Lifetime Data Analysis (2023) 29:735–751
https://doi.org/10.1007/s10985-023-09600-z

Volume under the ROC surface for high-dimensional independent screening with ordinal competing risk outcomes

Yang Qu[1] · **Yu Cheng**[1]

Received: 3 June 2022 / Accepted: 8 April 2023 / Published online: 9 May 2023

Abstract
We propose a screening method for high-dimensional data with ordinal competing risk outcomes, which is time-dependent and model-free. Existing methods are designed for cause-specific variable screening and fail to evaluate how a biomarker is associated with multiple competing events simultaneously. The proposed method utilizes the Volume under the ROC surface (VUS), which measures the concordance between values of a biomarker and event status at certain time points and provides an overall evaluation of the discrimination capacity of a biomarker. We show that the VUS possesses the sure screening property, i.e., true important covariates can be retained with probability tending to one, and the size of the selected set can be bounded with high probability. The VUS appears to be a viable model-free screening metric as compared to some existing methods in simulation studies, and it is especially robust to data contamination. Through an analysis of breast-cancer gene-expression data, we illustrate the unique insights into the overall discriminatory capability provided by the VUS.

Keywords Biomarker evaluation · Kendall's tau · Model-free screening · Sure screening property · U statistic

1 Introduction

Variable screening becomes increasingly necessary with the availability of ultra-high dimensional data. The goal of variable screening is to reduce the number of covariates to a moderate size, which would then allow the use of traditional or high-dimensional variable selection methods (Fan and Lv 2008). Fan and Lv (2008) proposed the sure independent screening (SIS) method, simply keeping covariates that

✉ Yu Cheng
yucheng@pitt.edu

[1] Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA

are independently highly correlated with the outcome of interest, and introduced the sure screening property, which means that the screening method would keep all important covariates with probability tending to one.

Various screening methods have been developed for survival outcomes, assuming non-informative right censoring. Tibshirani (2009), Fan et al. (2010), Zhao and Li (2012) proposed screening procedures assuming that event times follow Cox's model. Gorst-Rasmussen and Scheike (2013) justified their method for single-index hazard rate models. These methods may not be adequate when the model assumption is violated, and model-free methods have thus been developed; see Song et al. (2014), Li et al. (2016), Hong et al. (2018) and Pan et al. (2018) for instance. All of these methods were shown to possess the sure screening property.

With the existence of competing risks, subjects are exposed to more than one event. Screening methods designed for survival outcomes may not be directly applied, as the competing events cannot be treated as non-informative censoring when dealing with the event of primary interest. Several methods have been introduced recently for variable screening under the competing or semi-competing risk setting. Lu et al. (2020), Peng and Xiang (2021), and Liu et al. (2021) proposed screening methods for semi-competing risks data. These methods require two time values being observed, one is the minimum of non-terminal event, terminal event, and censoring time, and the other is the minimum of terminal event and censoring time, and thus semi-competing risk methods may not be applicable to the competing risk setting where only the minimum of the censoring time and all event times is observed. For competing risks data, Li Erqian (2018) developed a feature screening method for the proportional subdistribution hazard model by maximizing the log-likelihood function with constraints using Taylor's expansion and then selecting variables based on a penalized log-likelihood function. Chen et al. (2022) proposed a sure independent screening procedure by ranking the marginal correlation between each single covariate and the estimated cumulative incidence function of the event of interest (crCRS). Tian et al. (2022) considered a feature screening method for a varying coefficient proportional subdistribution hazard model via maximizing a constrained local partial likelihood. Wang et al. (2022) developed a safe feature elimination (SAFE) algorithm for Lasso-type variable screening based on a proportional subdistribution hazard model, which eliminates features that are guaranteed to have zero coefficients in the Lasso estimator. These methods designed for competing risks data mainly evaluate whether biomarkers are related to the occurrence of one particular event with the presence of competing risks. In practice, we sometimes encounter situations where biomarkers may be related to several competing events, and continuous changes in biomarkers, either increasing or decreasing, are associated with changes in event statuses, which results in a naturally ordered sequence of competing events. When the goal is to find such covariates that can predict different event statuses at a given time, existing methods may not be adequate, since they cannot evaluate biomarkers' ability to discriminate subjects from multiple competing event groups simultaneously, but mainly focus on their ability to distinguish the event group of interest from the event-free group. Therefore, in this article, we aim to fill this gap and consider variable screening for covariates that are related to ordinal competing risks simultaneously.

We utilize the volume under the ROC surface (VUS) proposed by Zhang et al. (2022) as our screening metric. Details are provided in Sect. 2. In Sect. 3, we show that VUS possesses the sure screening property under certain conditions. In Sect. 4, we run simulation studies to examine the performance of VUS. A real data analysis is carried out in Sect. 5.

2 Method

2.1 Volume under the ROC surface

We focus on cases with two ordered competing events, and cases with more than two competing events can be generalized. For convenience we assume that cause-1 event is the most severe condition, such that as the value of biomarker decreases, the event status would be likely to evolve from event-free to cause-2 event, and then to cause-1 event, forming a natural order.

