A Bayesian Network Interpretation of the Cox's Proportional Hazard Model

Jidapa Kraisangka[1] and Marek J. Druzdzel[1,2]

[1] *Decision System Laboratory, School of Computing and Information, 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 hazards (CPH) model is quite likely the most popular modeling technique in survival analysis. While the CPH model is able to represent a relationship between a collection of risks and their common effect, Bayesian networks have become an attractive alternative with an increased modeling power and far broader applications. Our paper focuses on a Bayesian network interpretation of the CPH model (BN-Cox). We provide a method of encoding knowledge from existing CPH models in the process of knowledge engineering for Bayesian networks. This is important because in practice we often have CPH models available in the literature and no access to the original data from which they have been derived.

We compare the accuracy of the resulting BN-Cox model to the original CPH model, Kaplan-Meier estimate, and Bayesian networks learned from data, including Naive Bayes, Tree Augmented Naive Bayes, Noisy-Max, and parameter learning by means of the EM algorithm. BN-Cox model came out as the most accurate of all BN approaches and very close to the original CPH model.

We study two approaches for simplifying the BN-Cox model for the sake of representational and computational efficiency: (1) parent divorcing and (2) removing less important risk factors. We show that removing less important risk factors leads to smaller loss of accuracy.

Keywords: Bayesian network, Cox's proportional hazard model

Preprint submitted to International Journal of Approximate Reasoning September 19, 2018

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 [1]. In medical sciences, survival analysis is primarily used to predict events such as death, relapse, or development of a new disease. One of the most popular methods used in survival analysis is called the Cox's proportional hazards (CPH) model [2]. The CPH model is similar to a multiple linear regression technique that explores the relationship between a hazard and related independent explanatory variables over a period of time. It describes the impact of a risk factor on a treatment of patients through a parameter called hazard ratio [3]. Hazard ratio between two groups, e.g., treatment and control group in a clinical trial, represents the relative likelihood of survival at any time in the study and is usually assumed to be constant over time.

The CPH model has been widely used for predicting patient survival rate. For example, the Seattle Heart Failure Model [4] uses the CPH model to predict 1-, 2-, and 3-year survival of heart failure patients. The Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension (PAH) Disease Management (REVEAL) [5] also uses the CPH model to derive the Risk Score Calculator to determine probability of a PAH patient survival within an enrolled year.

This paper focuses on an alternative tool for survival analysis: Bayesian networks. Compared to the CPH model, a Bayesian network can model explicitly the structure of the relationships among explanatory variables [6]. Researchers can intuitively design and build a Bayesian network from expert knowledge or available data. The network can depict a complex structure of a problem and provide a way to infer probability distributions that are suitable for prognosis and diagnosis, for example in medical decision support systems [7]. However, building Bayesian networks based purely on expert knowledge can be a time-consuming and costly task. Luckily, many CPH models can be found in the

literature. They are typically published as a set of numerical coefficients along with their significance levels. No original data are usually available. To use the knowledge encoded in these CPH models, an interpretation of the CPH parameters is needed. We propose such a Bayesian network interpretation of the CPH model (BN-Cox), discuss its advantages and potential challenges, including those related to its representational and computational complexity.

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 rearrest of felons who have been freshly released from prisons. Survival analysis can be conducted to estimate time-to-event for a group, compare time-to-event between several groups, or just to study the relationship between variables and the predicted events.

The probability of an individual surviving beyond a given time t, i.e., the survivor function, is defined as

$$S(t) = Pr(T > t) \ . \tag{1}$$

T is a variable denoting the time of occurrence of an event of interest. The survival probability at the beginning, i.e., at time t_0, may be equal to 1 or to some baseline survival probability, which will drop down to zero over time. While the survivor function represents the probability of survival, the hazard function, given by

$$\lambda(t) = \lim_{\triangle t \to 0} \frac{Pr(t \leq T < t + \triangle t \mid T \geq t)}{\triangle t} \ , \tag{2}$$

where T is also a time variable, represents the risk of event occurrence at time t. The hazard is a measure of risk at a small time interval $\triangle t$ and is sometimes called a hazard rate, expressing the number of events per interval of time [1].

The relationship between the hazard function and the survivor function (see more details in Allison's textbook [1]) is described as

$$\lambda(t) = -\frac{d}{dt} \log S(t) \tag{3}$$

or as

$$S(t) = \exp \int_0^t \lambda(u)\, du \,. \tag{4}$$

Hence, we can estimate the survival probability from the hazard function. In survival analysis, the hazard function can be represented by any probability distribution (e.g., exponential distribution) or can be modeled by regression techniques. One of the most popular survival analysis techniques is the Cox's proportional hazard (CPH) model [2]. The CPH model provides assessment of survival based on risk factors associated with the events indicated in the model. The simplest CPH model consists of time-independent risk factors. The hazard function in such a CPH model is expressed as

$$\lambda(t) = \lambda_0(t) \exp(\beta' \cdot \mathbf{X}) \,. \tag{5}$$

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 determines the risks at an underlying level of explanatory variables, i.e., when all risk factors are absent. According to Cox [2], this $\lambda_0(t)$ can be unspecified or follow any distribution, which makes the CPH model semi-parametric. The βs are coefficients corresponding to the risk factors, $\mathbf{X_i}$.

Application of the CPH model relies on the assumption that the hazard ratio of two observations is constant over time [2]. This enables us to infer the rate of risk of the treatment. For example, in the study of patients with pulmonary arterial hypertension (PAH) [5], the hazard ratio of a group of PAH patients having renal insufficiency to a group of patients without renal insufficiency (control/baseline group) is reported as 1.90. This means that those patients with renal insufficiency have a 90% higher risk of dying from PAH than patients without renal insufficiency. The ratio of two hazards, γ, is defined as:

$$\gamma = \frac{\lambda_2(t)}{\lambda_1(t)} = \frac{\exp(\beta' X_2)}{\exp(\beta' X_1)} \,. \tag{6}$$

If the risk factors X are binary, their values are typically expressed as *presence* ($X = 1$) or as *absence* or *baseline* ($X = 0$) of the risk factor. Once we know the

hazard ratio, we can estimate the survival probability [8] from

$$S\left(t\right) = S_0\left(t\right)^{\gamma} \ . \tag{7}$$

$S_0\left(t\right)$ is the baseline survival probability, i.e., when all risk factors are absent or at their baseline value ($X = 0$) at any time t, while γ is the hazard ratio of the group of interest to the baseline group. In other words, the survival probability of an individual can be estimated from

$$S\left(t\right) = S_0\left(t\right)^{\exp\left(\beta' \cdot \mathbf{X}\right)} \ . \tag{8}$$

CPH models can be used to predict patient prognosis. For example, the Seattle Heart Failure Model [4] uses a CPH model to predict 1-, 2-, and 3-year survival of heart failure patients. The Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension (PAH) Disease Management (REVEAL) [5] uses also a CPH model as the foundation of its Risk Score Calculator, that determines the probability of a PAH patient survival.

CPH models are prevalent in medical literature. While it is conceivable that their parameters can be elicited from experts, we have never encountered this and all CPH models that we have seen were build from data. The data themselves are almost never available.

3. Bayesian networks

Bayesian networks [9] are probabilistic graphical models capable of modeling the joint probability distribution over a finite set of random variables. The structure of a Bayesian network is an acyclic directed graph in which nodes are variables and directed arcs denote dependencies among them. A conditional probability table (CPT) of a variable X contains probability distributions over the states of X for all combinations of states of X's parents. The joint probability distribution over all variables of the network can be calculated by taking the product of all prior and conditional probability distributions:

$$\Pr(\mathbf{X}) = \Pr(X_1, \ldots, X_n) = \prod_{i=1}^{n} \Pr(X_i | Pa(X_i)) \ . \tag{9}$$

The structure of a Bayesian network and all its numerical probabilities can be obtained from experts or learned from data. Although Bayesian networks may take significant effort to construct, they are widely used in many areas, such as medical and engineering diagnosis or prognosis.

Bayesian networks have become an alternative approach to survival analysis. They are well-structured, intuitive, while also being theoretically sound [7]. They have the ability to capture expert knowledge, handle model complexity, and offer more flexibility in model interpretation [6]. Researchers can explicitly model dependencies among risk factors. Bayesian networks naturally allow for estimating the survival probability based on partial observations, while the CPH model is not designed for that, even though one could extend it along the lines of BN inference. If we know the *Age* and *SBP* of a patient (Figure 1), we can make a prediction of survival without observing the remaining risk factors. Researchers can also combine an equivalent of multiple CPH models into the same network. For example, Figure 1 shows an example of a Bayesian network that combines two risk models (*Heart-Related Deaths*, with risk factors *6 Minute Walking Distance*, *Age* and $SBP > 110$ *mmHg* and *PAH-Related Deaths* with the above risk factors and $PVR > 32$ *WoodUnit*) to determine the risk of dying of patients suffering from heart disease and pulmonary arterial hypertension (PAH). Not only we have an equivalent of two CPH models but the BN relaxes the assumption of mutual independence of risk factors (e.g., *Age* influences both *6 Minute Walking Distance* and $SBP > 110$ *mmHg*; $SBP > 110$ *mmHg* influences $PVR > 32$ *WoodUnit*). Moreover, we can use Bayesian network for reasoning from survival nodes to their causes, e.g., when testing a model or predicting values of the risk factors given survival and possibly other risk factors.

