Supplementary Materials for "Leveraging two-phase data for improved prediction of survival outcomes with application to nasopharyngeal cancer" by Eun Jeong Oh, Seungjun Ahn, Tristan Tham, and Min Qian

Web Appendix A

The fact that a small subset of patients are being tested for expensive biomarkers may suggest that there is a systematic reason for missing data on those biomarkers. Possible explanations could be that only a subset of patients are eligible or willing to participate, or certain characteristics (e.g., age, cancer stage) may increase the likelihood of undergoing the test. If the missingness is related to these observed characteristics but not to the missing values themselves, the missing data would be classified as Missing at Random (MAR). For instance, in the NPC study, patients nearing the end of their lives at terminal stages typically do not undergo HPV testing, as such tests would not improve their condition. Therefore, it is reasonable to assume MAR, as the missingness is related to an observable characteristic (cancer stage). Missing Not at Random (MNAR) may occur when the missing data is related to the missing values themselves. However, identifying MNAR is challenging because it depends on unobserved information, making it difficult to infer without deep insights or auxiliary data.

As part of a cost-effective two-phase sampling design, if a sub-sample of patients is randomly selected to undergo additional testing, this would represent Missing Completely at Random (MCAR). In this case, the missing data on the additional variables is assumed to be unrelated to both observed factors and the values that would have been observed had the data not been missing, as the selection for further testing is purely random and not influenced by any specific characteristics of the patients. Conversely, if the selection is impacted by observed data but not by the missing values themselves, the missing data would be classified as MAR. An example of this occurs when patients with insurance coverage are more likely to undergo further biomarker testing, such that the missingness of biomarker data is dependent on the insurance status (an observed variable) but not on the missing biomarker values themselves.

Web Appendix B

Here we describe the complete process for generating two phase data under different missing mechanisms as described in Section 3.1. We consider three settings for the missing mechanism: (i) missing completely at random (MCAR) where the probability of $\boldsymbol{V}$ being missing is independent of all covariates; (ii) missing at random (MAR) where the probability of $\boldsymbol{V}$ being missing depends on the observed data; and (iii) MAR with a mild-to-moderate violation, where the probability of $\boldsymbol{V}$ being missing remains dependent on the observed data, but is slightly modified based on the missing data. For each desired ratio $r \approx n'/n$, the probability of $\boldsymbol{V}$ being missing is (i) $1-r$ for MCAR; (ii) $(1-r)/\Phi(-\Phi^{-1}(r/3))$ if $U_1 > \Phi^{-1}(r/3)$ and 0 otherwise for MAR; and (iii) $(1-r-0.1I(\boldsymbol{V} \leq 0))/\Phi(-\Phi^{-1}(r/3)) + 0.1I(\boldsymbol{V} > 0)$ if $U_1 > \Phi^{-1}(r/3)$ and 0.1 otherwise for MAR with a mild-to-moderate violation, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. This ensures approximately $n' = r \cdot n$ across all three settings.

Web Appendix C

Hereby we differentiate discrimination, calibration, and overall performance of a survival model.

Discrimination. A model should be able to accurately discriminate different risk categories. Discrimination indicates how well a model can distinguish between patients who will die earlier and those who would die later. The Harrell's c-index (Harrell et al., 1982, 1984, 1996), also known as concordance index or c-statistic, is a commonly used measure to evaluate a discriminate ability of survival models. The closer the c-index is to 1, the better the model discriminates between low-risk and high-risk patients. A value of 0.5 indicates that the model is no better out predicting outcomes than random chance.

Calibration. The calibration of a model is a measure of an agreement between the observed and predicted outcomes. The commonly used calibration metric is calibration slope proposed by Van Houwelingen (2000). A model that calibrates well would result in a calibration slope close to value 1. An overfitted model would have a slope >1, whereas an underfitted model would have a slope <1. Overfitting is more frequently observed, while underfitting occurs when a model is excessively simple.

Overall performance. With regards to measuring overall performance, the commonly used metric is the Brier score proposed by Brier (1950). It incorporates both discrimination and calibration aspects of a model, taking values between 0 and 1. The Brier score is similar to the mean squared error in linear regressions. The integrated Brier score (IBS), introduced by Graf et al. (1999), integrates multiple scores obtained at all follow-up times. A score closer to 0 implies a better predictive performance.

The predictive model performance is mainly assessed using the c-index, calibration slope, and IBS. Variable selection performance on $(\boldsymbol{\beta}_U, \boldsymbol{\beta}_V)$ is evaluated by the Matthews correlation coefficient (MCC) proposed by Matthews (1975), defined as

$$\mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP}) \cdot (\mathrm{TP}+\mathrm{FN}) \cdot (\mathrm{TN}+\mathrm{FP}) \cdot (\mathrm{TN}+\mathrm{FN})}},$$

where TP, TN, FP, and FN are true negatives, true negatives, false negatives, and false positives, respectively.

Web Appendix D

In this section, we repeat the analysis in Section 3.1 by applying domain knowledge to the comparison methods as well. When the comparison methods were also applied with domain knowledge in the variable selection process, they demonstrated some benefits in general, as indicated by an increased c-index, lower IBS, improved calibration, and higher MCC (Web Tables 1–3). However, despite noticeable improvements in calibration for the competing methods, the standard deviation of the calibration slope by the EG method was relatively lower than that of the other methods, suggesting that our method provided a more consistent estimate approaching the ideal value of 1. Overall, our proposed EG method still outperformed the alternatives in most cases.

Additionally, it is worth noting that the MCC for the comparison methods decreased in Scenario III. This was expected, as the first two variables of $\boldsymbol{U}$ were included in the model based on domain knowledge, which is suboptimal when U_2 had no effect. Thus, in this context, where all methods incorporated partly misaligned or incorrect domain knowledge, our proposed method, which initially had a lower MCC, turned out perform better in terms of variable selection compared to the other methods. Furthermore, it maintained good and consistent performance across various missing mechanisms in terms of c-index, calibration, and IBS, highlighting the robustness and reliability of the proposed method.

Web Table 1: Simulation results under the MCAR setting when domain knowledge is also applied to the comparison methods. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.69 (0.11)	0.81 (0.72)	2.19 (0.71)	0.58 (0.07)
	NI	0.69 (0.12)	1.53 (3.60)	2.08 (0.61)	0.50 (0.12)
	MI-Wood	0.67 (0.11)	1.50 (3.61)	2.13 (0.57)	0.47 (0.07)
	MI-Bartlett	0.68 (0.11)	1.26 (1.39)	2.13 (0.58)	0.51 (0.14)
	EG	**0.71** (0.11)	**0.94** (0.86)	**2.07** (0.67)	**0.62** (0.11)
II	CCA	0.80 (0.10)	0.83 (0.52)	1.85 (1.01)	0.82 (0.07)
	NI	**0.82** (0.08)	1.15 (0.44)	**1.69** (0.76)	0.71 (0.13)
	MI-Wood	0.81 (0.08)	1.14 (0.47)	1.71 (0.77)	0.67 (0.09)
	MI-Bartlett	0.81 (0.08)	1.15 (0.47)	1.71 (0.77)	0.69 (0.11)
	EG	**0.82** (0.09)	**0.96** (0.54)	1.74 (0.98)	**0.85** (0.09)
III	CCA	0.69 (0.11)	0.67 (0.71)	2.14 (0.82)	0.43 (0.12)
	NI	0.70 (0.12)	1.18 (1.02)	**1.99** (0.62)	0.38 (0.24)
	MI-Wood	0.68 (0.12)	**1.10** (0.94)	2.04 (0.57)	0.34 (0.21)
	MI-Bartlett	0.70 (0.11)	1.26 (0.93)	2.00 (0.60)	0.45 (0.22)
	EG	**0.72** (0.12)	0.84 (0.65)	2.01 (0.78)	**0.56** (0.18)

