

Paper 065-2007

MARCH SANITY: A SAS® APPLICATION FOR ADMINISTERING AN NCAA TOURNAMENT POOL

Michael P. Walkup, Wake Forest University Health Sciences, Winston-Salem, NC

Timothy E. Craven, Wake Forest University Health Sciences, Winston-Salem, NC

Gregory B. Russell, Wake Forest University Health Sciences, Winston-Salem, NC

ABSTRACT

The annual NCAA men's basketball championship tournament is a sporting event that enjoys widespread popularity even among those who are not necessarily regular season fans of the sport. Each year hundreds of thousands of people participate in group tournament pools where each participant picks the outcome of every tournament game from the opening round through the championship by completing a 64-team bracket. The person with the completed bracket that most closely matches game results wins the pool. For large groups like business offices, administering an NCAA pool by paper can be a daunting task. Web-based applications are available online for administering group pools, but they have other pitfalls such as requiring users to register and providing only limited scoring options. We have devised an application for Microsoft Windows environment that uses widely available software (Microsoft Excel) for bracket completion, and Base SAS (PROC IMPORT and the data step) and SAS Macro Language for processing, analyzing and presenting results. The application is flexible (it can be used for a variety of pool types and scoring schemes), has a simple bracket completion scheme (participants complete their own brackets and provide all data in pre-designed spreadsheets), and can be used to distribute results at any point during the tournament, even before rounds are completed.

Keywords: PROC IMPORT, SAS Macro, Excel, gaming

INTRODUCTION

In recent years, the annual NCAA men's and women's basketball championship tournaments have become nearly a month-long national obsession. A driving force behind this popularity is the "office pool". Persons who are otherwise oblivious to college basketball games played during the regular season are enthusiastic participants in office tournament pools. This popularity has been well documented in the national media. Business consultants have written white papers on loss of productivity during the tournament (Challenger, 2006) while others have criticized such estimates as "overblown" (Clark, 2006).

Participation in office pools has been further enhanced by the availability of web sites that allow groups to register and enter brackets then score results. Computer applications have been written to administer NCAA tournament office pools (see Woo, 2000 for a description of such an application built with SAS/AF® software). The application presented here, which combines features of both Microsoft Excel and SAS, grew out of our dissatisfaction with the inflexibility of other widely-available methods.

NCAA TOURNAMENT FORMAT

The men's NCAA basketball tournament is an annual event involving 65 teams which are invited to compete in a single-elimination format. The 64th and 65th teams compete in a "play-in" game two days before the first full round of tournament play begins. The winner of the play-in game advances to play the tournament's top seeded team. Due to time constraints required to distribute and collect brackets, and general lack of interest in the play-in game, most tournament pools begin with the round of 64 as the first round of the tournament pool.

Our application uses a series of Excel spreadsheets and SAS programs to administer a tournament pool beginning with the round of 64. It is designed to promote group-wide participation, is easily coordinated through email groups, and provides the pool administrator with a flexible and powerful tool to score entries and display results. A typical pool will have one or more completed brackets entered per participant with the winner being the person with the highest point total. An advantage of our application is that it allows for the implementation of virtually any user-defined scoring system. Beginning with the round of 8 (known as the "elite 8") through the final four and tournament championship potential pool winners and entries eliminated from contention can be easily determined and displayed to pool participants.

ADMINISTERING THE POOL

Figure 1 shows a schematic representation of the step-by-step process to administer an NCAA basketball tournament pool using our application. Following is a detailed description of each step.

STEP 1 – EXCEL SPREADSHEET SETUP

A Microsoft Excel template containing a 64-team NCAA bracket is available for download from that company's website (<http://office.microsoft.com/en-us/templates/TC101032841033.aspx> is a link to the 2006 field). A copy of this spreadsheet was downloaded and modifications made to create a spreadsheet that would provide uniform data in a format amenable to importing and manipulating using SAS. The following modifications were made:

- (1) A cell was added to each entry to note team seeding.
- (2) Code was added to advance the team name once its seed was selected to win a given round.
- (3) An auto-advance field was added to perform a "quick" select for a team to advance to the final four.
- (4) A select box was added for the semi-final and final games since seeds are no longer relevant in the final two rounds.
- (5) A field to enter the pool participant's name was added.
- (6) A field was added to enter the pool participant's value(s) for tiebreaker(s).
- (7) Code to check for the participant's name and completion of each pick was added (this displays a red text message in the center of the sheet until all picks have been made and the participant's name is entered).
- (8) All cells (except those where seeds are entered) were protected to prevent formulas from being overwritten and to streamline tabbing through the fields.

