

A SAS Macro to Compute Added Predictive Ability of New Markers in Logistic Regression

Kevin F Kennedy, St. Luke's Hospital-Mid America Heart Institute, Kansas City, MO
Michael J Pencina, Dept. of Biostatistics, Boston University, Harvard Clinical Research Institute, Boston, MA

ABSTRACT

Prediction of dichotomous events is an important statistical concept. In clinical research, development of a disease or complication after a procedure are often endpoints of interest. However, this concept is multi-disciplinary. Many published models exist to predict a dichotomous outcome, including a model to predict winners in the NCAA basketball tournament [1] and a recently published risk model to predict the likelihood of a bleeding complication after percutaneous coronary intervention [2]. It is important to note, however, these models are dynamic, as changes in population and technology lead to discovery of new and better methods to predict outcomes. An important consideration is when to add new markers to existing models. While a significant p-value is an important condition, it does not necessarily imply an improvement in model performance. Traditionally, receiver operating characteristic (ROC) curves and its corresponding area underneath the curve (AUC) are used to compare models, specifically Delong's comparison for two correlated AUCs. However, the clinical relevance of this metric has been questioned by researchers [3, 6, 7]. To address this issue, Pencina and D'Agostino [3] have proposed two statistics to evaluate the significance of novel markers. The Integrated Discrimination Improvement (IDI) measures the new model's improvement in average sensitivity without sacrificing average specificity. The Net Reclassification Improvement (NRI) measures the correctness of reclassification of subjects based on their predicted probabilities of events using the new model with the option of imposing meaningful risk categories. A related approach based on the concept of net benefit has been proposed by Vickers et al. [4,5] and allows for a graphical representation of the gain incurred using the new model. This paper presents a simple SAS macro to obtain output of traditional comparisons (AUC) and novel measures (IDI, NRI, and Vickers' Decision Curve).

INTRODUCTION

Consider comparing 2 logistic regression models to predict a dichotomous outcome with the main difference being model 2 includes a variable for a new marker as a predictor:

$$\text{Model1} : \text{logit}(y) = \alpha + \beta_1 x_1 + \dots + \beta_n x_n$$

$$\text{Model2} : \text{logit}(y) = \alpha + \beta_1 x_1 + \dots + \beta_n x_n + \beta_{n+1} x_{n+1}$$

The output of each model will be a predicted probability of experiencing an event, and each individual in the dataset will have 2 probabilities, one from each model.

AUC

Difference in the area under the receiver operating characteristic curve (AUC) is a common method to compare two models. AUC is a measure of model discrimination (i.e., how well the model separates subjects who did and did not experience an event). It essentially depicts a tradeoff between the benefit of a model (true positive or sensitivity) vs. its costs (false positive or 1-specificity). AUC is computed by comparing all possible pairs in a dataset between individuals experiencing and not experiencing an event. For example, a dataset with 100 events and 1000 non-events would have $100 \times 1000 = 100,000$ pairs. In each pair the predicted probability is compared. If the individual experiencing an event has a higher predicted probability, that pair would be labeled 'concordant' and assigned a value of 1. Conversely, if the individual experiencing an event has a lower probability the pair would be labeled 'discordant' and assigned a value of 0. AUC would be an average of the 1s and 0s (and 0.5s for identical values).

AUC ranges from 0.5 (no discrimination) to 1 (perfect discrimination); however, values close to 1 are very rarely seen in contemporary data [6]. The value of AUC is important, but the real interest in this context lies on how much the AUC increases with the addition of the new marker. To address this scenario, Delong [9] presented a method for comparing 2 correlated AUCs. Prior to SAS version 9.2 you could run the %roc macro located on the SAS website [8]. However, to perform this comparison in version 9.2 two new statements were added to proc logistic (roc and roccontrast) to bypass running the %roc macro.

```

Sample code for SAS 9.2:
proc logistic data=auc;
model eve=x1 x2 x3 x4 x5;
roc 'first' x1 x2 x3 x4;
roc 'second' x1 x2 x3 x4 x5;
roccontrast reference('first')/estimate;
run;

```

The above code will return a comparison of model 1 (with the first 4 covariates) with model 2 (containing the additional covariate). For a pre-version 9.2 example go to the SAS website [8].