Web Appendix E

To understand the impact of the sample size ratio in two-phase data, we consider additional simulations to further illustrate the impact of the sample size ratio in two-phase data, i.e., n'/n. Due to the characteristics of two-phase data, n' is expected to be low, as it typically involves expensive biomarkers. To investigate this, we increase the sample size to $n = 1,000$, allowing n' to vary within a reasonable range that ensures both sufficient number of samples and number of events in each fold of cross-validation, where $n'/n \approx \{0.10, 0.15, 0.25\} = r$. When $r = 0.10$, our EG method still performed the best (Web Table 4). The benefit of our proposed method as compared to the

Web Table 2: Simulation results under the MAR setting when domain knowledge is also applied to the comparison methods. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.70 (0.11)	0.77 (0.55)	2.01 (0.90)	0.57 (0.06)
	NI	0.68 (0.12)	1.46 (3.77)	1.99 (0.69)	0.51 (0.13)
	MI-Wood	0.67 (0.12)	1.47 (3.76)	2.00 (0.68)	0.48 (0.10)
	MI-Bartlett	0.67 (0.12)	1.14 (1.03)	2.00 (0.68)	0.52 (0.14)
	EG	**0.71** (0.12)	**0.88** (0.75)	**1.94** (0.84)	**0.63** (0.12)
II	CCA	0.84 (0.08)	1.14 (2.23)	1.45 (0.69)	0.82 (0.07)
	NI	**0.85** (0.07)	1.22 (0.63)	1.40 (0.63)	0.69 (0.13)
	MI-Wood	0.84 (0.08)	1.26 (0.87)	1.43 (0.68)	0.67 (0.08)
	MI-Bartlett	0.84 (0.08)	1.27 (0.88)	1.42 (0.68)	0.68 (0.11)
	EG	**0.85** (0.08)	**1.05** (0.76)	**1.36** (0.64)	**0.85** (0.09)
III	CCA	0.71 (0.13)	0.68 (0.68)	1.94 (0.88)	0.41 (0.10)
	NI	0.72 (0.13)	1.20 (1.13)	**1.78** (0.72)	0.39 (0.23)
	MI-Wood	0.71 (0.14)	1.21 (1.13)	1.83 (0.71)	0.34 (0.20)
	MI-Bartlett	0.72 (0.13)	1.27 (1.16)	1.78 (0.70)	0.44 (0.23)
	EG	**0.73** (0.13)	**0.81** (0.70)	1.81 (0.84)	**0.55** (0.18)

Web Table 3: Simulation results under the MAR setting with a mild-to-moderate violation when domain knowledge is also applied to the comparison methods. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.70 (0.12)	0.79 (0.71)	2.04 (0.90)	0.56 (0.05)
	NI	0.68 (0.12)	1.43 (3.54)	1.99 (0.66)	0.49 (0.12)
	MI-Wood	0.67 (0.12)	1.42 (3.55)	1.99 (0.64)	0.47 (0.09)
	MI-Bartlett	0.67 (0.12)	1.50 (3.55)	1.97 (0.63)	0.52 (0.13)
	EG	**0.71** (0.12)	**0.91** (0.88)	**1.93** (0.80)	**0.63** (0.12)
II	CCA	0.84 (0.08)	0.96 (0.64)	1.49 (0.70)	0.82 (0.06)
	NI	**0.85** (0.07)	1.27 (0.79)	1.43 (0.67)	0.68 (0.12)
	MI-Wood	0.84 (0.08)	1.29 (0.85)	1.44 (0.68)	0.67 (0.09)
	MI-Bartlett	0.84 (0.08)	1.25 (0.66)	1.44 (0.68)	0.69 (0.11)
	EG	**0.85** (0.08)	**1.06** (0.77)	**1.40** (0.66)	**0.85** (0.09)
III	CCA	0.71 (0.13)	4.79 (41.3)	2.05 (1.00)	0.40 (0.11)
	NI	0.72 (0.13)	**1.18** (1.12)	1.87 (0.75)	0.38 (0.22)
	MI-Wood	0.71 (0.13)	1.23 (1.08)	1.90 (0.74)	0.33 (0.21)
	MI-Bartlett	0.71 (0.13)	1.23 (1.13)	1.86 (0.73)	0.44 (0.22)
	EG	**0.73** (0.14)	0.75 (0.67)	**1.90** (0.95)	**0.56** (0.18)

method that discards individuals with missing information (e.g., CCA) could be especially evident when the number of the target samples is substantially limited. As r increased to 0.15 (Web Table 5) and 0.25 (Web Table 6), the difference between our method and the competing methods gradually decreased. However, our method consistently outperformed the alternative approaches in terms of c-index, calibration slope, IBS, and MCC. Additionally, it had a calibration slope very close to 1, regardless of the value of r. Thus, the prognostic index contributes to well-calibrated risk predictions, a feature that none of the competing methods achieved.

Web Table 4: Simulation results for $r = 0.10$ under various settings for missing mechanism. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Setting	Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
MCAR	I	CCA	0.59 (0.10)	1.49 (1.29)	2.11 (0.43)	0.24 (0.26)
		NI	0.75 (0.06)	1.63 (0.57)	1.76 (0.50)	0.86 (0.10)
		MI-Wood	0.74 (0.06)	1.65 (0.58)	1.81 (0.45)	0.81 (0.05)
		MI-Bartlett	0.74 (0.07)	1.67 (0.60)	1.81 (0.45)	0.83 (0.07)
		EG	**0.77** (0.07)	**0.93** (0.36)	**1.73** (0.48)	**0.96** (0.08)
	II	CCA	0.78 (0.10)	1.52 (0.69)	1.71 (0.59)	0.67 (0.25)
		NI	0.84 (0.05)	1.45 (0.38)	1.47 (0.47)	0.82 (0.06)
		MI-Wood	0.84 (0.05)	1.44 (0.38)	1.49 (0.48)	0.81 (0.00)
		MI-Bartlett	0.84 (0.05)	1.44 (0.38)	1.49 (0.48)	0.81 (0.02)
		EG	**0.85** (0.05)	**0.95** (0.32)	**1.39** (0.46)	**0.99** (0.03)
	III	CCA	0.63 (0.12)	2.02 (2.29)	2.10 (0.54)	0.41 (0.34)
		NI	0.78 (0.06)	1.66 (0.50)	1.76 (0.54)	**0.83** (0.06)
		MI-Wood	0.78 (0.06)	1.64 (0.48)	1.77 (0.52)	0.81 (0.00)
		MI-Bartlett	0.78 (0.06)	1.64 (0.47)	1.77 (0.52)	0.81 (0.02)
		EG	**0.80** (0.06)	**1.00** (0.35)	**1.64** (0.58)	**0.83** (0.02)
MAR	I	CCA	0.59 (0.11)	1.39 (1.46)	2.14 (0.59)	0.20 (0.25)
		NI	0.76 (0.07)	1.62 (0.57)	1.71 (0.53)	0.85 (0.09)
		MI-Wood	0.75 (0.07)	1.66 (0.58)	1.76 (0.52)	0.81 (0.05)
		MI-Bartlett	0.75 (0.07)	1.67 (0.62)	1.76 (0.52)	0.82 (0.06)
		EG	**0.78** (0.06)	**0.92** (0.37)	**1.67** (0.54)	**0.96** (0.08)
	II	CCA	0.81 (0.10)	1.83 (1.99)	1.43 (0.67)	0.62 (0.25)
		NI	0.87 (0.05)	1.48 (0.38)	1.25 (0.51)	0.82 (0.03)
		MI-Wood	0.87 (0.05)	1.50 (0.41)	1.26 (0.52)	0.81 (0.00)
		MI-Bartlett	0.87 (0.05)	1.49 (0.40)	1.25 (0.51)	0.81 (0.00)
		EG	**0.88** (0.05)	**0.99** (0.38)	**1.18** (0.52)	**0.99** (0.03)
	III	CCA	0.67 (0.14)	1.62 (1.45)	1.92 (0.64)	0.45 (0.33)
		NI	0.80 (0.07)	1.68 (0.50)	1.68 (0.56)	0.82 (0.05)
		MI-Wood	0.80 (0.07)	1.66 (0.49)	1.68 (0.55)	0.81 (0.00)
		MI-Bartlett	0.80 (0.07)	1.66 (0.49)	1.68 (0.55)	0.81 (0.00)
		EG	**0.82** (0.06)	**0.96** (0.38)	**1.56** (0.62)	**0.83** (0.02)
MARviol	I	CCA	0.58 (0.11)	1.55 (1.39)	2.17 (0.57)	0.20 (0.24)
		NI	0.76 (0.06)	1.63 (0.60)	1.76 (0.54)	0.84 (0.09)
		MI-Wood	0.75 (0.06)	1.66 (0.60)	1.79 (0.53)	0.81 (0.05)
		MI-Bartlett	0.75 (0.07)	1.65 (0.59)	1.79 (0.54)	0.83 (0.06)
		EG	**0.78** (0.06)	**0.93** (0.37)	**1.69** (0.54)	**0.96** (0.08)
	II	CCA	0.78 (0.12)	1.81 (1.10)	1.55 (0.72)	0.56 (0.28)
		NI	0.87 (0.05)	1.45 (0.39)	1.30 (0.55)	0.81 (0.03)
		MI-Wood	0.87 (0.05)	1.49 (0.41)	1.30 (0.55)	0.81 (0.00)
		MI-Bartlett	0.87 (0.05)	1.48 (0.40)	1.30 (0.55)	0.81 (0.02)
		EG	**0.88** (0.05)	**0.98** (0.42)	**1.22** (0.54)	**0.99** (0.03)
	III	CCA	0.65 (0.13)	1.61 (2.09)	1.98 (0.61)	0.40 (0.33)
		NI	0.80 (0.07)	1.69 (0.53)	1.68 (0.55)	**0.83** (0.05)
		MI-Wood	0.79 (0.07)	1.64 (0.50)	1.67 (0.54)	0.81 (0.00)
		MI-Bartlett	0.79 (0.07)	1.64 (0.48)	1.67 (0.54)	0.81 (0.00)
		EG	**0.82** (0.06)	**0.99** (0.45)	**1.55** (0.60)	**0.83** (0.02)