Figure 2 shows an uncompleted copy of the revised Excel spreadsheet for the men's 2006 NCAA tournament field (beginning with the round of 64). Copies of this spreadsheet and all programs described below for administering the pool are available from the authors upon request.

To begin the pool, distribute the modified spreadsheet to all prospective pool participants via email.

Participants should be instructed to complete the entire bracket as described above, and to rename the spreadsheet using a standard naming convention (we found first and last name followed by a number to index multiple files to be a reasonable convention, e.g., "JohnSmith1.xls"). Participants who wish to place multiple entries in the pool should save each entry in a separate file and change the last number to index entries.

STEP 2 – PROCESSING ENTRIES

Open returned spreadsheets for a quick visual inspection and save completed entries in a dedicated directory. Open the program "GET_FILES.SAS" and modify the "%let" statement to reflect the directory where spreadsheets are stored, then run the file. This program will utilize SAS's %SYSFUNC and CALL EXECUTE macro tools to open the directory, iteratively loop through each file in it and import them into SAS data sets. Once imported, each file will be stripped of unnecessary fields and manipulated to create a "long" data set consisting of a separate row for each game

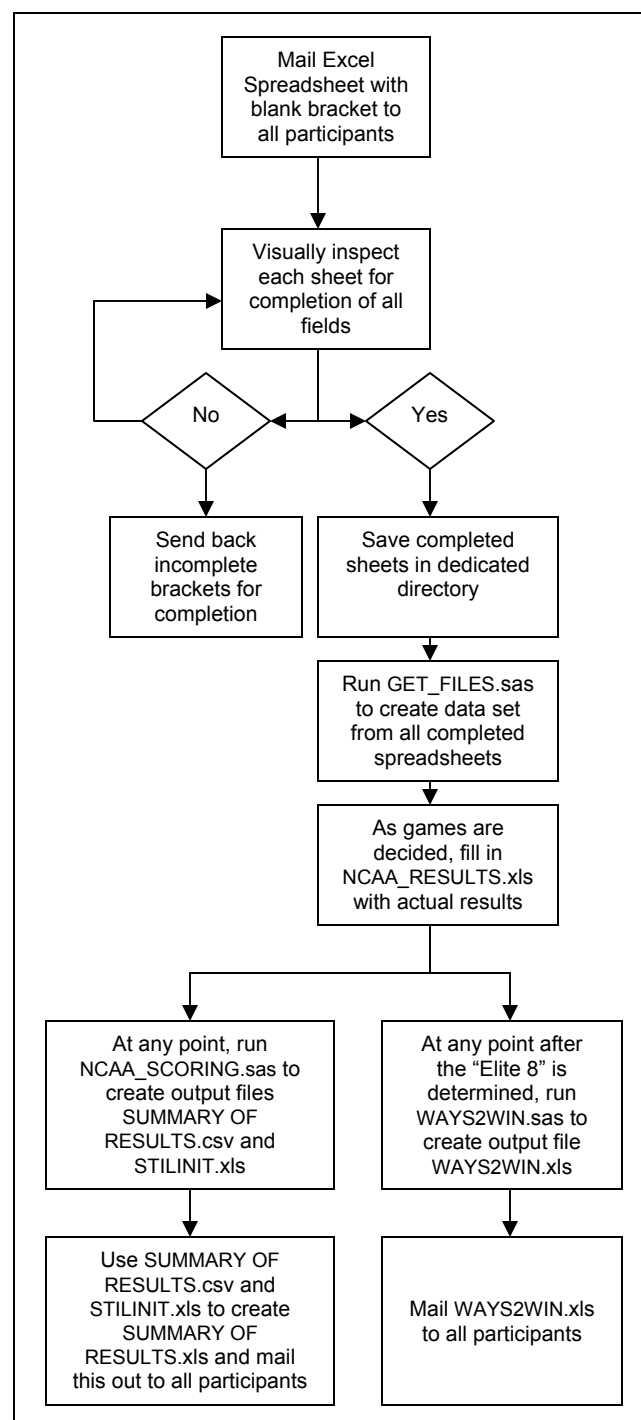


Figure 1. Process for administering an NCAA Tournament pool.

