Journal of Modern Applied Statistical Methods

Volume 18 | Issue 2 Article 19

9-1-2020

JMASM 54: A Comparison of Four Different Estimation Approaches For Prognostic Survival Oral Cancer Model

Wan Muhamad Amir
Universiti Sains Malaysia, wmamir@usm.my

Muhammad Azeem
Universiti Sains Malaysia, dr.axeem.sr@gmail.com

Masitah Hayati Harun
Universiti Sains Malaysia, masitahh@usm.my

Zalila Ali
Universiti Sains Malaysia, zalila_ali@usm.my

Mohamad Shafiq
International Islamic University Malaysia, shafiq@iium.edu.my

Recommended Citation

Amir, W. M., Azeem, M., Harun, M. H., Ali, Z., & Shafiq, M. (2019). JMASM 54: A comparison of four different estimation approaches for prognostic survival oral cancer model. Journal of Modern Applied Statistical Methods, 18(2), eP2823. doi: 10.22237/jmasm/1604189760

JMASM 54: A Comparison of Four Different Estimation Approaches For Prognostic Survival Oral Cancer Model

Cover Page Footnote

The authors would like to express their gratitude to Universiti Sains Malaysia (USM) for providing the research funding (Grant No. 1001/PPSG/8012278, School of Dental Sciences, Kampus Kesihatan).

Journal of Modern Applied Statistical Methods
November 2019, Vol. 18, No. 2, eP2823.
doi: 10.22237/jmasm/1604189760

בס"ד

ISSN 1538 − 9472

JMASM 54: A Comparison of Four Different Estimation Approaches for Prognostic Survival Oral Cancer Model

Wan Muhamad Amir
Universiti Sains Malaysia
Kelantan, Malaysia

Muhammad Azeem
Universiti Sains Malaysia
Kelantan, Malaysia

Masitah Hayati Harun
Universiti Sains Malaysia
Kelantan, Malaysia

Zalila Ali
Universiti Sains Malaysia
Kelantan, Malaysia

Mohamad Shafiq
International Islamic University Malaysia
Pahang, Malaysia

Four types of estimation approaches for prognostic survival oral cancer model building are considered via a SAS algorithm: Efron's Method, Exact Method, Breslow's Method, and Discrete Method. Each method is illustrated separately and compared according to their coefficient parameter. An approach is considered by adding a bootstrapping technique for each handling ties method and a complete SAS algorithm is supplied for each proposed method, including methods for handling ties.

Keywords: Prognostic survival oral cancer model, Efron's method, Exact method, Breslow's method, Discrete method, bootstrapping, SAS

Introduction

The proportional hazards model, Cox (1972) regression (also known as survival model or prognostic survival cancer model) estimates the effects of different covariates influencing the times to the failures of a system. It is used extensively in biomedicine, and reliability engineering. Survival model or prognostic survival oral cancer model is powerful tools that are used frequently in studies of clinical outcomes. These models can use a mixture of categorical and continuous variables and can handle partially observed (censored) responses. However, uncritical application of modeling techniques can result in models that poorly fit the dataset at hand, or, even more likely, inaccurately predict outcomes in new subjects.

doi: 10.22237/jmasm/1604189760 | Accepted: July 9, 2020; Published: September 1, 2020.
Correspondence: Wan Muhamad Amir, wmamir@usm.my

Measurement of predictive accuracy can be difficult for survival time data in the presence of censoring (Harrell et al., 1996; Fisher & Lin, 1999). The proportional hazards model was used to develop a prognostic model of metastatic hormone-refractory prostate cancer patients (HRPC) from 1991 to 2001 which is consist of 1,101 patients. Calibration of the survival model predictions was assessed by comparing the predicted probability with the actual survival probability (Halabi et al., 2003).

