JMLR: Workshop and Conference Proceedings vol 52, 252-263, 2016 PGM 2016

Making Large Cox's Proportional Hazard Models Tractable in Bayesian Networks

Jidapa Kraisangka[1] JIK41@PITT.EDU

Marek J. Druzdzel[1,2] MAREK@SIS.PITT.EDU

[1] *Decision System Laboratory, School of Information Sciences and Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, 15260, USA*

[2] *Faculty of Computer Science, Białystok University of Technology, Wiejska 45A, 15-351 Białystok, Poland*

Abstract

Cox's proportional hazard (CPH) model is a statistical technique that captures the interaction between a set of risk factors and an effect variable. While the CPH model is popular in survival analysis, Bayesian networks offer an attractive alternative that is intuitive, general, theoretically sound, and avoids CPH model's restrictive assumptions. Existing CPH models are a great source of existing knowledge that can be reused in Bayesian networks. The main problem with applying Bayesian networks to survival analysis is their exponential growth in complexity as the number of risk factors increases. It is not uncommon to see complex CPH models with as many as 20 risk factors. Our paper focuses on making large survival analysis models derived from the CPH model tractable in Bayesian networks. We evaluate the effect of two complexity reduction techniques: (1) parent divorcing, and (2) removing less important risk factors based on the accuracy of the resulting models.

Keywords: Bayesian networks; Cox's proportional hazard model; approximation.

1. Introduction

Survival analysis is a set of statistical methods that aim at modeling the relationship between a set of predictor variables and an outcome variable and, in particular, prediction of the time when an event occurs (Allison, 2010). One of the most popular survival analysis techniques is the Cox's Proportional Hazard (CPH) model (Cox, 1972), a set of regression methods used in the assessment of survival based on its risk factors or explanatory variables.

While the CPH model has been popular in survival analysis, it comes with several limitations. One of these is that the model focuses only on the interaction between the risk factors and the survival variable, in separation from the rest of the world. Practical models ask for embedding both the risk factors and the survival variable in interactions with other variables. Other assumptions include a special type of multiplicative combination of the risk factors, typically binary risk factors that are independent of one another. While it is not strictly necessary in theory, CPH models are typically learned from data.

Bayesian networks (BNs) offer a general, theoretically sound formalism for an intuitive representation of the joint probability distribution over a set of random variables. The CPH model is clearly a restricted subset of the BN model. This means in practice that BNs allow for combining an equivalent of multiple CPH models, allow for risk factors with multiple states, and for dependencies between them.

Building Bayesian networks may be time-consuming because of the quantities of numerical parameters necessary to capture practical interactions among variables. In (Kraisangka and Druzdzel, 2014), we proposed a methodology for translating a CPH model into a BN model. This resulting BN model, which we called the BNCox model, captures the predictive ability of the CPH model in a Bayesian network. The main application of this work is that it allows for reusing the rich body of knowledge published in medicine that describes the effect of various risk factors on survival.

One problem with this approach that has not been addressed is the exponential growth of the conditional probability tables corresponding to the survival variables. The tables double in size with every additional binary risk factor. When the number of risk factors is high, this table may plainly become intractable.

The goal of this paper is to address this problem. We test two approaches to complexity reduction: (1) modification of the Bayesian network structure through parent divorcing, and (2) simplifying the network structure by removing least influential risk factors. While (parent divorcing is a well known and widely applied technique, it is not straightforward in BNCox models because the CPH model is not decomposable. We demonstrate that approximate decompositions lead to sizeable loss in model accuracy. Finally, we show that removal of least influential risk factors leads to reasonable approximations and may offer a viable solution to the problem.

The remainder of this paper is structured as follows. Section 2 introduces the CPH model and two practical examples of the CPH model, Recidivism (Section 2.1, and the Pulmonary Arterial Hypertension (PAH) model (Section 2.2). Section 3 reviews the BNCox model. Section 4 contains the core of the paper, notably description of our efforts to improve the tractability of the BNCox model by means of structural decomposition of large CPTs through parent divorcing (Section 4.1) and removal of weak risk factors (Section 4.2). Finally, Section 5 summarizes our findings and outlines avenues for further research.

