Paper 42 -2025

# Using SAS ® Macro Statement to Combine and Merge Datasets for Temperature Study

Abbas S. Tavakoli, DrPH, MPH, ME[1], Isabelle Bennett, BS Student, [2] Robin B. Dail, PhD, RN, FAAN[1],

[1]University of South Carolina, College of Nursing
[2]University of South Carolina, College of Arts and Sciences, Department of Statistics

**Abstract**
Monitoring of physiological vital signs from body temperature and heart rate is crucial to guide treatment for preterm infants to optimize health outcomes. Monitoring continuous measurement of both central (abdominal) and peripheral (foot) skin temperatures will give a thermal gradient, which may predict the onset of infection. Studies have shown that heart rate characteristic (HRC) scores can be used to predict infection; however, a special HeRO monitor must be used in addition to standard care. Our team compared thermal gradients to HRCs to predict infection in early preterm infants (born < 32 weeks gestational age). A total of 364 preterm infants were enrolled in the study (NIH/NINR: 1R01NR017872) following Institutional Review Board approval and parental consent. Each infant had a skin temperature probe (thermistor) attached to their abdomen and another to the sole of one foot. Temperature and HRC data were downloaded from a research datalogger, and HeRO monitor as text files, then merged into Excel format, with each infant's data containing approximately 41,002 rows. Data were collected from five different sites: Duke (n=25), ECU (n=137), MUSC (n=31), Prisma (n=149), and UNC (n=11). SAS datasets were created for each infant, and a macro was used to combine the data from all infants at each site. Demographic variables were then merged with the combined site data, and finally, the datasets from all sites were merged. Several macro statements were written to facilitate the combination and merging of datasets using SAS, version 9.4.

**Keywords:** SAS, body temperature, infant

## Background

Monitoring of physiological vital signs from body temperature and heart rate is crucial to guide treatment for preterm infants to optimize their health outcomes. One method for monitoring thermal gradients in a preterm infant is through continuous measurement of both central (abdominal) and peripheral (foot) skin temperatures, indicating blood perfusion across the infant's body. (1) Abnormal thermal gradients, indicated by either an increased central peripheral temperature difference (CPtd >2ºC) or a negative CPtd (CPTd < 0ºC , where the foot temperature exceeds abdominal temperature), are associated with the onsent of infection and stressful events in premature infants.[2] Another physiologic vital sign which may alert clinicians of impending infection are heart rate characteristic scores (HRC) obtained from a HeRO monitor (Predictive Medical Science Corp, Charlottesville, VA) which are calculated from heart rate variability from the infant's standard cardiopulmonary monitoring. (3) This study, funded by NIH/NINR: 1R01NR017872, (4) is a continuation of Dail's program of research examining thermal gradients, which are indicative of autonomic nervous system dysregulation, to predict infection in very preterm infants. HeRO scores were used to compare to thermal gradients over each preterm infant's first 28 days of life.

## Purpose

The purpose of this paper is to demonstrate the use of SAS macros for combining and merging datasets from different sites in a temperature study.

## Methodology

A total of 364 premature infants, born at 32 weeks gestational age or less and having birthweights of less than 1500 grams, were enrolled for study after Institutional Review Board approval and parental consent for participation in this study. Physiological temperature data were measured and stored for the first 28 days of

life using research dataloggers, then downloaded to a USB drive. Each infant had one skin temperature probe (thermistor) attached to their abdomen and one skin temperature probe attached to the sole of one foot. Standard care was to maintain infant temperature in a heated incubator between 36.5°-37.0°C, using servo control with feedback of a separate standard care thermistor measuring abdominal temperature. Research temperatures were measured every minute for 28 days. A HeRO monitor connected to the infant's cardiopulmonary monitor measured HRCs every hour for 28 days. Research nurses downloaded temperature data and HRC data and sent the data to a data manager. The data manager merged these data into an Excel file for each infant containing the following variables: date, time, minutes since birth (MSB), abdominal temperature (ABD), foot temperature (FT), date of temperature measurements (DLD), time of temperature measurements (DT), Hero Monitor Date (DH), HRC score, and the time of the HeRO monitor measurement (TH). Additionally, demographic data were collected for each infant from the electronic medical record and entered a REDCap database. Data were gathered from five different sites: Duke (n=25), ECU (n=137), MUSC (n=31), Prisma (n=149), and UNC (n=11). Each infant's data was stored in a separate Excel file, with each variable having approximately 41,002 measurements per infant. SAS datasets were created for each infant, and a macro was used to combine all infants' data from each site. Demographic variables were then merged with the combined site data, and finally, the datasets from all sites were merged. A macro statement was written to facilitate the combination and merging of datasets using SAS/STAT® statistical software, version 9.4.[3]