Web Table 5: Simulation results for $r = 0.15$ under various settings for missing mechanism. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Setting	Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
MCAR	I	CCA	0.63 (0.10)	1.42 (1.29)	2.05 (0.45)	0.38 (0.29)
		NI	0.76 (0.05)	1.66 (0.47)	1.73 (0.42)	0.89 (0.10)
		MI-Wood	0.74 (0.05)	1.66 (0.49)	1.81 (0.39)	0.81 (0.05)
		MI-Bartlett	0.74 (0.05)	1.65 (0.50)	1.80 (0.39)	0.85 (0.08)
		EG	**0.77** (0.05)	**0.99** (0.29)	**1.66** (0.41)	**0.96** (0.08)
	II	CCA	0.81 (0.07)	1.48 (0.64)	1.56 (0.48)	0.79 (0.21)
		NI	0.84 (0.04)	1.41 (0.28)	1.44 (0.40)	0.83 (0.06)
		MI-Wood	0.84 (0.04)	1.42 (0.26)	1.46 (0.40)	0.81 (0.00)
		MI-Bartlett	0.84 (0.04)	1.42 (0.26)	1.46 (0.40)	0.81 (0.03)
		EG	**0.85** (0.04)	**0.98** (0.23)	**1.37** (0.41)	**0.99** (0.03)
	III	CCA	0.69 (0.09)	1.65 (1.07)	1.93 (0.43)	0.62 (0.29)
		NI	0.78 (0.05)	1.62 (0.41)	1.70 (0.37)	**0.84** (0.08)
		MI-Wood	0.78 (0.05)	1.61 (0.40)	1.72 (0.35)	0.81 (0.00)
		MI-Bartlett	0.78 (0.05)	1.60 (0.40)	1.72 (0.36)	0.82 (0.03)
		EG	**0.80** (0.05)	**0.99** (0.26)	**1.57** (0.41)	0.83 (0.02)
MAR	I	CCA	0.64 (0.11)	1.55 (1.19)	2.01 (0.46)	0.33 (0.27)
		NI	0.76 (0.06)	1.65 (0.49)	1.70 (0.41)	0.85 (0.09)
		MI-Wood	0.75 (0.06)	1.67 (0.52)	1.73 (0.39)	0.81 (0.05)
		MI-Bartlett	0.75 (0.06)	1.67 (0.51)	1.72 (0.39)	0.84 (0.08)
		EG	**0.79** (0.05)	**0.95** (0.32)	**1.61** (0.42)	**0.96** (0.08)
	II	CCA	0.84 (0.07)	1.66 (0.95)	1.31 (0.55)	0.74 (0.22)
		NI	0.87 (0.04)	1.44 (0.33)	1.22 (0.46)	0.82 (0.05)
		MI-Wood	0.87 (0.04)	1.49 (0.37)	1.23 (0.45)	0.81 (0.00)
		MI-Bartlett	0.87 (0.04)	1.46 (0.34)	1.23 (0.45)	0.82 (0.03)
		EG	**0.88** (0.04)	**0.97** (0.28)	**1.14** (0.44)	**0.99** (0.03)
	III	CCA	0.71 (0.11)	1.80 (2.28)	1.85 (0.56)	0.60 (0.32)
		NI	0.80 (0.05)	1.62 (0.39)	1.62 (0.50)	0.82 (0.06)
		MI-Wood	0.80 (0.05)	1.67 (0.41)	1.62 (0.49)	0.81 (0.00)
		MI-Bartlett	0.80 (0.05)	1.63 (0.38)	1.62 (0.49)	0.82 (0.04)
		EG	**0.82** (0.05)	**1.00** (0.32)	**1.50** (0.52)	**0.83** (0.02)
MARviol	I	CCA	0.63 (0.11)	1.43 (0.92)	1.99 (0.42)	0.31 (0.27)
		NI	0.76 (0.06)	1.67 (0.51)	1.70 (0.42)	0.85 (0.10)
		MI-Wood	0.75 (0.06)	1.68 (0.52)	1.74 (0.40)	0.82 (0.05)
		MI-Bartlett	0.75 (0.06)	1.71 (0.55)	1.73 (0.41)	0.84 (0.09)
		EG	**0.79** (0.05)	**0.96** (0.32)	**1.62** (0.43)	**0.96** (0.08)
	II	CCA	0.83 (0.08)	1.59 (0.71)	1.36 (0.58)	0.71 (0.24)
		NI	0.87 (0.04)	1.45 (0.32)	1.23 (0.45)	0.83 (0.06)
		MI-Wood	0.87 (0.04)	1.47 (0.33)	1.23 (0.44)	0.81 (0.00)
		MI-Bartlett	0.87 (0.04)	1.47 (0.33)	1.24 (0.44)	0.81 (0.02)
		EG	**0.88** (0.04)	**0.97** (0.28)	**1.15** (0.42)	**0.99** (0.03)
	III	CCA	0.71 (0.12)	1.87 (1.79)	1.86 (0.50)	0.59 (0.32)
		NI	0.80 (0.05)	1.63 (0.39)	1.62 (0.49)	0.82 (0.06)
		MI-Wood	0.80 (0.05)	1.66 (0.40)	1.62 (0.48)	0.81 (0.00)
		MI-Bartlett	0.80 (0.05)	1.66 (0.40)	1.62 (0.48)	0.81 (0.02)
		EG	**0.82** (0.04)	**1.01** (0.32)	**1.51** (0.53)	**0.83** (0.02)