Many advantages exist for reporting AUC. First, it is a well known statistic, commonly seen in any model predicting a dichotomous outcome. Second, there are recommend ranges for an excellent/good/poor value for AUC making it easier to evaluate the model. Third, it is the default output in most all statistical software including proc logistic in SAS.

There are also disadvantages with AUC. First, it is a rank based statistic. A pair of individuals with probabilities 0.5 and 0.51 would be treated the same as the probabilities 0.1 and 0.9. Second, it focuses only on discrimination which is not the only and may not be the most important in the context of risk prediction. For example, a model that assigns all events a probability of 0.51 and all non-events a probability of 0.5 would have an AUC of 1; however, it would be difficult to label this a useful model. Third, it is very difficult for a new marker to significantly change the value of AUC [7]. Fourth, the magnitude of improvement in AUC is not nearly as meaningful as the value of AUC itself. For this reason it is very difficult to use increase in the AUC to justify adding new markers to a model even though these markers can have major impact at the individual level [6].

ALTERNATIVE MEASURES

AUC is a helpful measure, but does have shortcomings in comparing these models. To address these Pencina and D'Agostino [3] presented 2 additional statistics, Integrated Discrimination Improvement (IDI) and Net Reclassification Improvement (NRI). Vickers [REF?] presented a method for graphically comparing two models.

IDI

IDI is a measure of the new model's improvement in average sensitivity (true positive rate) without sacrificing its average specificity (true negative rate). In comparing the models, IDI measures the increment in the predicted probabilities for the subset experiencing an event and the decrement for the subset not experiencing an event:

$$Absolute_IDI = (\bar{p}_{event_2} - \bar{p}_{event_1}) + (\bar{p}_{nonevent_1} - \bar{p}_{nonevent_2}) \quad (1)$$

$$Relative_IDI = \frac{\bar{p}_{event_2} - \bar{p}_{nonevent_2}}{\bar{p}_{event_1} - \bar{p}_{nonevent_1}} - 1 \quad (2)$$

where :

\bar{p}_{event_2} is the mean predicted probability of the events in the 2nd model

IDI can also be described with the following graph.

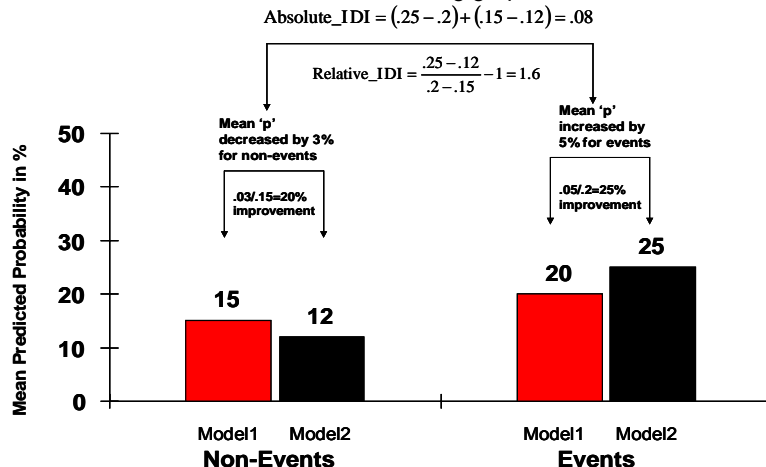


Figure 1: IDI

One caveat is that absolute IDI does not provide guidelines for what represents a clinically important value, mostly due to the fact that this number will be heavily driven by the event rate in the data. If the predicted event only has a rate of 3% then the absolute IDI value will be a relatively small value. This is where relative IDI is useful—it returns a percentage (no units). Pencina and D’Agostino provide an example where the absolute IDI is only 0.009 but represented a 13% improvement on the relative scale. In Figure1 the relative IDI value of 1.6 represents a 160% improvement.