Considering a three-dimensional ROC surface formed by the correct classification probabilities for subjects from cause-1 event, cause-2 event, and the event-free groups, the VUS is the volume under this ROC surface, and it corresponds to a concordance index (Mossman 1999; Nakas and Yiannoutsos 2004) as described below. Let T be the time to any ordinal competing events and let $\epsilon = 1, 2$ be the cause indicator. Let $\mathbf{Z}$ be a p-dimensional vector of covariates. For three randomly selected subjects i, j and k such that $T_i \le t_0, \epsilon_i = 1, T_j \le t_0, \epsilon_j = 2$, and $T_k > t_0$, the VUS at t_0 for a single covariate Z_l is

$$VUS_l(t_0) = P(Z_{il} < Z_{jl} < Z_{kl} | T_i \le t_0, \epsilon_i = 1, T_j \le t_0, \epsilon_j = 2, T_k > t_0),$$

where Z_{il}, Z_{jl} and Z_{kl} are the lth covariate for subjects i, j and k. Therefore, VUS measures the concordance probability of a biomarker and a sequence of events (Zhang et al. 2022).

Let C denote the time of censoring and assume that C is independent of T and covariates $\mathbf{Z}$. Define $X = \min(T, C)$, and $\eta = I(T \le C)\epsilon$, and the observed data consist of independent and identically distributed triplets $\{(X_i, \eta_i, \mathbf{Z}_i)\}, i = 1 \dots n$. Let G be the survival function of C and denote its Kaplan-Meier estimator as $\hat{G}$. Then from Zhang et al. (2022), an estimator of the VUS at a pre-specified time t_0 for the l-th covariate based on inverse probability of censoring weighting is given by

$$\widehat{VUS}_l(t_0) = \frac{\sum_{i \ne j \ne k} \frac{I(X_i \le t_0, \eta_i = 1, X_j \le t_0, \eta_j = 2, X_k > t_0, Z_{il} < Z_{jl} < Z_{kl})}{\hat{G}(X_i -)\hat{G}(X_j -)\hat{G}(t_0)}}{\sum_{i \ne j \ne k} \frac{I(X_i \le t_0, \eta_i = 1, X_j \le t_0, \eta_j = 2, X_k > t_0)}{\hat{G}(X_i -)\hat{G}(X_j -)\hat{G}(t_0)}}.$$

2.2 Screening procedure

Zhang et al. (2022) showed that VUS provides an assessment of the overall discrimination capacity of a continuous marker for multi-level categorical outcomes with a natural order. Due to its nice property and interpretation, we consider using $VUS(t)$ as a metric for screening variables predictive of competing risk outcomes. Define $\mathscr{M}^*$ as the set of true active variables such that

$$\mathscr{M}^* = \{l : P(T \le t, \epsilon = 1 \text{ or } 2|\mathbf{Z}) \text{ depends on } Z_l\}.$$

In practice, the conditional distribution given a marker is unknown, and a working distribution is needed to estimate the true active set $\mathscr{M}^*$. To avoid imposing specific model assumptions, we use the VUS to select important biomarkers whose continuous changes would lead to changes in event statuses. We thus implicitly restrict $\mathscr{M}^*$ to be a set of biomarkers with an ordinal relationship with the event status.

When a covariate Z_l has no association with event time, its true VUS would be 1/6, as for three subjects randomly selected from three distinct groups, there are six possible rankings of their biomarkers with equal probabilities, and $Z_{il} < Z_{jl} < Z_{kl}$ appears with 1/6 probability; see also Nakas and Yiannoutsos (2004) and Li and Fine (2008). Ties in biomarkers can be handled following Zhang et al. (2022). We therefore select a set of important covariates by comparing their VUS estimates $\widehat{VUS}(t_0)$ with 1/6. Suppose γ_n is a pre-specified threshold. The selected set is denoted by

$$\hat{\mathscr{M}} = \{l : |\widehat{VUS}_l(t_0) - 1/6| \ge \gamma_n\}.$$

Basically, t_0 can be selected as any time points of scientific interest satisfying certain conditions, which will be mentioned in later sections. By the definition of VUS and the ordinal relationship between true active covariates and the sequence of events, $\hat{\mathscr{M}}$ provides an estimator of $\mathscr{M}^*$.

Note that there are several cases where the VUS falls below 1/6. In constructing the ROC surface, we assumed that a smaller covariate value would indicate a more severe condition. This assumption may be violated in practice. For example, the important covariates can still be related to the outcomes in an ordinal manner, but with larger values indicating more severe conditions. Although these covariates have a strong association with the events, they are less likely to be selected than variables satisfying the assumption because their VUS values are below 1/6, and their absolute deviations from 1/6 are bounded by 1/6. In this case, we would reverse their signs to satisfy the assumption. After taking reverse, $|VUS - 1/6|$ would largely increase, and these important variables are much more likely to be captured. However, if the relationship between a covariate and the outcomes is not in the assumed order or its opposite, the value of the VUS would be relatively small, even after taking inverse, and our metric is less likely to pick such covariates. Additional simulation results have been provided in Supplementary Material showing where VUS would locate under different possible relationships between the biomarker and the event status.

3 Sure screening property

In this section, we establish the sure screening property of the VUS method. The following conditions are required.

Condition 1 There exists a $\nu > 0$ such that $P(C = \nu) > 0$ and $P(C > \nu) = 0$.

Condition 2 $\min_{l \in \mathscr{M}^*} |VUS_l(t_0) - 1/6| \geq c_0 n^{-\kappa}$ for some $0 < \kappa < 1/2$ and $c_0 > 0$.

