

## Using SAS® to Validate Clinical Prediction Models

Xiaoting Wu, Department of Cardiac Surgery, Michigan Medicine, Ann Arbor, MI;

Chang He, The Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative;

Donald S. Likosky, Department of Cardiac Surgery, Michigan Medicine, Ann Arbor, MI; The Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative

for the PERForm Registry and the Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative

### ABSTRACT

Model validation is an important step in establishing a clinical prediction model. Model validation process quantifies how well the model predicts outcomes for future patients. However, there were very few SAS programming examples showing the validation process. We previously developed a generalized mixed effect model that predicts peri-operative blood transfusion from patient characteristics. In this paper, we demonstrate the SAS® techniques that we used to validate such a model. These validation methods include calibration, discrimination and sensitivity analysis using bootstrapping method.

### INTRODUCTION

Information of patients' blood transfusion risk prior to cardiac surgery may help clinicians assess a patient's condition and facilitate informed decision making [1]. For example, given the prediction of transfusion risk, blood conservation modalities may be undertaken prior to surgery to reduce a patient's risk of receiving potentially unnecessary blood transfusions [2], which are independently associated with adverse outcomes [3, 4]. To provide a risk assessment tool for blood transfusions, we developed a prediction model based on a patient's preoperative risk factors from a multi-hospital dataset of more than 20,000 coronary artery bypass grafts procedures. The transfusion rate was 36.8% in our data. This model identified 16 preoperative predictors, and accounted for hospitals as random effects [5].

After model development, model validation is a critical step to assess the model performance [6]. The model performance in aspects of model calibration, discrimination and sensitivity analysis were assessed using SAS®. Previously the data was randomly split into model development dataset and model validate dataset. We used a development dataset for variable selection and functional form assessment. We used a validation dataset to assess the model performance. We finally used the large dataset (i.e., development combined with validation datasets) to obtain more stable parameter estimates. The following procedures can be applied to the internal or external validation dataset to validate the model performance.

### CALIBRATION

Calibration demonstrates the agreement between observed outcomes and predictions. Here, the calibration plot has the predicted probability by deciles on the x-axis, and the observed rates by deciles on the y-axis. A good calibration should lie close to a 45-degree line.

First, we used OUTPUT statement to obtain the prediction from our mixed effect model. Option (NOBLUP) is used to exclude the predictors of the random effects when calculating the predicted probability for each patient. In this model, we have 16 predictors. These predictors were chosen by model selection as well as their clinical relationship with blood transfusion. In this model, subject=STS\_hospnpi fits the random hospital effect. We also used STORE statement to obtain the model estimate to "parameter\_dat" dataset.

```

/*****output prediction from the mixed effect model *****/
proc glimmix data=mix_model;
class bsa4c (ref="LT1.6") albumin_3c (ref=">3.5") female (ref="0") ef4cat
(ref="60%+") crealst4c (ref="LT0.8") race3c (ref="White") status3c
(ref="Elective") vd3 (ref="No") chf_ (ref="No") pvd_ (ref="No") cvd_
(ref="No") dialysis_ (ref="No") prior_cv(ref="No") STS_hospnpi;

model rbc = year age bsa4c albumin_3c hct_ hct_gt36_ hct_gt39_ hct_gt43_
female ef4cat crealst4c race3c status3c vd3 chf_ pvd_ cvd_
dialysis_ prior_cv /link=logit dist=bin solution ;
random int/ subject=STS_hospnpi;
store parameter_dat;
output out=pre pred(noblup ilink)=p; run;