Web Table 6: Simulation results for $r = 0.25$ under various settings for missing mechanism. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Setting	Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
MCAR	I	CCA	0.70 (0.08)	1.61 (1.65)	1.87 (0.37)	0.60 (0.27)
		NI	0.77 (0.04)	1.65 (0.42)	1.68 (0.35)	0.93 (0.10)
		MI-Wood	0.74 (0.04)	1.64 (0.39)	1.78 (0.34)	0.82 (0.07)
		MI-Bartlett	0.75 (0.04)	1.62 (0.42)	1.77 (0.34)	0.88 (0.10)
		EG	**0.78** (0.04)	**1.00** (0.24)	**1.63** (0.36)	**0.96** (0.08)
	II	CCA	0.84 (0.03)	1.48 (0.32)	1.42 (0.46)	0.91 (0.13)
		NI	0.84 (0.03)	1.36 (0.20)	1.42 (0.45)	0.86 (0.08)
		MI-Wood	0.84 (0.03)	1.41 (0.21)	1.44 (0.44)	0.81 (0.02)
		MI-Bartlett	0.84 (0.03)	1.39 (0.19)	1.44 (0.44)	0.83 (0.06)
		EG	**0.85** (0.03)	**0.98** (0.16)	**1.36** (0.39)	**0.99** (0.03)
	III	CCA	0.76 (0.05)	1.63 (1.13)	1.68 (0.41)	**0.87** (0.16)
		NI	0.78 (0.03)	1.55 (0.32)	1.62 (0.35)	**0.87** (0.09)
		MI-Wood	0.78 (0.03)	1.60 (0.29)	1.66 (0.33)	0.81 (0.02)
		MI-Bartlett	0.78 (0.03)	1.58 (0.30)	1.66 (0.33)	0.83 (0.06)
		EG	**0.80** (0.03)	**1.00** (0.20)	**1.47** (0.35)	0.83 (0.02)
MAR	I	CCA	0.70 (0.08)	1.58 (1.21)	1.88 (0.38)	0.55 (0.27)
		NI	0.76 (0.05)	1.62 (0.38)	1.68 (0.35)	0.89 (0.10)
		MI-Wood	0.75 (0.05)	1.66 (0.40)	1.74 (0.34)	0.82 (0.07)
		MI-Bartlett	0.75 (0.04)	1.64 (0.39)	1.73 (0.35)	0.86 (0.09)
		EG	**0.78** (0.04)	**0.98** (0.22)	**1.57** (0.35)	**0.96** (0.08)
	II	CCA	0.85 (0.05)	1.48 (0.39)	1.26 (0.46)	0.86 (0.16)
		NI	0.86 (0.03)	1.42 (0.20)	1.23 (0.40)	0.83 (0.06)
		MI-Wood	0.86 (0.03)	1.44 (0.21)	1.24 (0.40)	0.81 (0.00)
		MI-Bartlett	0.86 (0.03)	1.43 (0.21)	1.24 (0.40)	0.82 (0.04)
		EG	**0.87** (0.03)	**0.99** (0.17)	**1.16** (0.35)	**0.99** (0.03)
	III	CCA	0.77 (0.06)	1.57 (0.77)	1.65 (0.48)	0.83 (0.20)
		NI	0.79 (0.04)	1.60 (0.35)	1.57 (0.43)	**0.84** (0.07)
		MI-Wood	0.79 (0.04)	1.63 (0.32)	1.59 (0.42)	0.81 (0.00)
		MI-Bartlett	0.79 (0.04)	1.61 (0.33)	1.59 (0.42)	0.82 (0.04)
		EG	**0.82** (0.04)	**1.00** (0.22)	**1.43** (0.43)	0.83 (0.02)
MARviol	I	CCA	0.70 (0.08)	1.50 (0.81)	1.90 (0.40)	0.57 (0.24)
		NI	0.76 (0.05)	1.61 (0.42)	1.67 (0.36)	0.89 (0.10)
		MI-Wood	0.75 (0.04)	1.65 (0.42)	1.74 (0.34)	0.81 (0.06)
		MI-Bartlett	0.75 (0.04)	1.65 (0.45)	1.72 (0.35)	0.88 (0.09)
		EG	**0.78** (0.04)	**0.98** (0.22)	**1.58** (0.36)	**0.96** (0.08)
	II	CCA	0.85 (0.04)	1.52 (0.59)	1.27 (0.43)	0.87 (0.14)
		NI	0.86 (0.03)	1.39 (0.19)	1.24 (0.41)	0.83 (0.06)
		MI-Wood	0.86 (0.03)	1.42 (0.21)	1.26 (0.40)	0.81 (0.00)
		MI-Bartlett	0.86 (0.03)	1.41 (0.21)	1.26 (0.40)	0.83 (0.05)
		EG	**0.87** (0.03)	**0.98** (0.17)	**1.18** (0.35)	**0.99** (0.03)
	III	CCA	0.77 (0.06)	1.66 (1.01)	1.66 (0.48)	0.82 (0.19)
		NI	0.79 (0.04)	1.58 (0.34)	1.57 (0.43)	**0.83** (0.07)
		MI-Wood	0.79 (0.04)	1.62 (0.31)	1.59 (0.41)	0.81 (0.00)
		MI-Bartlett	0.79 (0.04)	1.59 (0.33)	1.59 (0.41)	0.82 (0.05)
		EG	**0.82** (0.04)	**1.00** (0.22)	**1.44** (0.44)	**0.83** (0.02)

Web Appendix F

Additional simulations are conducted to evaluate our proposed method in comparison to other methods with a continuous missing covariate. The data-generating distribution and setup resembles Section 3.1 except for $\boldsymbol{V}$, where $\boldsymbol{V} \sim N(0.4|U_1| - 0.1, 0.2^2)$, and the types of models to be used for imputing variables, including mean imputation for NI and the default imputation methods for MI approaches. Under this simulation design, the results remained largely the same as in the binary missing covariate case, demonstrating that our proposed method outperformed its alternatives in terms of a higher c-index, calibration slope closer to 1, lower IBS, and higher MCC in a majority of cases (Web Tables 7–9). Specifically, our proposed EG method outperformed its competitors in terms of achieving a higher c-index, better calibration, and higher MCC, along with a lower standard deviation. In contrast, the CCA method had a low discriminatory ability, and the other methods, such as NI and the two MI approaches, showed poor calibration, indicating inaccurate risk estimates. The MI-Bartlett method performed slightly better than the MI-Wood method in terms of c-index, IBS, and MCC. However, the MI-Bartlett method experienced overfitting, characterized by a calibration slope far from 1 with increased variability. The superiority of the proposed EG method was most evident in Scenario I, followed by a lesser, yet still notable, benefit in Scenario II. In Scenario III, the MI-Bartlett method showed a similar c-index to the EG method, with better IBS and/or MCC; this is expected, as the first two variables of $\boldsymbol{U}$ were included in the model based on domain knowledge, making it less suitable for the proposed method when U_2 had no effect. Nevertheless, the proposed EG method consistently yielded the best calibration, with significantly lower variability, across all scenarios by incorporating the prognostic index and making effective utilization of two-phase data.