Condition 3 There exists $\delta > 0$, such that $P(T_i \leq t_0, \epsilon_i = 1, T_j \leq t_0, \epsilon_j = 2, T_k > t_0) > \delta$.

Condition 1 is used to show asymptotic properties in Song et al. (2014) for survival outcomes. Condition 2 shows that true active covariates can be distinguished from pure noise by the definition of VUS, and further links the selected set $\hat{\mathcal{M}}$ with $\mathscr{M}^*$. Condition 3 requires that there are subjects in each of the event groups. It is used to show asymptotic properties and can be easily satisfied with a properly selected t_0.

Theorem 1 *Under Conditions* 1–3, *for any positive constant* c_6, *there exist positive constants* c_3, c_4 *and* c_5 *such that* $c_6 = \frac{1}{\delta}(2c_3 + 2)c_4 + \frac{1}{\delta}c_5$, *and for any single covariate* Z,

$$\begin{aligned} P(|\widehat{VUS} - VUS| \geq c_6 n^{-\kappa}) \leq\ & 10n^3 \exp\{-\frac{1}{36}\epsilon^8 n\} + 4\exp\{-\frac{2}{3}c_3^2\epsilon^6 n\} \\ & + 4\exp\left\{-\frac{2}{27}c_4^2\epsilon^6\delta^2(1 + c_4 n^{-\kappa})^{-2}n^{1-2\kappa}\right\} \\ & + 2.5n^3 \exp\left\{-\frac{1}{9}c_4^2\delta^2\epsilon^8(3 + 3c_4 n^{-\kappa} + c_4 n^{-\kappa}\delta)^{-2}n^{1-2\kappa}\right\} \\ & + 4\exp\{-\frac{2}{27}c_5^2\epsilon^6 n^{1-2\kappa}\} \\ & + 2.5n^3 \exp\{-\frac{1}{9}c_5^2\epsilon^8(3 + c_5 n^{-\kappa})^{-2}n^{1-2\kappa}\}. \end{aligned} \tag{1}$$

Taking $\gamma_n = cn^{-\kappa}$ *with* $c = c_0 - c_6$, *we have*

$$P(\mathscr{M}^* \subset \hat{\mathcal{M}}) \geq 1 - sP(|\widehat{VUS} - VUS| \geq (c_0 - c)n^{-\kappa}),$$

where $s = |\mathscr{M}^*|$ *is the cardinality of* $\mathscr{M}^*$.

Theorem 2 *Under the conditions of Theorem* 1, *with* $p = o(\exp(n^{1-2\kappa}))$ *and assuming* $\sum_{l=1}^{p} |VUS_l - 1/6| = O(n^{\xi})$ *for some* $\xi > 0$, *we have*

$$\begin{aligned} P(|\hat{\mathscr{M}}| \leq O(n^{\xi+\kappa})) &\geq P(\max_{1 \leq l \leq p} |\widehat{VUS}_l - VUS_l| \leq \frac{1}{2}c_6 n^{-\kappa}) \\ &\geq 1 - pP(|\widehat{VUS}_l - VUS_l| \geq \frac{1}{2}c_6 n^{-\kappa}). \end{aligned}$$

The detailed proofs of Theorems 1 and 2 can be found in the Supplementary Material. We assume the sparsity of true active covariates; thus, Theorem 1 shows the sure screening property of the VUS method. Theorem 2 demonstrates that the

size of selected important set $\hat{\mathcal{M}}$ can be controlled when $p = o(\exp(n^{1-2\kappa}))$ and $\sum_{l=1}^{p} |VUS_l - 1/6| = O(n^{\xi})$.

4 Simulations

We evaluate the finite sample performance of the VUS-based screening under different scenarios. The VUS is compared with three existing methods, PSIS (Zhao and Li 2012), Kendall's τ (Song et al. 2014) and crCRS (Chen et al. 2022). crCRS can screen variables that are related to each specific cause of events under the competing risk setting. Therefore, for the crCRS method, we show variables selected for each cause of event separately, and we also consider a combination of variables selected for each cause to obtain the set of variables that are related to multiple events. For the PSIS method originally designed for survival outcomes using Cox model, we make adaptions to competing risk setting. For each cause of event, we combine subjects who had competing events and those who were censored, and treat them as independent censoring. Then we fit a cause-specific hazard model for each cause of event based on each single biomarker, and look for important biomarkers associated with each cause-specific hazard. Again, we show variables selected for each cause of event, and also the union of important biomarkers from the two events. Kendall's τ method can only handle typical survival outcomes with independent censoring, and cause-1 and cause-2 events are typically not independent given covariates. Thus, subjects who have experienced either event 1 or event 2 are combined together as the overall event group in implementing this method to satisfy the independence assumption. By comparing VUS with existing methods or their adaptions, we would like to investigate different behaviors of possible ways of variable screening under competing risk settings.

We considered the following four scenarios. Under each scenario, we simulated 200 datasets with the number of subjects $n = 200$ and the number of covariates $p = 5000$. Covariates $\mathbf{Z} = (Z_1, \ldots, Z_p)'$ were generated from a multivariate normal distribution with mean 0 and correlation $0.5^{|i-j|}$ between Z_i and Z_j. For each scenario, there are four true covariates Z_1, Z_2, Z_3, Z_4. Censoring times under each scenario were generated from uniform distributions to achieve 20% or 40% censoring.