A survival model was developed using the following predictor variables: diagnosis, age, number of days in the hospital before study entry, presence of cancer, neurologic function, and 11 physiologic measures recorded on day 3 after study entry. Physicians were interviewed on day 3. Patients were followed for survival for 180 days after study entry (Knaus et al., 1995; Harrell et al., 1996). Chen and George (1985) investigated the stability of a stepwise selection procedure in the framework of the Cox proportional hazard regression model based on bootstrap resampling procedure. They developed a bootstrap-model selection procedure, combining with existing selection techniques for the best variable selection and illustrate the proposed strategy using data from two cancer clinical trials featuring two different situations (Sauerbrei & Schumacher, 1992). Chen and George (1985) described the use of the bootstrap in prognostic survival model for acute lymphocytic leukemia patients using computer-based statistical methodology. To validate the accuracy of the prognostic survival oral cancer model, they used a bootstrap resampling technique (100 bootstrap samples) to select the important prognostic factors via a stepwise regression. At the second stage, it involved 400 bootstrap samples for the estimate the corresponding regression parameters. The bootstrap result suggested the model constructed from the training set is reasonable (Chen & George, 1985).

In order to enhance the efficiency of calculation, the combination of bootstrap with prognostic survival oral cancer model methodology will be the main focus of this study. The bootstrap does not rely on a theoretical sampling distribution as in statistical significance testing (Efron, 1979; Efron & Tibshirani, 1993). It begins with an original sample taken from the population, then it takes place with replacement, the combinations of samples are limitless and are driven by random number generators from Monte Carlo. The first step in the bootstrap method is to copy the original sample several times (uses the empirical density function (EDF)) and create a pseudo-population. From the pseudo-population, bootstrap draws several samples with replacement (Efron & Tibshirani, 1993). The strength of bootstrap's method is its ability to develop a sample that is the same size of the

original sample that may include an observation several times while omitting other observations.

Data and Algorithm for Prognostic Survival Oral Cancer Model

Data from a medical record unit and related information were extracted for this demonstration. The sampling frame was the list of patients diagnosed with oral cancer admitted to Hospital University Sains Malaysia (HUSM). The details of the studied variables are shown in Table 1.

Flow Chart for Prognostic Survival Oral Cancer Model

Figure 1 is the flow chart of four different methods: Efron's Method, Exact Method, Breslow's Method, and Discrete Method, for prognostic survival oral cancer model using SAS algorithm. The result for each method is given by Table 2 to Table 5.

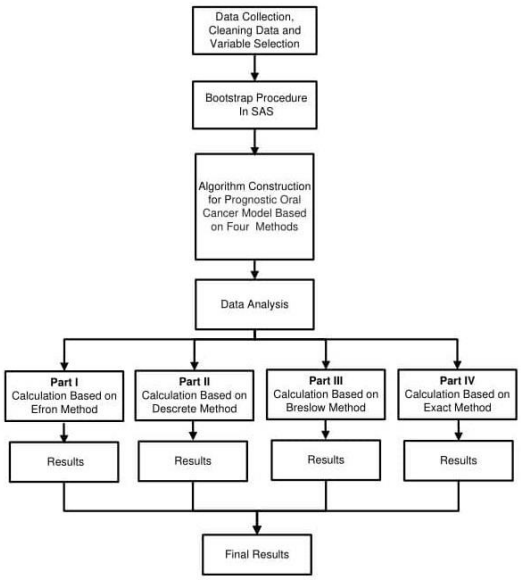

Figure 1. Flow chart for prognostic survival oral cancer model based on four methods

Table 1. Description of data

Num.	Variables	Explanation of user variables
1	Age	Age in years
2	Gender	Gender patients; 1=Male, 2=Female
3	Smoking	Smoking status; 0=Never, 1=Yes
4	Alcohol	Alcohol consumption; 0=Never, 1=Yes
5	Betel	Betel Quid; 0=Never, 1=Yes
6	Size	Tumor size; 1=Less than 4 cm, 2=Greater than 4cm
7	Nerve	Nerve invasion; 0=No, 1=Yes
8	Time	Time in months