## Data Steps

Table 1 provides an example of a portion of data for a single infant in an Excel file. We use the UNC site to demonstrate the process of combining and merging data for all infants.

**Table1 Example of data in Excel**

| id | msb | Date | Time | dld | dt | abd | ftp | dh | Th | HRC |
|---|---|---|---|---|---|---|---|---|---|---|
| 211_003 | 258 | 8/30/2019 | 11:34:00 | 8/30/2019 | 11:34:00 | 35.82 | 35.79 | . | . | . |
| 211_003 | 259 | 8/30/2019 | 11:35:00 | 8/30/2019 | 11:35:00 | 35.89 | 35.76 | . | . | . |
| 211_003 | 260 | 8/30/2019 | 11:36:00 | 8/30/2019 | 11:36:00 | 35.96 | 35.75 | . | . | . |
| 211_003 | 261 | 8/30/2019 | 11:37:00 | 8/30/2019 | 11:37:00 | 36.02 | 35.75 | . | . | . |
| 211_003 | 262 | 8/30/2019 | 11:38:00 | 8/30/2019 | 11:38:00 | 36.03 | 35.73 | . | . | . |
| 211_003 | 263 | 8/30/2019 | 11:39:00 | 8/30/2019 | 11:39:00 | 36.01 | 35.71 | . | . | . |
| 211_003 | 264 | 8/30/2019 | 11:40:00 | 8/30/2019 | 11:40:00 | 36.1 | 35.72 | . | . | . |
| 211_003 | 265 | 8/30/2019 | 11:41:00 | 8/30/2019 | 11:41:00 | 36.19 | 35.73 | . | . | . |
| 211_003 | 266 | 8/30/2019 | 11:42:00 | 8/30/2019 | 11:42:00 | 36.24 | 35.75 | . | . | . |
| 211_003 | 267 | 8/30/2019 | 11:43:00 | 8/30/2019 | 11:43:00 | 36.48 | 35.76 | . | . | . |
| 211_003 | 268 | 8/30/2019 | 11:44:00 | 8/30/2019 | 11:44:00 | 36.6 | 35.78 | . | . | . |
| 211_003 | 269 | 8/30/2019 | 11:45:00 | 8/30/2019 | 11:45:00 | 36.67 | 35.8 | . | . | . |
| 211_003 | 270 | 8/30/2019 | 11:46:00 | 8/30/2019 | 11:46:00 | 36.69 | 35.84 | . | . | . |
| 211_003 | 271 | 8/30/2019 | 11:47:00 | 8/30/2019 | 11:47:00 | 36.69 | 35.87 | . | . | . |
| 211_003 | 272 | 8/30/2019 | 11:48:00 | 8/30/2019 | 11:48:00 | 36.77 | 35.89 | . | . | . |
| 211_003 | 273 | 8/30/2019 | 11:49:00 | 8/30/2019 | 11:49:00 | 36.69 | 35.91 | . | . | . |
| 211_003 | 274 | 8/30/2019 | 11:50:00 | 8/30/2019 | 11:50:00 | 36.73 | 35.93 | . | . | . |
| 211_003 | 275 | 8/30/2019 | 11:51:00 | 8/30/2019 | 11:51:00 | 36.76 | 35.95 | . | . | . |

Table 2 presents the SAS macro program that generates an ID number and includes a note for Infant 1. This macro creates a unique ID for each infant, with the first three digits representing the site information and the last three digits indicating the infant's number. The

program also combines all infant data into a single data set. We use this program to generate and merge data for all infants across other sites.

