Hines IL, 60141, USA

**ABSTRACT**

We developed a SAS [1] macro to standardize and streamline the generation of randomization lists for treatment assignment in randomized clinical trials using permuted blocks randomization. The macro uses the CALL RANUNI [2] function and PROC PLAN [3] to generate such lists. It prompts the user for a set of parameters that define the trial's randomization scheme specifications and another set of parameters that control the macro execution.

The design of this macro along with sample output generated for a hypothetical randomized clinical trial will be presented.

The macro was developed using SAS Proprietary Software Release 8.2 (TS2M0) running on ALPHASERVER Model 1200 5/533 4MB Serial Number 80000000 and executed on the OpenVMS 7.3-1 platform, using eXcursion Server.

**INTRODUCTION**

Randomization schemes for randomized clinical trials depend on the type of the randomization selected (e.g. simple, random block, urn, adaptive), trials' stratification factors, sample sizes, treatment arms, and whether treatment arms are balanced or not. Multi-site trials are usually stratified on participating sites, and often other factors in addition. Various calculations, based on the trial's randomization scheme, have to be performed beforehand to determine the nature and size of the required randomization lists.

Before this macro was developed, we used a set of five SAS programs executed consecutively to generate the randomization lists using permuted blocks randomization. Although these programs take advantage of SAS/MACRO for some automation, manual calculations based on the trial's randomization scheme were still needed to determine the nature and size of the needed randomization lists. In addition, various places in each program have to be modified to reflect the various parameters of the randomization scheme and other control variables. Table 1 lists these five programs and the outputs they generate.

These two reasons were the main motivation for developing this macro. Essentially, the macro combines steps from the first four programs, but they are modified so that the macro determines the nature and size of the needed randomization lists, and eliminates the need for program editing to encompass the randomization scheme of a new trial. These two tasks were accomplished by defining various input and control parameters.

Table 1 the original five SAS programs.	
Program	Output
Drug.sas	A text file of random numbers for treatment assignments.
Random.sas	A text file of random numbers for randomization numbers.
Sites.sas	A text file of sequentially assigned participating sites and their generated treatment assignments file [from Drug.sas] and randomization numbers file [from Random.sas].
Output.sas	A structured text file of the formatted participating sites and their generated randomization numbers and treatment assignments [from Sites.sas].  The randomization lists.
Envelop.sas	A formatted output list of drug codes envelopes for un-blinding.

## THE MACRO DESCRIPTION

The macro prompts the user to enter the parameters needed to generate the randomization lists. This is done through a %WINDOW statement that displays these input parameters in the Display Manager as shown in Figure 1.

SAS: MAIN  
Command ==>

```

Number of sites:      ->      |
Patients/site/stratum: ->    |
First block size:    ->    |
Second block size:   ->    |
Seed for random block sizes: -> |
Seed for random treatment drug: -> |
Number of treatment arms: -> |

Number of rand. groups (9999): -> |
Number of rand.id per site (9999): -> |
Sites (#####*):-> |
-----|
Stratum output list file: -> |
Stratum output flat file: -> |
List title: -> |
Stratum title: -> |
QUIT? (y/n)-----> |

```

Figure 1 The macro input parameters in the display manager.

Table 2 lists these parameters and their usage in the macro. Those of the trial's randomization scheme are identified by a "\*" under their macro variables. The others control the macro execution.

<b>Table 2 the macro input parameters</b>	
<b>Macro Var.</b>	<b>Prompt Description</b>
Sites *	Number of sites: The number of participating sites
Perh *	Patients/site/stratum: The sample size for each site within each stratum.
Blka *	First block size: The number represents the first block size.
Blkb *	Second block size: The number represents the second block size.
bseed	Seed for block sizes: The initial seed used in "CALL RANUNI" to randomly generate block sizes.
rseed	Seed for treatment: The initial seed used in PROC PLAN to randomly generate treatment assignments.
Drug *	Number of arms: Number of treatment arms.
rdrug	Number of randomization groups: The maximum number of treatment assignments that the macro generates.
rperh	Number of randomization groups: The maximum number of randomization numbers that the macro generates.
Slist *	Sites (###*###*###*): The list of site numbers separated by *.
sof	Stratum list file: The name of the output file that contains the formatted randomization lists.
soff	Stratum flat file: The name of the output file that contains the flat records of the site numbers, randomization numbers, and treatment assignment.
ttitle	List title: The title for the randomization lists.
strtm	Stratum title: The title for the current stratum.