Scenario 1: Latent event times were generated from log-logistic models

$$\log(T_j) = \beta_j' \mathbf{Z} + \sigma e, \quad j = 1, 2,$$

with $\sigma = 0.2$, e following standard logistic distribution, $\beta_1' = (1, 0.9, 0.8, 0.5, 0, \ldots, 0)$ for the cause-1 event and $\beta_2' = (0.5, 0.3, 0.2, 0.1, 0, \cdots, 0)$ for the cause-2 event. If $T_1 < T_2$, the time to the first event T was set as T_1, and the event indicator ϵ was set as 1; otherwise, the first event time was set as T_2 with ϵ being 2. As only the time to the first event is recorded, T_1 and T_2 cannot be observed simultaneously, and they are thus referred to as latent event times. The VUS was estimated at $t_0 = 1.5$.

Scenario 2: Event times were generated from a proportional subdistribution hazard model (Fine and Gray 1999) with the cumulative incidence function (CIF) for cause-1 being

$$F_1(t|\mathbf{Y}) = 1 - \left[1 - 0.8\left\{1 - \exp\left(-(t/20)^5\right)\right\}\right]^{\exp(\beta_1'\mathbf{Y})}.$$

The conditional distribution for T given the cause-2 event occurring first was set as

$$P(T \le t|\epsilon = 2, \mathbf{Y}) = 1 - \exp\left(-\exp(\beta_2'\mathbf{Y})(t/20)^5\right).$$

$\mathbf{Y}$ was a long vector containing all discretized biomarkers. Each observed continuous biomarker Z_k was discretized into three categories and denoted by three dummy variables: $Y_{k1} = I(Z_k < -0.5)$, $Y_{k2} = I(-0.5 \le Z_k < 0.5)$, and $Y_{k3} = I(Z_k \ge 0.5)$. The associated coefficients for the three corresponding Ys of Z_1, Z_2, Z_3, Z_4 were $\beta_{1k}' = (\log 3, \log 1/3, \log 1/6)$ for cause-1 and $\beta_{2k}' = (\log 9, 0, \log 1/2)$ for cause-2, $k = 1, 2, 3, 4$, and zero for all other Ys. The VUS was estimated at $t_0 = 17$.

Scenario 3: Event times were generated from Gerds' multinomial logistic regression model (Gerds et al. 2012) with the cause-1 CIF

$$F_1(t|\mathbf{Y}) = \frac{\exp(a_1 t + b_1 + \beta_1'\mathbf{Y})}{\exp(a_1 t + b_1 + \beta_1'\mathbf{Y}) + \exp(a_2 t + b_2 + \beta_2'\mathbf{Y}) + 1}$$

and the cause-2 CIF

$$F_2(t|\mathbf{Y}) = \frac{\exp(a_2 t + b_2 + \beta_2'\mathbf{Y})}{\exp(a_1 t + b_1 + \beta_1'\mathbf{Y}) + \exp(a_2 t + b_2 + \beta_2'\mathbf{Y}) + 1},$$

where $a_1 = a_2 = 2$ and $b_1 = b_2 = -15$. $\mathbf{Y}$ was the same as in scenario 2, and the associated coefficients for $\mathbf{Y}$ were $\beta_{1k}' = (\log 0.9, \log 0.1, \log 0.05)$ and $\beta_{2k}' = (\log 0.1, \log 0.9, \log 0.45)$, $k = 1, 2, 3, 4$ for Z_1, Z_2, Z_3, Z_4 and zero otherwise. The cause indicator was generated from a Bernoulli distribution with probability $F_1/(F_1 + F_2)$ where F_1 and F_2 were calculated at the simulated event time for each subject. The VUS was estimated at $t_0 = 9$.

Scenario 4: Event times were generated from the cause-specific Cox proportional hazard model with constant hazard rates for both events. We first created a risk score $\boldsymbol{\beta}'\mathbf{Z}$ with $\boldsymbol{\beta} = (3, 2.5, 2, 1.5, 0, \ldots, 0)'$. Two cut points were selected with $c_1 = -2.5$ and $c_2 = 3$. For subjects whose risk scores were below c_1, the hazard rates were 0.9 for cause-1 event and 0.1 for cause-2 event. For subjects whose risk scores were between c_1 and c_2, the hazard rates were 0.1 and 0.9 for event 1 and event 2, respectively. For subjects having risk scores above c_2, the hazard rates were 0.05 and 0.45. We estimated VUS at $t_0 = 1$.

For VUS and Kendall's τ, we summarize how many true variables can be captured on average when we select 8, 20, 40, 60, 80 variables. For the PSIS and crCRS methods, first, for each cause of events we show the number of true active variables captured when 8, 20, 40, 60, 80 variables are selected for that particular cause (PSIS1, PSIS2, crCRS1 and crCRS2 in tables). Then we also estimate a set of important covariates that may provide overall predictive accuracy by selecting 4, 10, 20, 30, 40 variables for each cause of events and then taking the union (PSIS and crCRS in tables). Following the investigation in Song et al. (2014), under each

setting, we also examine the performance of four metrics when observed covariates $\mathbf{Z}$ are contaminated; with a probability of 0.1, each covariate could be contaminated by a t distribution with mean 0 and 1 degree of freedom.