2. Cox's Proportional Hazard Model

Survival analysis basically focuses on modeling time-to-event occurrences. For example, we may focus on time-to-death of patients with a specific disease, failure time of machines, or time to possible rearrest of individuals who have been released from prison. One of the most popular techniques in survival analysis is the Cox's proportional hazard (CPH) model (Cox, 1972), which is a set of regression methods used in the assessment of survival based on its risk factors or explanatory variables. As defined originally by Cox (1972), the hazard regression model is expressed as

$$\lambda(t) = \lambda_0(t) \exp^{\beta' \cdot \mathbf{X}} \ .$$

This hazard model is composed of two main parts: The baseline hazard function, $\lambda_0(t)$, and the set of effect parameters, $\beta' \cdot \mathbf{X} = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n$. The baseline hazard function, estimated from data, determines the risks when all explanatory variables are absent. According to Cox (1972), $\lambda_0(t)$ can be unspecified or follow any distribution, which makes the CPH model a semi-parametric model. The βs are the coefficients corresponding to the risk factors $\mathbf{X}$. To compute the survival probability, we can use the survivor function as follows

$$S\left(t\right) = S_0\left(t\right)^{\gamma} \ ; \ \ \gamma = \frac{\lambda(t_2)}{\lambda(t_1)} = \frac{\exp\left(\beta' \mathbf{X_2}\right)}{\exp\left(\beta' \mathbf{X_1}\right)} \ .$$

The CPH model relies on the assumption that the hazard ratio of two observations, e.g., treatment and control group in a clinical trial, is constant over time (Cox, 1972). If the hazards at two points

in time, t_1 and t_2, are $\lambda(t_1)$ and $\lambda(t_2)$ respectively, their ratio γ is a constant. The CPH model is typically used to estimate the probability that an individual will survive (event s) at time t given a set of risk factors $\mathbf{X_i}$, i.e., $P(s|\mathbf{X_i}, t)$.

2.1 Recidivism Model

Recidivism of prisoners (Rossi et al., 1980) CPH model is a frequently used example in survival analysis. The data set has been made available to researchers and was used as an illustration in survival analysis examples by Allison (2010) using SAS and Fox using R (Fox, 2002). The Recidivism data describe 432 male prisoners who were under one year observation after being released from prison. The event of interest in this analysis is re-arrest, i.e., whether the prisoner is re-arrested during the period of study or not. For the purpose of simplicity, only seven binary risk factors (listed in Table 1) from the total of nine risk factors available in the data set were used to create a CPH model.

Risk factor (X_i)	β	$exp(\beta)$	p-value
X_1: fin	-0.40415	0.6675	0.0339
X_2: race	0.22931	1.2577	0.4549
X_3: wexp	0.41055	1.5076	0.0403
X_4: mar	-0.49926	0.6070	0.1874
X_5: paro	-0.06721	0.9350	0.7288
X_6: prio	0.28708	1.3325	0.2654
X_7: educ	-0.80736	0.4460	0.0557

Table 1: A list of seven binary risk factors, their corresponding coefficients β, hazard ratio $exp(\beta)$, and p-value estimated from the Recidivism data set.

The Recidivism CPH model is captured by the following survival function:

$$S(t) = S_0(t)^{e^{\beta_1 X_1 + \ldots + \beta_7 X_7}} .$$

$S_0(t)$ is the baseline survival probability function estimated from the model. In this case, $S_0(t)$ consisted of baseline survival probabilities from the 1^{st} week to the 52^{th} week for individuals with all risk factors absent (baseline) i.e., at 52^{th} week, $S_0(t = 52) = 0.7527$ (see Figure 1). We can calculate the survival probabilities conditional on any combination of risk factors from the above equation. The resulting survival curve will be proportional to the baseline survival curve (Figure 1) due to the CPH assumptions.

2.2 The PAH Model

CPH models, reported usually as lists of risk factors along with their $exp(\beta)$ coefficients are prevalent in medical literature. One such model is the Pulmonary Arterial Hypertension (PAH) CPH model, created for the purpose of predicting the probability of one year survival of patients suffering from Pulmonary Arterial Hypertension. The model is based on a data set from the Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management Predicting Survival in Pulmonary Arterial Hypertension (REVEAL) (Benza et al., 2010). The model includes 19 binary risk factors (reproduced from the original paper in Table 2) and the baseline probability of survival, $S_0(1) = 0.9698$.