There are two general approaches to building Bayesian networks for the purpose of risk assessment. Researchers can implement static models that predict risk or survival at a snap-shot of time. For example, Loghmanpour et al. [10] created Bayesian networked-based risk assessment models for patient data with the left ventricular assist devices (LVADs). Bayesian networks may estimate the risk at specific points in time, e.g., 30 days, 90 days, 6 month, 1 year, or 2 years,

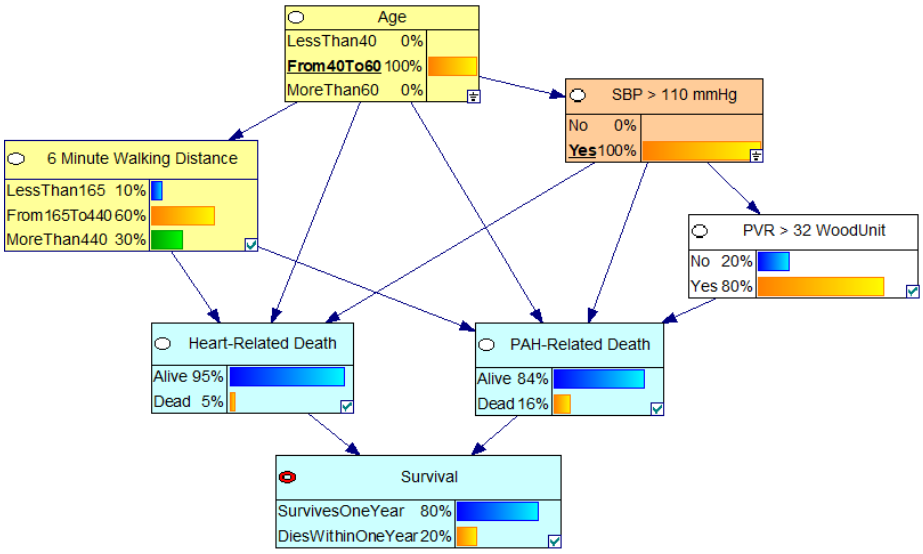


Figure 1: An example of Bayesian network predicting 1-year survivals based on two risk models of patients with partial observations (only *Age* and *SBP* are observed)

with high accuracy, outperforming traditional risk scores methods. A more complex approach uses dynamic Bayesian networks (DBNs) [11]. Van Gerven et al. [12] implemented a DBN for prognosis of patients who suffer from low-grade midgut carcinoid tumor. Instead of analyzing each time point separately, the DBN model calculates how the state of a patient changes over time under the influence of various therapy choices. This allows for modelling temporal nature of medical problems throughout the course of care and provides detailed prognostic predictions. However, it requires significantly more effort during model construction, i.e., require expertise to define the causal structure and temporal interactions, large amount of data, and is generally time-consuming [12].

4. Bayesian network interpretation of the CPH model

As we mentioned earlier, the process of building Bayesian networks can take a significant effort, especially when little or no data are available. In this section, we discuss how to use parameters from existing CPH models to create Bayesian

networks. This approach is especially useful when very little or no data are available. We assume that the CPH model's assumptions are not violated and the risk factors or random variables $\mathbf{X}$ are time-independent discrete/binary variables [13].

To create a Bayesian network, first we create its structure by designating the random variables representing risk factors as parents ($\mathbf{X}$) of the survival node (S). The number of states of each random variable is the same as in the CPH model.

Unlike the CPH model, static Bayesian networks capture a snapshot of a system at a certain time. We need, thus, to represent time explicitly. We achieve this by adding an indexing variable (T) for *time*, capturing each discrete point in time that is of interest, e.g., every day, every two weeks, etc. This time variable can be omitted if we are interested in the prediction at one point in time, e.g., at one year. Figure 5 show an example of such a model (we will call it the BN-Cox model), showing the relationship between risk factors ($\mathbf{X}$), the time variable (T), and the survival node (S).

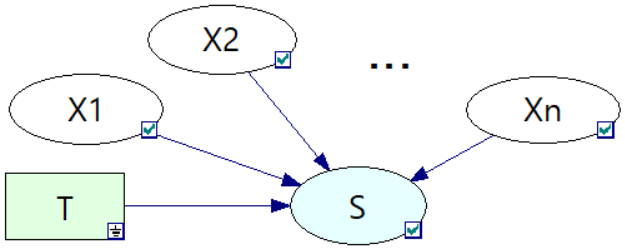


Figure 2: An example Bayesian network representing interactions among variables $X_1, X_2, \ldots, X_n$ representing risk factors, parents of the survival node S, and T, the time variable capturing each point in time.

In the next step, we create the conditional probability table for the survival node (S). Recall that we can obtain the survival probabilities from Equation 7 in the CPH model. For each time snapshot captured by the variable T, we assess a set of survival probabilities, $S(t)$ from the CPH model. A set of survival probabilities here means that we configure the hazard ratio γ according to the

combination of the parent states. γ is equal to hazard ratio of the conditioning case $\mathbf{X_i}$ to the baseline case $\mathbf{X_b}$, i.e., case in which all risk variables are *absent*, i.e.,

$$\gamma = \frac{exp(\beta' \mathbf{X_i})}{exp(\beta' \mathbf{X_b})} = \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_n X_{in} \ . \tag{10}$$

Equation 10 allows us to assess the survival probabilities directly from the parameters of the CPH model. First, we configure all risk factor cases in Equation 10 to find all hazard ratio values. Then we obtain the baseline survival probability at the first point in time from the CPH model ($S_0(t = 1)$) and use Equation 7 (or Equation 8) to find the survival probability. The survival probability calculated for each combination of risk factors corresponds to the conditional probability of survival. Hence, the conditional probability to be encoded in the CPT can be estimated by

$$Pr(s \mid \mathbf{X}, T = t) = S_0 \left(t\right)^{exp(\beta' \mathbf{X})} \ , \tag{11}$$

where s corresponds to the state *survived* in the survival node S, $\mathbf{X}$ are risk factors, and T is the time point. This allowed us to reproduce fully the CPH model by means of a Bayesian network.

CPH models, reported usually as lists of risk factors along with their parameters, are prevalent in medical literature. One such model is the CPH model created for the purpose of predicting the probability of one year survival of patients suffering from Pulmonary Arterial Hypertension [5]. The model includes 19 binary risk factors (reproduced from the original paper in Table 1) and the baseline probability of survival, $S_0(1) = 0.9698$. By following the method outlined above, we created a BN-Cox model shown in Figure 3, which is equivalent to the CPH model reported in [13].

5. Evaluation of the BN-Cox model

In this section, we provide an empirical evaluation of the BN-Cox model by comparing its predictive precision to the baseline survival analysis models like

Table 1: A list of 19 binary risk factors, their corresponding coefficients β, hazard ratios $exp(\beta)$ and p-values reported for the CPH model from Benza et al. [5].

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

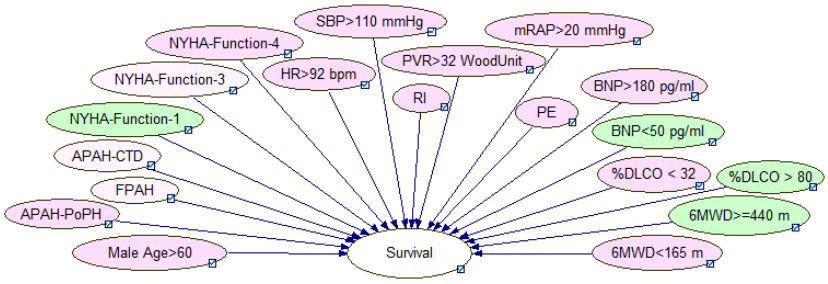


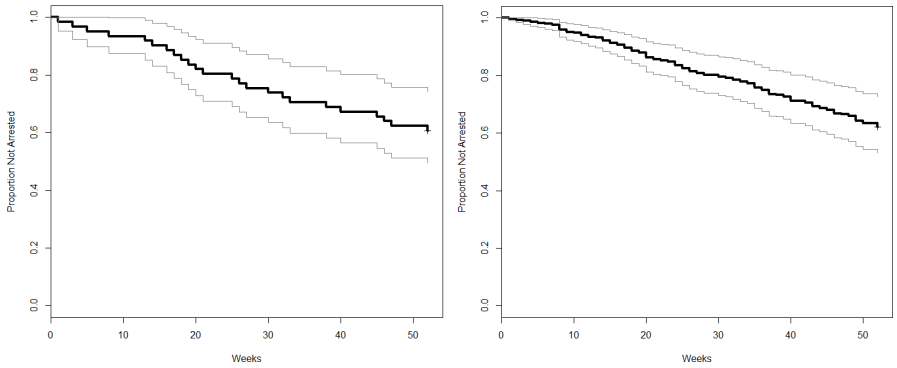
Figure 3: A Bayesian network representing the interaction among variables for the PAH CPH model.

the CPH model and the Kaplan-Meier (K-M) estimator [14] and to Bayesian networks learned from data. We chose a classical example application of the CPH model, the Recidivism data set [15]. The data set was collected in the course of an experimental study of 432 male prisoners, who were under observation for one year 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. The Recidivism data set is quite likely the most widely used example data set for survival analysis [1, 16], especially for the CPH model. We use the Recidivism data set to compare the precision of the proposed BN-Cox model, the CPH model, the K-M model, and the Bayesian network models learned directly from data. We will explain how we built the models for the purpose of evaluation in Section 5.1 and show the result of the predictive comparison in Section 5.2.