```

We then rank the predicted probability into deciles using PROC RANK. Variable "p" is the individual transfusion probability calculated by the fixed effect from the model.

```

* create probability deciles, and the ranks of probability <rank_p>;
proc rank data=pre out=ranky descending groups=10; var p; ranks rank_p;
run;

*output the median probability by deciles ('rank_p');
proc means data=ranky median mean;
var p ;
by rank_p;
output out=median_pr median=median_predict_p mean=mean_predict; run;

*output the observed transfusion rates by deciles, which is the number of
rbc events divided by the total number of observations in each decile;
proc sql;
create table observe_pr as
select sum(rbc) as no_events, count (*) as no_obs, calculated
no_events/ calculated no_obs as observe_pr, rank_p
from ranky
group by rank_p;quit;

* create the merge dataset that include the median probability and observed
transfusion rate in each decile;

data mergel;
merge observe_pr median_pr;by rank_p;run;

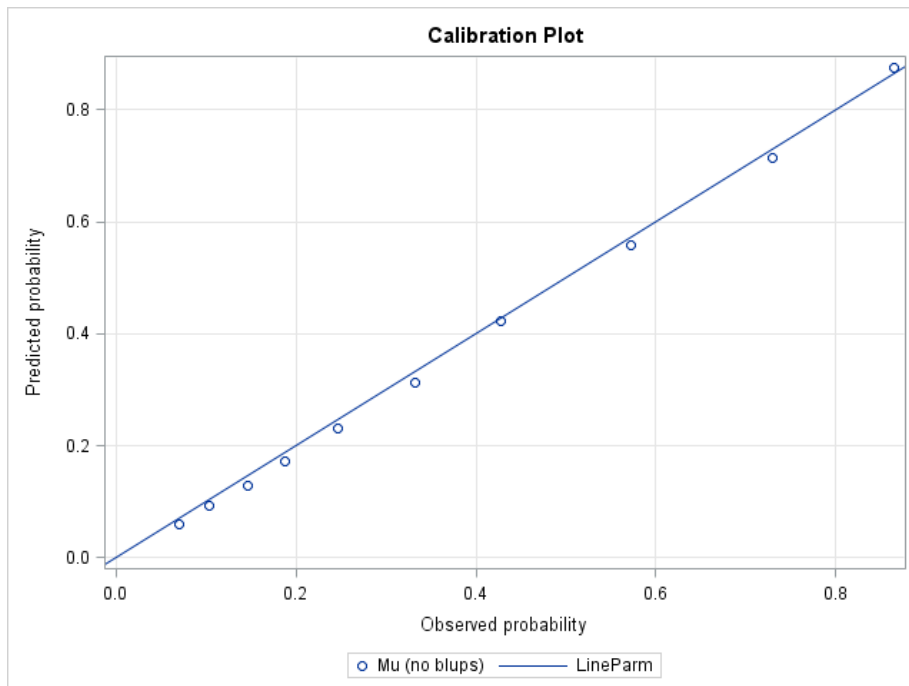
```

In each decile, we calculated the observed rate and the median prediction probability. We then plot the calibration plot with the observed transfusion rate in each decile on x axis, and the median prediction probability on y axis. The 45-degree line was added as a reference line. The data points which are

represented by circles fall close to the reference line (figure 1). This indicates that the model fits the data well.

```
*Plot the calibration Graph;  
proc sgplot data=merge1;  
  scatter x=observe_p y=median_predict_p;  
  lineparm x=0 y=0 slope=1; /** plot the reference line **/  
  xaxis grid; yaxis grid;  
run;
```

Figure 1.



## DISCRIMINATION

The common measure for model discrimination is the area under the receiver operating characteristic (ROC) curve [7]. This is often called a c-statistic, which can be interpreted as the probability that a subject with an observed outcome would have higher probability of predicted outcome than a subject without the observed outcome.