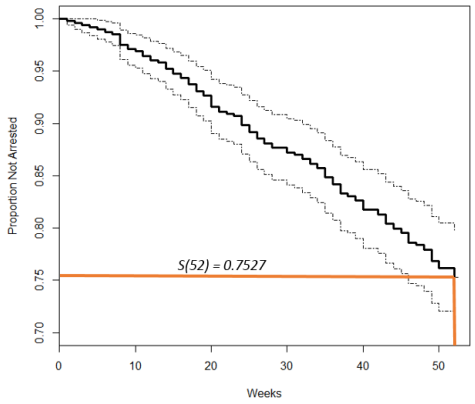


Figure 1: The survival probability curve for individuals with all risk factors absent. The thick black line shows the baseline survival probability, while the two grey lines show the upper and the lower bound of the 5% confidence interval.

Risk factors X_i	β	$exp(\beta)$	p-value
APAH-CTD	0.7737	1.59	<0.001
FPAH	1.2801	3.60	<0.001
APAH-PoPH	0.4624	2.17	0.012
Male >60 years age	0.7779	2.18	<0.001
Renal insufficiency	0.6422	1.90	<0.001
FC I	-0.8740	0.42	0.039
FC III	0.3454	1.41	0.008
FC IV	1.1402	3.13	<0.001
SBP <110 mmHg	0.5128	1.67	<0.001
Heart Rate >92bmp	0.3322	1.39	0.005
6MWD ≥440 m	-0.5455	0.58	0.006
6MWD <165 m	0.5210	1.68	<0.001
BNP <50 pg/ML	-0.6922	0.50	0.003
BNP >180 pg/ML	0.6791	1.97	<0.001
Pericardial effusion	0.3014	1.35	0.014
% DLCO ≥80%	-0.5317	0.59	0.031
% DLCO ≤32%	0.3756	1.46	0.018
mRAP > 20 mmHg	0.5816	1.79	0.043
PVR >32 Wood units%	1.4062	4.08	<0.001

Table 2: A list of 19 binary risk factors, their corresponding coefficients β, hazard ratios $exp(\beta)$ and p-values reported in the PAH REVEAL system (Benza et al., 2010).

3. Bayesian network interpretation of the CPH models

In (Kraisangka and Druzdzel, 2014), we proposed a Bayesian network interpretation of the CPH model by (1) creating a simple network structure in which every risk factor had a directed arc coming into the survival node, and (2) subsequently populating the CPT for the survival node with probabilities of survival conditional on every combination of values of the risk factors using the following formula:

$$Pr(s|\mathbf{X_i}, T = t) = S_0(t)^{e^{\beta' \mathbf{X_i}}} .$$

The BNCox model captured the temporal characteristics of the CPH model by representing time explicitly as an indexing variable. The time variable is discretized into the number of time snapshots used in the original CPH model. The resulting BNCox model produces survival probabilities for each of these time snapshots that are identical to those of the CPH model. An immediate advantage of the BNCox model is that it is capable of predicting survival when only a subset of risk factors is known. The CPH model requires that values of all risk factors are known.

For the sake of simplicity, in the remainder of this paper, we will focus on static BNCox models, which derive the probabilities of survival at one point in time. We will fix the value of the time variable to 52^{th} week for the Recidivism model, and to one year for the PAH model.

4. Making CPH model tractable in BN

We investigate two approaches to reducing the complexity of the survival node's CPT in the BNCox model: (1) decomposition through parent divorcing, and (2) simplification through removing least important variables. The following sections describe both approaches and the resulting loss of accuracy.

4.1 BNCox Decomposition

The size of a CPT in a node is exponential in the number of parents of the node. When the number of parents is large, the CPT of the node becomes too complex for both knowledge engineering and for inference. One way of reducing the size of the CPT and the resulting complexity of inference is through structural decomposition known as parent divorcing. Parent divorcing amounts to adding intermediate nodes that each take a distinct subset of the original parents and each has a directed arc ending at the original child node.

Parent divorcing can lead to substantial reduction of the number of parameters and, effectively, higher efficiency of Bayesian updating (Zagorecki et al., 2006). Figure 2 shows an example decomposition of a Bayesian network node *S* with four binary parents.

(a) Simple Bayesian network (b) Decomposed Bayesian network

Figure 2: An example of a Bayesian network structural decomposition known as parent divorcing. By adding two intermediate nodes, I1 and I2, the node *S* with four parents will have only two parents resulting in a smaller CPT and a smaller total number of parameters.