Algorithm for Prognostic Survival Oral Cancer Model

```
/* PROGRAMMING FOR ORAL CANCER MODEL BASED ON FOUR METHODS */
%MACRO bootstrap(data=_last_, booted=booted, boots=10, seed=1234);
DATA &booted;
** randomly picks an integer from 1 to n;
pickobs = INT(RANUNI(&seed)*n)+1;
** POINT tells SAS to read value pickobs
** NOBS sets n to number of obs in &Data;
** when the point option is used SAS will loop through the data
step forever;
SET &data POINT = pickobs NOBS = n;
** saves number of current bootstrap;
REPLICATE=int(i/n)+1;
      i+1;
** stop will leave data set when n*&boots obs have been created;
IF i > n*&boots THEN STOP;
RUN;
%MEND bootstrap;

/* INPUT DATA */
Data Cancer;
input Age Gender Smoking Alcohol Betel Size Nerve Time;
cards;
-     1     0     0     2     0     87
66    2     0     0     2     0     18
50    2     1     0     1     0     65
48    2     1     0     1     0     69
65    2     1     0     1     0     42
```

```
66    2     1     0     2     0     44
58    1     0     1     2     1     13
49    1     0     1     2     1     15
49    1     0     1     1     0     19
55    1     0     1     2     0     18
50    1     0     1     2     0     77
69    2     1     0     1     0     11
54    2     1     0     2     0     37
52    1     0     1     2     1     16
62    2     1     0     2     1     9
62    1     0     1     2     1     7
;
run;
ods rtf file='abc.rtf' style=journal;

/**GENERATE BOOTSTRAP SAMPLE**/
%bootstrap(data= Cancer, boots=10);
run;
/**PRINT DATA **/
proc print data=booted;
run;
/**SURVIVAL ANALYSIS**/
Proc lifetest data=booted plots= (s);
Title 'Survival by Treatment';
Time Time*Nerve(1);
Strata Gender;
run;

proc lifetest data= booted plots=(s,ls,lls) censoredsymbol=none;
time Time*Nerve(1);
strata Gender;
run;

/****PROCEDURE EFRON****/
PROC PHREG DATA=booted;
MODEL Time*Nerve(1) = Age Gender Smoking Alcohol Betel Size /
TIES=EFRON ;
BASELINE OUT=set1 SURVIVAL=st LOGSURV=lst LOGLOGS=llst ;
```

```
OUTPUT OUT=resid1 DFBETA=dfgred RESSCH=scgred RESDEV=deres
            RESMART=mares XBETA=linpred STDXBETA=cipred;
RUN;
PROC PRINT DATA=set1;
RUN;
PROC PRINT DATA=resid1;
RUN;
PROC GPLOT DATA=resid1;
PLOT dfgred*Time;
RUN;

/*****PROCEDURE DISCRETE******/
Proc phreg data= booted;
model Time*Nerve(1) = Age Gender Smoking Alcohol Betel Size /ties
=discrete;
BASELINE OUT=set2 SURVIVAL=st LOGSURV=lst LOGLOGS=llst;
OUTPUT OUT=resid2 DFBETA=dfgred RESSCH=scgred RESDEV=deres
            RESMART=mares XBETA=linpred STDXBETA=cipred;
RUN;
PROC PRINT DATA=set2;
RUN;
PROC PRINT DATA=resid2;
RUN;
PROC GPLOT DATA=resid2;
PLOT dfgred*time;
RUN;

/*****PROCEDURE BRESLOW********/
Proc phreg data= booted;
model Time*Nerve(1) = Age Gender Smoking Alcohol Betel Size /ties
= breslow;
BASELINE OUT=set3 SURVIVAL=st  LOGSURV=lst LOGLOGS=llst;
OUTPUT OUT=resid3 DFBETA=dfgred RESSCH=scgred RESDEV=deres
            RESMART=mares XBETA=linpred STDXBETA=cipred;
RUN;
PROC PRINT DATA=set3;
RUN;
PROC PRINT DATA=resid3;
```

```
RUN;
PROC GPLOT DATA=resid3;
PLOT dfgred*Time;
RUN;
/********PROCEDURE EXACT*********/
Proc phreg data= booted;
model Time*Nerve(1) = Age Gender Smoking Alcohol Betel Size /ties
=exact;
BASELINE OUT=set4 SURVIVAL=st LOGSURV=lst LOGLOGS=llst;
OUTPUT OUT=resid4 DFBETA=dfgred RESSCH=scgred RESDEV=deres
            RESMART=mares XBETA=linpred STDXBETA=cipred;
RUN;
PROC PRINT DATA=set4;
RUN;
PROC PRINT DATA=resid4;
RUN;
PROC GPLOT DATA=resid4;
PLOT dfgred*Time;
RUN;
ods rtf close;
```

Results

Shown in Figure 2 are the survival probabilities for nerve invasion scenario according to gender. The plot shows the survival probability is about lower for females compared to male at all times point to develop nerve invasion among oral cancer patient which registered in Hospital University Sains Malaysia (HUSM).