The macro uses a format library of site numbers and treatment regimens to incorporate the actual site names and treatment regimen names in the randomization lists. Figure 2 shows the code that creates these formats.

After the macro generates the randomization lists and the flat file, it displays another prompt in the display manager for the user to indicate whether to generate more lists or not. This is shown in Figure 3. If the user enters 1 for "More", the macro displays the previous display manager parameter input view, otherwise the macro stops executing.

```
Libname library '[abdellat.quiq_rand.hypo]';

PROC FORMAT library=library;

VALUE $HOSPFMT '105'='105 Site 1'
              '215'='215 Site 2'
              '217'='217 Site 3'
              '324'='324 Site 4'
              '445'='445 Site 5'
              '611'='611 Site 6'
              '623'='623 Site 7'
              '655'='655 Site 8'
              '665'='665 Site 9'
              '729'='729 Site 10'
              '830'='830 Site 11';

VALUE TRMTFMT 1='Drug A'
              2='Drug B'
              3='Placebo'; run;
```

Figure 2 the site and drug formats.

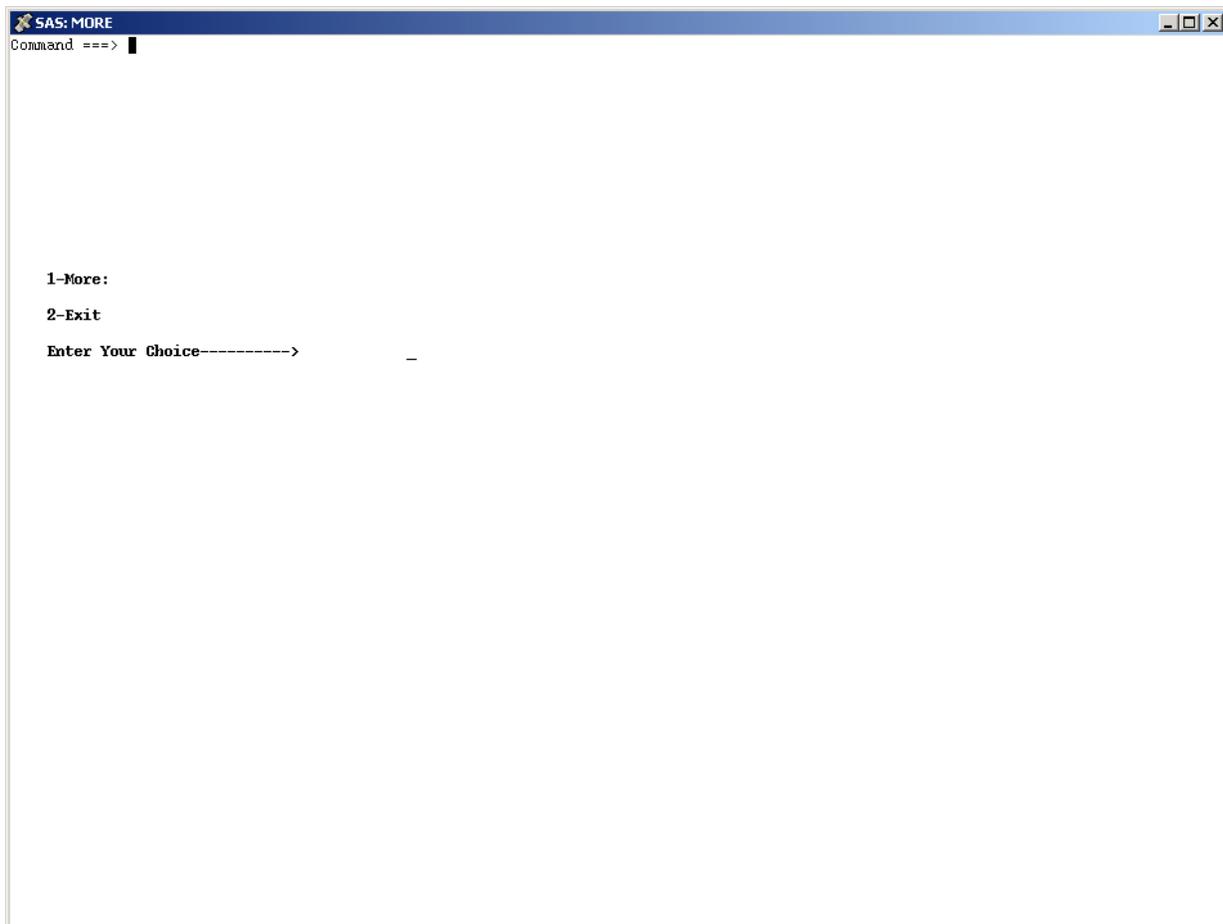


Figure 3 the "More" prompt in the display manager.

The two main building blocks of the macro are the SAS uniform distribution function CALL RANUNI, which is used to create blocks of randomly assigned sizes and PROC PLAN, which is used to generate treatment assignments based on the created blocks. The macro creates enough blocks to ensure the generation of the required randomization numbers and treatment assignments for the desired randomization list.

In addition to the code that constructs and displays the input parameters, the macro consists of five main parts. Table 3 lists these parts and their functions.

<b>Part</b>	<b>Function</b>
Part 0	Housekeeping: The major macro variables are defined and initialized.
Part 1	Treatment Assignments: Generates the treatment assignments data set "DRUGNUM".
Part 2	Randomization Numbers: Generates the randomization numbers data set "RANNUM".
Part 3	Site Assignment: Assigns the generated treatment assignments and randomization numbers to the participating sites by merging "DRUGNUM" and "RANNUM" data sets to create the "STRATUM" data set.
Part 4	Final Output Generation: Generates the randomization lists and the flat file for all participating sites from the "STRATUM" data set.