Consistently shown in Tables 1, 2, 3 and 4, generally, the number of true variables captured by VUS decreases when we have a higher rate of censoring, as do Kendall's τ, PSIS and crCRS, except that under the Gerds model when very few true variables can be captured by Kendall's τ and crCRS, a lower rate of censoring may not indicate a better performance of these two methods.

Regarding the average number of true active variables captured, overall VUS has comparable performance to the other three under latent log-logistic and Fine and Gray models and performs much better than Kendall's τ and crCRS under Gerds and Cox models. When data are not contaminated, PSIS1 and the union of PSIS perform consistently better than the VUS method. Yet the PSIS method doesn't work well when trying to capture important features for cause-2 events, and the good performance of the union of PSIS mainly relies on PSIS1. A possible explanation is that, when fitting cause-2 Cox models, the cause-1 event is treated as independent censoring, as what is normally done for cause-specific Cox models. However, under our assumption, the cause-1 event is the most "severe" event, and subjects in this group tend to have the smallest values of a biomarker. When mixing up the cause-1 group and the event-free group, for the latter of which subjects tend to have the largest values of the biomarker, it becomes much harder for the Cox model to distinguish the cause-2 group from the combination of the other two groups based on the ranking of the biomarker. Therefore, the PSIS method may not provide a thorough evaluation of the overall predictive accuracy across all events under the ordinal competing risks setting.

Kendall's τ method completely fails under the Gerds multinomial regression model. It is because in this model, after collapsing cause-1 and cause-2 events for Kendall's τ, the overall risk is the same for the two categories $Z < -0.5$ and $-0.5 \leq Z < 0.5$, although the overall risk for subjects with large covariates is small. Due to the inability to distinguish these two categories, Kendall's τ isn't able to fully capture the relationship between the covariate and the ordinal outcome, while VUS still works. For the crCRS method, we observe that true variables can rarely be captured under the Gerds and the Cox model. The crCRS measures the linear correlation between the covariate and the cumulative incidence function, and fails due to the discretization of covariates and risk scores, and also the form of cumulative incidence functions under these two models.

With data contamination, all methods are negatively affected. The PSIS and the crCRS method perform poorly under all four models. In comparison, VUS is more robust to contamination across all four models. Kendall's τ works well under the log-logistic and Fine-Gray models, but becomes worse than PSIS under Gerds' and Cox' models.

In summary, compared to possible adaptions of existing screening methods to competing-risk outcomes and the crCRS method, VUS provides a different aspect of the association between biomarkers and the competing events, as it measures the overall ability of the biomarkers to distinguish subjects from

Table 1 Number of true variables captured under latent log-logistic models

	Non-contaminated								Contaminated							
	20%				40%				20%				40%			
Size	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS
8	3.920	4	4	3.985	3.740	3.990	4	3.920	3.845	3.980	2.420	2.280	3.525	3.965	2.295	1.860
20	3.970	4	4	3.995	3.860	3.995	4	3.975	3.900	3.990	2.615	2.465	3.670	3.995	2.510	2.110
40	3.980	4	4	4	3.885	4	4	3.985	3.940	3.990	2.690	2.615	3.780	4	2.610	2.280
60	3.985	4	4	4	3.890	4	4	3.990	3.955	3.990	2.770	2.670	3.795	4	2.685	2.355
80	3.995	4	4	4	3.905	4	4	3.995	3.965	3.990	2.815	2.710	3.815	4	2.780	2.395
Size	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2
8	4	2.075	3.995	3.965	4	1.770	3.975	3.635	2.490	0.755	2.325	2.055	2.395	0.470	2.030	0.870
20	4	2.230	3.995	3.990	4	1.960	3.985	3.770	2.585	0.855	2.520	2.260	2.495	0.610	2.255	1.075
40	4	2.340	4	3.995	4	2.080	3.995	3.825	2.690	0.970	2.605	2.370	2.635	0.705	2.375	1.285
60	4	2.405	4	3.995	4	2.160	3.995	3.855	2.755	1.050	2.660	2.430	2.755	0.835	2.455	1.425
80	4	2.455	4	3.995	4	2.225	3.995	3.875	2.825	1.125	2.740	2.475	2.780	0.910	2.525	1.570

Table 2 Number of true variables captured under the Fine-Gray model

	Non-contaminated								Contaminated							
	20%				40%				20%				40%			
Size	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS
8	3.990	3.980	3.995	3.685	3.960	3.955	3.990	2.720	3.905	3.940	2.155	1.170	3.825	3.765	2.070	0.670
20	3.995	3.995	4	3.885	3.970	3.985	4	3.290	3.970	3.980	2.375	1.475	3.915	3.880	2.320	0.860
40	4	4	4	3.925	3.995	3.990	4	3.555	3.985	3.980	2.500	1.640	3.935	3.935	2.445	1.050
60	4	4	4	3.955	4	3.990	4	3.546	3.985	3.980	2.555	1.745	3.965	3.955	2.525	1.135
80	4	4	4	3.960	4	3.990	4	3.675	3.990	3.990	2.620	1.855	3.975	3.970	2.575	1.235
Size	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2
8	4	0.385	3.735	3.860	4	0.230	2.725	3.170	2.345	0.035	1.210	1.370	2.275	0.025	0.660	0.775
20	4	0.655	3.870	3.925	4	0.405	3.155	3.545	2.485	0.060	1.480	1.610	2.435	0.040	0.905	1.000
40	4	0.920	3.920	3.960	4	0.645	3.425	3.670	2.590	0.110	1.680	1.820	2.540	0.090	1.070	1.185
60	4	1.095	3.950	3.975	4	0.810	3.600	3.770	2.665	0.150	1.810	1.920	2.630	0.120	1.205	1.310
80	4	1.260	3.960	3.980	4	0.955	3.660	3.815	2.720	0.225	1.945	2.010	2.695	0.165	1.310	1.410