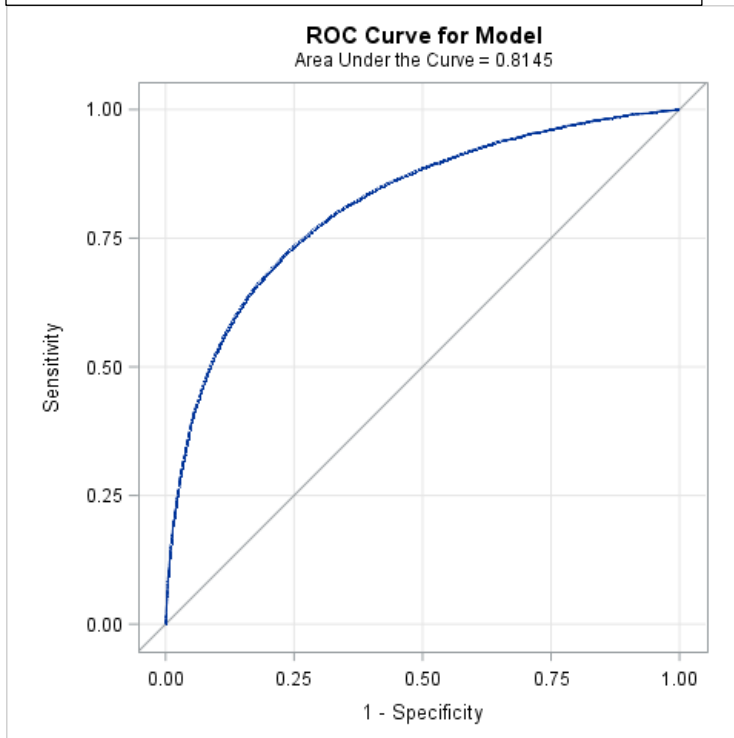
We first used the STORE Statement in GLIMMIX to store the parameters from the model development. ILINK was used so that the prediction is at the scale of probability. While GLIMMIX does not have ROC function, we used the predicted probability (variable "Predicted") and the ROC options in PROC LOGISTIC to generate the ROC curves.

```

proc plm restore=parameter_dat;
  score data=mix_model out=out/ilink;run;

proc logistic data=out descending ;
  model rbc = Predicted;
  roc;
  ods output ROCassociation=roc;
run;

```



## BOOTSTRAP SAMPLING

To assess model performance in different clinical subgroups, bootstrapping sampling can be used [8, 9]. In this example, we generated bootstrap samples by sampling patients with replacement from a defined clinical subgroup. There're many ways to create bootstrap samples in SAS, including PROC SURVEYSELECT, do loops. After generation of bootstrap samples, we calculated the C-Statistics in each bootstrapping sample, and estimate the bootstrapping mean and variance.

For example, one of the important clinical subgroups for blood transfusion is defined by a patient's admission status (i.e., elective, urgent or emergent). For each admission status, we created 100 bootstrapping samples from the original data. We applied the model estimates to these bootstrapping samples using PROC PLM RESTORE. We then calculated the C-Statistics within each sample. From the bootstrapping samples, we could assess the standard deviation of the C-Statistics. From this sensitivity analysis, we were able to validate how robust our model performance is among different clinical groups.

```

/*****create boot samples***/
%macro bootsample(b);
data sub1 (where=(status3c="Elective"))
    sub2 (where=(status3c="Urgent"))
    sub3 (where=(status3c="Emergent")); /* Create one data set for each
subgroup */
    set mix_model;
run;

data boot_subgroup;
%do t=1 %to 3;
    do sample=1 to &b;
        do i = 1 to nobs;
            pt = round(ranuni(&t)*nobs) ; /* ranuni returns a random number from the
uniform distribution on (0,1) interval */
            set sub&t nobs = nobs point=pt;
            output;
        end;
    end;
%end;
stop;
run;

%mend;

%bootsample(100);

