SAS® GLOBAL FORUM 2019

USERS PROGRAM

APRIL 28 – MAY 1 | DALLAS, TX

Xiaoting (Ting) Wu

Xiaoting (Ting) Wu, Ph.D. MS. , is a biostatistician at the Department of Cardiac Surgery at University of Michigan, US. She has SAS experience for more than 4 years, and is one of the winners for New SAS Professional Award SAS Global Forum 2019. Ting has many year of experience in data and statistical analysis including prediction models, survival analysis, mixed effect models and causal inference. She has broad interests in statistical methodology, statistical consulting, data visualization and SAS applications.

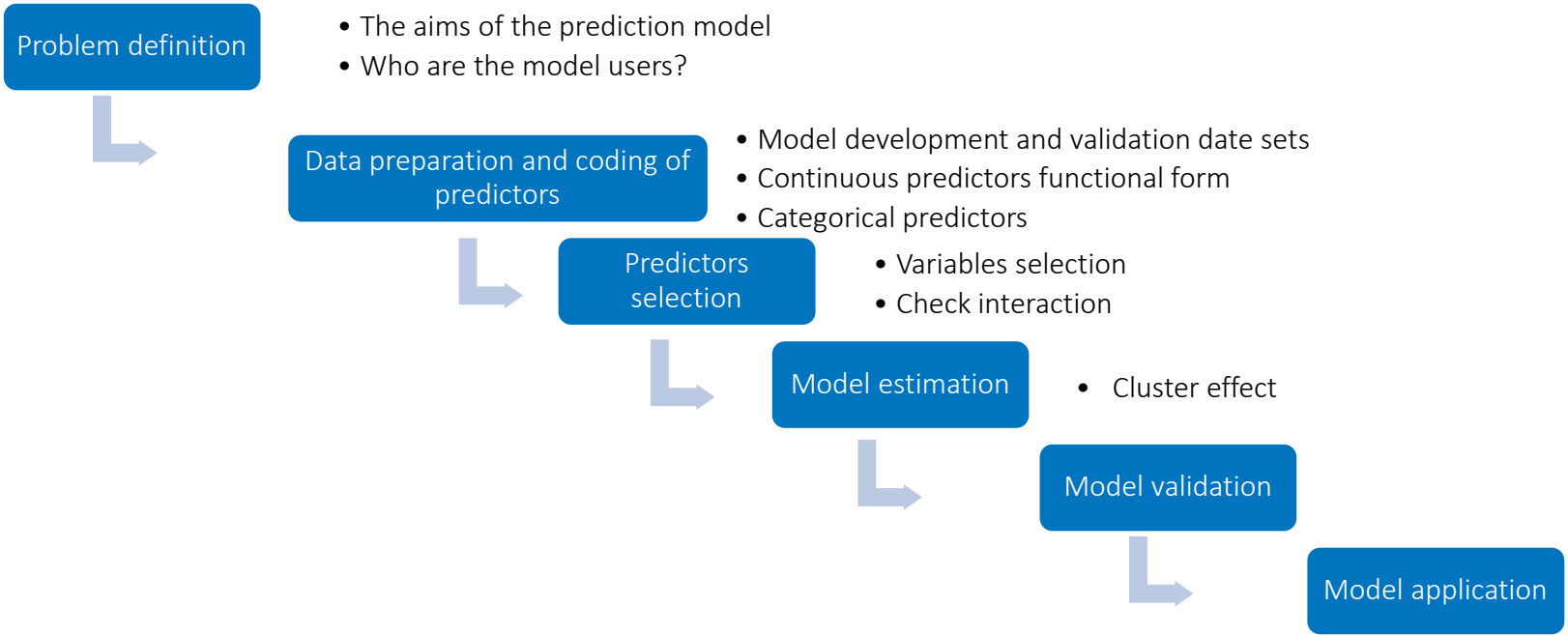
Using SAS® to Validate Prediction Models

Xiaoting (Ting) Wu

Department of Cardiac Surgery, University of Michigan, Ann Arbor, MI

Background

The process of establishing prediction models



Background

Example: Blood Transfusion Prediction Model

- Information of patients' blood transfusion risk prior to cardiac surgery may help clinicians assess a patient's condition and facilitate informed decision making
- Prediction outcome: blood transfusion risk
- Data: a multi-hospital dataset of more than 20,000 coronary artery bypass grafts procedures
- Transfusion rate was 36.8 %

REF: Likosky, D.S., et al., *Prediction of Transfusions After Isolated Coronary Artery Bypass Grafting Surgical Procedures.* Ann Thorac Surg, 2017.