NRI

NRI is very similar to IDI but differs as it treats the predicted probabilities on a categorical scale. Using these categories NRI will evaluate the number of ‘net’ reclassified individuals using model2 over model1. This is done by calculating how many individuals experiencing an event increased in risk category (e.g., from medium to high) and how many individuals not experiencing an event decreased in risk category (e.g., medium to low).

$$NRI = \left[\left(\frac{\# \text{Events moving up}}{\# \text{ of Events}} \right) - \left(\frac{\# \text{Events moving down}}{\# \text{ of Events}} \right) \right] + \left[\left(\frac{\# \text{Non - events moving down}}{\# \text{ of Non - events}} \right) - \left(\frac{\# \text{Non - events moving up}}{\# \text{ of Non - events}} \right) \right] \tag{3}$$

There are 3 potential methods for defining categories. First, any movement. Using this classification if an individual’s probability using model 1 was 0.17 and using model 2 it became 0.171 this would be considered an ‘up’ movement. Second, a threshold to define movement. For example, a threshold of 3% would mean that to be considered an ‘up’ movement the model 2 probability would need to be 0.03 higher than model 1 (e.g., 0.17 to 0.21). Using this definition the ‘up’ movement in case1 above would be considered a ‘no’ movement. Third, a user defined classification. This would require more intimate knowledge about the data and population but one could define 3 groups: low (probability<10%), moderate (Probability 10-20%), and high (probability >20%). Here if an individual was a ‘low’ in model 1 but became a ‘medium’ in model 2 it would be considered an ‘up’ movement.

NRI COMPUTATION EXAMPLE:

A dataset with 100 events and 200 non-events, using 3 user defined categories: <10% (low), 10-20% (moderate), >20% (high). By making 2 separate (event and non-event) crosstabs one can calculate NRI:

Table1: Crosstab for Events (n=100)

		Model 2			
		Low	Mod	High	
Model 1	Low	10	8	2	20 Events moving up 10 Events moving down 70 Events not moving Net of 10/100 (10%) of events getting reclassified correctly
	Mod	3	30	10	
	High	2	5	30	

Table2: Crosstab for Non-events (n=200)

		Model 2			
		Low	Mod	High	
Model 1	Low	50	5	0	15 Non-events moving up 150 Non-events moving down 35 Non-events not moving Net of 20/200 (10%) of non-events getting reclassified correctly
	Mod	20	40	10	
	High	5	10	60	

$$NRI = \left[\left(\frac{20}{100} \right) - \left(\frac{10}{100} \right) \right] + \left[\left(\frac{35}{200} \right) - \left(\frac{15}{200} \right) \right] = .2$$

It is important to remember that NRI is dependent on the groups chosen. For example, the results could be different had the groups been <15%, 15-30%, and >30%, by defining 4 groups instead of 3, or using any up/down or threshold definitions. Due to this issue, defining your NRI strategy before data analysis is recommended.

VICKERS DECISION CURVE

The last measure of incremental predictive ability is Vickers Decision Curve [4, 5]. This is an important measure if a specific threshold was used to classify an individual (e.g., if an individual’s predicted probability is greater than 10% a different treatment strategy will be used). In this situation we would want to compare model performance at the specific threshold. By using a formula for net benefit, which was first attributed to Peirce in 1884, a graphical comparison of models 1 and 2 can be made

$$NetBenefit = \frac{True\ Positives}{n} - \left(\frac{False\ Positives}{n} \cdot \frac{p_t}{(1-p_t)} \right)$$

VICKERS DECISION CURVE EXAMPLE

Consider evaluating net benefit from both models in a sample of 1000, with a threshold of 10%.