5.1. Model construction

For the sake of simplicity, we selected only four risk factors from the seven risk factors in the original Recidivism data set and preprocessed them into binary variables. The selected variables included the financial aid status *fin* (*no*=0, *yes*=1), prisoner's *race* (*other*=0, *black*=1), having prior full-time work experience *wexp* (*yes*=0, *no*=1), and number of prior convictions *prio* (*five and below*=0, *more than five*=1). The time variable in this data set is *week*, which is the week when a prisoner was rearrested during the observation period of one year (52 weeks). The survival variable is *arrest* indicating the rearrest status of a prisoner (*rearrested*=1, *never-rearrested*=0).

To implement the K-M and the CPH models, we used the R programming environment with the *Survival* library [16]. We learned the K-M model directly from the data set. The result of the K-M model is a set of 16 survival curves estimated from the data, each for one combination of the risk factors. We show one of these 16 survival curves for the group of prisoners with $fin = 0, race = 1, wexp = 1$, and $prio = 0$ in Figure 4a.



(a) K-M estimators (b) CPH model

Figure 4: Survival curves along with their 95% confidence intervals from the K-M model (a) and the CPH model (b) show the survival probability of a group of prisoners when *fin*=0, *race*=1, *wexp*=1, and *prio*=0.

For the CPH model, we obtained the following survival function:

$$S\left(t\right) = S_0\left(t\right)^{\exp\left(-0.3899fin+0.2591race+0.5249wexp+0.3330prio\right)}, \qquad (12)$$

where $S_0(t)$ is a vector of baseline probabilities estimated from the data set from the beginning of the observation period until the end of the 52^{nd} week. The baseline survival probability, $S_0(t)$, is the probability measured when all risk factors are absent (*fin*=0, *race*=0, *wexp*=0, and *prio*=0) at time t. For example, the baseline survival probability for the first week, $S_0(1)$, is 0.9973. Figure 4b shows the predicted survival curve from the CPH model of the same group of prisoners with *fin*=0, *race*=1, *wexp*=1, and *prio*=0.

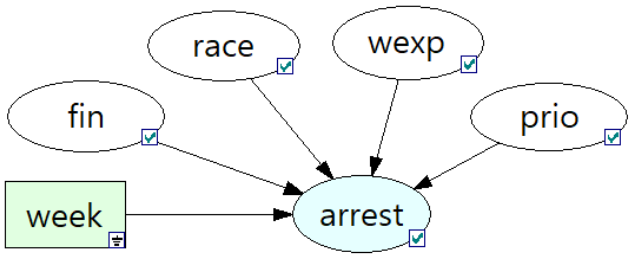


Figure 5: The structure of a BN-Cox model for the Recidivism data set.

To construct a BN-Cox model from the created CPH model for this evaluation, we followed the two steps outlined in Section 4. We used GeNIe[1] to implement its structure, and obtained survival probabilities calculated for each combination of risk factors corresponding to the conditional probability of survival. However, for the purpose of simplicity, we reduced the number of states for the time variable *week* from 52 to 13, which amounts to analyzing the system at 4-week steps. Other random variables (risk factors) have the same states as in the CPH model. We show the resulting structure of the Bayesian network in Figure 5. Table 2 shows selected probabilities of survival for all combinations of states of the risk variables.

For the Bayesian network learning from the data set (BN-Learn), we built the model in GeNIe using the same structure as in the BN-Cox model (Figure 5). The BN-Learn model learned directly the numerical parameters from data using the Expectation-Maximization (EM) algorithm [17, 18], which belongs to one of the standard functions of GeNIe. We also learned Bayesian network models using Naive Bayes (BN-NB), Tree Augmented Naive Bayes (BN-TAN), and Noisy-Max (BN-NoisyMax) approaches. The Noisy-Max model was learned using the method published in [19].

In summary, we created all seven models (K-M, CPH, BN-Cox, BN-Learn, BN-TAN, BN-NB and BN-NoisyMax) with four risk factors: *fin*, *race*, *wexp* and *prio*. These four risk factors are binary variables resulting in $2^4 = 16$ combinations of risk factors. We compare the prediction accuracy of each model in the following section.

5.2. Recidivism prediction with four risk factors

As mentioned previously, the simplified Recidivism model produces 16 combinations of risk factors. We plotted the distribution of the number of records corresponding to these 16 cases in Figure 6. We compared the accuracy of prediction of models for all 16 cases. However, in this paper, we selected only

[1] Available at `https://www.bayesfusion.com/`.

Table 2: Conditional probabilities of survival for all cases at each snapshot of time. γ is calculated from Equation 10 and $S(1)$, $S(2)$, ..., $S(13)$ are calculated from Equation 7 at 4-week steps. s is the survival variable *arrest*.

$Pr(s \mid X_i)$	γ	$S(1)$	$S(2)$	...	$S(12)$	$S(13)$
$\mathbf{Pr}(s \mid f=0, r=0, w=0, p=0)$	0.0000	0.998	0.992	...	0.830	0.804
$\mathbf{Pr}(s \mid f=0, r=0, w=0, p=1)$	0.3330	0.998	0.989	...	0.771	0.738
$\mathbf{Pr}(s \mid f=0, r=0, w=1, p=0)$	0.5249	0.997	0.986	...	0.729	0.692
$\mathbf{Pr}(s \mid f=0, r=0, w=1, p=1)$	0.8579	0.996	0.981	...	0.644	0.599
$\mathbf{Pr}(s \mid f=0, r=1, w=0, p=0)$	0.2591	0.998	0.990	...	0.785	0.754
$\mathbf{Pr}(s \mid f=0, r=1, w=0, p=1)$	0.5921	0.997	0.986	...	0.714	0.675
$\mathbf{Pr}(s \mid f=0, r=1, w=1, p=0)$	0.7840	0.997	0.983	...	0.665	0.621
$\mathbf{Pr}(s \mid f=0, r=1, w=1, p=1)$	1.1117	0.995	0.976	...	0.565	0.514
$\mathbf{Pr}(s \mid f=1, r=0, w=0, p=0)$	-0.3899	0.999	0.995	...	0.881	0.863
$\mathbf{Pr}(s \mid f=1, r=0, w=0, p=1)$	-0.0569	0.998	0.992	...	0.838	0.814
$\mathbf{Pr}(s \mid f=1, r=0, w=1, p=0)$	-0.1350	0.998	0.991	...	0.808	0.779
$\mathbf{Pr}(s \mid f=1, r=0, w=1, p=1)$	0.4680	0.997	0.987	...	0.742	0.706
$\mathbf{Pr}(s \mid f=1, r=1, w=0, p=0)$	-0.1308	0.999	0.993	...	0.849	0.826
$\mathbf{Pr}(s \mid f=1, r=1, w=0, p=1)$	0.2022	0.998	0.990	...	0.796	0.766
$\mathbf{Pr}(s \mid f=1, r=1, w=1, p=0)$	0.3941	0.998	0.988	...	0.758	0.724
$\mathbf{Pr}(s \mid f=1, r=1, w=1, p=1)$	0.7271	0.997	0.983	...	0.680	0.637

four cases as samples, including one with the highest number of records (102 records), one with a medium-to-high number of records (61 records), one with a medium-to-small number of records (9 records), and one with a small number of records (2 records). We used dark grey color to distinguish the selected cases in Figure 6.