Web Appendix G

Additional simulations are conducted to evaluate our proposed method in comparison to other methods with two missing covariates. The data-generating distribution and setup resembles Sections 3.1, except for $\boldsymbol{V} = (V_1, V_2)$, where V_1 follows Section 3.1, V_2 follows Web Appendix D. Additionally, the probability of entire $\boldsymbol{V}$ being missing is determined based on V_1 for MAR with a mild-to-moderate violation. Web Tables 10–12 report the performance metrics for different missing

Web Table 7: Simulation results under the MCAR setting for a continuous missing covariate. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.56 (0.11)	1.56 (3.63)	2.38 (0.70)	0.15 (0.22)
	NI	0.61 (0.12)	2.36 (8.69)	2.36 (0.90)	0.33 (0.25)
	MI-Wood	0.60 (0.12)	1.83 (2.23)	2.36 (0.84)	0.32 (0.27)
	MI-Bartlett	0.62 (0.12)	3.30 (5.52)	2.33 (0.90)	0.39 (0.26)
	EG	**0.68** (0.11)	**0.78** (1.12)	**2.16** (0.86)	**0.65** (0.13)
II	CCA	0.71 (0.14)	1.67 (1.44)	2.12 (0.75)	0.44 (0.30)
	NI	0.83 (0.07)	1.59 (0.81)	1.68 (0.67)	0.79 (0.16)
	MI-Wood	0.83 (0.08)	1.80 (0.91)	1.66 (0.70)	0.71 (0.14)
	MI-Bartlett	0.83 (0.07)	1.82 (1.03)	1.65 (0.70)	0.78 (0.15)
	EG	**0.84** (0.07)	**0.93** (0.39)	**1.64** (0.70)	**0.87** (0.09)
III	CCA	0.61 (0.13)	1.40 (1.46)	2.31 (0.74)	0.25 (0.29)
	NI	0.72 (0.13)	3.01 (10.6)	2.09 (0.95)	0.65 (0.28)
	MI-Wood	0.72 (0.12)	2.72 (6.26)	**2.02** (0.91)	0.63 (0.24)
	MI-Bartlett	**0.73** (0.12)	3.50 (8.95)	**2.02** (0.93)	**0.68** (0.24)
	EG	**0.73** (0.10)	**1.07** (2.34)	2.06 (0.85)	0.61 (0.19)

Web Table 8: Simulation results under the MAR setting for a continuous missing covariate. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.55 (0.09)	1.25 (2.18)	2.34 (0.67)	0.11 (0.17)
	NI	0.62 (0.13)	1.63 (1.95)	2.15 (0.77)	0.34 (0.28)
	MI-Wood	0.61 (0.12)	2.08 (2.24)	2.21 (0.75)	0.32 (0.26)
	MI-Bartlett	0.62 (0.13)	1.18 (15.9)	2.24 (0.76)	0.38 (0.26)
	EG	**0.68** (0.12)	**0.85** (0.73)	**2.07** (0.73)	**0.65** (0.13)
II	CCA	0.75 (0.16)	1.88 (1.43)	1.85 (0.94)	0.45 (0.27)
	NI	0.84 (0.09)	1.67 (1.02)	1.56 (0.95)	0.74 (0.14)
	MI-Wood	0.84 (0.09)	1.85 (1.26)	1.57 (0.97)	0.72 (0.12)
	MI-Bartlett	0.84 (0.09)	1.82 (1.19)	1.56 (0.97)	0.78 (0.14)
	EG	**0.85** (0.08)	**0.96** (0.52)	**1.54** (0.88)	**0.89** (0.09)
III	CCA	0.62 (0.14)	1.59 (1.32)	2.12 (0.73)	0.27 (0.31)
	NI	0.71 (0.12)	1.94 (1.66)	**1.99** (0.93)	0.64 (0.25)
	MI-Wood	0.72 (0.12)	2.37 (3.28)	2.03 (1.11)	0.65 (0.20)
	MI-Bartlett	0.72 (0.12)	2.18 (2.26)	**1.99** (1.10)	**0.69** (0.21)
	EG	**0.73** (0.12)	**0.91** (1.10)	2.07 (0.98)	0.61 (0.18)

mechanisms under this simulation design. The results remained largely consistent, with our proposed method outperforming its alternatives in terms of a higher c-index, a calibration slope closer to 1, lower IBS, and higher MCC in the majority of cases.

Web Table 9: Simulation results under the MAR setting with a mild-to-moderate violation for a continuous missing covariate. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.55 (0.10)	**1.09** (1.96)	2.35 (0.66)	0.13 (0.18)
	NI	0.62 (0.13)	1.62 (2.48)	2.16 (0.76)	0.33 (0.27)
	MI-Wood	0.60 (0.12)	2.38 (4.22)	2.23 (0.73)	0.32 (0.25)
	MI-Bartlett	0.62 (0.12)	3.52 (6.93)	2.24 (0.77)	0.39 (0.28)
	EG	**0.68** (0.12)	0.84 (0.70)	**2.06** (0.71)	**0.65** (0.13)
II	CCA	0.76 (0.15)	1.93 (1.65)	1.85 (0.95)	0.47 (0.26)
	NI	0.84 (0.08)	1.67 (1.02)	1.55 (0.94)	0.75 (0.14)
	MI-Wood	0.84 (0.09)	1.78 (1.19)	1.55 (0.96)	0.73 (0.12)
	MI-Bartlett	**0.85** (0.09)	1.80 (1.15)	1.55 (0.96)	0.77 (0.14)
	EG	**0.85** (0.08)	**0.95** (0.50)	**1.52** (0.87)	**0.90** (0.09)
III	CCA	0.61 (0.14)	1.57 (1.85)	2.15 (0.73)	0.27 (0.31)
	NI	0.71 (0.12)	1.89 (1.59)	**2.00** (0.92)	0.64 (0.25)
	MI-Wood	0.72 (0.11)	2.18 (2.33)	2.05 (1.10)	0.65 (0.20)
	MI-Bartlett	**0.73** (0.12)	2.22 (2.44)	2.02 (1.10)	**0.69** (0.22)
	EG	**0.73** (0.12)	**0.89** (0.95)	2.08 (0.97)	0.61 (0.18)

Web Table 10: Simulation results under the MCAR setting for two missing covariates. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.56 (0.11)	1.56 (3.63)	2.38 (0.70)	0.15 (0.22)
	NI	0.61 (0.12)	2.36 (8.69)	2.36 (0.90)	0.33 (0.25)
	MI-Wood	0.60 (0.12)	1.83 (2.23)	2.36 (0.84)	0.32 (0.27)
	MI-Bartlett	0.62 (0.12)	3.30 (5.52)	2.33 (0.90)	0.39 (0.26)
	EG	**0.68** (0.11)	**0.78** (1.12)	**2.16** (0.86)	**0.65** (0.13)
II	CCA	0.71 (0.14)	1.67 (1.44)	2.12 (0.75)	0.44 (0.30)
	NI	0.83 (0.07)	1.59 (0.81)	1.68 (0.67)	0.79 (0.16)
	MI-Wood	0.83 (0.08)	1.80 (0.91)	1.66 (0.70)	0.71 (0.14)
	MI-Bartlett	0.83 (0.07)	1.82 (1.03)	1.65 (0.70)	0.78 (0.15)
	EG	**0.84** (0.07)	**0.93** (0.39)	**1.64** (0.70)	**0.87** (0.09)
III	CCA	0.61 (0.13)	1.40 (1.46)	2.31 (0.74)	0.25 (0.29)
	NI	0.72 (0.13)	3.01 (10.6)	2.09 (0.95)	0.65 (0.28)
	MI-Wood	0.72 (0.12)	2.72 (6.26)	**2.02** (0.91)	0.63 (0.24)
	MI-Bartlett	**0.73** (0.12)	3.50 (8.95)	**2.02** (0.93)	**0.68** (0.24)
	EG	**0.73** (0.10)	**1.07** (2.34)	2.06 (0.85)	0.61 (0.19)

Web Appendix H

Additional simulations are conducted to investigate the impact of the violation of the proportional hazards assumption on the performance of our method compared to other existing methods. The data-generating distribution and setup resembles Sections 3.1, except that we introduce time-