Table3: Vickers Model1

		Outcome	
		Yes	No
Predicted Probability	≥10%	80	200
	<10%	20	700

$$NetBenefit = \frac{80}{1000} - \left(\frac{200}{1000} \cdot \frac{.1}{(1-.1)} \right) = .0578$$

Table4: Vickers Model2

		Outcome	
		Yes	No
Predicted Probability	≥10%	85	180
	<10%	15	720

$$NetBenefit = \frac{85}{1000} - \left(\frac{180}{1000} \cdot \frac{.1}{(1-.1)} \right) = .065$$

There are 2 additional strategies that can be computed through Vickers analysis: 1) “treat all” which assumes all individuals are positive and 2) “treat none” which assumes all individuals are negative.

$$NetBenefit_{treatall} = \frac{100}{1000} - \left(\frac{900}{1000} \cdot \frac{.1}{(1-.1)} \right) = 0$$

$$NetBenefit_{treatnone} = \frac{0}{1000} - \left(\frac{0}{1000} \cdot \frac{.1}{(1-.1)} \right) = 0$$

The first comparisons to make are the models vs. treat all and treat none. This comparison determines if the models are useful at the particular threshold. If the net benefit of the models are the same as treat all or treat none it is unwise to even use the model for decision making. In this case, both models are preferable to both treat all and treat none. The second comparison is model1 vs. model2. Here the net benefit is higher for the 2nd model making it preferable at this threshold.

However, simply determining that the net benefit is higher for one model over another is probably not enough to justify its use. Vickers [5] describes bootstrapping methods to obtain 95% CIs for the difference in net benefit between the 2 models.

Vickers Decision Curve is created by computing net benefit across a range of thresholds. By using do loops in SAS we can compute net benefit for both models for each integer threshold (0—100%). The following is an example of Vickers Decision Curve:

Vickers Decision Curve

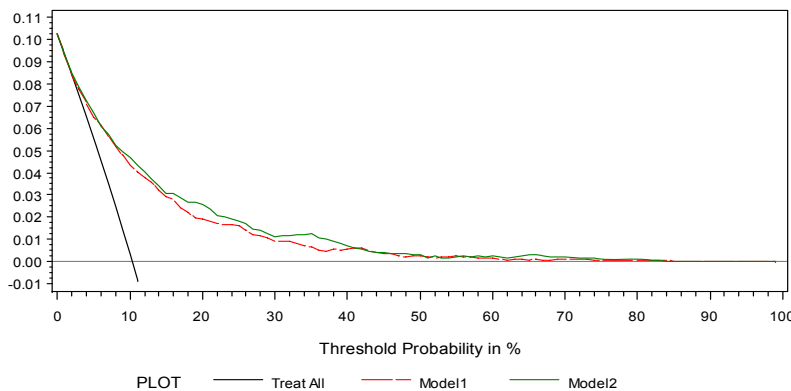


Figure 2: Sample Vickers Decision Curve

This plot allows us to see thresholds at which the model is useful. The top curve represents the most useful model. At the threshold from 0—5% the 3 curves are nearly indistinguishable. The black line is the treat all strategy which assumes every individual “positive”, and this line has nearly identical net benefit as Model1 and Model2. If the threshold is in this range it is unwise to use either model as there is no benefit over assuming all individuals are positive. Next, the threshold range from

50%—100% the three lines are also nearly identical. The Gray line is the treat none strategy which assumes every individual is “negative”, and this line has nearly identical net benefit as Model1 and Model2. If the threshold is in this range it is unwise to use either model as there is no benefit over assuming all patients are negative. Last, the threshold range from 5%—50% represents the threshold at which Model2 (green) will be of more benefit than Model1, since both models are above the treat all and treat non” lines.

SAS MACRO TO OBTAIN OUTPUT

Now sample output and code will be shown from the %added_pred macro. Unfortunately, due to the length of the macro and calls to other macros only portions of the code will be shown. The corresponding author can provide the full code on request.

```
%macro added_pred(data= ,id= , y= , model1cov= , model2cov=,
nripoints=ALL,nriallcut=%str( ), vickersplot=FALSE,maxvicker=100,
vickerpoints=%str( ));
```

This macro only has 5 required user inputs. By entering just the data, id variable, outcome(y), model1 covariates, and model2 covariates, the user will obtain output on AUC, IDI, and NRI using the groups of any up/down movement. Below are the optional inputs:

- 1) If the user wants defined NRI groups (i.e. <10%, 10-20%, and >20%) enter: **nripoints=.1 .2**
- 2) If the user wants to use a threshold for NRI (i.e. 3%) enter **nriallcut=.03**
- 3) If the user wants Vickers Decision Curve enter **vickersplot=TRUE**
- 4) If the user wants specific thresholds tested for Vickers curve (say threshold=10%) enter **vickerpoints=.1**
- 5) if the user doesn't want to show the entire range of thresholds on Vickers Plot (say, 0-50%) enter **maxvicker=50**

The user can also request multiple NRI thresholds (nriallcut=.03 .05 .07) and multiple thresholds for Vickers curve (vickerpoints=.1 .2 .3). Also, the user can request NRI done all 3 ways: any up/down, threshold, and user defined.

