

Manipulating Clinical Research Data with the TRANSPOSE Procedure in SAS/ASSIST® Software

Stanley Cron, Center for AIDS Research, Baylor College of Medicine, Houston, Texas

INTRODUCTION

The analysis of clinical research data occasionally presents problems due to the arrangement of variables in the data set. One common dilemma involves obtaining a frequency distribution for a multiple response item on a data collection form. This type of variable is typically represented by several variables in a data set, i.e., one variable for each response blank on the form. A complete frequency distribution of all responses or values for the multiple response item necessitates that these variables be combined into one variable.

Another common problem involves analyzing repeated measurements that are contained in a single variable. Comparisons between time points usually require that each group of time point measurements be a separate variable. The Transpose a SAS Data Set option in SAS/ASSIST software can be used to solve both of these problems. Variables representing multiple responses or observations can be collapsed into one variable containing all of the observations. In addition, a single variable containing repeated measurements can be divided into multiple variables.

This poster will demonstrate the use of the TRANSPOSE Procedure in SAS/ASSIST to rearrange data from two studies of patients with HIV infection. It is most appropriate for those persons in clinical research who are novice users of SAS® software. One example will focus on a data set used for a fungal culture results form, in which ten variables are used to enter all of the fungal organisms identified for a particular subject. The second example consists of a small data set of laboratory values that are gathered over a series of time intervals.

COLLAPSING VARIABLES

In our ongoing study of oral manifestations in HIV infected pediatric patients, a subject's laboratory culture results with all fungal organisms identified at a particular visit are recorded on a data collection form. This form has ten blanks in which the fungal organism names and their corresponding numeric codes are recorded. A single data set was created for this form, with one variable for each blank. In order to obtain a complete frequency distribution for what is essentially one variable (all organisms identified), the ten variables are collapsed into one variable using the following procedure:

- 1.) Select Transpose a SAS Data Set in the Data Management/UTILITIES menu
- 2.) Select the fungal culture data set and name the output data set
- 3.) Select FUNGAL_A - FUNGAL_J as the variables to be transposed
 - This step inverts the data set, causing the transposed variables to become observations
- 4.) Group the transposed variables by STUDYID (Data set must be sorted by this variable)
 - This step essentially creates one new variable, with FUNGAL_A - FUNGAL_J for each subject placed in one column. Without this step, one variable will be created for each subject's observation in the original data set.

EXAMPLE 1. (Only 2 variables used to simplify example)

<u>STUDYID</u>	<u>FUNGAL_A</u>	<u>FUNGAL_B</u>		<u>STUDYID</u>	<u>NAME</u>	<u>COL1</u>
001	2	3	→	001	FUNGAL_A	2
002	3			001	FUNGAL_B	3
003	3	7		002	FUNGAL_A	3
				002	FUNGAL_B	
				003	FUNGAL_A	3
				003	FUNGAL_B	7

DIVIDING ONE VARIABLE INTO MULTIPLE VARIABLES

In a smaller study of HIV infected adults, laboratory values were collected at four different time points. The variable of interest, number of oocysts per 24 hours, was entered as a single variable in the data set. In order to compare the oocyst count at different time points, the oocysts variable was split into four variables (one for each time point) using the following procedure:

- 1.) Select Transpose a SAS Data Set in the Data Management/UTILITIES menu
- 2.) Select the laboratory values data set and name the output data set
- 3.) Select OOCYSTS as the variable to be transposed
 - This step inverts the data set, causing the observations to become variables
- 4.) Group the transposed variable by STUDYNO (Data set must be sorted by this variable)
 - The result of this step is to have one record (with 4 measurement variables) for each subject. Without this step, one variable will be created for each observation in the original data set
- 5.) Select WEEK as the variable to label the transposed variable columns
 - This step labels each new variable with the appropriate value from the WEEK variable

EXAMPLE 2. (Only 2 time points used to simplify example)

<u>STUDYNO</u>	<u>WEEK</u>	<u>OOCYSTS</u>		<u>STUDYNO</u>	<u>NAME</u>	<u>0</u>	<u>2</u>
2	0	656		2	OOCYSTS	656	58.5
2	2	58.5	→	3	OOCYSTS	190	31.0
3	0	190					
3	2	31.0					

CONCLUSION

Researchers are often presented with the problem of having their data in a configuration that is not appropriate for their analysis. The use of the TRANSPOSE Procedure in SAS/ASSIST software can help solve many of these data management problems by giving the researcher control over the arrangement of the data set variables. In addition, the point and click interface allows the researcher to perform these tasks without having to write a program.

ACKNOWLEDGEMENTS

The author is grateful to Dr. Claudia Kozinetz for editorial advice and Ms. Kelly Trott for assistance with the manuscript. I also wish to thank Dr. Nathaniel Smith for the use of his data.

This poster was produced in collaboration with the Baylor Oral Manifestations of Pediatric HIV Infection Study Group in the Department of Pediatrics at Baylor College of Medicine.

SAS and SAS/ASSIST are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Stanley Cron
 Center for AIDS Research, Baylor College of Medicine
 6621 Fannin, Suite 710 MC 3-2316
 Houston, Texas 77330
 scron@bcm.tmc.edu