The prognostics survival oral cancer model using Efron's Method is given by

$$HR = \exp\left[0.03178(\text{Age}) + 1.04331(\text{Smoking}) + 1.37850(\text{Betel Quid}) - 0.60054(\text{Tumor Size})\right] \quad (1)$$

Shown in Table 2 are the results of Efron's Method estimation for prognostics cancer. There are three factors were associated to the survival of oral cancer towards nerve invasion. Three factors (smoking (β_2 = 1.04331, p = 0.0298), Betel (β_3 = 1.37850, p = 0.0217), and tumor size (β_4 = –0.60054, p = 0.0283)) were

significant at $\alpha = 0.05$ and one factor (age ($\beta_1 = 0.03178$, $p = 0.0808$)) is quite significant at $\alpha = 0.05$.

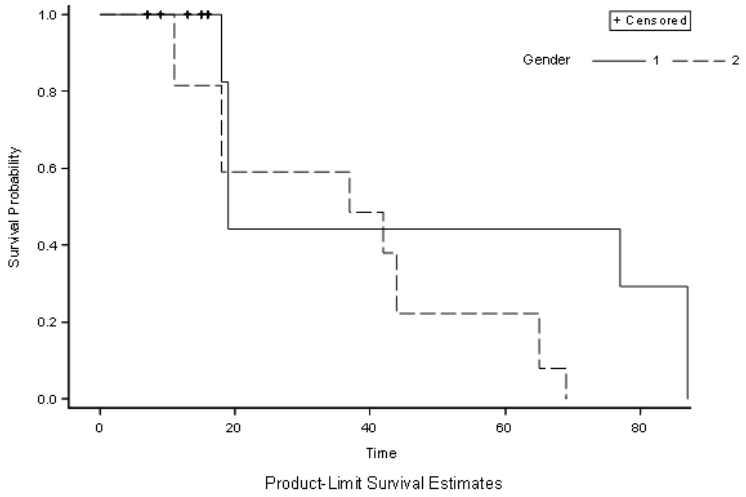

Figure 2. Survival probabilities for nerve invasion scenario

Table 2. Efron's method for prognostics survival oral cancer estimation

Analysis of maximum likelihood estimates						
Parameter	**DF**	**Parameter estimate**	**Standard error**	**Chi-square**	***p*-value**	**Hazard ratio**
Age	1	0.03178	0.01820	3.0494	0.0808	1.032
Smoking	1	1.04331	0.48026	4.7193	0.0298	2.839
Betel	1	1.37850	0.60059	5.2682	0.0217	3.969
Tumor size	1	–0.60054	0.27376	4.8124	0.0283	0.549

Note: Significant at $p < 0.05$

Table 3. Breslow's method for prognostics survival oral cancer estimation

Analysis of maximum likelihood estimates						
Parameter	**DF**	**Parameter estimate**	**Standard error**	**Chi-square**	***p*-value**	**Hazard ratio**
Age	1	0.02674	0.01741	2.3607	0.1244	1.027
Smoking	1	0.82533	0.44844	3.3872	0.0657	2.283
Betel	1	1.11424	0.56325	3.9135	0.0479	3.047
Tumor size	1	–0.54203	0.26882	4.0656	0.0438	0.582

Note: Significant at $p < 0.05$

The prognostics survival oral cancer model using Breslow's Method is given by

$$HR = \exp\left[0.02674(\text{Age}) + 0.82533(\text{Smoking}) + 1.11424(\text{Betel Quid}) - 0.54203(\text{Tumor Size})\right] \quad (2)$$

Shown in Table 3 are the results of Breslow's Method estimation for prognostics cancer. There are three factors were associated to the survival of oral cancer towards nerve invasion. Two factors (Betel (β_3 = 1.11424, p = 0.0479) and tumor size (β_4 = –0.54203, p = 0.0438)) were significant at α = 0.05 and two factors (age (β_1 = 0.02674, p = 0.1244) and smoking (β_2 = 0.82533, p = 0.0657)) are not significant at α = 0.05.