AUC OUTPUT

SAS version 9.2 makes AUC comparisons easy (see sample code presented in AUC section). By entering only the required inputs the macro will return the following comparisons:

Model1 AUC	Model2 AUC	Difference in AUC	Standard Error of Difference in AUC	P-value for AUC Difference	95% CI for Difference in AUC
0.77907	0.79375	0.0147	.0042	0.0005	(0.00646,0.0229)

Here you can evaluate the individual AUC values (0.779, 0.794) for both models along with a comparison between the models by using the methods of Delong (difference=0.0147, p=0.005).

IDI OUTPUT

To compute IDI the macro runs two separate logistic models and outputs predicted probabilities to an output dataset:

```
Model1:
proc logistic data=&data;
model &y=&model1cov;
output out=old pred=p_old;
run;

Model2:
proc logistic data=&data;
model &y=&model2cov;
output out=new pred=p_new;
run;
```

By merging the two output data sets together we can then use proc means or proc sql to obtain the 4 needed mean predicted probabilities. Below is sample code to obtain the count and mean predicted probability for both models (in the subset experiencing an event) and it will turn them into macro variables by using the (into:) statement in proc sql.

```

proc sql noprint;
select count(*),avg(p_old), avg(p_new) into :num_event, :p_event_old,
:p_event_new
from mergeddata
where &y=1;
quit;

```

Once the counts, mean probabilities, and standard errors are calculated we can calculate absolute and relative IDI, and by using only the required input the following will be output:

Integrated Discrimination Improvement	IDI Standard Error	Z-value for IDI	P-value for IDI	IDI 95% CI	Probability change for Events	Probability change for Nonevents	Relative IDI
0.020711	.006441029	6.30234	<.0001	(0.0143,0.0272)	0.018586	-.002125	0.16679

In this sample output we have a statistically significant absolute IDI value of 0.0207 (0.018586+0.002125), we also have a relative IDI value of 0.167 representing a ~17% improvement.

NRI OUTPUT

The macro computes NRI in 3 ways: 1) any up/down movement, 2) a threshold, and 3) user defined groups. Recall, the user can test multiple thresholds and any user defined group. Because of this, it is imperative to count the number of groups for the user and threshold strategies. This will be performed by using the %words macro:

```

%macro words(list);
%local count;
%let count=0;
%do %while(%qscan(&list,&count+1,%str( )) ne %str());
%let count=%eval(&count+1);
%end;
&count
%mend words;

```

This macro will simply return the number of words in a character string (separated by spaces). For example:

```

%let string=This is a string;
%put %words(&string);

```

The log will output the number '4'.

If, for example, the user wishes to define the groups (<5%, 5-10%, >10%) then enter into the macro: {nrpoints=.05 .1}. Then by using the %words macro we can figure out the number of groups by:

```

%let numgroups=%eval(%words(&nrpoints)+1);

```

This will return the number 3. Once we know the number of groups we can compare the predicted probabilities to determine what group an individual falls in by using the following code:

```

{1} if 0<=p_old<=%scan(&nrpoints,1,' ') then group_pre=1;
{2} if 0<=p_new<=%scan(&nrpoints,1,' ') then group_post=1;

%let i=1;
{3} %do %until(&i>%eval(&numgroups-1));
{4} if %scan(&nrpoints,&i,' ')<p_old then do;
group_pre=&i+1;
end;
{5} if %scan(&nrpoints,&i,' ')<p_new then do;
group_post=&i+1;
end;
%let i=%eval(&i+1);
%end;

```

Steps {1} and {2} compare the old and new predicted probabilities with the first number in &nrpoints. If the predicted probabilities are less than this number they will be in group 1 (group_pre for model1, group_post for model2)

Steps {3}-{5} start a do loop for the other groupings. Once each individual has their model1 and model2 groups a comparison can be made to determine movement: up, down, or no movement. If group_post>group_pre there was an 'up' movement, if group_post=group_pre there was 'no' movement, and if group_post<group_pre there was a 'down'

movement.