Background

Example: Blood Transfusion Prediction Model

- Dataset: model development and validation dataset
- Variable selection and functional form assessment
- Generalized linear mixed effect model
- Final model: 16 preoperative predictors as fixed effect, and hospital as random effect

$$\text{Blood transfusion Probability} = \frac{\exp\left(\beta_0 + \beta_1 X_1 + \cdots + \beta_\kappa X_\kappa\right)}{1 + \exp\left(\beta_0 + \beta_1 X_1 + \cdots + \beta_\kappa X_\kappa\right)}$$

where, β_0: intercept in model, $\beta_1, \cdots \beta_\kappa$: regression

Background

Model validation techniques

Validation techniques	Measure	Description
Calibration	Calibration plot	Compares median observed with median predicted in deciles
Discrimination	C-statistics ROC curve	Interpretation as the probability of correct classification for a pair of subjects with and without the outcome
Bootstrapping resampling	Clinical subgroup c-statistics bootstrap mean and standard deviation	Determine the discriminative ability in the bootstrapping samples

Calibration

- Calibration demonstrates the agreement between observed and predicted outcomes.
- Option 1: Uses a smooth curve to compare predicted and empirical probabilities
- Option 2: Splits the data into risk deciles, and compare the predicted and empirical probabilities in each risk decile

Calibration

```
/*******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;
```

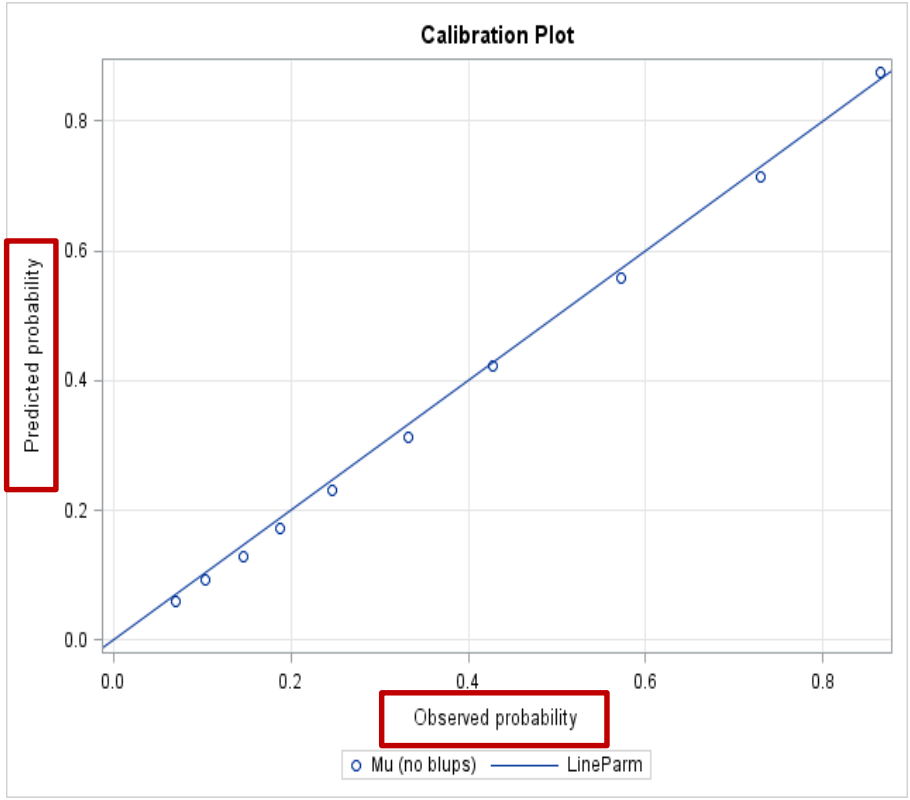
- subject=STS_hospnpi fits the random hospital effect.
- STORE statement to obtain the model estimate to “parameter_dat” dataset.
- OUTPUT statement to obtain the prediction from our mixed effect model.
- Option (NOBLUP) is used to exclude the predictors of the random effects when calculating the predicted probability for each patient