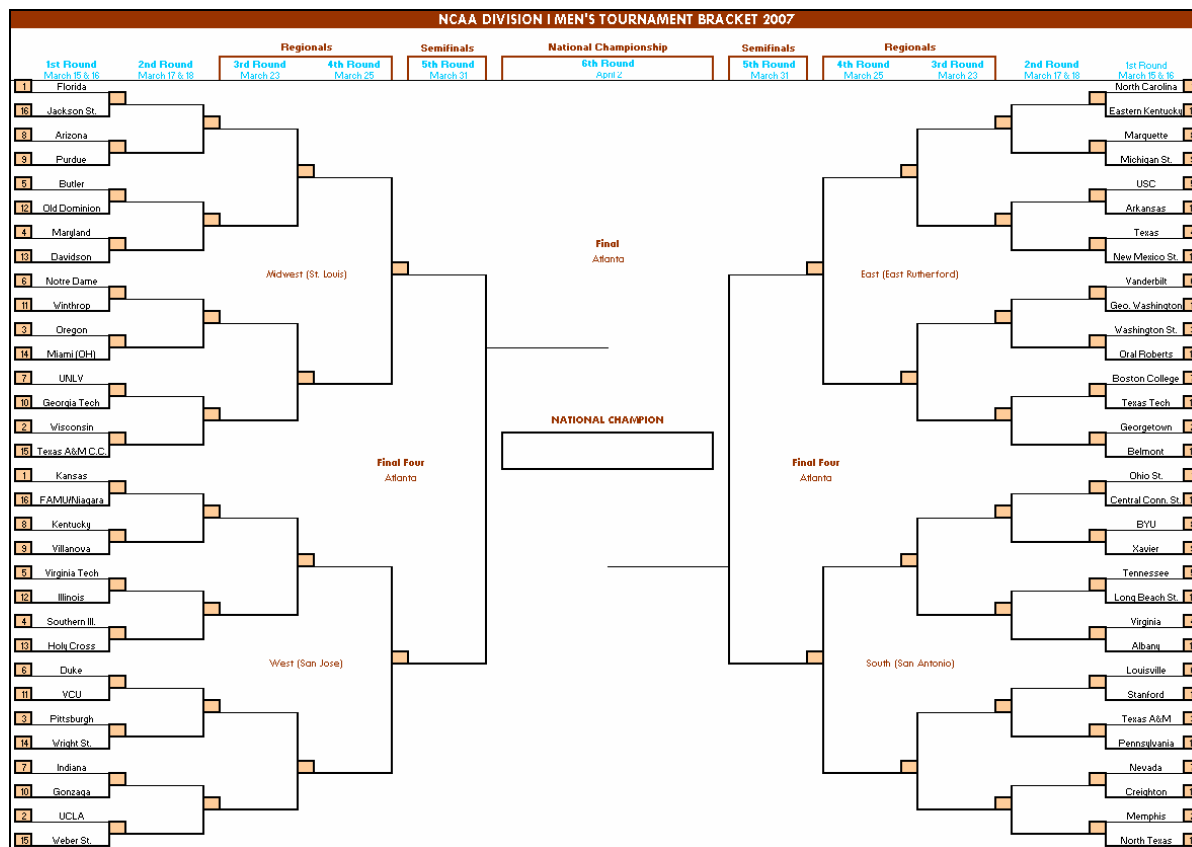


Figure 2. Excel spreadsheet bracket ready for completion.

in the tournament (63 observations) and variables for participant name, a round/game identifier variable, and the team picked to win that game. These individual data sets are then iteratively stacked onto a master data set called "Allsheets" which will be used for scoring results.

At this point the pool administrator would wait until the tournament begins and outcomes of the match-ups are known. The administrator should complete a master scoring spreadsheet called "NCAA_RESULTS.xls" as games are completed. Once this has been done, scoring (or partial scoring) can be performed at any time.

STEP 3 – SCORING RESULTS

To score entries, run the program "NCAA_SCORING.sas", making sure that the appropriate directory reference has been specified in the first "%let" statement. This will produce a results file called "SUMMARY_OF_RESULTS.csv" containing the name of each entrant with the following information: current total points (the scoring program can be run whenever "NCAA_RESULTS.xls" has been updated); potential total points (conditional on results to date and future picks); entrant tie-breaker value(s); and entrant picks for the elite 8, final 4, championship game and eventual champion.