Web Table 11: Simulation results under the MAR setting for two missing covariates. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.55 (0.09)	1.25 (2.18)	2.34 (0.67)	0.11 (0.17)
	NI	0.62 (0.13)	1.63 (1.95)	2.15 (0.77)	0.34 (0.28)
	MI-Wood	0.61 (0.12)	2.08 (2.24)	2.21 (0.75)	0.32 (0.26)
	MI-Bartlett	0.62 (0.13)	1.18 (15.9)	2.24 (0.76)	0.38 (0.26)
	EG	**0.68** (0.12)	**0.85** (0.73)	**2.07** (0.73)	**0.65** (0.13)
II	CCA	0.75 (0.16)	1.88 (1.43)	1.85 (0.94)	0.45 (0.27)
	NI	0.84 (0.09)	1.67 (1.02)	1.56 (0.95)	0.74 (0.14)
	MI-Wood	0.84 (0.09)	1.85 (1.26)	1.57 (0.97)	0.72 (0.12)
	MI-Bartlett	0.84 (0.09)	1.82 (1.19)	1.56 (0.97)	0.78 (0.14)
	EG	**0.85** (0.08)	**0.96** (0.52)	**1.54** (0.88)	**0.89** (0.09)
III	CCA	0.62 (0.14)	1.59 (1.32)	2.12 (0.73)	0.27 (0.31)
	NI	0.71 (0.12)	1.94 (1.66)	**1.99** (0.93)	0.64 (0.25)
	MI-Wood	0.72 (0.12)	2.37 (3.28)	2.03 (1.11)	0.65 (0.20)
	MI-Bartlett	0.72 (0.12)	2.18 (2.26)	**1.99** (1.10)	**0.69** (0.21)
	EG	**0.73** (0.12)	**0.91** (1.10)	2.07 (0.98)	0.61 (0.18)

Web Table 12: Simulation results under the MAR setting with a mild-to-moderate violation for two missing covariates. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.55 (0.10)	**1.09** (1.96)	2.35 (0.66)	0.13 (0.18)
	NI	0.62 (0.13)	1.62 (2.48)	2.16 (0.76)	0.33 (0.27)
	MI-Wood	0.60 (0.12)	2.38 (4.22)	2.23 (0.73)	0.32 (0.25)
	MI-Bartlett	0.62 (0.12)	3.52 (6.93)	2.24 (0.77)	0.39 (0.28)
	EG	**0.68** (0.12)	0.84 (0.70)	**2.06** (0.71)	**0.65** (0.13)
II	CCA	0.76 (0.15)	1.93 (1.65)	1.85 (0.95)	0.47 (0.26)
	NI	0.84 (0.08)	1.67 (1.02)	1.55 (0.94)	0.75 (0.14)
	MI-Wood	0.84 (0.09)	1.78 (1.19)	1.55 (0.96)	0.73 (0.12)
	MI-Bartlett	**0.85** (0.09)	1.80 (1.15)	1.55 (0.96)	0.77 (0.14)
	EG	**0.85** (0.08)	**0.95** (0.50)	**1.52** (0.87)	**0.90** (0.09)
III	CCA	0.61 (0.14)	1.57 (1.85)	2.15 (0.73)	0.27 (0.31)
	NI	0.71 (0.12)	1.89 (1.59)	**2.00** (0.92)	0.64 (0.25)
	MI-Wood	0.72 (0.11)	2.18 (2.33)	2.05 (1.10)	0.65 (0.20)
	MI-Bartlett	**0.73** (0.12)	2.22 (2.44)	2.02 (1.10)	**0.69** (0.22)
	EG	**0.73** (0.12)	**0.89** (0.95)	2.08 (0.97)	0.61 (0.18)

dependent effect of $\boldsymbol{U}$ on the survival times. Web Tables 13–15 show the performance metrics for different missing mechanisms under the non-proportionality of the hazards. The results showed a reduced c-index and biased risk estimates across all methods, as expected, although the overall impact on our method was relatively small. For example, the c-index reduction in Scenarios II and III was considerably smaller for our method compared to the other methods. In addition, the variable selection performance, demonstrated by MCC, was most significantly impacted for all methods except our method. Our proposed method experienced only a minimal decrease in MCC, due to the benefit of incorporating domain knowledge and prognostic index.

Web Table 13: Simulation results under the MCAR setting when the proportional hazards assumption is violated. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.53 (0.09)	1.18 (2.58)	2.24 (0.68)	0.08 (0.18)
	NI	0.52 (0.07)	0.25 (9.87)	2.19 (0.64)	0.05 (0.16)
	MI-Wood	0.51 (0.04)	1.49 (1.87)	2.17 (0.60)	0.03 (0.11)
	MI-Bartlett	0.55 (0.10)	8.81 (19.1)	2.19 (0.63)	0.12 (0.19)
	EG	**0.66** (0.12)	**0.86** (1.04)	**2.05** (0.72)	**0.57** (0.06)
II	CCA	0.56 (0.12)	**0.86** (5.35)	2.11 (0.69)	0.19 (0.28)
	NI	0.60 (0.14)	2.09 (2.36)	2.01 (0.63)	0.31 (0.31)
	MI-Wood	0.58 (0.12)	2.14 (3.27)	2.06 (0.65)	0.27 (0.29)
	MI-Bartlett	0.61 (0.13)	2.90 (6.08)	2.07 (0.69)	0.38 (0.28)
	EG	**0.74** (0.12)	1.24 (2.58)	**1.80** (0.76)	**0.82** (0.05)
III	CCA	0.53 (0.10)	**1.12** (2.79)	2.22 (0.67)	0.11 (0.23)
	NI	0.53 (0.09)	4.35 (11.4)	2.14 (0.57)	0.13 (0.25)
	MI-Wood	0.52 (0.08)	1.59 (3.42)	2.17 (0.60)	0.10 (0.21)
	MI-Bartlett	0.56 (0.11)	7.67 (28.2)	2.16 (0.66)	0.24 (0.25)
	EG	**0.65** (0.14)	0.82 (0.95)	**2.00** (0.67)	**0.42** (0.11)

Web Table 14: Simulation results under the MAR setting when the proportional hazards assumption is violated. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.53 (0.09)	**0.98** (9.14)	2.18 (0.67)	0.07 (0.18)
	NI	0.51 (0.06)	1.73 (12.3)	2.08 (0.53)	0.05 (0.14)
	MI-Wood	0.50 (0.04)	1.19 (2.99)	2.07 (0.51)	0.01 (0.07)
	MI-Bartlett	0.54 (0.09)	2.61 (8.14)	2.10 (0.64)	0.11 (0.18)
	EG	**0.66** (0.13)	0.86 (0.90)	**2.03** (0.84)	**0.57** (0.06)
II	CCA	0.58 (0.13)	**1.44** (8.63)	1.88 (0.88)	0.21 (0.30)
	NI	0.61 (0.15)	2.60 (2.81)	1.78 (0.68)	0.27 (0.31)
	MI-Wood	0.60 (0.14)	2.12 (2.55)	1.79 (0.69)	0.25 (0.29)
	MI-Bartlett	0.65 (0.15)	5.82 (8.49)	1.71 (0.71)	0.37 (0.31)
	EG	**0.77** (0.13)	2.37 (10.9)	**1.56** (0.76)	**0.82** (0.05)
III	CCA	0.55 (0.11)	1.52 (3.65)	2.00 (0.63)	0.14 (0.26)
	NI	0.54 (0.11)	3.77 (7.98)	1.98 (0.60)	0.15 (0.27)
	MI-Wood	0.53 (0.09)	2.66 (4.33)	1.94 (0.53)	0.08 (0.19)
	MI-Bartlett	0.57 (0.11)	11.4 (40.8)	1.96 (0.61)	0.25 (0.28)
	EG	**0.67** (0.16)	**1.04** (2.16)	**1.86** (0.68)	**0.42** (0.11)

Web Table 15: Simulation results under the MAR setting with a mild-to-moderate violation when the proportional hazards assumption is violated. For each performance metric, the mean is reported with the standard deviation in parentheses. The best results are highlighted in boldface.