### **Part 0: Housekeeping**

In the housekeeping section, the major macro variables are defined, and initialized.

```
%global sites perh blk a blk b bseed rseed;

%global drug max size c tot obs block;
%global rdrug rmax rsize rc rtot robs rblock;

%global hds slist solf soff;
%let max=%eval(&sites*&perh);

%let size=0;
%let c=0;
%let tot=0;
%let obs=0;
%let block=0;

%let rmax=%eval(&sites*&rperh);

%let rsize=0;
%let rc=0;
%let rtot=0;
%let robs=0;
%let rblock=0;
```

In addition, the filenames for the randomization list and flat files are assigned. The assignment assumes these files are saved in a fixed directory. This can be modified by adding an additional input parameter to specify the desired output directory

```
filename l "[abdellat.quiq_rand]&self";
filename flat "[abdellat.quiq_rand]&soff";
```

### Part 1: Treatment Assignments

This is where the macro generates the treatment assignments data set. The following is the pseudo-code for this section:

- A. Create a permanent empty PLAN data set.
- B. Figure out the number of needed blocks based on the two specified sizes.
- C. Create a data set of the needed blocks with sizes randomly assigned to either of the two specified sizes.
- D. Print the created blocks data set.
- E. Generate treatment arms from each generated block by stepping through the following loop iterations:
  1. Fetch the next block.
  2. Calculate the number of generated drug assignments from the block.
  3. Run PROC Plan.
  4. Add the generated PLAN dataset to the permanent PLAN dataset.
  5. Hash a new seed for the “rseed” macro variable.
  6. Re-evaluate the number of generated treatment assignments accumulated in the “size” macro variable with reference to what is needed, which is stored in the “max” macro variable. If the value of “size” is less than the value of “max”, go to 1. Otherwise go to 7.
  7. Create the treatment assignment data set “DRUGNUM” from the permanent PLAN data set assigning a unique ID “ID” to each treatment assignment observation.

### Part 2: Randomization Numbers

The steps of this section are similar to those of Part 1. However, the macro here generates the randomization numbers data set “RANNUM” using the “rdrug” and “rperh” input parameters. It re-evaluates the number of generated randomization numbers accumulated in the “rsize” macro variable with reference to what is needed, which is stored in the “rmax” macro variable. As in Part 1, unique ID “ID” is assigned to each randomization number observation.