Two auxiliary nodes reduce the size of the CPT at *S*. The total number of parameters decreases from $2^4 = 16$ to $2^2 + 2^2 + 2^2 = 12$. The savings become even more dramatic when the number of

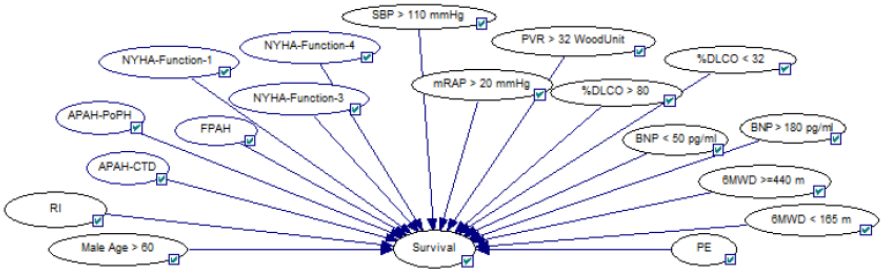
parents is large. With 100 binary parents, for example, a single simple decomposition can reduce the number of parameters from 2^{100} to $2^{50} + 2^2 + 2^{50} \cong 2^{51}$.

For an interaction between a node and its parents to be amenable to exact parent divorcing, the function expressing the interaction has to be decomposable in the following way

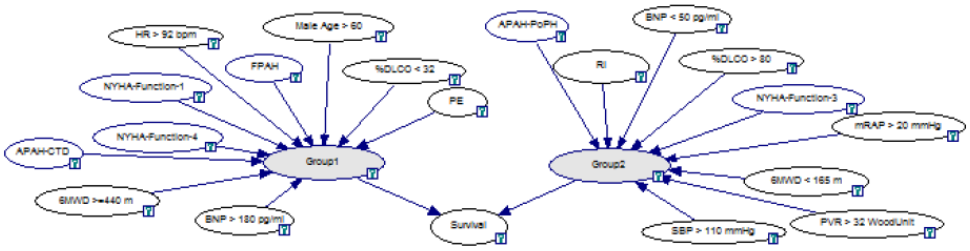
$$f(x_1, x_2, \ldots, x_n) = g(f_1(x_1, x_2, \ldots, x_k), f_2(x_{k+1}, x_{k+2}, \ldots, x_n)) \ .$$

In case of the noisy-OR gates (e.g., Díez and Druzdzel (2006)), the combination function is a logical OR, which is commutative, i.e., $OR(X_1, ..., X_n)$ function is equivalent to $OR(X_1, OR(X_2, OR(...OR(X_{n1}, X_n)...)))$. Other functions, such as *AND*, *MIN*, and *MAX* are also decomposable.

The main question is whether the survivor function can be similarly decomposed. For the sake of simplicity, we will use the structure of the BNCox model from Figure 2 as an example for demonstration. The original survivor function from the model is as follows:



(a) The structure of PAH BNCox created from the CPH model



(b) The structure of the decomposed PAH BNCox model

Figure 3: A PAH BNCox model decomposition from (a) to (b) with 19 risk factors

$$S(t) = S_0(t)^{e^{(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)}} \ .$$

In order to decompose the BNCox model, we have to find a function f that is capable of expressing $S(t)$ in the following way:

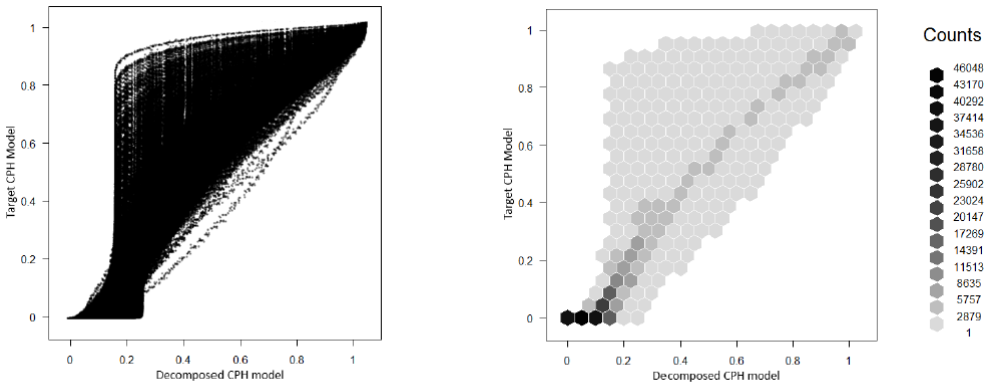
$$S\left(t\right) = f(S_1(t)^{e^{(\beta_1 X_1 + \beta_2 X_2)}}, S_2(t)^{e^{\beta_3 X_3 + \beta_4 X_4}}) \ . \tag{1}$$

The survivor function describes an interaction between states of risk factors (PRESENT and ABSENT) and the probability of survival. This is different from an OR functions which describe interaction between states. While we cannot offer a proof of this, we believe that the BNCox model

is not decomposable in a closed form and will, therefore, require an approximation for the sake of decomposition.