**Table 2. Combine and create id numbers for all infants for UNC site**

**SAS Program**

```
libname r24 'd:\r24\data\unc\sasdata\';

%macro data (d=d2 , s=r24.unc2 , i=211002);

data &d;
set &s;
iid=&i;
run;
%mend data;

run;

%data (d=d2 , s=r24.unc2 , i=211002);
%data (d=d3 , s=r24.unc3 , i=211003);
%data (d=d4 , s=r24.unc4 , i=211004);
%data (d=d5 , s=r24.unc5 , i=211005);
%data (d=d6 , s=r24.unc6 , i=211006);
%data (d=d7 , s=r24.unc7 , i=211007);
%data (d=d8 , s=r24.unc8 , i=211008);
%data (d=d9 , s=r24.unc9 , i=211009);
%data (d=d10, s=r24.unc10, i=211010);
%data (d=d11, s=r24.unc11, i=211011);
%data (d=d12, s=r24.unc12, i=211012);

run;

data final;
set d2-d12;
by id;
if first.id then idn=1;
 else idn+1;
run;

libname r24 'd:\r24\data\unc\sasdata\';
(keep= id iid DATE TIME msb dld dt abd ftp dh th hrc) ;
retain id iid DATE TIME msb dld dt abd ftp dh th hrc;
;
 set final;

 format abd ftp hrc 8.2;
 run;
```

Table 3 presents the SAS program used to merge the combined data from the UNC site with the demographic file. This program is also applied to merge the combined files and demographic files for all infants from other sites.

**Table 3. SAS Program to merge combined files with demographic file.**

**SAS Program**

```
libname r24 'd:\r24\data\unc\sasdata\';

data one;
 set r24.uncall;
 site=5;
```

```
proc sort data=one;
by id; run;

libname r24 'd:\r24\data\demo\';


data two;
 set r24.mdemo;
proc sort data=two;
by id; run;

data all;
merge one (in=a) two (in=b);
by id; run;

libname r24 'd:\r24\data\unc\combine\';

data r24.uncdemo;
set all;  run;
```

Table 4 shows the SAS program to merge all sites

**Table 4 SAS program to merge all sites.**

| SAS Program |
|---|
| proc format; |
| value sitef 1="Duke"  2="ECU"  3="MUSC"  4="Prisma"  5="UNC"  ; |
| libname \r24 'd:\r24\data\duke\combine\'; |
| data d1;  set \r24.dukedemo; run; |
| libname \r24 'd:\r24\data\ecu\combine\'; |
| data d2;  set \r24.ecudemo; run; |
| libname \r24 'd:\r24\data\musc\combine\'; |
| data d3;  set \r24.muscdemo; run; |
| libname \r24 'd:\r24\data\prisma\combine\'; |
| data d4;  set \r24.prismademo; run; |
| libname \r24 'd:\r24\data\unc\combine\'; |
| data d5; |
| set \r24.uncdemo; run; |
| data all; set d1-d5; format site sitef.; run; |
| libname \r24 'd:\r24\sasdata\'; |
| libname library  'd:\r24\sasdata\'; |
| data \r24.\r24all; set all; run; |

Table 5 shows the SAS program used to run descriptive statistics. The PROC FREQ and PROC MEANS procedures are employed to generate descriptive statistics by site and ID.