Calibration

- Obtain predicted probability and observed rate by deciles

```
* 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 merge1;
   merge observe_pr median_pr;by rank_p;run;
```

Calibration




```
*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;
```

Pearson's correlation coefficient = 1.00

Discrimination

- The common measure for model discrimination is the area under the receiver operating characteristic (ROC) curve (AUC).

Discrimination

- ROC curves plot sensitivity against 1-specificity
- AUC is equivalent to c-statistics
- C-statistics 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.
- A rough rule on c-statistics
  .80-1 = very good
  .70-.80 = good
  .50-.70 = weak

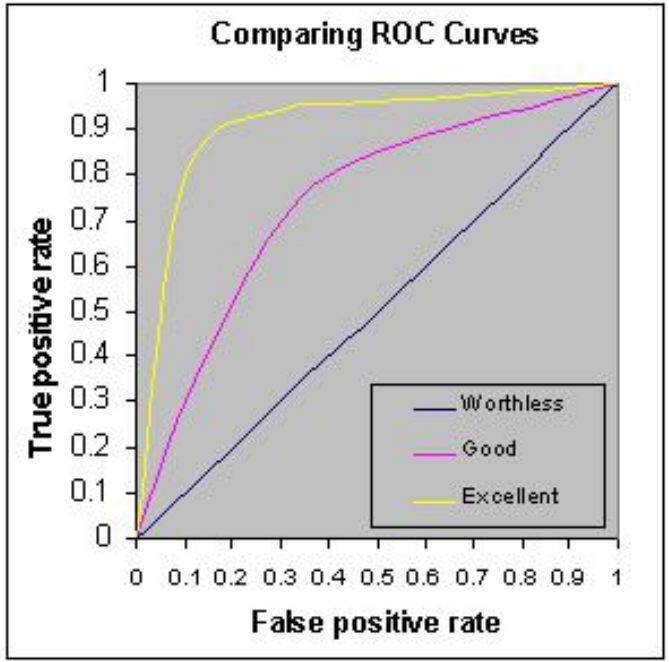
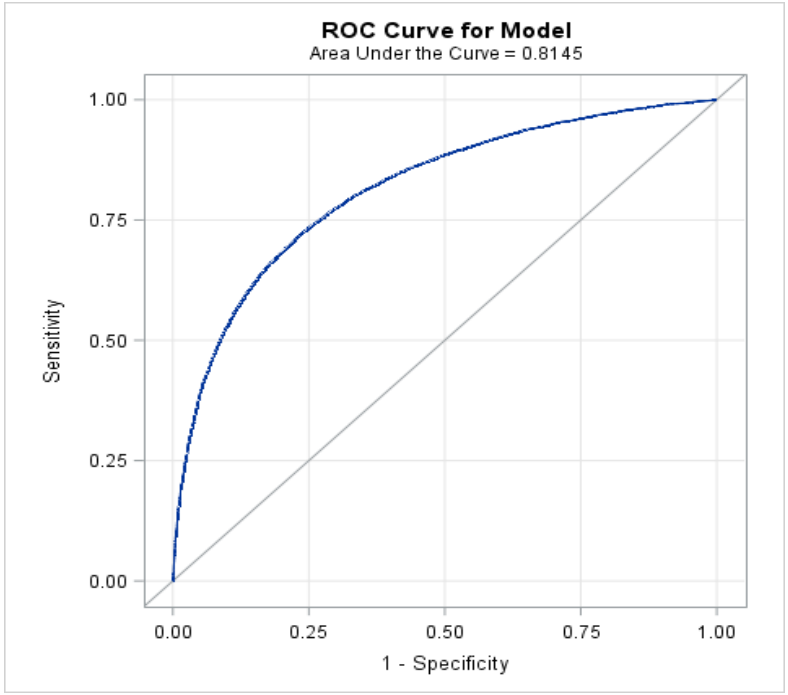