To test this idea, we used the PAH model of Benza et al. (2010). We created the BNCox version of the PAH model and derived the CPT of the *survival* node following the interpretation of the CPH model (Kraisangka and Druzdzel, 2014). All prior probabilities of each parent node followed the uniform distribution, since the prior probability distributions over the risk factors were not reported in the original reference. The structure of PAH BNCox is shown in Figure 3-a.

After creating the PAH BNCox model, we generated a data set from the original CPH model by means of stochastic sampling. We generated at least 5 records for each combination of risk factors. Once we have generated the data set, we randomly divided all risk factor nodes into two groups. Figure 3-b shows the structure of the decomposed PAH BNCox. Then, we learned the decomposed network's CPTs from the data set using the EM algorithm with intermediate nodes being unobserved (i.e., absent in the data file). Figure 4-a shows a scatterplot of the probabilities of survival from the original PAH BNCox model (our baseline) against the probabilities of survival generated by the decomposed model. As we can see, the approximation resulting from the PAH BNCox model is rather poor.



(a) Scatterplot (b) Scatterplot with Hexagonal Binning

Figure 4: Probabilities from the decomposed model against the original CPH model shown in (a) a scatterplot and (b) a scatterplot with hexagonal binning

Due to the high density in the scatterplot, it is not clear where most of the points are. We used an alternative visualization technique called Hexagon binning plot (Lewin-Koh, 2011) to show the same data in Figure 4-b. Each hexagon is colored-coded according to the number of points falling in that region. Figure 5 shows the Euclidean distance between the survival probabilities calculated by the original CPH and the decomposed BNCox model for all possible combinations of values of risk factors sorted from smallest to largest distance. We see an overall poor fit between the decomposed and the original structures.

Although, we have not tested all version of the network decomposition, we tried other decompositions with different number of groups including 4 groups, 6 groups, 9 groups. However, all those resulting decomposed networks confirm poor approximation of the original model.

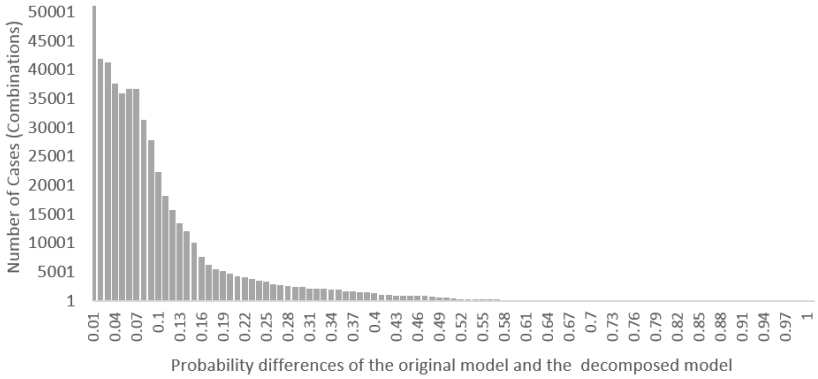


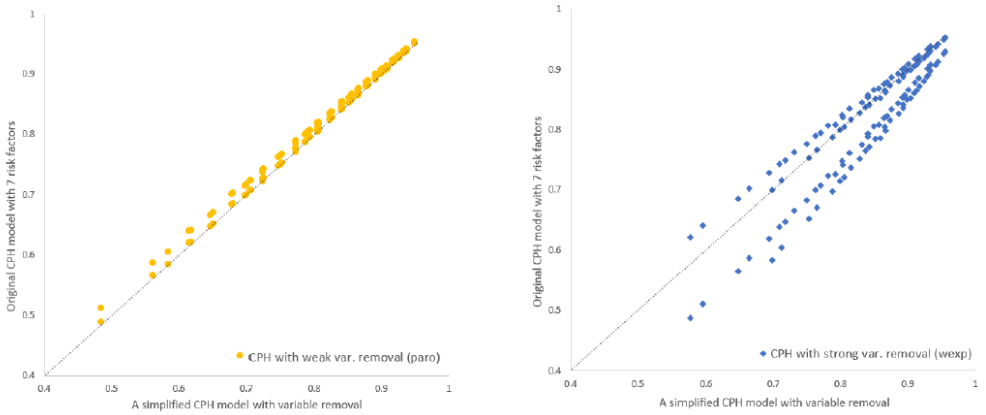
Figure 5: The histogram showing the Euclidean distance between the survival probabilities produced by the original PAH BNCox model and the decomposed model sorted from the smallest to largest distance.