The prognostics survival oral cancer model using the Exact Method is given by

$$HR = \exp\left[0.03183(\text{Age}) + 1.04480(\text{Smoking}) + 1.38103(\text{Betel Quid}) - 0.60037(\text{Tumor Size})\right] \quad (3)$$

Shown in Table 4 are the results of the Exact Method estimation for prognostics survival oral cancer. The finding shows that there are three factors were associated to the survival of oral cancer towards nerve invasion. Three factors (smoking (β_2 = 1.04480, p = 0.0296), Betel (β_3 = 1.38103, p = 0.0215), and tumor size (β_4 = -0.60037, p = 0.0283)) were significant at α = 0.05 and one factor (age (β_1 = 0.03183, p = 0.0804)) is quite significant at α = 0.05.

Table 4. Exact Method for prognostics survival oral cancer estimation

Analysis of maximum likelihood estimates						
Parameter	**DF**	**Parameter estimate**	**Standard error**	**Chi-square**	***p*-value**	**Hazard ratio**
Age	1	0.03183	0.01820	3.0575	0.0804	1.032
Smoking	1	1.04480	0.48028	4.7324	0.0296	2.843
Betel	1	1.38103	0.60085	5.2829	0.0215	3.979
Tumor size	1	–0.60037	0.27371	4.8112	0.0283	0.549

Note: Significant at $p < 0.05$

Table 5. Discrete Method for prognostics survival oral cancer estimation

Analysis of maximum likelihood estimates						
Parameter	**DF**	**Parameter estimate**	**Standard error**	**Chi-square**	***p*-value**	**Hazard ratio**
Age	1	0.03168	0.01949	2.6431	0.1040	1.032
Smoking	1	0.99543	0.49682	4.0144	0.0451	2.706
Betel	1	1.34334	0.62708	4.5891	0.0322	3.832
Tumor size	1	–0.66706	0.30193	4.8812	0.0272	0.513

Note: Significant at $p < 0.05$

The prognostics survival oral cancer model using the Discrete Method is given by

$$HR = \exp\left[0.03168(\text{Age}) + 0.99543(\text{Smoking}) + 1.34334(\text{Betel Quid}) - 0.66706(\text{Tumor Size})\right] \quad (4)$$

Shown in Table 5 are the results of the Discrete Method estimation for prognostics survival oral cancer. There are three factors were associated to the survival of oral cancer towards nerve invasion. Three factors (smoking ($\beta_2 = 0.99543$, $p = 0.0451$), Betel ($\beta_3 = 1.34334$, $p = 0.0322$), and tumor size ($\beta_4 = -0.66706$, $p = 0.0272$)) were significant at $\alpha = 0.05$ and one factor (age ($\beta_1 = 0.03168$, $p = 0.1040$)) is quite significant at $\alpha = 0.05$.

Table 6. Comparing the *p*-values of Exact Method vs. Efron's Method

	Exact Method (reference)		**Efron's Method**		
Parameter	**Param. est.**	**Pr > ChiSq**	**Param. est.**	**Pr > ChiSq**	**Differences (%)**
Age	0.03183	0.0804	0.03178	0.0808	0.04
Smoking	1.04480	0.0296	1.04331	0.0298	0.02
Betel	1.38103	0.0215	1.37850	0.0217	0.02
Tumor size	–0.60037	0.0283	–0.60054	0.0283	0.00

Table 7. Comparing the *p*-values of Exact Method vs. Breslow's Method

	Exact Method (reference)		**Breslow's Method**		
Parameter	**Param. est.**	**Pr > ChiSq**	**Param. est.**	**Pr > ChiSq**	**Differences (%)**
Age	0.03183	0.0804	0.02674	0.1244	4.40
Smoking	1.04480	0.0296	0.82533	0.0657	3.61
Betel	1.38103	0.0215	1.11424	0.0479	2.64
Tumor size	–0.60037	0.0283	–0.54203	0.0438	1.55