Coding for 'any' up/down and 'threshold' is similar and also requires use of the %words macro. By default, output for 'any' up/down will be displayed. This is performed by defining a threshold of '0'. If the user wishes to test additional thresholds (say 3%) enter nrilcut=.03 {note can enter more than one just separate by spaces}. The following code will count the number of thresholds:

```
{6} %let nricuts=0 &nriallcut;
{7} %let numcut=%words(&nricuts);
```

Line 6 will add '0' to the thresholds which will be used for the 'any' up/down movement, and Line 7 will count the number of thresholds (in this example '2' will be returned for .03 and 0). Now by using the following code up/down movement can be determined:

```
{8} %do i=1 %to &numcut;
{9} %let nricut=%scan(&nricuts,&i,' ');
      data nricut&i;
      set probs;
{10} if &y=1 then do;
      if pdiff<=&nricut then down_event=1;
      if pdiff>&nricut then up_event=1;
      end;
{11} if &y=0 then do;
      if pdiff<=&nricut then down_nonevent=1;
      if pdiff>&nricut then up_nonevent=1;
      end;
      {additional statement follow}
```

This code will create 2 datasets: one for any up/down and one for a threshold of 3%. Steps {10} and {11} are dependent upon a variable defined earlier in the macro 'pdiff' which is simply the difference between the predicted probability in model 2 vs. model 1 (p_new-p_old). This value will be compared to the threshold being used at the time in the do loop (here, either 0 or .03).

In these particular examples the code below will display the output:

```
%added_pred(..Required input..,nripoints=.05 .1,nriallcut=.03);
```

Group	Net Reclassification Improvement	NRI Standard Error	Z-Value for NRI	NRI P-Value	NRI 95% CI	% of Events correctly reclassified	% of Nonevents correctly reclassified
ALL	0.45478	0.089899	9.70704	<.0001	(0.3649,0.5447)	(10%)	56%
cut_.03	0.10110	0.080563	2.45678	0.0140	(0.0205,0.1817)	4%	6%
User	0.12699	0.050045	4.95013	<.0001	(0.0769,0.177)	5%	8%

All three strategies of NRI (All, threshold, and user) result in statistically significant results. However, there are vast differences between 'All' and the other two strategies. Notice, 56% of the nonevents were reclassified correctly under the 'All' strategy but only 6% using a threshold of 0.03. This indicates most nonevents went down in probability between model 1 and 2; however, a vast majority of those did not decrease by 3% or more. Also, note that there was a loss of (10%) reclassification for the events with the 'all' strategy but a 4% improvement using the threshold of 3%. This is indicating most events went down in probability using any up/down; however, if one uses a threshold to define up/down there was a benefit for the events (meaning most that went down in probability did not go down by a substantial amount). It is important to keep in mind that any movement in the 'all' strategy will be classified an up/down, so caution should be used in interpreting these results.

VICKERS DECISION CURVE OUTPUT

Recall that Vickers Decision Curve is produced by computing 'net benefit' at a range of thresholds. This can be done in SAS by do loops:

```
%if &vickersplot=TRUE %then %do;
      %do i=0 %to 100;
      data vicker&i;
```

```

set probs(keep=&iid p_new p_old &y);
if p_new>&i/100 then positive=1;
else positive=0;
if p_old>&i/100 then positive_old=1;
else positive_old=0;
run;
{additional statements follow}

```

Here, if the predicted probabilities are above the threshold (&i/100) they will be labeled as positive. Now, through proc freq we can obtain counts of true and false positives. A sample plot is found in Figure 2.