```

```

/*****example: model application to the bootstrapping samples of
emergent status *****/

%macro combine;
%do i=1 %to 100;
  proc plm restore=parameter_dat;
    score data=boot_subgroup(where=(sample=&i and status3c="Emergent"))
out=out&i/ilink;run;

  proc logistic data=out&i descending ;
    model rbc = Predicted;
    roc;
    ods output ROCassociation=roc&i;
%end;
run;

data roc_test;
  set %do i=1 %to 100;roc&i %end;
  where ROCModel='Model';
run;
%mend;

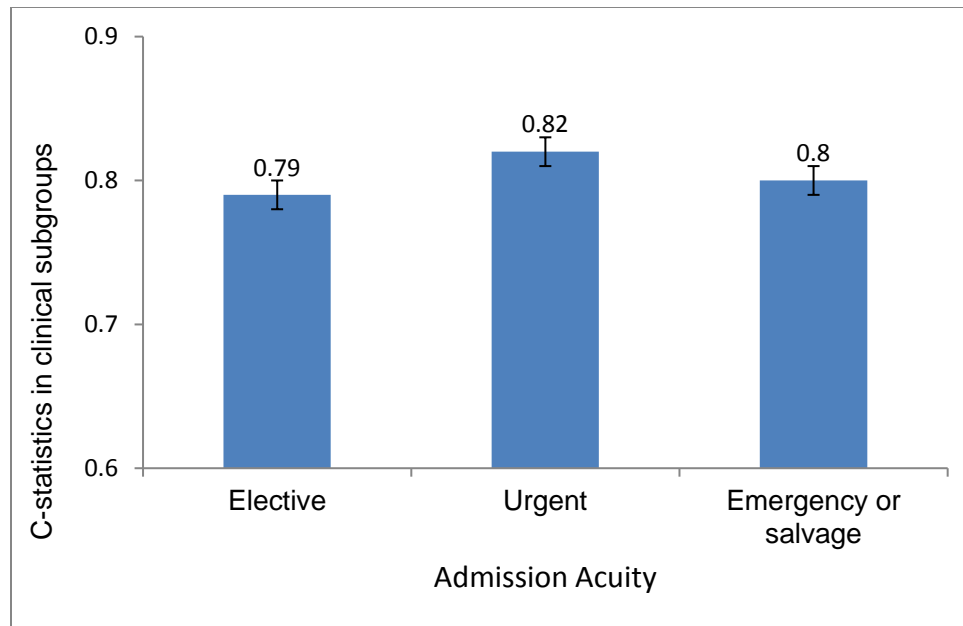
%combine;

/**** obtain mean and variance for c-statistics of modeling for emergent
status*****/
proc means data=roc_test mean std;
  var area;
run;

```

The model c-statistics was then calculated in different clinical subgroup. For example, here is the result of AUC in emergent admission patients from the bootstrap samples.

Analysis Variable : Area Area under the Curve	
Mean	Std Dev
0.8005692	0.0119594



## CONCLUSION

This paper covers some common techniques for validating the performance of a generalized mixed effect model. We demonstrated SAS applications in model calibration, discriminations and sensitivity analysis.

## REFERENCES

1. Alghamdi, A.A., et al., *Development and validation of Transfusion Risk Understanding Scoring Tool (TRUST) to stratify cardiac surgery patients according to their blood transfusion needs*. *Transfusion*, 2006. **46**(7): p. 1120-9.
2. Ranucci, M., et al., *Predicting transfusions in cardiac surgery: the easier, the better: the Transfusion Risk and Clinical Knowledge score*. *Vox Sang*, 2009. **96**(4): p. 324-32.
3. Speiss, B.D., *Transfusion and outcome in heart surgery*. *Ann Thorac Surg*, 2002. **74**(4): p. 986-7.
4. Paone, G., et al., *Transfusion of 1 and 2 units of red blood cells is associated with increased morbidity and mortality*. *Ann Thorac Surg*, 2014. **97**(1): p. 87-93; discussion 93-4.
5. Likosky, D.S., et al., *Prediction of Transfusions After Isolated Coronary Artery Bypass Grafting Surgical Procedures*. *Ann Thorac Surg*, 2017. **103**(3): p. 764-772.
6. EW, S., *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer, 2009.
7. Duchnowski, M., *Predictive Models: Storing, Scoring and Evaluating SAS*, 2017.
8. MR, C., *Bootstrap Methods: A Practioner's Guide*. John Wiley & Sons, 1999.
9. Nancy Barker, O.P.S., Wallingford, UK, *A Practical Introduction to the Bootstrap Using the SAS System*. Semantic Scholar, 2005.

## ACKNOWLEDGMENTS

Thanks to the support from The Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative.

## CONTACT INFORMATION

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

Xiaoting Wu (Ting), PhD, MS  
Department of Cardiac Surgery  
[1500 E Medical Center Drive](#)  
[Ann Arbor, MI 48109](#)  
[734.936.7731](#)  
[xiaotinw@med.umich.edu](mailto:xiaotinw@med.umich.edu)

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.