Figure 7 shows the survival probabilities predicted by each of the seven models: K-M model (round-dotted line), CPH model (square-dotted line), BN-Cox (diamond), BN-Learn (triangle), BN-TAN (red dash) BN-NB (dark blue dash), and BN-NoisyMax (orange dash). We observe an almost perfect match between the CPH and the BN-Cox model in all 16 cases. Both BN-TAN and BN-NB are close for every case, while BN-NoisyMax falls in-between. The K-M

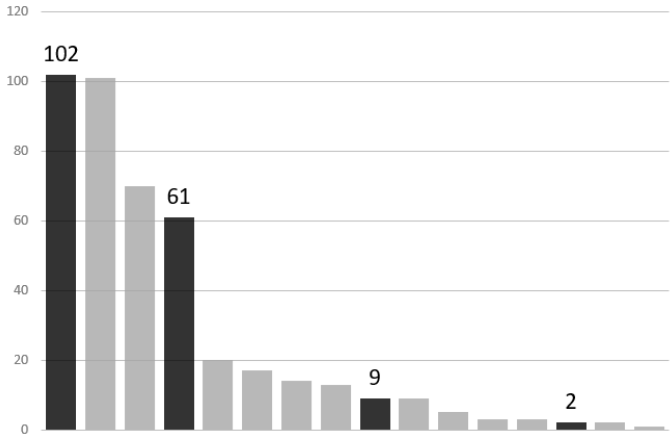


Figure 6: Distribution of the number of records in the Recidivism data set with four risk factors for each of the 16 combinations of risk factors (sorted in descending order). We selected four combinations of risk factors (marked as dark grey) for our comparison.

model and BN-Learn are also close, although they both depart from the CPH model significantly as the number of records gets smaller (Figures 7c and 7d).

BN-Cox, BN-Learn, BN-TAN, BN-NB, and BN-NoisyMax models are simplified and produce 13 survival probabilities for each case while the K-M and CPH model produced 52 survival probabilities. We found that when we have enough data to learn, e.g., more than a hundred records, there is a remarkable agreement among all seven models. However, when there are fewer data points, we found that the curves produced by the K-M estimate and the Bayesian network learned from data (BN-Learn), while in agreement with one another, depart from the CPH model significantly. The BN-Cox model and the CPH model, which again agree perfectly, produce smoother curves. We also observed agreement between the BN-TAN and the BN-NB models producing smoother curves for cases with few or no data records. The BN-NoisyMax model predicts probabilities in-between but not so similar to the remaining models.

In addition to the simplified, four-risk-factor model, we also created a complete Recidivism model with all eight risk factors (see details in [13]), seven of which are binary and one is categorical. The results were similar to those of the

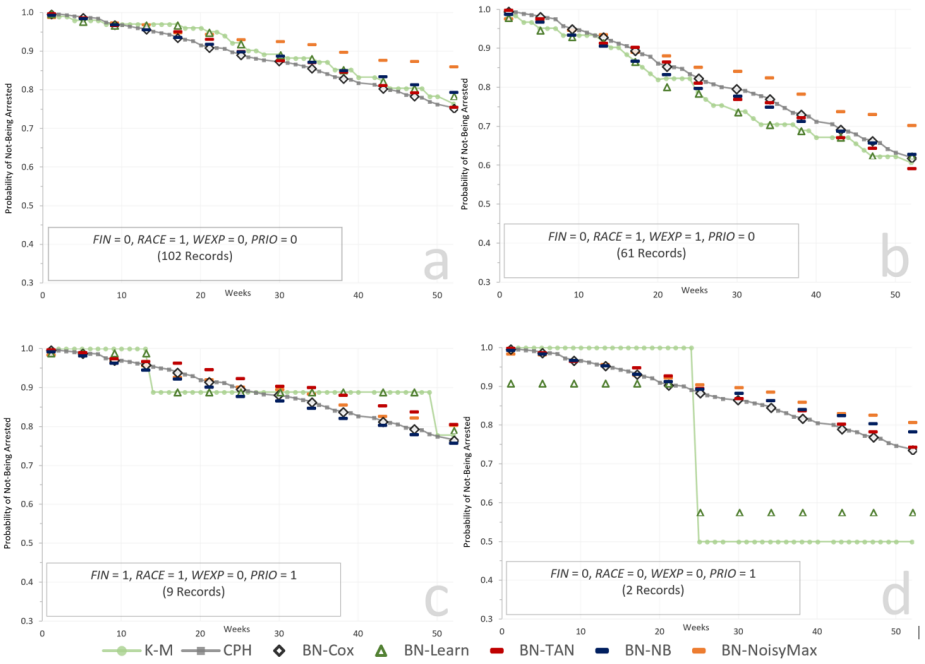


Figure 7: The predicted survival curves generated by each model for four ases with different number of records.

simplified models. The survival probabilities predicted by the BN-Cox model were identical to those of the CPH model. The BN-Learn, the K-M model, the BN-TAN, the BN-NB, and the BN-NoisyMax models produced similar trends, but the BN-Learn had an overall lower predicted survival probability. We can see larger differences in the predicted probability when there are few data records to learn from. The K-M model, BN-Learn, and BN-TAN produce different results only when the number of data records is small or zero. In this case, the CPH and the BN-Cox models agree perfectly.

We compared the accuracy (ACC) and the area under the receiver operating characteristic (ROC) curve (AUC) for the BN-Cox, the BN-Learn, the BN-TAN, the BN-NB, and the BN-NoisyMax models (for four-variable models and all-variables models) using 10-fold cross validation (Table 3). For four-variables models, each model produced very similar accuracy (ACC) and the area under

ROC (AUC). Both BN-Cox and BN-Learn performed similarly. BN-TAN and BN-NoisyMax are unable to correctly predict the re-arrest. We also observed similar performance for all-variables model. BN-Learn offered the best accuracy among all methods for the four-variables and all-variables models, while BN-NoisyMax was the least accurate. However, the differences in accuracy among the models are not significant (McNemar's test $p > 0.05$). In summary, our results show that when we do not have any data to learn from but only have an existing model, i.e., the CPH model, we can create a BN-Cox model to get similar performance.

Table 3: Performance of Bayesian network models with four variables and all variables

Performance	BN-Cox	BN-Learn	BN-TAN	BN-NB	NoisyMax
(Four-variables models)					
Accuracy (ACC)	0.8759	0.8769	0.8769	0.8761	0.8769
Area under ROC (AUC)	0.7605	0.7609	0.7536	0.7514	0.7421
(All-variables models)					
Accuracy (ACC)	0.8797	0.8803	0.8764	0.8748	0.8769
Area under ROC (AUC)	0.8322	0.8345	0.7926	0.7635	0.5646

6. Application of the BN-Cox model to risk calculation

As we mentioned earlier, CPH models are widely used in medical risk assessment and are often reported in the literature. Recently, we proposed a BN-Cox-based risk score calculator to the existing Pulmonary Arterial Hypertension (PAH) risk calculator [20]. The core of the original PAH risk score calculator by Benza et al. [21], the electronic mobile calculator app developed by the United Therapeutics Europe Limited and available at `http://www.pah-app.com/`, was based on the CPH model. Hence, we replaced the CPH model by a BN-Cox model constructed from the CPH parameters reported in Benza et al. [21] (Table 1). Figure 3 shows the structure of the BN-Cox model for the BN-Cox-based calculator. In this case, we omitted the *time* variable, as the purpose of the PAH Risk Calculator is to capture the risk at one point in time (one year). We cre-

ated the conditional probability table of the survival node from Equation 11. We configured all risk factors cases (all binary risk factors generated 2^{19} cases) and created the CPT of the survival node from the 2^{19} cases. This allowed us to reproduce fully the PAH CPH model by means of a Bayesian network (see more details in [20]). This by itself offers no advantages over a CPH model-based calculator but we view it as the first step toward a better calculator that relaxes some of the CPH assumptions and is capable of representing a generalized structure of interactions between risk factors and the survival variables.

With the PAH BN-Cox model, we created a risk score calculator using an approach similar to [21]. Equation 11 captures the survival probabilities s given the states of risk factors. We can extract a hidden hazard ratio of each variable by configuring states of other risk factors to be absent. For example, the hazard ratio of a risk factor x_j can be estimated from

$$\gamma = \frac{\log(Pr(s\ |\bar{x}_1, \ldots, \bar{x}_{j-1}, \mathbf{x_j}, \bar{x}_{j+1}, \ldots, \bar{x}_n))}{\log(Pr(s\ |\bar{x}_1, \ldots, \bar{x}_{j-1}, \bar{\mathbf{x}}_\mathbf{j}, \bar{x}_{j+1}, \ldots, \bar{x}_n))}\ . \tag{13}$$

The term $\log(Pr(s\ |\bar{x}_1, \ldots, \bar{x}_{j-1}, \bar{\mathbf{x}}_\mathbf{j}, \bar{x}_{j+1}, \ldots, \bar{x}_n))$ is similar to the baseline survival probability in the CPH model ($S_0(1) = 0.9698$). Hence, with this equation, we can track back all hazard ratios. Then, we use the same criteria as the original PAH Risk Calculator to convert the hazard rate to a score. Score of 2, for example, indicates at least two-fold increase in risk of mortality compared to the baseline risk.

Figure 8 shows a screen shot of the graphical user interface (GUI) of our prototype of the Bayesian network risk calculator. The left-hand side pane allows for entering risk factors for a given patient. The right-hand side pane shows the calculated score and survival probabilities. Currently, the numerical risks produced by the BN-Cox calculator are identical to those of the original CPH-based PAH Risk Calculator [21]. However, the BN-Cox model makes CPH's assumptions explicit and will allow to relax them in the future. One immediate advantage of the BN-Cox representation is that BNs make it possible to refine the parameters with additional data records.