Also, the user can request specific threshold analysis. For example, if you know a specific threshold is important (say 15%) you can enter into the macro: {vickerpoints=.15} to get the following output.

cutpoint	mean(95%CI) for difference in Net Benefit(new-old)	mean(95%CI) for net benefit of Old Model	mean(95%CI) for difference in Net Benefit (old-treatall)	mean(95%CI) for net benefit of New Model	mean(95%CI) for difference in Net Benefit (new-treatall)
15	0.0013(-0.0026,0.005)	0.0295(0.0229,0.0362)	0.0852(0.0783,0.0922)	0.0308(0.0244,0.0374)	0.0866(0.0799,0.093)

Notice in the 2nd column there is a comparison between model1 and model2, which is accomplished through bootstrapping techniques. Notice that '0' is contained in the 95% CI indicating no real difference between the models at this threshold. The 3rd and 5th columns are the net benefit for Model1 and Model2, respectively. The 4th and 6th columns are a comparison between model1 (4th column), model2 (6th column) vs. the strategy of "treat all" which assumes all patients are positive. By inspecting the CIs there is positive net benefit to both models, and each model is better than "treat all"; however, there is no real difference between Model1 and Model2.

CONCLUSION

Predicting dichotomous events will continue to be an important aspect of research. Additionally, improving already usable models will be just as important. This paper presented 4 useful measures in determining model improvement by the addition of new markers. Pencina and D'Agostino [3] state, "the choice of the improvement metric should take into account the question to be answered". If the primary goal is to determine if an individual falls above or below a specific cut-off then NRI or Vickers analysis will probably be the preferred metrics. On the other hand, if no cut-offs exist it may be best to look at IDI and AUC. Also, the example from Pencina and D'Agostino [3] shows a case where the addition of a new variable to a model results in a very small change in AUC but a marked improvement in IDI and NRI. The small change in AUC is explained more fully by Pepe et al. [7] and may justify not solely using AUC as the guideline for model improvement but rather a combination of metrics. It is also important to note that other factors will play a role in determining to add a marker. For example, if the new marker is more difficult and expensive to obtain guidelines on what constitutes a significant improvement in model performance will likely be more strict. It is important to keep in mind all of these factors in assessing model performance.

This paper also presented an easy to use SAS macro to obtain output of all measures. All sample output would be produced with the following code. You may obtain the entire macro and additional examples by contacting the corresponding author.

```
%added_pred(data=dataset, id=idvar, y=outcome, model1cov=x1 x2 x3 x4, model2cov=x1 x2 x3 x4 x5, nrpoints=.05 .1, nrallcut=.03, vickersplot=TRUE, vickerpoints=.15);
```

AWKNOWLEGMENTS

We thank Josh Murphy for his help in review of the paper.

REFERENCES

- 1) Kvam P, Sokol J. A Logistic Regression/Markov Chain Model for NCAA Basketball. Naval Research Logistics 2006;53:788-803.
- 2) Mehta SK, Frutkin AD, Lindsey J, Spertus JA, Rao SV, Ou FS, Roe MT, Peterson ED, Marso SP. Bleeding in Patients Undergoing Percutaneous Coronary Intervention: The Development of a Clinical Risk Algorithm from the National Cardiovascular Data Registry. Circ Cardiovasc Intervent. 2009;2:222-229.

- 3) Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med* 2008; **27**:157-72.
- 4) Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*. 2006 Nov-Dec;26(6):565-74
- 5) Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, A novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Medical Informatics and Decision Making*. 2008 Nov 26;8(1):53.
- 6) Cook NR. Use and Misuse of the receiver operating characteristics curve in risk prediction. *Circulation* 2007; **115**:928-935.
- 7) Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004;159:882-890.
- 8) <http://support.sas.com/kb/25/017.html>
- 9) E.R. DeLong, D.M. DeLong, and D.L. Clarke-Pearson (1988), "Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics*, 44, 837-845.

CONTACT INFORMATION

Your comments are very encouraged, and if you wish to use the %added_pred macro please contact me at

Kevin Kennedy
St. Luke's Hospital-Mid America Heart Institute
4401 Wornall Rd
Kansas City, MO, 64111
816-932-1799
kfkennedy@saint-lukes.org or kfk3388@gmail.com

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.