### Part 3: Site Assignment

This section assigns the generated treatment assignments and randomization numbers to the participating sites. Its steps are summarized in the following pseudo-code:

- A. Construct a macro variable to hold a dataset name for each site.
- B. Create a randomization number dataset for each site from the randomization numbers data set “RANNUM”.
- C. Append all site randomization number datasets into “STRATUM” data set.
- D. Merge “RANNUM” and “STRATUM” data sets by “ID” giving the “RANDDRUG” dataset.

#### Part 4: Final Output Generation

This section outputs the randomization lists and corresponding flat file from the “RANDDRUG” dataset to the file names specified in the “solf” and “soff” macro variables. These are done through PUT statements.

#### A HYPOTHETICAL TRIAL

The hypothetical trial named the “Hypo Trial” is a multi-site randomized placebo controlled trial where ten sites will participate. The calculated sample size for the trial is 600 patients. Patients will be randomized to one of three treatment regimens: Drug A, Drug B, or Placebo. Patients will be stratified by participating sites and age of  $\leq 50$  or  $> 50$  years. Randomization will be balanced across treatment regimens and strata. Table 4 shows the input parameter derived values based on this design.

Figure 4 shows the display manager parameter input view with the entered values. Figures 5 and 6 show a sample page of the generated randomization lists from the first stratum, and a sample from the generated flat file respectively.

As seen in Figure 5 below, the randomization list shows the trial’s name, the stratum, the site number/name, and the page number as the page header. The randomization information includes the randomization number and treatment assignment.

For each randomization number, the list includes spaced-columns to record the randomized patient number, date of randomization, the initials of the staff giving the randomization number, and code break information including the break date and comments.

**Table 4 the input parameters’ derived values.**

Macro Var.	Value
Sites	10
Perh	30
Blka	3
Blkb	6
bseed	Any
rseed	Any
Drug	3
rdrug	8888
rperh	8888
Slist	101*...110*
solf	HIP01L
soff	HIP01F
t11	The Hypo Trial
strtm	Stratum: Age $\leq 50$

Note:

By convention, the “Rdrug” and “rperh” macro variables are assigned the same value for each stratum.

They control the maximum number of treatment assignment and the maximum number of randomization numbers that the macro generates, consecutively.

SAS: MAIN  
Command ==>

```

Number of sites:      ->      10
Patients/site/stratum: ->    30
First block size:    ->      3
Second block size:   ->      6
Seed for random block sizes: -> 456354
Seed for random treatment drug: -> 675653
Number of treatment arms: -> 3

Number of rand. groups (9999): -> 8888
Number of rand.id per site (9999): -> 8888
Sites (#####*): -> 105*215*217*324*445*611*623*655*665*729*830*

Stratum output list file: -> HYP01L
Stratum output flat file: -> HYP01F
List title: -> THE HYPO TRIAL
Stratum title: -> STRATUM: Age <=50

QUIT? (y/n)-----> n

```

Figure 4 the entered input parameters.