Figure is adapted from http://gim.unmc.edu/dxtests/roc3.htm

Discrimination

```
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;
```

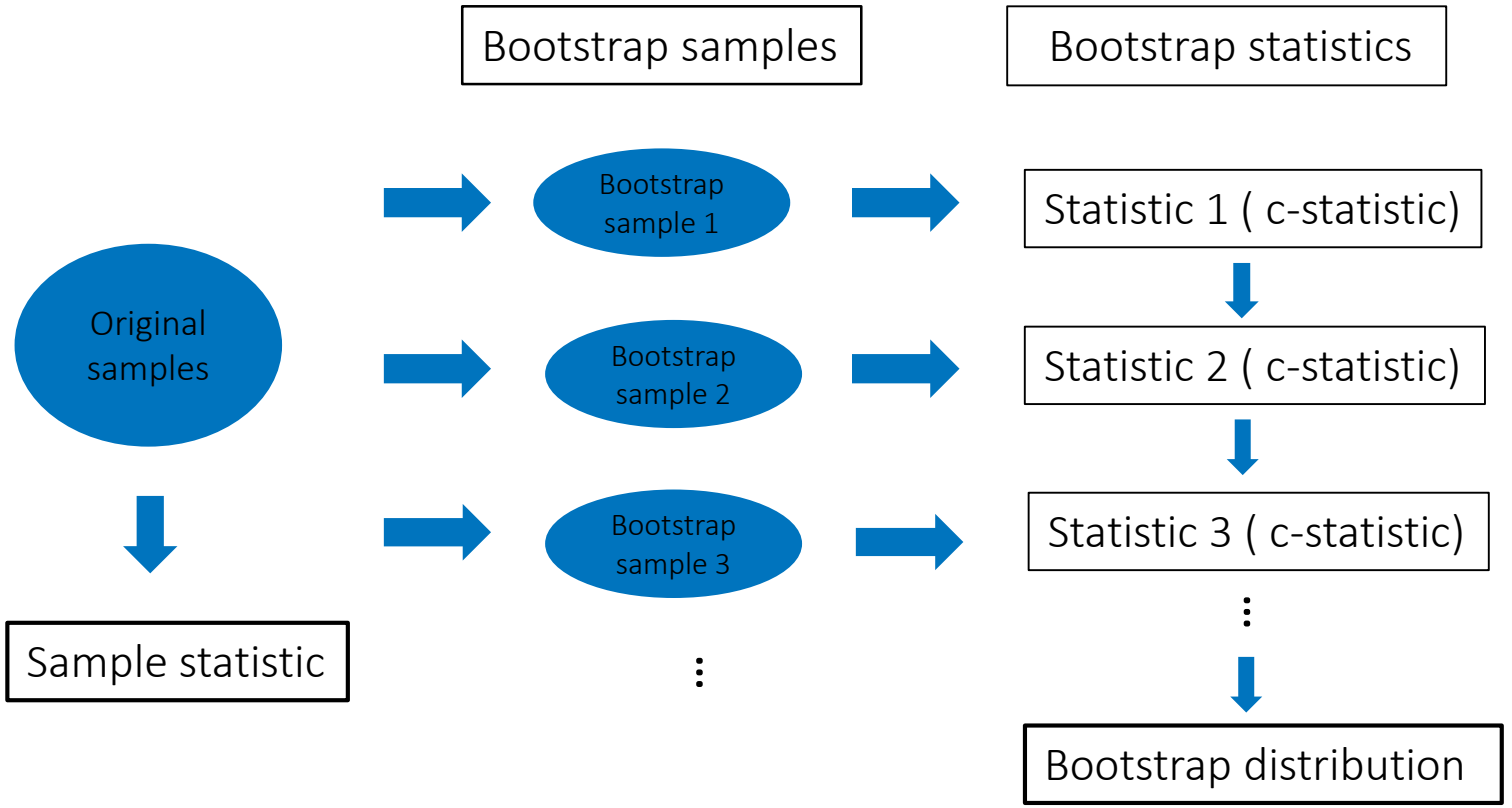
- While GLIMMIX does not have ROC function, we used the predicted probability (variable “Predicted”) generated from PROC PLM and ROC options in PROC LOGISTIC to generate the ROC curves.