Scenario	Method	c-index	Calibration slope	IBS $\times 10^1$	MCC
I	CCA	0.52 (0.09)	3.99 (18.0)	2.17 (0.64)	0.06 (0.15)
	NI	0.51 (0.07)	**1.05** (11.0)	2.10 (0.55)	0.05 (0.14)
	MI-Wood	0.50 (0.04)	-0.02 (3.39)	2.09 (0.53)	0.02 (0.08)
	MI-Bartlett	0.55 (0.11)	-2.89 (36.7)	2.17 (0.71)	0.14 (0.18)
	EG	**0.66** (0.13)	0.83 (1.02)	**2.05** (0.82)	**0.57** (0.06)
II	CCA	0.56 (0.12)	**1.89** (4.70)	1.96 (0.87)	0.16 (0.27)
	NI	0.61 (0.14)	2.64 (2.82)	1.82 (0.70)	0.28 (0.31)
	MI-Wood	0.59 (0.13)	2.40 (3.82)	1.83 (0.70)	0.21 (0.27)
	MI-Bartlett	0.64 (0.15)	6.87 (11.9)	1.80 (0.72)	0.35 (0.32)
	EG	**0.77** (0.12)	2.09 (8.56)	**1.63** (0.80)	**0.82** (0.05)
III	CCA	0.55 (0.11)	1.83 (3.55)	1.99 (0.68)	0.12 (0.23)
	NI	0.54 (0.11)	3.41 (6.51)	1.95 (0.60)	0.13 (0.25)
	MI-Wood	0.52 (0.08)	**1.20** (1.22)	1.94 (0.55)	0.06 (0.18)
	MI-Bartlett	0.57 (0.12)	3.65 (12.8)	1.98 (0.63)	0.25 (0.27)
	EG	**0.67** (0.15)	4.42 (35.5)	**1.89** (0.73)	**0.42** (0.11)

Web Appendix I

A total of $1,762$ NPC patients were included in this analysis with a median follow-up time of 11 months, in which 266 of them died due to NPC. There were $1,245$ NPC patients whose HPV status was unknown, leaving only 517 with known HPV status, of whom 180 were tested positive for HPV. Web Table 16 reports descriptive statistics for the study sample of all patients ($n = 1,762$), stratified by whether HPV status was missing or observed. Significant differences were found between these two groups in terms of histologic type ($p < .001$) and AJCC-7 stage ($p = 0.027$). Web Table 17 presents descriptive statistics for a subgroup of patients with observed HPV status only ($n' = 517$), stratified by HPV+ and HPV− status. Significant differences between the HPV+ and HPV− groups were observed for age ($p < .001$), race ($p < .001$), and AJCC-7 M stage ($p = 0.048$).

Web Figure 1 shows the Kaplan-Meier curves in the target samples by the three risk groups identified using the proposed expert-guided method for cause-specific survival; there was a significant difference in cause-specific survival probabilities across these groups ($p < .001$) based on the log-rank test. The estimated 2-year cause-specific survival (95% CI) was 94.1% (87.6%, 100.0%) for the low-risk group, 85.1% (79.2%, 91.5%) for the medium-risk group, and 59.6% (47.6%, 74.5%) for the high-risk group. A pairwise log-rank test with the Bonferroni-Holm method of adjustment indicated that there were significant pairwise differences between all three groups.

Web Table 16: Descriptive statistics of study samples for all patients stratified by whether HPV status was missing or observed.

	No. (%) or mean (SD)			
Variable	Overall, N = 1762	Missing, N = 1245	Observed, N = 517	p-value[a]
Gender				0.827
Male	1246 (70.7%)	878 (70.5%)	368 (71.2%)	
Female	516 (29.3%)	367 (29.5%)	149 (28.8%)	
Age				0.051
<25	73 (4.1%)	54 (4.3%)	19 (3.7%)	
25–49	468 (26.6%)	330 (26.5%)	138 (26.7%)	
50–74	1044 (59.3%)	721 (57.9%)	323 (62.5%)	
75+	177 (10.0%)	140 (11.2%)	37 (7.2%)	
Martial status				0.276
Married	996 (56.5%)	698 (56.1%)	298 (57.6%)	
Single	380 (21.6%)	262 (21.0%)	118 (22.8%)	
Others[b]	386 (21.9%)	285 (22.9%)	101 (19.5%)	
Race				0.062
White	810 (46.0%)	547 (43.9%)	263 (50.9%)	
Black	212 (12.0%)	156 (12.5%)	56 (10.8%)	
East Asian[c]	691 (39.2%)	513 (41.2%)	178 (34.4%)	
Others[d]	49 (2.8%)	29 (2.3%)	20 (3.9%)	
Histologic type				<.001
Keratinizing[e]	577 (32.7%)	373 (30.0%)	204 (39.5%)	
Diff/nonkera[f]	434 (24.6%)	272 (21.8%)	162 (31.3%)	
Undiff/nonkera[g]	251 (14.2%)	200 (16.1%)	51 (9.9%)	
Others[h]	500 (28.4%)	400 (32.1%)	100 (19.3%)	
AJCC-7 stage				0.027
I	147 (8.3%)	105 (8.4%)	42 (8.1%)	
II	299 (17.0%)	204 (16.4%)	95 (18.4%)	
III	459 (26.0%)	304 (24.4%)	155 (30.0%)	
IV[i]	857 (48.6%)	632 (50.8%)	225 (43.5%)	
AJCC-7 T stage				0.107
Early stage[j]	1035 (58.7%)	747 (60.0%)	288 (55.7%)	

Advanced stage[k]	727 (41.3%)	498 (40.0%)	229 (44.3%)	
AJCC-7 N stage				0.828
N = 0	403 (22.9%)	287 (23.1%)	116 (22.4%)	
N > 0	1359 (77.1%)	958 (76.9%)	401 (77.6%)	
AJCC-7 M stage				0.221
M0	1572 (89.2%)	1103 (88.6%)	469 (90.7%)	
M1	190 (10.8%)	142 (11.4%)	48 (9.3%)	
Sequence number				0.989
One primary only	1505 (85.4%)	1064 (85.5%)	441 (85.3%)	
Others[l]	257 (14.6%)	181 (14.5%)	76 (14.7%)	
Tumor size	41.9 (57.2)	42.7 (61.5)	40.1 (45.1)	0.328

Abbreviations: Diff/nonkera, differentiated/nonkeratinizing; Undiff/nonkera = undifferentiated/nonkeratinizing. [a]Pearson's Chi-squared test; Welch two-sample t-test.

[b]Others include divorced, separated, unmarried or domestic partner, widowed, and unknown.

[c]East Asian includes Chinese, Japanese, Korean (1988+), and Vietnamese (1988+).

[d]Others include American Indian/Alaska Native, Asian Indian (2010+), Asian Indian or Pakistani-NOS (1988+), Black, Fiji Islander (1991+), Filipino, Guamanian-NOS (1991+), Hawaiian, Hmong (1988+), Kampuchean (1988+), Laotian (1988+), Micronesian-NOS (1991+), Other, Other Asian (1991+), Pacific Islander-NOS (1991+), Pakistani (2010+), Polynesian-NOS (1991+), Samoan (1991+), Thai (1994+), and Tongan (1991+).

[e]Keratinizing squamous cell carcinoma includes 8070 and 8071.

[f]Differentiated non-keratinizing carcinoma includes 8072 and 8073.

[g]Undifferentiated non-keratinizing carcinoma includes 8020, 8021, 8082.

[h]Others include 8000, 8010, 8032, 8041, 8046, 8051, 8074, 8075, 8083, 8090, 8121, 8123, 8140, 8200, 8240, 8246, 8260, 8310, 8430, 8480, 8525, 8560, 8562, 8800, 8801, 8802, 8805, 8890, 8900, 8910, 8920, 8941, 8982, 9364, 9370, 9371, and 9500.

[i]IV includes IVA, IVB, IVC, and IV NOS (Not Otherswise Specified).

[j]Early stage includes T1 and T2.

[k]Advanced stage includes T3, T4, T4a, and T4b.