HYP01L - Microsoft Word

File Edit View Insert Format Tools Table Window Help Acrobat

100% Courier New 10 B I U

RANDOMIZATION LIST FOR The HYPO TRIAL PAGE: 1

\*\*\*\*\* Stratum: Age <=50 \*\*\*\*\*

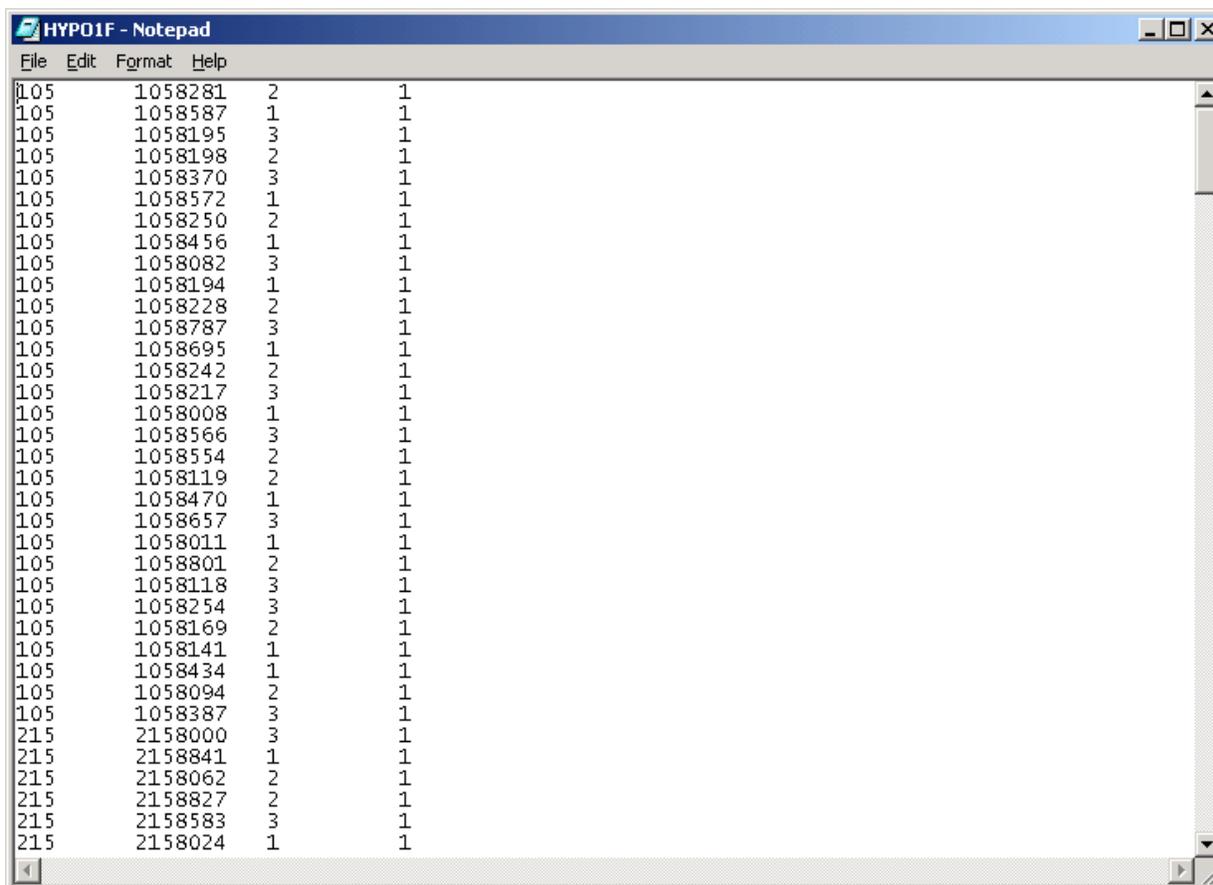
HOSPITAL 105.; 105 Site 1

PATIENT NUM.	RAND.#	TREATMENT	CODE BREAK INFORMATION			
			DATE OF RAND. MM/DD/YY	YOUR INIT.	DATE MM/DD/YY	COMMENT
_____	1058281	Drug B	___/___/___	___	___/___/___	_____
_____	1058587	Drug A	___/___/___	___	___/___/___	_____
_____	1058195	Placebo	___/___/___	___	___/___/___	_____
_____	1058198	Drug B	___/___/___	___	___/___/___	_____
_____	1058370	Placebo	___/___/___	___	___/___/___	_____
_____	1058572	Drug A	___/___/___	___	___/___/___	_____
_____	1058250	Drug B	___/___/___	___	___/___/___	_____
_____	1058456	Drug A	___/___/___	___	___/___/___	_____
_____	1058082	Placebo	___/___/___	___	___/___/___	_____
_____	1058194	Drug A	___/___/___	___	___/___/___	_____

Page 1 Sec 1 1/16 At 1" Ln 1 Col 1 REC TRK EXT OVR

Figure 5 a partial randomization list.

The flat file contains the site number, randomization number and drug number of the specified stratum for all the sites.



Site Number	Randomization Number	Drug Number
105	1058281	2
105	1058587	1
105	1058195	3
105	1058198	2
105	1058370	3
105	1058572	1
105	1058250	2
105	1058456	1
105	1058082	3
105	1058194	1
105	1058228	2
105	1058787	3
105	1058695	1
105	1058242	2
105	1058217	3
105	1058008	1
105	1058566	3
105	1058554	2
105	1058119	2
105	1058470	1
105	1058657	3
105	1058011	1
105	1058801	2
105	1058118	3
105	1058254	3
105	1058169	2
105	1058141	1
105	1058434	1
105	1058094	2
105	1058387	3
215	2158000	3
215	2158841	1
215	2158062	2
215	2158827	2
215	2158583	3
215	2158024	1

Figure 6 the flat file for the first stratum.