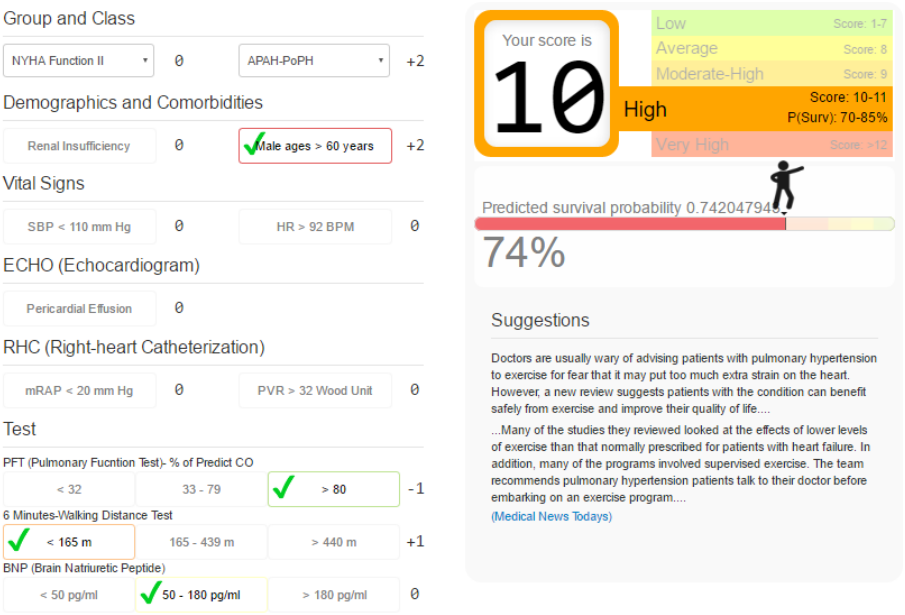


Figure 8: A prototype GUI for our Bayesian network risk score calculator for a 1-year PAH prognosis model. The left-hand pane allows for entering risk factors for a given patient case. The right-hand pane shows the calculated score and the survival probability.

7. Making the BN-Cox model computationally efficient

One of the challenges to applying the BN-Cox model is an exponential growth of the conditional probability tables (CPT) corresponding to the survival variables, as the number of risk factors increases [22]. When the number of risk factors is high, this table becomes intractable. We evaluated two approaches to mitigate this problem: (1) decomposition of the underlying Bayesian network known as parent divorcing, and (2) simplifying the network structure by removing least influential risk factors.

7.1. BN-Cox decomposition

In Bayesian networks, one way of reducing the complexity when the CPT of a node becomes too complex is through decomposition. This process can lead to substantial efficiency improvements in Bayesian updating [23]. In case of the

noisy-OR gates [24], for example, the combination function can be decomposed into a series of binary OR functions. For example, the $OR(X_1, \ldots, X_n)$ function is equivalent to $OR(X_1, OR(X_2, OR(\ldots OR(X_{n1}, X_n)\ldots)))$. Other functions, such as *AND*, *MIN*, and *MAX* can be decomposed similarly.

Decomposition of the CPH model amounts to finding a function f that is capable of expressing the survival function $S(t)$ in the following way:

$$S\left(t\right) = f\left(S_1\left(t\right)^{e^{(\beta_1 X_1 + \beta_2 X_2)}}, S_2\left(t\right)^{e^{(\beta_3 X_3 + \beta_3 X_3)}}\right) . \tag{14}$$

However, the survivor function describes an interaction between states of risk factors (PRESENT and ABSENT) and the probability of survival. This is different from the OR function, which describes interaction between states of variables. The following theorem, with an elegant proof offered by NAME, states that there is no universal decomposition function for the BN-Cox model.

Theorem 1. *There exists no universal decomposition function for parent-divorcing a BN-Cox model.*

Proof:. By contradiction for the simplest case with two binary risk factors X and Y, parents of the survival node S. The probability of survival is in this case expressed by the following function:

$$P(S = 1) = S_0^{exp(\beta_X X + \beta_Y Y)} . \tag{15}$$

We will attempt decomposition of the survival function into $P(A_X = 1) = S_0^{exp(\beta_X X)}$, the survival probability considering X as the only risk factor, and $P(A_Y = 1) = S_0^{exp(\beta_Y Y)}$, the survival probability considering Y as the only risk factor. Decomposition using parent divorcing requires two auxiliary nodes, A_X and A_Y, parents of S, with the conditional probabilities $P(S|A_X, A_Y)$

$$\begin{aligned} c_1 &= P(S = 1|A_X = 0, A_Y = 0) \\ c_2 &= P(S = 1|A_X = 0, A_Y = 1) \\ c_3 &= P(S = 1|A_X = 1, A_Y = 0) \\ c_4 &= P(S = 1|A_X = 1, A_Y = 1) . \end{aligned}$$

In order to decompose the BN-Cox model using the parent divorcing method, the following must hold for all values of $X \in [0,1]$ and $Y \in [0,1]$

$$\begin{aligned} S_0^{exp(\beta_X X+\beta_Y Y)} &= c_1(1-S_0^{\beta_X X})(1-S_0^{exp(\beta_Y Y)}) + c_2(1-S_0^{\beta_X X})(S_0^{exp(\beta_Y Y)}) \\ &\quad +c_3(S_0^{\beta_X X})(1-S_0^{exp(\beta_Y Y)}) + c_4(S_0^{\beta_X X})(S_0^{exp(\beta_Y Y)}) \ . \quad (16) \end{aligned}$$

By substituting $(X,Y) = (0,0)$ we get

$$S_0 = c_1(1-S_0)(1-S_0) + c_2(1-S_0)(S_0) + c_3(S_0)(1-S_0) + c_4(S_0) * (S_0) \ . \quad (17)$$

For the function to be universal, i.e., independent of the actual values of S_0, β_X, and β_Y, it must hold that $c_1 = 1$, $c_2 + c_3 = 1$, and $c_4 = 0$. If we substitute $c_4 = 0$ and $c_1 = 1$ into Equation 16 for $(X,Y) = (1,1)$, we get:

$$\begin{aligned} S_0^{exp(\beta_X+\beta_Y)} &= (1-S_0^{\beta_X})(1-S_0^{exp(\beta_Y)}) + c_2(1-S_0^{\beta_X})(S_0^{exp(\beta_Y)}) \\ &\quad +c_3(S_0^{\beta_X})(1-S_0^{exp(\beta_Y)}) \\ &= S_0^{exp(\beta_X+\beta_Y)} + c_2 S_0^{exp(\beta_Y)} - (c_2+c_3)(S_0^{exp(\beta_X)})(S_0^{exp(\beta_Y)}) \\ &\quad +c_3 S_0^{exp(\beta_X)} \ , \end{aligned}$$

which for the function to be universal requires $c_2 = c_3 = 0$. This contradicts the condition $c_2 + c_3 = 1$ from Equation 17 and concludes the proof. □

While BN-Cox model cannot be decomposed by means of the parent divorcing method, one suggestion offered by NAME was studying other decompositions. Complexity of such decompositions requires studying the rank of the CPH model along the lines of analysis for several popular canonical models [25, 26, 27]. While we leave the search for other possible decomposition methods outside of this paper, we offer a quick experimental analysis of the possible approximate decompositions. To check the quality of possible approximate decompositions, we performed several experiments that consisted of manually decomposing the BN-Cox model and refitting its probabilities from data. Figure 9 shows an example of the structured decomposition of the original BN-Cox model shown in Figure 3. Unfortunately, all our attempts resulted in poor numerical fit and models of clearly inferior quality than the original BN-Cox model.

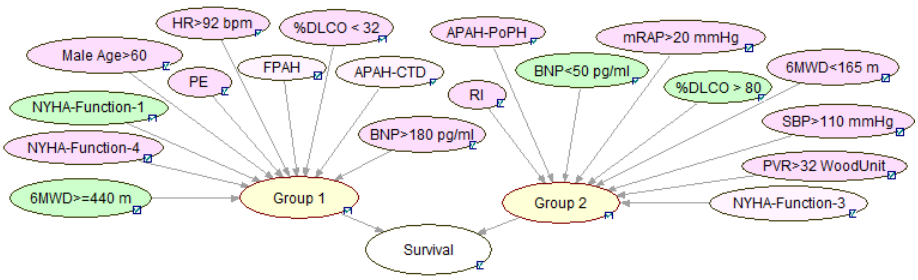


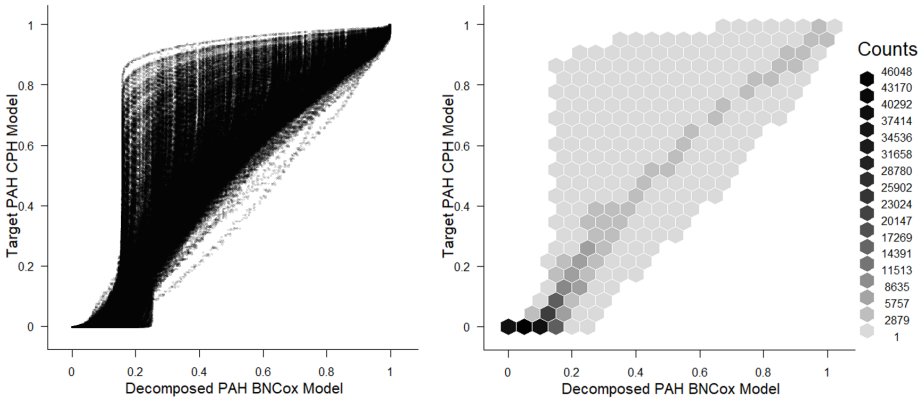
Figure 9: An example of a BN-Cox model decomposition of the a BN-Cox model from Figure 3