Table 8. Comparing the *p*-values of Exact Method vs. Discrete Method

	Exact Method (reference)		Discrete Method		
Parameter	**Param. est.**	**Pr > ChiSq**	**Param. est.**	**Pr > ChiSq**	**Differences (%)**
Age	0.03183	0.0804	0.03168	0.1040	2.36
Smoking	1.04480	0.0296	0.99543	0.0451	1.55
Betel	1.38103	0.0215	1.34334	0.0322	1.07
Tumor size	–0.60037	0.0283	–0.66706	0.0272	0.11

Tables 6-8 summarize differences (%) between the studied methods according to the Pr > ChiSq point of view. The smallest differences which gained from the pairs of calculation will indicate the most appropriate model obtained.

Prognostic Survival Oral Cancer Model

Factors influencing the oral cancer were investigated by using prognostic survival oral cancer model. There are three factors associated to the survival of oral cancer towards nerve invasion. Table 6 to Table 8 show the *p*-values which are obtained from the different methods. The comparison is made based on the Exact Method. All the variables in the Exact Method show the most significant results. According to the analysis, smoking, betel quid, and tumor size are the significant factors. Using the Exact Method of estimation, the prognostic oral cancer model can be written as follows:

$$HR = \exp\left[0.03183(\text{Age}) + 1.04480(\text{Smoking}) + 1.38103(\text{Betel Quid}) - 0.60037(\text{Tumor Size})\right]$$

The results are summarized in Table 4. The Age factor shows there is an increasing in hazard rate (*HR* = 1.032). It means that, the oral cancer patients with an increase in one-year age, will have an increase 1.03 times the odd to develop the nerve invasion. Smoking factor shows that (*HR* = 2.843). It means the oral cancer patients who smoke have 3 times the odd to develop the nerve invasion than those who do not smoke. The next factor is betel quid, which shows that (*HR* = 3.979). It means the oral cancer patients who consult betel quid have 4 times the odd to develop the nerve invasion than those who do not consult betel quid. Oral cancer patients who are having tumor size less than 4 cm decrease odds to nerve invasion by 45.1% than those who having tumor size greater than 4 cm.

Conclusion

An algorithm was presented, and procedures were compared for modeling using prognostics survival oral cancer model through SAS language. The Exact Method was most accurate, likely due to explicitly of time event data.

Acknowledgements

The authors would like to express their gratitude to Universiti Sains Malaysia (USM) for providing the research funding (Grant No. 1001/PPSG/8012278, School of Dental Sciences, Kampus Kesihatan).

References

Chen, C. H., & George, S. L. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statistics in Medicine, 4*(1), 39-46. doi: 10.1002/sim.4780040107

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187-220. doi: 10.1111/j.2517-6161.1972.tb00899.x

Efron, B. (1979). Bootstrap method: Another look at the jackknife. *Annals of Statistics, 7*(1), 1-26. doi: 10.1214/aos/1176344552

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall. doi: 10.1201/9780429246593

Fisher, L. D., & Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health, 20*(1), 145-157. doi: 10.1146/annurev.publhealth.20.1.145

Halabi, S., Small, E. J., Kantoff, P. W., Kattan, M. W., Kaplan, E. B., Dawson, N. A., Levine, E. G., Blumenstein, B. A., & Vogelzang, N. J. (2003). Prognostic model for predicting survival in men with hormone-refractory metastatic prostate cancer. *Journal of Clinical Oncology, 21*(7), 1232-1237. doi: 10.1200/JCO.2003.06.100

Harrell, F. E., Jr., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine, 15*(4), 361-

387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Knaus, W. A., Harrell, F. E., Lynn, J., Goldman, L., Phillips, R. S., Connors, A. F., Dawson, N. V., Fulkerson, W. J., Califf, R. M., Desbiens, N., Layde, P., Oye, R. K., Bellamy, P. E., Hakim, R., & Wagner, D. P. (1995). The SUPPORT prognostic model: objective estimates of survival for seriously ill hospitalized adults. *Annals of Internal Medicine, 122*(3), 191-203. doi: 10.7326/0003-4819-122-3-199502010-00007

Sauerbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building: application to the Cox regression model. *Statistics in Medicine, 11*(16), 2093-2109. doi: 10.1002/sim.4780111607