## DISCUSSION

The macro limits the number of defined block sizes to two, which are entered in the “blka” and “blkb” input parameters. Values selected for the block sizes should be divisible by the number of treatment arms. For example, if there are two treatment arms, the numbers 2 and 4 can be selected; for three treatment arms, the numbers 3 and 6 can be selected. The fixed block size can be forced by entering the same block size for both parameters.

The drug format can be used to control for unbalanced randomization schemes. For example, if the number of treatment arms is two, but we want to randomize twice as many patients to the first treatment arm as to the second, we enter 6 in the “Drug” parameter so that the macro generates 6 treatment arms. Then in the treatment regimen format statement, we can assign the drugs 1,2,3, and 4 to the first treatment arm and the drugs 5 and 6 to the second treatment arm.

The generated randomization number is a combination of two parts. The first part is the three digits representing the site number. The second part is the randomly generated four-digit number. The first digit of this number always starts with the first digit of the number entered in the “rperh” input parameters. The number 8888 was entered for the first stratum, so the four-digit randomly assigned number always starts with 8 for this stratum. For the second stratum, the number 9999 would be a good choice. By convention, the formula (10-stratum numbers) is used to determine the values of “rperh”. For example, if there are 5 strata, the input values of these parameters would be 5555, 6666, 7777, 8888, and 9999 for the first, second, third, fourth, and fifth stratum respectively.

The list contains pages for the given stratum for all participating sites. The randomization lists and the flat file for the second stratum can be generated with a second iteration of the macro or by a new execution. The advantage of the former is that the input parameters remain populated with the earlier entered values of the first stratum. Thus the user would only need to enter values for the ones that are stratum-dependent.

These lists are kept at the Coordinating Center (CC). When a Site Coordinator (SC) calls the CC to randomize a patient, if the patient is qualified, the next randomization number from the site’s specified stratum is given and relevant information are recorded on the list.

The design of the randomization numbers minimizes the error of giving randomization numbers from the wrong site or the wrong stratum. SCs know their site numbers and are told the 4-digit number range of each stratum at the beginning of the trial. For example, if the randomization number of 1058281 is inadvertently given to a SC from site 115, the SC should catch that error and inform the CC staff member who gave the number. Similarly, if the SC from site 105 was randomizing a patient to the first stratum and the number 1059876 was given, the SC should catch that error. Of course this can be avoided by the use of automated randomization systems. However, there will always be the need for the person-to-person randomization method as backup.

## CONCLUSIONS

Development time has been drastically reduced. While it used to take an average of six hours to generate the randomization lists for a new trial, it takes an average of twenty minutes using this macro. It took only ten minutes to generate the lists and the flat files for the two strata of the hypothetical trial.

The macro has recently been developed. We anticipate further enhancements to increase its flexibility and functionality. Future efforts will focus on generating drug code envelopes.

This homogeneous approach is much more effective and straightforward than the previous approach. It limits the editing required by each trial to the PROC FORMAT program.

## REFERENCES

1. SAS/MACRO is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® Indicates USA registration.
2. SAS/BASE is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® Indicates USA registration.
3. SAS/STAT is a registered trademark or trademark of SAS Institute Inc. in USA and other countries. ® Indicates USA registration.

## ACKNOWLEDGMENT

I would like to acknowledge Dr. Domenic J. Reda, Hines VA CSPCC, and Dr. Robert Anderson, Hines VA CSPCC for their insightful review of this paper.

## CONTACT INFORMATION

Your questions and comments are valued and encouraged.  
Contact the author at:

[abdellat@research.hines.med.va.gov](mailto:abdellat@research.hines.med.va.gov)

Send an e-mail message to the author indicating your name, job title, and affiliation if you would like an electronic file of the macro.

SAS and all other SAS Institute, Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® Indicates USA registration. Other brand and product names are trademarks of their respective companies.