Figure 10 illustrates the poor quality of approximation of the decomposed model. We used the scatterplot of the survival probabilities from the decomposed BN-Cox model against the ones produced by the original model (Figure 10a). In case of perfect fit, the plot would be a perfect diagonal line from $(0, 0)$ to $(1, 1)$. Figure 10b shows the same scatterplot transformed by hexagon binning techniques [28]. Each hexagon is color-coded according to the number of points falling in that region. While many probabilities are similar, we see a large off-diagonal cloud that indicates poor fit. Figure 11 shows the histogram of Euclidean distances between the survival probabilities calculated by the original CPH and the decomposed model for all possible combinations of values of risk factors sorted from the smallest to the largest distance. We clearly see an overall poor fit between the decomposed and the original model.

Although, we have not tested all versions of network decomposition, we tried other decompositions with different number of groups including 4 groups, 6 groups, and 9 groups. All these decompositions confirmed poor approximation of the original model.

7.2. BN-Cox simplification by removing least influential risk factors

Another method of reducing the complexity of the BN-Cox model is to simplify the CPH model itself by removing the least influential risk factors. It can be expected that some of the risk factors will have minimal effect on the result and omitting them altogether will not lead to much loss of precision. On the other hand, removing each of these least influential factors will cut the size of the



(a) Scatterplot (b) Scatterplot with Hexagonal Binning

Figure 10: Survival probabilities produced by the decomposed model against survival probabilities produced by the original CPH model.

survival node's CPT by at least half. In practice, there are several techniques of variable selection in survival analysis [29]. We started out by evaluating the effect of removing the weakest variable. The "weakest" means the variable with the highest p value and possibly the smallest value of the β coefficient. The larger the value of p, the less certain we are that the risk factor is really affecting survival, the smaller the value of β, the weaker the effect, even if there is any.

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

We performed simplification experiments on the Recidivism CPH model consisting of seven binary risk factors listed in Table 4. First, we compared the effect

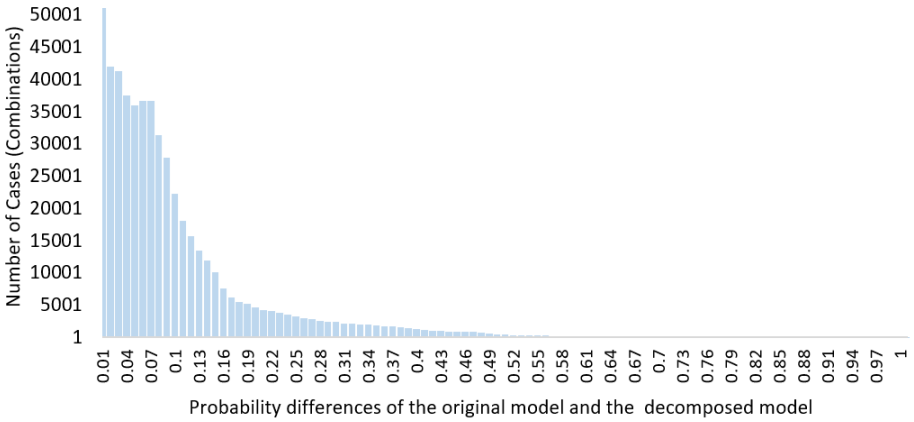
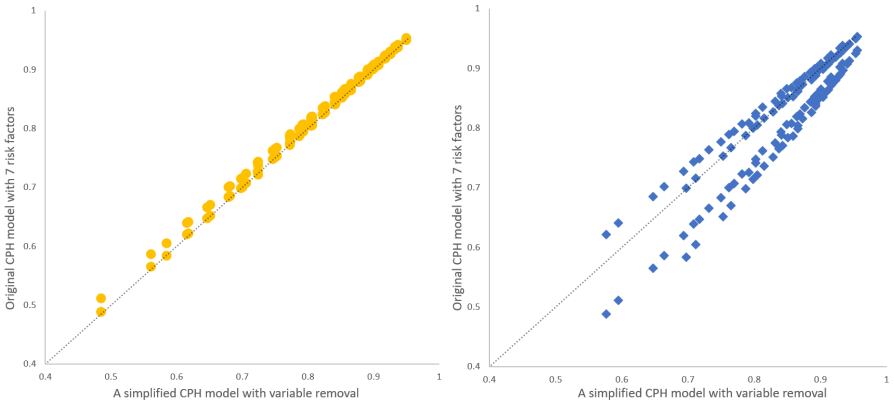


Figure 11: The histogram showing the Euclidean distance between the survival probabilities produced by the original PAH BN-Cox model and the decomposed model sorted from the smallest to largest distance.

of removing the least significant variable against the effect of removing the most significant one. The weakest variable in Table 4 seems *paro* with $\beta = -0.06721$ and $p = 0.7288$, while the strongest variable seems *wexp* with $\beta = 0.41055$ and $p = 0.0403$. To create a simplified model, we remove the selected variables from the data set and refit the CPH model. Hence, we have two refitted models: (1) one with the variable *paro* (weakest) removed, and (2) one with the variable *wexp* (strongest) removed. We compared the predicted survival probabilities against the original model shown in Figure 12. 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. For example, the survival probabilities produced by the original model when *fin*=0, *race*=0, *wexp*=0, *mar*=0, *prio*=0, *educ*=0, *paro*=0 and when *fin*=0, *race*=0, *wexp*=0, *mar*=0, *prio*=0, *educ*=0, *paro*=1 are mapped to the survival probability produced by the *paro*-removed model when *fin*=0, *race*=0, *wexp*=0, *mar*=0, *prio*=0, *educ*=0.

The results obtained by removing the least significant variable (Figure 12a)

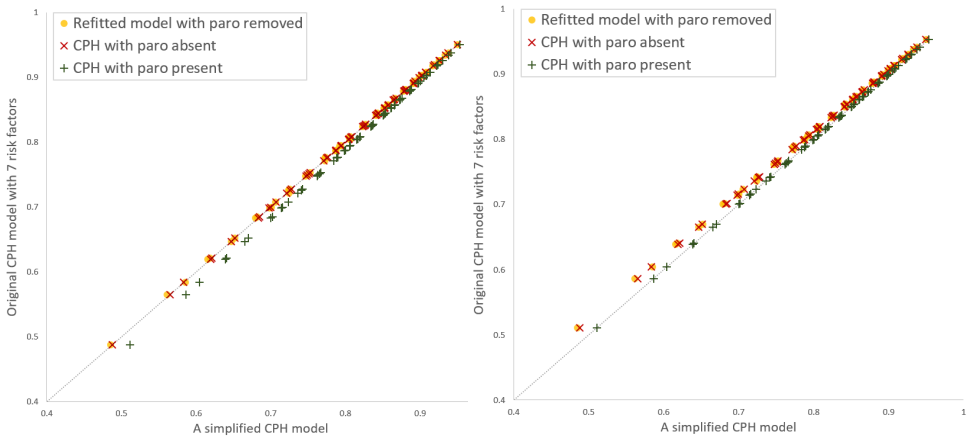


(a) Weak variable (*paro*) removed (b) Strong variable (*wexp*) removed

Figure 12: 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 in prediction between the original and the modified models.

are closer to the original model than the results obtained by removing the strongest variable (Figure 12b). In this experiment, we identified the least/most significant variables by their β and p-values in the original CPH model. However, one can use any variable selection method here [29].

Removing a weak variable and refitting works only when we have the original data set. In practice, however, we often have only the CPH parameters and not the data from which they were obtained. For those variables with small influences, it can be expected that setting those variables to be *absent* will be similar to removing those variables in the simplified refitted model. In a follow-up experiment, we evaluated the effect of fixing state of the weakest variable (*paro*) to *absent* against the simplified refitted model. We used the original CPH model (Table 4) and simplified the model by fixing the state of *paro*. As a result, we have two sets of predicted probabilities from the fixed-state model: cases when *paro* is fixed to *absent* ($paro = 0$) and cases when *paro* is fixed to *present* ($paro = 1$). Then, we compared those results to the original CPH models and the model with *paro* removed (Figure 13).



(a) Cases when $paro = 0$ (b) Cases when $paro = 1$

Figure 13: The scatterplots 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.