Table 3 Number of true variables captured under Gerds model

	Non-contaminated								Contaminated							
	20%				40%				20%				40%			
Size	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS
8	2.930	0.010	3.700	0.000	2.560	0.005	3.670	0.000	2.190	0.005	1.180	0.000	1.790	0.005	1.125	0.000
20	3.425	0.020	3.980	0.010	3.090	0.010	3.890	0.010	2.705	0.010	1.585	0.010	2.450	0.015	1.345	0.005
40	3.640	0.045	3.990	0.025	3.335	0.025	3.955	0.040	3.015	0.025	1.735	0.035	2.800	0.025	1.515	0.025
60	3.765	0.065	3.995	0.030	3.550	0.050	3.965	0.055	3.190	0.050	1.840	0.050	3.005	0.045	1.670	0.035
80	3.785	0.085	3.995	0.040	3.565	0.075	3.975	0.065	3.320	0.055	1.950	0.055	3.120	0.080	1.745	0.055
Size	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2
8	3.880	2.170	0.000	0.005	3.815	1.655	0.005	0.005	1.340	0.325	0.000	0.000	1.115	0.230	0.005	0.000
20	3.960	2.655	0.025	0.005	3.915	2.115	0.030	0.020	1.530	0.450	0.030	0.015	1.365	0.315	0.020	0.010
40	3.970	2.945	0.035	0.025	3.980	2.565	0.045	0.040	1.745	0.635	0.045	0.035	1.580	0.435	0.040	0.030
60	3.975	3.120	0.045	0.055	3.985	2.780	0.075	0.080	1.935	0.775	0.060	0.050	1.730	0.580	0.055	0.040
80	3.975	3.245	0.065	0.080	3.990	2.960	0.085	0.095	2.045	0.845	0.085	0.065	1.920	0.655	0.075	0.055

Table 4 Number of true variables captured under Cox model

	Non-contaminated								Contaminated							
	20%				40%				20%				40%			
Size	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS	VUS	τ	PSIS	crCRS
8	3.675	2.005	3.930	0.275	3.560	1.215	3.805	0.130	3.515	1.725	2.255	0.070	3.385	1.045	2.005	0.040
20	3.890	2.725	4	0.400	3.855	1.940	3.965	0.245	3.805	2.555	2.855	0.110	3.705	1.735	2.780	0.075
40	3.945	3.340	4	0.595	3.920	2.745	3.995	0.350	3.930	3.120	3.250	0.140	3.885	2.445	3.135	0.100
60	3.970	3.635	4	0.690	3.950	3.260	3.995	0.415	3.955	3.535	3.440	0.155	3.940	2.985	3.415	0.140
80	3.985	3.825	4	0.815	3.980	3.645	4	0.465	3.980	3.765	3.575	0.235	3.970	3.560	3.580	0.165
Size	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2	PSIS1	PSIS2	crCRS1	crCRS2
8	3.995	1.765	0.265	0.375	3.945	1.585	0.130	0.205	2.555	0.680	0.075	0.090	2.405	0.705	0.040	0.060
20	4	2.510	0.440	0.575	3.995	2.540	0.245	0.340	3.015	1.435	0.125	0.135	2.890	1.375	0.085	0.085
40	4	3.170	0.595	0.805	3.995	3.235	0.355	0.460	3.315	2.110	0.190	0.210	3.255	2.115	0.135	0.150
60	4	3.575	0.740	0.985	4	3.550	0.455	0.605	3.545	2.820	0.235	0.260	3.510	2.765	0.170	0.200
80	4	3.780	0.885	1.135	4	3.790	0.535	0.725	3.740	3.430	0.280	0.335	3.705	3.390	0.210	0.240

multiple groups simultaneously, which cannot be fully captured by existing screening methods, especially when distinguishing event-1 and event-2 groups.

5 Data analysis

We applied our proposed screening method to a gene-expression dataset from a breast cancer study (van de Vijver et al. 2002). This dataset is obtained from the R package "cancerdata" (Budczies and Kosztyla 2021) and contains 295 women with breast cancer and expression values of 24,481 genes in tumor samples of each woman. Two events of interest are distant metastasis and death. Among 295 patients, five (1.7%) patients died without metastasis, 101 (34.2%) experienced metastasis, and 74 (25.1%) died after metastasis. The overall censoring rate, i.e, patients survived without metastasis, is 64.1%. The objective of our analysis is to capture the genes that are associated with the progression from breast cancer to distant metastasis and/or death.

For our simulations, we've tried different t_0s (not shown here), and found that for the VUS, what really matters is the number of subjects falling in each disease category at the time of prediction. Therefore, t_0 needs to be chosen carefully to guarantee enough samples in each category, which can be easily examined. In this data example, we looked at patients' survival at $t_0 = 5$ years and considered the most severe event each patient experienced before t_0. Particularly, death was treated as cause-1 event and metastasis was treated as cause-2 event. At five years, 48 patients died either with or without metastasis, 32 patients were alive but with metastasis, 207 patients were alive and metastasis-free, and 8 patients were censored before t_0.