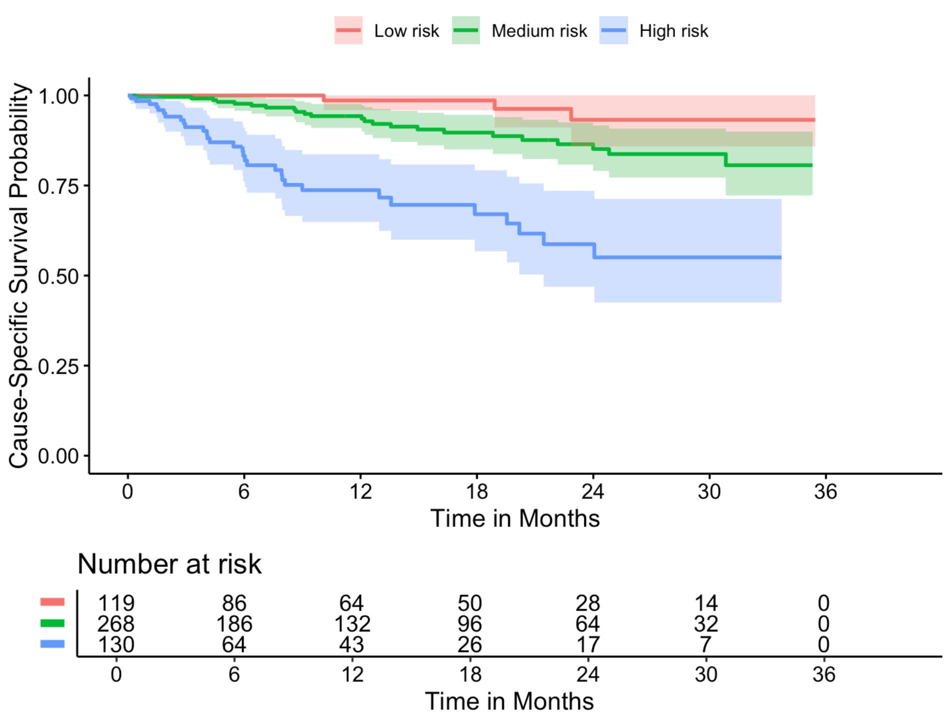
[l]Others include 1st of 2 or more primaries, 2nd of 2 or more primaries, 3rd of 3 or more primaries, and 4th of 4 or more primaries.

Web Table 17: Descriptive statistics of study samples for a subgroup of patients with observed HPV only stratified by whether a patient was HPV+ or HPV–.

	No. (%) or mean (SD)			
Variable	Overall, N = 517	HPV+, N = 180	HPV–, N = 337	p-value[a]
Gender				0.741
Male	368 (71.2%)	126 (70.0%)	242 (71.8%)	
Female	149 (28.8%)	54 (30.0%)	95 (28.2%)	
Age				<.001
<25	19 (3.7%)	12 (6.7%)	7 (2.1%)	
25–49	138 (26.7%)	39 (21.7%)	99 (29.4%)	
50–74	323 (62.5%)	123 (68.3%)	200 (59.3%)	
75+	37 (7.2%)	6 (3.3%)	31 (9.2%)	
Martial status				0.463
Married	298 (57.6%)	98 (54.4%)	200 (59.3%)	
Single	118 (22.8%)	42 (23.3%)	76 (22.6%)	
Others[b]	101 (19.5%)	40 (22.2%)	61 (18.1%)	
Race				<.001
White	263 (50.9%)	113 (62.8%)	150 (44.5%)	
Black	56 (10.8%)	18 (10.0%)	38 (11.3%)	
East Asian[c]	178 (34.4%)	40 (22.2%)	138 (40.9%)	
Others[d]	20 (3.9%)	9 (5.0%)	11 (3.3%)	
Histologic type				0.339
Keratinizing[e]	204 (39.5%)	78 (43.3%)	126 (37.4%)	
Diff/nonkera[f]	162 (31.3%)	58 (32.2%)	104 (30.9%)	
Undiff/nonkera[g]	51 (9.9%)	14 (7.8%)	37 (11.0%)	
Others[h]	100 (19.3%)	30 (16.7%)	70 (20.8%)	
AJCC-7 stage				0.334
I	42 (8.1%)	10 (5.6%)	32 (9.5%)	
II	95 (18.4%)	38 (21.1%)	57 (16.9%)	
III	155 (30.0%)	53 (29.4%)	102 (30.3%)	
IV[i]	225 (43.5%)	79 (43.9%)	146 (43.3%)	
AJCC-7 T stage				0.483
Early stage[j]	288 (55.7%)	96 (53.3%)	192 (57.0%)	
Advanced stage[k]	229 (44.3%)	84 (46.7%)	145 (43.0%)	
AJCC-7 N stage				>.999
N = 0	116 (22.4%)	40 (22.2%)	76 (22.6%)	

N > 0	401 (77.6%)	140 (77.8%)	261 (77.4%)	
AJCC-7 M stage				0.048
M0	469 (90.7%)	170 (94.4%)	299 (88.7%)	
M1	48 (9.3%)	10 (5.6%)	38 (11.3%)	
Sequence number				0.428
One primary only	441 (85.3%)	150 (83.3%)	291 (86.4%)	
Others[l]	76 (14.7%)	30 (16.7%)	46 (13.6%)	
Tumor size	40.1 (45.1)	39.5 (16.2)	40.4 (54.6)	0.780

Abbreviations: Diff/nonkera, differentiated/nonkeratinizing; Undiff/nonkera = undifferentiated/nonkeratinizing. [a]Pearson's Chi-squared test; Welch two-sample t-test.
[b]Others include divorced, separated, unmarried or domestic partner, widowed, and unknown.
[c]East Asian includes Chinese, Japanese, Korean (1988+), and Vietnamese (1988+).
[d]Others include American Indian/Alaska Native, Asian Indian (2010+), Asian Indian or Pakistani-NOS (1988+), Black, Fiji Islander (1991+), Filipino, Guamanian-NOS (1991+), Hawaiian, Hmong (1988+), Kampuchean (1988+), Laotian (1988+), Micronesian-NOS (1991+), Other, Other Asian (1991+), Pacific Islander-NOS (1991+), Pakistani (2010+), Polynesian-NOS (1991+), Samoan (1991+), Thai (1994+), and Tongan (1991+).
[e]Keratinizing squamous cell carcinoma includes 8070 and 8071.
[f]Differentiated non-keratinizing carcinoma includes 8072 and 8073.
[g]Undifferentiated non-keratinizing carcinoma includes 8020, 8021, 8082.
[h]Others include 8000, 8010, 8032, 8041, 8046, 8051, 8074, 8075, 8083, 8090, 8121, 8123, 8140, 8200, 8240, 8246, 8260, 8310, 8430, 8480, 8525, 8560, 8562, 8800, 8801, 8802, 8805, 8890, 8900, 8910, 8920, 8941, 8982, 9364, 9370, 9371, and 9500.
[i]IV includes IVA, IVB, IVC, and IV NOS (Not Otherswise Specified).
[j]Early stage includes T1 and T2.
[k]Advanced stage includes T3, T4, T4a, and T4b.
[l]Others include 1st of 2 or more primaries, 2nd of 2 or more primaries, 3rd of 3 or more primaries, and 4th of 4 or more primaries.



Web Figure 1: Kaplan-Meier curves in the target samples by the three risk groups identified using the expert-guided method for cause-specific survival.

References

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, *78*(1), 1–3.

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, *18*(17-18), 2529–2545.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama*, *247*(18), 2543–2546.

Harrell, F. E., Lee, K. L., Califf, R. M., Pryor, D. B., & Rosati, R. A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in medicine*, *3*(2), 143–152.

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, *15*(4), 361–387.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *405*(2), 442–451.

Van Houwelingen, H. C. (2000). Validation, calibration, revision and combination of prognostic survival models. *Statistics in medicine*, *19*(24), 3401–3415.