4.2 BNCox simplification by removing least influential risk factors

It can be expected that some of the risk factors included in a CPH model will have minimal effect on the probability of survival and omitting them altogether will not lead to much loss of precision. On the other hand, removing each of these uninfluential factors will reduce the size of the survival node's CPT. Our second approach to reducing the complexity of the BNCox model is, thus, removing the least influential risk factors. There exist several techniques for variable selection in survival analysis (Fan and Li, 2002). Some of them originate from multiple linear regression techniques, e.g, stepwise selection or best subset selection. Other modern techniques are designed for the CPH model, such as, penalized likelihood approach (Fan and Li, 2001) and LASSO techniques. These variable selection methods are used to determine the best subset of risk factors during model development by refitting the model in order to obtain the best accuracy with the smallest number of variables. Because our main application of the BNCox model is assimilating CPH parameters from medical literature, we assume that we already have the best CPH model. Hence, we will investigate the effect of identifying and omitting the least influential risk factors from among a reported set and focus on comparing those simplified models to the original CPH model.

We first performed experiments on the Recidivism CPH model. We removed the weakest risk factor and the strongest risk factor judging by their p-values. We define the "weakest" influences as those with the highest p-values and the smallest values of the coefficient β. The smaller the value of β, the weaker the absolute influence of the risk factor on the survival probability. The larger the value of p, the less certain we are that the risk factor is really affecting survival. The weakest variable in Table 1 seems *paro* with $\beta = -0.06721$ and $p = 0.7288$. The strongest variable, on the other hand, seems *wexp* with $\beta = 0.41055$ and $p = 0.0403$. We created two modifications of the Recidivism CPH model: (1) one with the variable *paro* removed (weakest), and (2) one with the variable *wexp* (strongest) removed. These new refitted models have different sets of β coefficients than the original model. We subsequently calculated the survival probabilities using all combinations of risk factors (a total of $2^6 = 64$ probabilities) from each of the two modified models.

Figure 6 shows the result of survivals of the simplified models against the original Recidivism model.



(a) Refitted model with weak variable ($paro$) removed

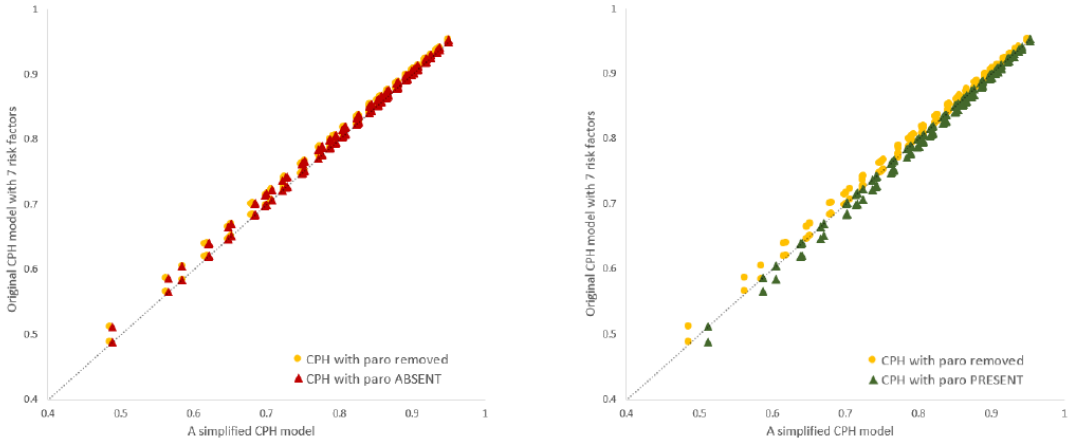
(b) Refitted model with strong variable ($wexp$) removed

Figure 6: The scatterplot of the survival probability produced by the simplified models against the survival probability produced by the original CPH model. The diagonal gray line shows the ideal scatterplot, one representing no difference between the original and the modified models.