Sensitivity analysis - Bootstrap

- The question: how well this model performs in different subgroup population? Age, admission acuity, etc.
- “Pull oneself up by one’s bootstraps” – “bootstrap” (Efron 1993)

Bootstrap



Bootstrap

The question: how well this model performs in different subgroup population?

Plan:

- Bootstrap samples patients with replacement from a defined clinical subgroup.
- Calculate the C-statistics in each bootstrap sample. Bootstrapping mean and variance of c-statistics can be obtained.

Bootstrap

```
/*******create boot samples, part of these codes adapted from Barker et al. ***/
%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);
```

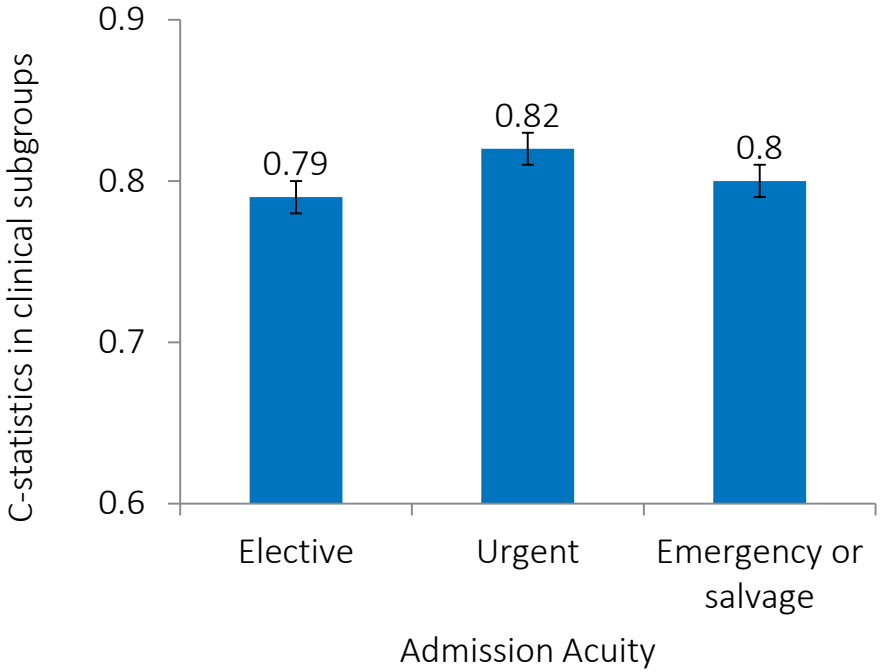
Bootstrap

Calculate c-statistics in each bootstrapping samples and combine the results.

```
/********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;
```

Bootstrap

C-statistics from the bootstrapping samples



Other thoughts

- Internal vs. External data validation
- Cross validation
- Bootstrapping resampling with PROC SURVEYSELECT procedure

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 using bootstrapping resampling.

Validation techniques	Measure	Description
Calibration	Calibration plot	Compares median observed with median predicted in deciles
Discrimination	C-statistics ROC curve	Interpretation as the probability of correct classification for a pair of subjects with and without the outcome
Bootstrapping resampling	Clinical subgroup c-statistics bootstrap mean and standard deviation	Determine the discriminative ability in the bootstrapping samples

Thank you!

Contact Information
Xiaoting Wu

xiaotinw@med.umich.edu

Reminder:

Complete your session survey in the conference mobile app.