A second prepared spreadsheet ("SUMMARY_OF_RESULTS.xls") is available to display results in a suitable format for distribution to pool participants. To complete this step, highlight and copy all the data from "SUMMARY_OF_RESULTS.csv". Open "SUMMARY_OF_RESULTS.xls", right-click on the first empty cell, choose "Paste Special" from the pop-up menu, and select "Values" from the "Paste" menu then select "OK". This will prevent the paste function from overwriting special formatting in the Excel spreadsheet and only the data will be updated.

The special formatting shows teams selected to advance to the elite 8 and beyond with lines drawn through those that have been eliminated (see Figure 3). To accomplish this, the program "NCAA_SCORING.sas" creates an Excel spreadsheet named "STILLINIT.xls" that is incorporated into a formula in the first column past the end of the data in "SUMMARY_OF_RESULTS.xls" (the user may click on any cell in this column to display the formula). Results are sorted by descending order of potential total points when distributed; however, because results are provided in an Excel spreadsheet format, the user can easily sort data by other fields (e.g., current total points).

Player Name	File Name	Total Points	Points Possible	Tiebreak 1 (Margin)	Tiebreak 2 (total pts)	Champion	Finals	Semis			
RESULTS	NCAA_results.xls	104	120					Florida	UCLA	Texas	Memphis
Greg Russell	grussell1.xls	61	72	9	153	Florida	Florida	Ohio-St.	Florida	UCLA	North-Carolina
Tim Craven	traven1.xls	57	68	6	150	Florida	Florida	Ohio-St.	Florida	Kansas	North-Carolina
Michael Walkup	mwalkup1.xls	58	63	8	160	Texas-A&M	Florida	Texas-A&M	Florida	UCLA	North-Carolina
Tim Craven	traven2.xls	52	52	7	154	Ohio-St.	Southern-III.	Ohio-St.	Florida	Southern-III.	Washington-St.

Figure 3. Sample portion of results report.

STEP 4 – SHOWING PROSPECTIVE WINNERS

Beginning with the elite 8 round of the tournament and thereafter, the prospective winners (and those eliminated from contention) can be found by running the program “WAYS2WIN.sas”. This program creates a file called “WAYS2WIN.csv” listing all possible outcomes for the remaining rounds of play with the name(s) of entrant(s) who will win the pool for the specified outcome, as well as those who place second or third if the pool has a payout for the top three finishers. Outcomes resulting in ties will be output once for each entrant with that particular outcome, along with the number of ties for that outcome. The information in this file should then be copied and pasted into a formatted Excel spreadsheet called “WAYS2WIN.xls” as described above (see Figure 4).

“WAYS2WIN.sas” is only run at the elite 8 round or after because that is the point where the number of potential outcomes (128) becomes small enough to be usefully displayed. The number of potential outcomes before any

games are played at the round of 16 is 32,768, and in general there are $2^{(2^k - 1)}$ possible outcomes to a tournament with k rounds. Furthermore, if m games have already been played in the k^{th} round, for $0 \leq m < 2^{(k-1)}$ the number of possible champions is $2^{(2^k - 1 - m)}$. (For details on the derivation of this result see the Appendix.)