The original CPH model consisted of 7 binary risk factors resulting in $2^7 = 128$ predicted survival probabilities. Since we removed one variable from the original CPH model, the total number of predicted probabilities in the simplified model is $2^6 = 64$. Two survival probabilities in the original CPH model correspond to one probability in the modified models. As expected, we observe loss of accuracy in the simplified models. Removing the weakest variable (Figure 6-a) leads to almost no loss in accuracy, while removing the strongest variable leads to a noticeable loss (Figure 6-b).

We can use any variable selection methods to obtain the best approximation when we have a data set available. More typical, however, is the situation in which we have no access to the original data but have a full set of parameters β (this is the case with our second example, the PAH CPH model). Instead of refitting the model to data as in the case of the preceeding experiment, one could fix the state of the weakest variable to either "ABSENT" or to "PRESENT". In our next experiment, we evaluated the effect of fixing the state of the weakest variable against the refitting simplified model from previous experiment. We used the original CPH model (Table 1) and fixed the state of the weakest variable, *paro*, producing two sets of survival probabilities (for $paro = 0$, i.e., absent, and for $paro = 1$, i.e., present). Figure 7 shows scatterplots for these two CPH models against the model with *paro* removed and the original CPH model.

All three models produce similar errors measured by Euclidean distance. Setting the weakest variable to *absent* produces results that are closest to the variable-removed model. In other words, we could approximate the simplified model by setting state of a risk factor to *absent* in the original model without refitting the model from the data set. To confirm this result, we further evaluated this effect by setting the states of several least influential variables to *absent* and comparing them against refitted variable-removed models. We chose four least influential risk factors: $paro$, $race$, $prio$, and



(a) Fixed *paro* variable to ABSENT (b) Fixed *paro* variable to PRESENT

Figure 7: The scatterplot shows probabilities produced by two fixed-variable models against one variable-removed model (*paro*-ABSENT, *paro*-PRESENT, and the *paro*-removed model). The diagonal gray line shows ideal probability as produced from the original CPH model.

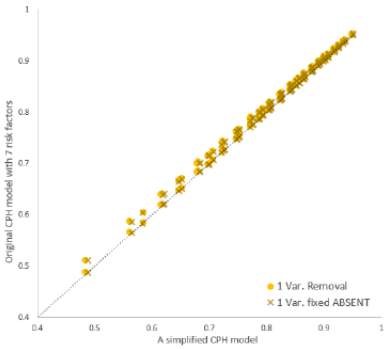
mar from the original model, using their p-value. From the original CPH model, we set *paro* to *absent* and iteratively set additional less influential risk factors (*race*, *prio*, and *mar* respectively). This resulted in four models that are samples of one-, two-, three-, and four-variable *absent* models. Figure 8 shows the resulting survival probabilities of the simplified models both fixed-*absent* and refitted models with variable removed against the original CPH model in different number of risk factors.

Figure 8 shows the effect of fixing variables to *absent* and refitting simplified models with variable removal against the original CPH model. As expected, the more we simplify the model, the higher loss in accuracy. Figure 8-d shows huge loss in accuracy due to a small number of remaining risk factors. With only three binary risk factors, 16 survival probabilities in the original CPH models were approximated by one survival probabilities in the simplified models formed in the eight vertical lines in the scatterplot. We also observed that the more we simplify models, the more difference between the model with fixing variables and the refitted model with variable removal.

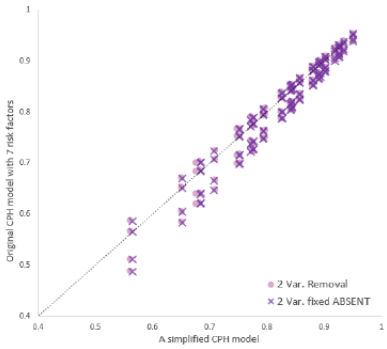
5. Discussion

A major challenge in making BNCox models practical is the exponential growth of the conditional probability tables at the survival node. We tested two approaches to handling this problem: (1) parent divorcing, and (2) removing least influential risk factors. The BNCox model does not seem to be decomposable and complexity reduction by parent divorcing leads to a high loss of accuracy. We suggest simplifying complex models BNCox models by removing the weakest risk factors from the model.