Figure 13a shows the predicted probabilities of all cases with *paro*=0. The diagonal grey-dotted line shows ideal probabilities with all *paro*=0 cases as produced from the original CPH model. It can be expected that all probabilities produced from the model fixing *paro* to *absent* (the red cross markers) are perfectly on the diagonal line, while, setting *paro* to *present* produced some errors (the plus markers). We observed that fixing *paro* to *absent* produced results very close to the results from the *paro*-removed model (the yellow circle markers). For those cases with *paro*=1 in the original CPH model, we also observed similar trends in Figure 13b. All probabilities from the model fixing *paro* to *present* (the green cross marker) lie perfectly on the diagonal grey-dotted line, which shows ideal probabilities with all *paro*=1 cases as produced from the original CPH model. In this case, setting *paro* to *absent* produced errors. In summary, 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. However, fixing state of the variable to *absent* still produces errors for those original cases with the risk factor *present*, and vise versa.

We also created four models with one, two, three, and four least significant

variables *absent*. We refitted corresponding four models by removing the least influential risk factors. Figure 14 shows the results for the simplified models with both fixed to be *absent* and refitted models against the original CPH model for different numbers of risk factors. We can see that removal of multiple variables, especially when their influence is larger, can lead to departure from the ideal precision (the diagonal line in the plots). We should add that removing four of the seven Recidivism variables was expected to make a large impact on the quality of the resulting model. We believe that the loss of precision will be much smaller when the number of variables removed is small.

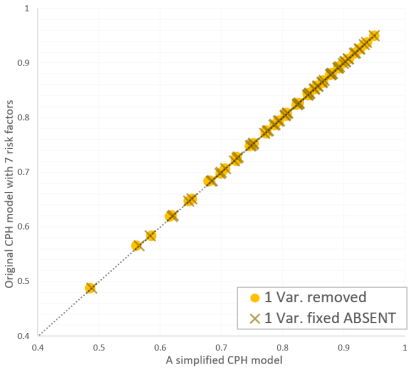
As we mentioned, fixing the small-influence risk factors to *absent* is similar to removing those risk factors from the model but still produces error when the risk factors are *present*. In Bayesian networks, we can use marginalization to simplify a model. Marginalization amounts to removing a risk factor X_i from consideration while preserving the joint probability distribution among the remaining variables and the effect of the remaining risk factor on the survival probability, s. Marginalization of a risk factor, X_i, amounts to:

$$Pr(s \mid \xi) = \sum_{i=1}^{n} Pr(s \mid \xi, x_i) \cdot Pr(x_i) \ , \tag{18}$$

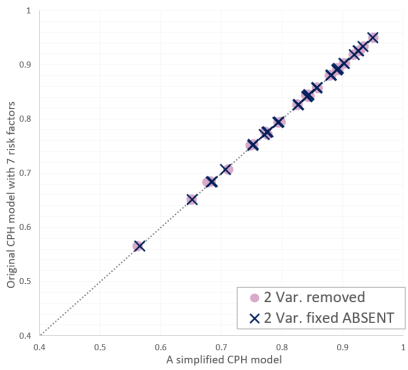
where x_i are states of, X_i, $Pr(s)$ is the survival probability, and ξ are all other risk factors.

We performed an experiment on the use of marginalization and compared the results against the results from previous experiments. We created a BN-Cox model from all CPH parameters in Table 4, then marginalized the variables *paro* and *wexp* out. Hence, we collected the predicted probabilities from each marginalized model and compared the results to the original CPH model, the model with variable removed, and the fixing-state model, case-by-case (Figure 15).

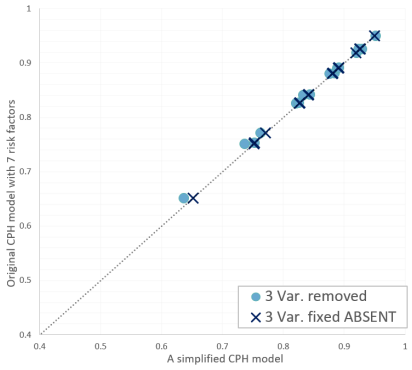
Figure 15a shows the result of the *paro*-marginalized model (the grey square markers) against the results from previous experiments, including the results from the no-*paro* refitted model (the yellow circle markers), the result from fixing-state model with *absent paro* (the red cross markers), the result from



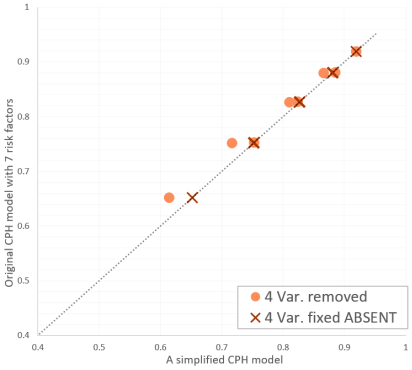
(a) One risk factors *absent* vs. *removed*



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



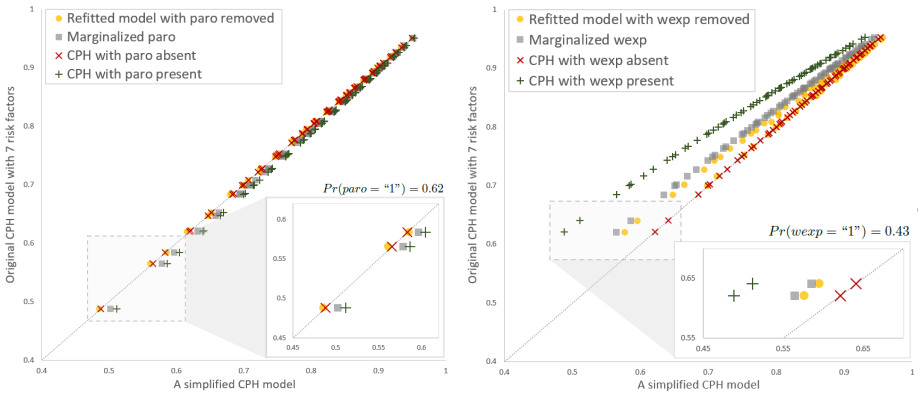
(c) Three risk factors *absent* vs. *removed*



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

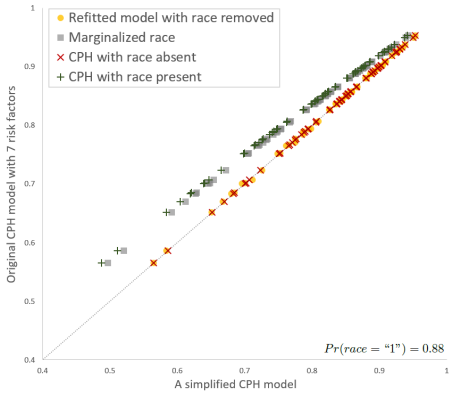
Figure 14: Effect of *absent* and *removed* risk factors in the simplified models against the original CPH model. The diagonal gray line shows the ideal probability as produced from the original CPH model. The predicted probabilities from the simplified models are compared only for the cases when those selected risk factors in the original CPH model are *absent*.

fixing-state model with *present paro* (the green plus markers), and the result from the original model when *paro*=0 (the diagonal grey-dotted line). As we expected, the marginalized model produced probabilities in-between the results from the *absent*- and *present*-fixed model, and also closer to the *present*-fixed model, since 68 percents of inmates in the Recidivism data set have been on parole before being released. We observed the same trends in other figures: The marginalized model produced the results by weighing out the effect of each state by its prior probability. We believe that marginalization is the correct way to

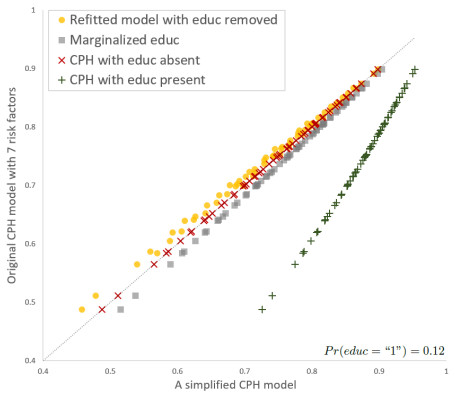


(a) Marginalized *paro*

(b) Marginalized *wexp*



(c) Marginalized *race*



(d) Marginalized *educ*

Figure 15: Effect of marginalized risk factors in the simplified models against the original CPH model, the refitted model, and the fixed-state models. The diagonal gray line shows the ideal probability as produced from the original CPH model.

remove the selected variable since it still preserves the effect of the risk variables.

8. Discussion

In this paper, we have introduced BN-Cox, a Bayesian network interpretation of the Cox's Proportional Hazard (CPH) model. We have shown that BN-Cox reproduces the CPH model precisely under perfect circumstances, i.e., when all CPH assumptions hold. Since the BN-Cox model is derived from the CPH

model, all predicted risks are the same for the two models.