Office pools can be scored in many different ways. The most common methods apply increasing weights to correct picks at each round, a method that rewards entries with teams advancing deep into the tournament. By default, the weighting scheme applied in “NCAA_SCORING.sas” and “WAYS2WIN.sas” assigns one point to each winner in the first round, two points per winner in the second round, and so on. However, the user may apply other weights by simply changing the value of the variable “weight” in each round.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Comb. #	Name	Winner(s)	Tiebreak1	Tiebreak2	Rank	Ties	Points	Semis				Finals		CHAMP
2	1	Mike	mwalkup1.xls	9	153	1	1	72	Florida	Southern III.	Texas	Ohio St.	Florida	Texas	Florida
3	1	Greg	grussell1.xls	6	150	1	1	72	Florida	Southern III.	Texas	Ohio St.	Florida	Texas	Florida
4	2	Mike	mwalkup1.xls	9	153	1	1	66	Florida	Southern III.	Texas	Ohio St.	Florida	Texas	Texas
5	2	Greg	grussell1.xls	6	150	1	1	66	Florida	Southern III.	Texas	Ohio St.	Florida	Texas	Texas
6	3	Mike	mwalkup1.xls	9	153	1	1	77	Florida	Southern III.	Texas	Ohio St.	Florida	Ohio St.	Florida
7	3	Greg	grussell1.xls	6	150	1	1	77	Florida	Southern III.	Texas	Ohio St.	Florida	Ohio St.	Florida
8	4	Mike	mwalkup1.xls	9	153	1	2	71	Florida	Southern III.	Texas	Ohio St.	Florida	Ohio St.	Ohio St.
9	4	Greg	grussell1.xls	6	150	1	2	71	Florida	Southern III.	Texas	Ohio St.	Florida	Ohio St.	Ohio St.
10	4	Tim	traven1.xls	7	154	1	2	71	Florida	Southern III.	Texas	Ohio St.	Florida	Ohio St.	Ohio St.
11	5	Tim	traven1.xls	7	154	1	0	65	Florida	Southern III.	Texas	Ohio St.	Southern III.	Texas	Southern III.
12	5	Mike	mwalkup1.xls	9	153	2	1	61	Florida	Southern III.	Texas	Ohio St.	Southern III.	Texas	Southern III.
13	5	Greg	grussell1.xls	6	150	2	1	61	Florida	Southern III.	Texas	Ohio St.	Southern III.	Texas	Southern III.
14	6	Tim	traven1.xls	7	154	1	0	65	Florida	Southern III.	Texas	Ohio St.	Southern III.	Texas	Texas
15	6	Mike	mwalkup1.xls	9	153	2	1	61	Florida	Southern III.	Texas	Ohio St.	Southern III.	Texas	Texas
16	6	Greg	grussell1.xls	6	150	2	1	61	Florida	Southern III.	Texas	Ohio St.	Southern III.	Texas	Texas
17	7	Tim	traven1.xls	7	154	1	0	70	Florida	Southern III.	Texas	Ohio St.	Southern III.	Ohio St.	Southern III.

Figure 4. Sample output of ways to win.

DISCUSSION

Several popular public web sites like *CBSSportsline.com* and *ESPN.com* have tools for doing group NCAA tournament pools. These sites provide brackets that can be completed online and that will show results for the group after each round. Typically users can access these features without having to pay a fee. However, online resources have the distinct disadvantage of requiring users to register by selecting usernames and passwords and by providing some level of personal information as part of the registration process. Because registrants must invariably provide a valid email address they usually become subject to email solicitations from the sponsoring site, or worse SPAM email from third parties if the online sponsor shares that information with other web sites.

Online pools typically lack the flexibility of using the different scoring schemes allowed by our SAS application. "WAYS2WIN.sas" also provides possible scenarios for pool winners and losers under any weighting scheme desired. This feature would be particularly advantageous in allowing each participant to see their highest score as well as all deviations from the maximum score under each possible combination beginning with the round of 8 teams.

Our application makes extensive use of features contained in Microsoft Excel, so it cannot be considered a purely "SAS-based" solution. However, utilizing Excel for bracket entry greatly increases its flexibility. While it would be rare for everyone in an office or other group setting to have a copy of SAS at his or her disposal, Microsoft Office enjoys almost universal popularity. SAS remains the "brains" behind the scoring and preparation of results. The SAS programs have been set up in a way to allow administrators to use the tool for many different pools with only slight modifications to the code.

The more advanced SAS user may choose to take advantage of other SAS features to further modify this application. For example, as written our application requires multiple entries to be submitted as separate files; however, with only minor modifications the SAS 9 EXCEL LIBNAME engine could be used to allow an alternative setup where entrants submit a single file containing multiple worksheets with completed brackets (see Fan, 2005). The application's greatest strength is the ease in which completed brackets may be imported into a SAS data set. Furthermore, features of the application can be easily adapted to other applications. For example, one could adapt the customized spreadsheet distribution and processing step to an email survey where individual responses are collected and scored. Once this is accomplished the vast range of SAS analytical tools are at the user's disposal to score, analyze and present results in a myriad of ways.

DISCLAIMER

The contents of this paper are solely the work of the authors and do not purport to represent opinions, recommendations or actions of Wake Forest University, its faculty or staff. Gambling on sporting events is illegal in many states and the reader is encouraged to observe all Federal, state and local laws. This application is provided solely for the purpose of entertainment.

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 registered trademarks or trademarks of their respective companies.