Different from simulation studies, in real data analysis, we are not sure how covariates are associated with the outcomes. As mentioned in Sect. 2, for covariates whose larger values are associated with more severe conditions, we can reverse the relationship. In this dataset, it is not clear which genes violate our assumption. Therefore, for each gene, we calculated two VUS estimates, one assuming that lower values are associated with more severe states, and the other assuming that the relationship is in the opposite direction. For the latter, we used two minus the observed value of gene expression as values of the covariates to estimate VUS, where two is the maximum value of all covariates. We kept the VUS estimate that was further away from 1/6, which could better reflect the ordinal association between the variable and event groups, as illustrated by simulation results in Supplementary Material.

Following Fan and Lv (2008), we selected $[n/\log(n)] = 51$ important variables for VUS and Kendall's τ. For the PSIS and the crCRS method, 26 variables were selected for each type of event, and the final important set contained the union of important variables of two events.

We show the top 51 genes selected by VUS in Table 5. Genes that are selected by both VUS and Kendall's τ are denoted by *, genes selected by both VUS and PSIS are denoted by &, and those selected by VUS and crCRS are denoted by #. Bold-faced genes are selected by all four methods, which include five genes. The same dataset was also analyzed in Song et al. (2014) and Lu et al. (2020). In Song

Table 5 Top 51 genes selected by VUS

U96131 &	NM_005480 *#	Contig29555_RC	**Contig31288_RC**
M96577	NM_003494	NM_003295	NM_000987
NM_004219	**NM_005733**	NM_004701 *#	Contig57173_RC
NM_001605 &#	NM_019059	NM_004217 #	NM_002466 *
Contig57584_RC * &	NM_007019 *#	**NM_006607**	NM_006579 *
NM_001809 #	NM_006845	Contig6498	D38553 *#
Contig38288_RC * &	NM_005804	NM_006623 &	D43950 &
Contig35629_RC	NM_018188 *	NM_001673	**D14678**
Contig41828_RC	NM_018410 *#	NM_002624	NM_001255
NM_020313	NM_014454	NM_018688	Contig56390_RC *#
Contig38901_RC	NM_003504	**NM_001333**	NM_007195
NM_019597 &	NM_017761 *	NM_018834	NM_003600 *#
Contig41977_RC	Contig43747_RC	NM_001168 *#	

Bold-faced genes are selected by all four metrics; '*' are selected by VUS and Kendall's τ; ' &' are selected by VUS and PSIS; '#' are selected by VUS and crCRS

et al. (2014), the event of interest was the overall survival time, and only the top 20 selected genes were shown. We compared the results and found that 13 genes were both selected by our VUS method and by Song et al. (2014).

Lu et al. (2020) treated the data as semi-competing risks outcomes. While explicitly handling semi-competing risks is beyond the scope of this work, the proposed VUS is still applicable by counting the most severe event that occurred. The authors selected 51 genes using the proposed method in Lu et al. (2020) and improved the results based on an adaptive threshold rule, for which 25 genes were selected. They showed that the adaptive threshold rule would perform better than the proposed method itself, so we compared the selected genes with their 25 selected genes. Among the 25 selected genes, 13 genes were selected by both our VUS method and Lu et al. (2020), and these 13 genes were partially different from those selected by VUS and Kenall's τ in Song et al. (2014). These comparisons imply that there is no one-size-fits-all metric, and our VUS method is a viable variable screening metric for competing risk outcomes as it focuses on a different aspect of the association between the covariates and the outcome than existing methods.

6 Discussion

In this paper, we have shown that VUS possesses the sure screening property. The VUS can provide an overall assessment of diagnostic accuracy of covariates in predicting ordinal outcomes and has a straightforward interpretation as the concordance probability between the value of covariates and the disease status. Simulation studies and data analysis have shown that compared to cause-specific screening methods, VUS can capture additional aspect of the association between biomarkers

and multiple competing events, therefore it can serve as an alternative for variable screening, especially with data contamination in covariates.

A limitation of VUS is that it is designed to pick covariates that have ordinal relationships with the outcomes, which means that as the value of the biomarker increase or decrease, the event status would change, and thus can be ordered by the values of biomarker. For convenience in this paper we have assumed that as the value of a covariate decreases, the disease status would evolve from the healthy condition to event-2 and then to the most severe case (cause-1 event). For each single marker, we can visually examine the ordinal relationship by checking the weighted density plot by event status, as what have been done in Zhang et al. (2022). The problem is that for variable screening, we have tens of thousands of variables. In practice, it may be more feasible to relax the ordinal assumption.

One possible solution is that instead of measuring the concordance between the value of a biomarker itself and the disease status, we may look at the concordance between some functions based on each single biomarker and the true event status, for example, the estimated CIF based on each single Z. Similarly, we may evaluate how a group of biomarkers can be associated with the competing risk outcomes by modeling CIF based on this group of biomarkers to handle correlated biomarkers and categorical biomarkers. However, this solution will rely on additional model assumptions of CIF.