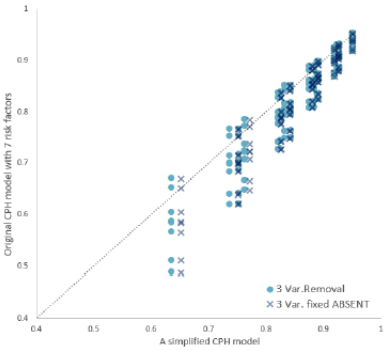
Any statistical variable selection methods (Fan and Li, 2002) can be used to simplify or reduce the number of risk factors in the CPH models when data are available. However, when data are not available, we can simplify the model by removing least influential risk factors, judging them by the



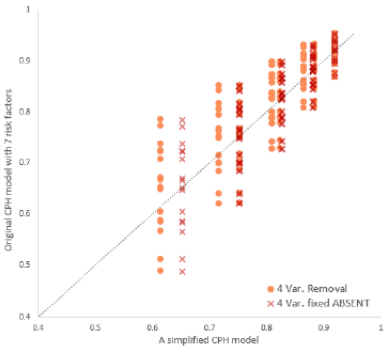
(a) One risk factor *absent* vs. *removed*



(b) Two risk factors *absent* vs. *removed*



(c) Three risk factors *absent* vs. *removed*



(d) Four risk factors *absent* vs. *removed*

Figure 8: Effect of *absent* and *removed* risk factors in the simplified models against the original CPH model. The diagonal gray line shows ideal probability as produced from the original CPH model.

β coefficient and the p-value. Removing less influential variables means that we set the states of those variables to *absent* in the CPH model. We evaluated the effect of this approach and observed a manageable loss of accuracy in the simplified models.

It is worth noting that Oniśko and Druzdzel (2013) found that in medical diagnostic systems, precision of parameters in Bayesian networks may not be as important for the diagnostic accuracy of models. Rounding parameters to even as few as two intervals yielded reasonable diagnostic performance. Removing least important risk factors may not harm much the predictive accuracy of the model.

Acknowledgments

We acknowledge the support the National Institute of Health under grant number U01HL101066-01 and the Faculty of Information and Communication Technology, Mahidol University, Thailand. Implementation of our work is based on GeNIe and SMILE, a Bayesian modeling environment available free of charge for academic research and teaching use at http://www.bayesfusion.com/. While we take full responsibility for any remaining errors and shortcomings of the paper, we would like to thank anonymous reviewers for their valuable suggestions.

References

P. D. Allison. *Survival Analysis Using SAS: A Practical Guide, Second Edition*. SAS Institute Inc., Cary, NA, 2010.

R. L. Benza, D. P. Miller, M. Gomberg-Maitland, R. P. Frantz, A. J. Foreman, C. S. Coffey, A. Frost, R. J. Barst, D. B. Badesch, C. G. Elliott, T. G. Liou, and M. D. McGoon. Predicting survival in pulmonary arterial hypertension: Insights from the Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management (REVEAL). *Circulation*, 122(2):164–172, March 2010. doi: 10.1161/CIRCULATIONAHA.109.898122.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, March 1972.

F. J. Díez and M. J. Druzdzel. Canonical probabilistic models for knowledge engineering. Technical report, UNED, Madrid, Spain, 2006.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

J. Fan and R. Li. Variable selection for Cox's proportional hazards model and frailty model. *Annals of Statistics*, 30(1):74–99, 2002.

J. Fox. *An R and S-Plus Companion to Applied Regression*. Sage Publication Inc., CA, 2002.

J. Kraisangka and M. J. Druzdzel. Discrete Bayesian network interpretation of the Coxs proportional hazards model. In *Probabilistic Graphical Models*, volume 8754 of *Lecture Notes in Computer Science*, pages 238–253. Springer International Publishing, 2014. ISBN 978-3-319-11432-3.

N. Lewin-Koh. Hexagon binning: an overview. *Online: http:// cran. r-project. org/ web/ packages/ hexbin/ vignettes/ hexagon_ binning. pdf*, 2011.

A. Oniśko and M. J. Druzdzel. Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems. *Artificial Intelligence in Medicine*, 57(3):197–206, Mar. 2013.

P. H. Rossi, R. A. Berk, and K. J. Lenihan. *Money, Work, and Crime - Experimental Evidence*. Academic Press, Inc., San Diego, CA, 1980.

A. Zagorecki, M. Voortman, and M. J. Druzdzel. Decomposing local probability distributions in Bayesian networks for improved inference and parameter learning. In G. Sutcliffe and R. Goebel, editors, *Recent Advances in Artificial Intelligence: Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS–2006)*, pages 860–865, Menlo Park, CA, 2006. AAAI Press.