**Table5 SAS Procedure to run descriptive statistics.**

| SAS Program |
|---|
| Ods rtf;<br>ods listing close;<br>proc sort data =one; by  id; |

```
%macro freqb (d,s,i,r,c,t);
proc freq data=&d;
where site=&s;

    tables  (&i)*(&r)*(&c)  ;
    title " frequency table/&t";

%mend freqb; run;
%freqb (one,1,id,day, cptdg group, duke by id day);
%freqb (one,1,id,week, cptdg group, duke by id week);
%freqb (one,2,id,day, cptdg group, ECU by id day);
%freqb (one,2,id,week, cptdg group, ECU by id week);
%freqb (one,3,id,day, cptdg group, MUSC by id day);
%freqb (one,3,id,week, cptdg group, MUSC by id week);
%freqb (one,4,id,day, cptdg group, Prisma by id day);
%freqb (one,4,id,week, cptdg group, Prisma by id week);
%freqb (one,5,id,day, cptdg group, UNC by id day);
%freqb (one,5,id,week, cptdg group, UNC by id week);
by  id;
run;
ods rtf close ; ods listing; quit; run;

Ods rtf;  ods listing close;
%macro avg (d,q,v,t);
proc means data=&d n mean std min max  maxdec=2;
  class &q;      var  &v  ;
    title " means /by  &t" ; run;
%mend avg;
%avg(one, site,cptd abd ftp hrc gagew gaged bwt, by site);
 %avg(one, id ,cptd abd ftp hrc gagew gaged bwt, by id);
%avg(one, site id ,cptd abd ftp hrc gagew gaged bwt, by site id);
%avg(one, site id day ,cptd abd ftp hrc gagew gaged bwt, by site id day);
%avg(one, site id week ,cptd abd ftp hrc gagew gaged bwt, by site id week );
ods rtf close; ods listing; quit; run;
```

## Conclusion

Data collected for this Infant Temperature study went through several steps to be prepared for one aspect of data analyses. Several programs were used to create id numbers, combine, and merge different datasets. SAS datasets were created for each infant, infants were assigned ID numbers, and a macro was used to combine all infants' data from each site. Demographic variables were then merged with the combined site data, and finally, the datasets from all sites were merged. Macro statement is a powerful tool in SAS to use. Future exploration should look at optimizing computing efficiency by writing the macros as loops; this will reduce overall time by reducing the number of times the macro step is called and ran. When creating and combining the ID numbers for the infants, each ID was listed out by hand. This could become quite long depending on how many individuals there are an increase's chance for human error. Writing a loop to loop through all datasets and assign an ID accordingly would improve this program. Conditional logic could be explored when assigning site IDs to the various sites, further streamlining the code and improving readability. Furthermore, to optimize the efficiency of the average macro in table 5, a variable list containing the various sites could be created and the macro updated to handle the list as opposed to calling the macro for each individual site for selected variables.

### REFERENCES


1. Knobel, R. B., Holditch-Davis, D., Schwartz, T. A., & Wimmer, J. E. Jr. (2009). Extremely low birth weight preterm infants lack vasomotor response in relationship to cold body temperatures at birth. Journal of Perinatology, 29(12), 814-821. PMCID: PMC2787712
2. Knobel-Dail RB, Sloane R, Holditch-Davis D, Tanaka DT. Negative Temperature Differential in Preterm Infants Less Than 29 Weeks Gestational Age: Associations With Infection and Maternal Smoking. *Nurs Res.* 2017;66(6):442-453.
3. Coggins, S. A., Weitkamp, J. H., Grunwald, L., Stark, A. R., Reese, J., Walsh, W., & Wynn, J. L. (2016). Heart rate characteristic index monitoring for bloodstream infection in an NICU: a 3-year experience. *Arch Dis Child Fetal Neonatal Ed, 101*(4), F329-332. https://doi.org/10.1136/archdischild-2015-309210
4. Dail, R. B., Everhart, K., Hardin, J., Chang, W., Kuehn, D., Iskersky, V. Fisher, K. & Murphy, H. J. (2020). Predicting Infection in Very Preterm Infants: A Study Protocol. Oct 14; Nursing Research. doi: 10.1097/NNR.0000000000000483. [Epub ahead of

print] PubMed PMID: 33060417; NIHMSID:NIHMS1636929.
3. SAS Institute Incorporated. (2013). *SAS for Windows 9.4*. Cary, NC: SAS Institute Inc.

## TRADEMARK NOTICE



## Contact Information

Abbas S. Tavakoli, DrPH, MPH, ME
Clinical Professor
College of Nursing
University of South Carolina
1601 Greene Street
Columbia, SC 29208-4001
Fax: (803) 777-5561
E-mail: abbas.tavakoli@sc.edu