APPENDIX – NUMBER OF POSSIBLE WINNERS FOR A SINGLE-ELIMINATION TOURNAMENT

In general, there will be $2^{(2^k - 1)}$ total possible outcomes for a single-elimination paired-match style tournament with k rounds. Half the teams are eliminated in each round until just one team is left (the tournament champion) after completion of the final round. To see this result consider the following:

- A tournament with k rounds ($k \geq 1$) will have 2^k teams (e.g., excluding the play-in game, the men's NCAA basketball tournament has 6 rounds and $2^6 = 64$ teams).
- In general, round j will have a total of $2^k / 2^j$ games played. That is for a 6 round bracket, the first round will have $64/2 = 32$ games, the second round will have $64/4 = 16$ games, down to the 6th round having $64/64 = 1$ game. A given game has 2 possible outcomes making the total number of possible outcomes for a round $2^{(2^k / 2^j)}$.
- The total number of possible outcomes for the entire tournament is the product of the possible number of outcomes for each round. This can be expressed as:

$$\begin{aligned} \prod_{j=1}^k 2^{(2^k / 2^j)} &= 2^{(2^k / 2^1)} \cdot 2^{(2^k / 2^2)} \cdot \dots \cdot 2^{(2^k / 2^k)} \\ &= 2^{[2^k (\frac{1}{2^1} + \frac{1}{2^2} + \dots + \frac{1}{2^k})]} \\ &= 2^{[2^k (\frac{2^k - 1}{2^k})]} \\ &= 2^{(2^k - 1)} \text{ total possible outcomes for the tournament.} \end{aligned}$$

This same logic can be used to derive the number of possible outcomes for the tournament conditional on results of previous rounds by treating each round as new, smaller tournament. The total number of possibilities for the tournament for each round of a 6 round bracket (after the previous rounds have been played) is as follows:

Round	Possible Outcomes (for tournament—not round)
1	$2^{(2^6 - 1)} = 9.2234 \times 10^{18}$
2 (treated as 5 round tournament)	$2^{(2^5 - 1)} = 2,147,483,648$
3 (treated as 4 round tournament)	$2^{(2^4 - 1)} = 32,768$
4 (treated as 3 round tournament)	$2^{(2^3 - 1)} = 128$
5 (treated as 2 round tournament)	$2^{(2^2 - 1)} = 8$
6 (treated as 1 round tournament)	$2^{(2^1 - 1)} = 2$

The result for partial rounds follows directly: If m games have already been played in the k^{th} round (for $0 \leq m < 2^k - 1$), then there will be $2^{(2^k - 1 - m)}$ total possible outcomes for the tournament capped at $2^k - 1$ since there are only $2^k - 1$ games in round k .

Note: On their web site, *Forbes.com* gives the number of ways to win a tournament with “64 games and two possible results of each game” as

18,446,744,073,7094,551,616

(see http://www.forbes.com/2006/03/10/cx_hc_0313marchmadnessslide_2.html). However, this is actually the number of possible outcomes for the entire 65-team field, or $2 \times 2^{(2^6 - 1)}$.

REFERENCES

Challenger, J. (2006), "Employers Not Eager for Hoops Tournament", *Challenger, Gray & Christmas, Inc* (press box), <http://www.challengergray.com/marchmadness.aspx>.

Clark, H. (2006), "The (Overblown) Cost of March Madness", *Forbes.com*, <http://www.msnbc.msn.com/id/11809691/>.

Fan, Z. (2006), "Make the Invisible Visible: A Case Study of Importing Multiple Worksheet Files by Using the SAS9 LIBNAME Engine in Microsoft Excel", *31st SAS Users Group International Proceedings*, <http://www2.sas.com/proceedings/sugi31/034-31.pdf>

SAS Institute Inc. 2004. *Base SAS 9.1.3 Procedures Guide, Volumes 1, 2, 3, and 4*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2004. *SAS 9.1 Macro Language: Reference*. Cary, NC: SAS Institute Inc.

Woo, W. (2000), "Office Pooling With SAS Software" *25th SAS Users Group International Proceedings*, <http://www2.sas.com/proceedings/sugi25/25/cc/25p090.pdf>.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors:

Michael P. Walkup
Wake Forest University Health Sciences
Department of Biostatistical Sciences
Medical Center Boulevard
Winston-Salem, NC 27157-1063
Telephone: (336) 713-4740
E-Mail: mwalkup@wfubmc.edu

Timothy E. Craven
Wake Forest University Health Sciences
Department of Biostatistical Sciences
Medical Center Boulevard
Winston-Salem, NC 27157-1063
Telephone: (336) 716-6109
E-Mail: tcraven@wfubmc.edu

Gregory B. Russell
Wake Forest University Health Sciences
Department of Biostatistical Sciences
Medical Center Boulevard
Winston-Salem, NC 27157-1063
Telephone: (336) 716-5449
E-Mail: grussell@wfubmc.edu