The CPH model, the Bayesian network learning from data, and the K-M estimates have a similar predictive ability when there are enough data to learn from. However, when there are few data records, the survival curves derived from the K-M estimate and Bayesian network learning from data models depart from the CPH survival curves. In most practical problems, data can include many combinations of risk factors with very small number of records for most of the combinations. In such cases, the CPH model helps to smoothen out the distributions similarly to other models of interactions among variables [30]. Hence, Bayesian networks interpreted from the CPH model may be more useful in practice than the K-M estimate or Bayesian networks learned from data. Their largest benefit over the CPH model, however, are in all those cases, when the CPH assumptions are violated. Bayesian networks have no problem with modeling such situations and offer a more flexible modeling tool that allows for a combination of expert knowledge with data.

We have presented one application of the BN-Cox model: A risk calculator for the pulmonary arterial hypertension. We created the BN-Cox model i.e., from the CPH parameters available in the literature without having access to the REVEAL Registry data [5]. Our calculator reproduced the results of the current PAH Risk Calculator exactly. We plan to refine this calculator by (1) learning the parameters of the BN model from the data captured in the REVEAL Registry, and (2) enhancing the resulting BN model with medical expert knowledge. The extended model will relax the assumption of the multiplicative character of interactions between the risk factors and the survival variable. We have little doubt that, with some modeling effort, we should be able to obtain a calculator producing higher accuracy of the risk estimate than the original CPH-based risk calculator.

Finally, we have studied two ways of making the BN-Cox model computationally efficient. Our main challenge to making BN-Cox more practical is an exponential growth of the conditional probability tables of the survival variable node. We tested two approaches: (1) parent divorcing, and (2) removing least

influential risk factors. The BN-Cox model turns out to be not decomposable and approximating of decomposition leads to high loss of accuracy. Hence, we suggest to simplify the network structure by removing the least influential risk factors.

We can use any statistical variable selection method [29] to simplify or reduce the number of risk factors in the CPH models when we have a data set to refit the simplified model. However, when data are not available, we can simplify the model by removing least influential risk factors based on both the value of β coefficients and the statistical significance. When removing risk factors, we suggest marginalization, as it leads to smallest error on the average. There are good reasons for not worrying too much about the precision of the BN parameters in practice. Oniśko and Druzdzel [31] found that in medical diagnostic systems based on Bayesian networks, precision of parameters may not be as important as popularly believed. Hence, approximating relationships may leave minimal effect on the practical accuracy of systems based on Bayesian networks.

Acknowledgments

We acknowledge the support the National Institute of Health under grants U01HL101066-01 and 1R01HL134673-01, Department of Defence under grant number W81XWH-17-1-0556, and the Faculty of Information and Communication Technology, Mahidol University, Thailand. Implementation of this work is based on GeNIe and SMILE, a Bayesian inference engine developed at the Decision Systems Laboratory, University of Pittsburgh. It is currently a commercial product but is still available free of charge for academic research and teaching from BayesFusion, LLC, at `https://www.bayesfusion.com/`. Parts of this paper were presented at the 2014 and 2016 International Conferences on Probabilistic Graphical Models [13, 22]. The BN-Cox Interpretation of the PAH risk calculator, discussed in details in [20], was developed in collaboration with Dr. Raymond L. Benza of Allegheny General Hospital, Pittsburgh, PA.

While we are taking full responsibility for any remaining errors and shortcomings of the paper, we would like to thank Dr. Jong-Hyeon Jeong of the School of Public Health, University of Pittsburgh, for offering us insight and literature pointers useful in our quest to decompose the CPH model and Krzysztof Nowak, for fitting of Noisy-MAX parameters to Recidivism data in our experiments. NAME offered the proof of Theorem 1. We also thank the anonymous reviewers for the PGM conferences and for IJAR for their valuable input that has greatly improved the quality of this paper.

9. References

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

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

[3] S. L. Spruance, J. E. Reid, M. Grace, M. Samore, Hazard ratio in clinical trials, Antimicrobial Agents and Chemotherapy 48 (2004) 2787–2892.

[4] W. C. Levy, D. Mozaffarian, D. T. Linker, S. C. Sutradhar, S. D. Anker, A. B. Croppan, I. Anand, A. Maggionl, P. Burton, M. D. Sullivan, B. Pitt, P. A. Poole-Wilson, D. L. Mann, M. Packer, The Seattle Heart Failure Model: Prediction of survival in heart failure, Circulation 113 (2006) 1424–1433.

[5] 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, 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 (2010) 164–172.

[6] A. A. Hanna, P. J. Lucas, Prognostic models in medicine – AI and statistical approaches, Methods of Information in Medicine 40 (2001) 1–5.

[7] D. Husmeier, R. Dybowski, S. Roberts, Probabilistic Modeling in Bioinformatics and Medical Informatics, Springer, 2005.

[8] L. D. Casea, G. Kimmickb, E. D. Pasketta, K. Lohmana, R. Tucker, Interpreting measures of treatment effect in cancer clinical trials, The Oncologist 7 (2002) 181–187.

[9] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[10] N. A. Loghmanpour, M. K. Kanwar, M. J. Druzdzel, R. L. Benza, S. Murali, J. F. Antaki, A new Bayesian network-based risk stratification model for prediction of short-term and long-term LVAD mortality, ASAIO Journal 61 (2015) 313–323.

[11] K. P. Murphy, Dynamic Bayesian networks: Representation, inference and learning, Doctoral dissertation, University of California, Berkeley, 2002.

[12] M. A. Van Gerven, B. G. Taal, P. J. Lucas, Dynamic Bayesian networks as prognostic models for clinical patient management, Journal of Biomedical Informatics 41 (2008) 515–529.

[13] J. Kraisangka, M. J. Druzdzel, Discrete Bayesian network interpretation of the Cox's proportional hazards model, in: L. C. van der Gaag, A. J. Feelders (Eds.), Probabilistic Graphical Models, volume 8754 of *Lecture Notes in Computer Science*, Springer International Publishing, 2014, pp. 238–253. URL: http://dx.doi.org/10.1007/978-3-319-11433-0_16. doi:10.1007/978-3-319-11433-0_16.

[14] E. L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, Journal of the American Statistical Association 53 (1958) 457–481.

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

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

[17] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the royal statistical society. Series B (methodological) (1977) 1–38.

[18] S. L. Lauritzen, The EM algorithm for graphical association models with missing data, Computational Statistics & Data Analysis 19 (1995) 191–201.

[19] K. Nowak, M. J. Druzdzel, Learning parameters in canonical models using weighted least squares, in: L. C. van der Gaag, A. J. Feelders (Eds.), Probabilistic Graphical Models, Springer Lecture Notes in Computer Science, Vol. 8754, Springer International Publishing, Heidelberg, 2014, pp. 366–381.

[20] J. Kraisangka, M. J. Druzdzel, R. L. Benza, A risk calculator for the pulmonary arterial hypertension based on a Bayesian network, in: Proceedings of the 13th UAI Bayesian Modeling Applications Workshop, 2016, pp. 1–6.

[21] R. L. Benza, M. Gomberg-Maitland, D. P. Miller, A. Frost, R. P. Frantz, A. J. Foreman, D. B. Badesch, M. D. McGoon, The REVEAL registry risk score calculator in patients newly diagnosed with pulmonary arterial hypertension, CHEST 141 (2012) 354–362.

[22] J. Kraisangka, M. J. Druzdzel, Making large Cox's proportional hazard models tractable in Bayesian networks, in: Proceedings of the Eighth International Conference on Probabilistic Graphical Models, 2016, pp. 252–263.

[23] A. Zagorecki, M. Voortman, M. J. Druzdzel, Decomposing local probability distributions in Bayesian networks for improved inference and parameter learning, in: G. Sutcliffe, R. Goebel (Eds.), Recent Advances in Artificial Intelligence: Proceedings of the Nineteenth International Florida

Artificial Intelligence Research Society Conference (FLAIRS–2006), AAAI Press, Menlo Park, CA, 2006, pp. 860–865.

[24] F. J. Díez, M. J. Druzdzel, Canonical Probabilistic Models for Knowledge Engineering, Technical Report, UNED, Madrid, Spain, 2006.

[25] F. J. Díez, S. F. Galan, Efficient computation for the noisy MAX, International Journal of Intelligent Systems 18 (2003) 165–177.

[26] P. Savicky, J. Vomlel, Exploiting tensor rank-one decomposition in probabilistic inference, Kybernetika (Special issue dedicated to the memory of Albert Perez) 43 (2007) 747–764.

[27] J. Vomlel, P. Tichavsky, Probabilistic inference with noisy-threshold models based on a CP tensor decomposition, International Journal of Approximate Reasoning 55 (2014) 1072–1092.

[28] N. Lewin-Koh, Hexagon Binning: an Overview, 2018. URL: `https://cran.r-project.org/web/packages/hexbin/vignettes/hexagon_binning.pdf`.

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

[30] A. Oniśko, M. J. Druzdzel, H. Wasyluk, Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates, International Journal of Approximate Reasoning 27 (2001) 165–182.

[31] A. Oniśko, M. J. Druzdzel, Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems, Artificial Intelligence in Medicine 57 (2013) 197–206.