Another possible solution to relax the ordinal assumption lies in the definition of VUS. VUS actually measures the probability of concordance between the predicted class membership and the true class membership, and subjects are classified based on the value of their covariates. The classification rule is based on our knowledge of the relationship between the covariates and the event status. For three subjects i, j and k randomly selected from the cause-1 group, cause-2 group and event-free group respectively, there are six possible ways of ranking their covariates, indicating six possible ways of general ordinal relationship between the biomarker and the competing events. Based on each possible relationship, we can develop a classification rule, and define the VUS accordingly, which would result in six different VUS values in total, and each VUS would be able to pick covariates that satisfy a particular relationship. When we are not sure how a covariate is associated with the competing events, we can estimate six different VUSs, and keep the one furthest away from 1/6, which can better capture the association between the covariates and the events. However, now the screening metric is the maximum of six VUS values that are based on different ordinal assumptions, thus the statistical property may be different from a single VUS metric. The detailed investigation is beyond the scope of this paper and warrants future research.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10985-023-09600-z.

Acknowledgements This research was supported in part by the National Science Foundation Division of Mathematical Sciences (1916001 to Cheng) and by the University of Pittsburgh Center for Research Computing through the resources provided. Specifically, this work used the H2P cluster, which is supported by NSF award number OAC-2117681.

References

Budczies J, Kosztyla D (2021) cancerdata: development and validation of diagnostic tests from high-dimensional molecular data: datasets. R package version 1.30.0

Chen X, Li C, Zhang T, Gao Z (2022) On correlation rank screening for ultra-high dimensional competing risks data. J Appl Stat 49(7):1848–1864. https://doi.org/10.1080/02664763.2021.1884209

Fan J, Lv J (2008) Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc Ser B (Stat Methodol) 70(5):849–911

Fan J, Feng Y, Wu Y (2010) High-dimensional variable selection for Cox's proportional hazards model, Collections, vol 6. Institute of Mathematical Statistics, Beachwood, pp 70–86

Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94(446):496–509

Gerds TA, Scheike TH, Andersen PK (2012) Absolute risk regression for competing risks: interpretation, link functions, and prediction. Stat Med 31(29):3921–3930

Gorst-Rasmussen A, Scheike T (2013) Independent screening for single-index hazard rate models with ultrahigh dimensional features. J R Stat Soc Ser B (Stat Methodol) 75(2):217–245

Hong HG, Chen X, Christiani DC, Li Y (2018) Integrated powered density: screening ultrahigh dimensional covariates with survival outcomes. Biometrics 74(2):421–429

Li J, Fine JP (2008) ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. Biostatistics 9(3):566–576

Li J, Zheng Q, Peng L, Huang Z (2016) Survival impact index and ultrahigh-dimensional model-free screening with survival outcomes. Biometrics 72(4):1145–1154

Li E, Mei B, Tian M (2018) Feature screening based on ultrahigh dimensional competing risks models. Sci Sin Math 48(8):1061. https://doi.org/10.1360/N012017-00069

Liu Y, Chen X, Wang H (2021) The fused Kolmogorov–Smirnov screening for ultra-high dimensional semi-competing risks data. Appl Math Model 98:109–120. https://doi.org/10.1016/j.apm.2021.04.031

Lu S, Chen X, Xu S, Liu C (2020) Joint model-free feature screening for ultra-high dimensional semi-competing risks data. Comput Stat Data Anal 147:106942

Mossman D (1999) Three-way rocs. Med Decis Mak 19(1):78–89

Nakas CT, Yiannoutsos CT (2004) Ordered multiple-class ROC analysis with continuous measurements. Stat Med 23(22):3437–3449

Pan J, Yu Y, Zhou Y (2018) Nonparametric independence feature screening for ultrahigh-dimensional survival data. Metrika 81(7):821–847

Peng M, Xiang L (2021) Correlation-based joint feature screening for semi-competing risks outcomes with application to breast cancer data. Stat Methods Med Res 30(11):2428–2446. https://doi.org/10.1177/09622802211037071

Song R, Lu W, Ma S, Jessie Jeng X (2014) Censored rank independence screening for high-dimensional survival data. Biometrika 101(4):799–814

Tian B, Liu Z, Wang H (2022) Non-marginal feature screening for varying coefficient competing risks model. Stat Probab Lett 190:109648. https://doi.org/10.1016/j.spl.2022.109648

Tibshirani RJ (2009) Univariate shrinkage in the cox model for high dimensional data. Stat Appl Genet Mol Biol 8(1):21

van de Vijver MJ, He YD, Van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R (2002) A gene-expression signature as a predictor of survival in breast cancer. New Engl J Med 347(25):1999–2009

Wang H, Shen Z, Tan Z, Zhang Z, Li G (2022) Fast lasso-type safe screening for fine-gray competing risks model with ultrahigh dimensional covariates. Stat Med 41(24):4941–4960. https://doi.org/10.1002/sim.9545

Zhang S, Qu Y, Cheng Y, Lopez OL, Wahed AS (2022) Prognostic accuracy for predicting ordinal competing risk outcomes using ROC surfaces. Lifetime Data Anal 28(1):1–22. https://doi.org/10.1007/s10985-021-09539-z

Zhao SD, Li Y (2012) Principled sure independence screening for cox models with ultra-high-dimensional covariates. J Multivar Anal 105(1):397–411. https://doi.org/10.1016/j.jmva.2011.08.002