FROM PARTIAL LIKELIHOOD TO PAIRWISE PREFERENCE: DEEP SURVIVAL ANALYSIS WITH BRADLEY–TERRY MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep learning has become central to modern survival analysis, with applications spanning clinical trials, operations research, and economics. A defining—and largely underexplored—feature of many such applications, most acutely in medical imaging, is that the covariate dimension d far exceeds the number of available subjects n ($d \gg n$): a single CT volume already contains on the order of 10^7 voxels, whereas a cohort rarely exceeds a few thousand patients. This regime forces very small training batches, which interact poorly with the Cox proportional hazards model: its partial likelihood couples all subjects through a shared risk-set denominator, making minibatch gradients biased and stochastic optimization unreliable exactly when batches are smallest. We make a simple observation with broad consequences: fitting a Bradley–Terry model in place of the Cox partial likelihood recasts survival training as a sum over independent pairwise preferences, yielding unbiased minibatch gradients while targeting the same risk ordering. Casting survival analysis as preference learning further lets us import the extensive toolbox developed around the Bradley–Terry model in RLHF—graded (fine-grained) labels that exploit the magnitude of survival-time gaps. On a 3D medical imaging dataset, our method matches or surpasses Cox-based deep survival baselines, with the largest gains in the small-batch, high-dimensional regime where partial-likelihood training degrades most.

1 INTRODUCTION

Survival analysis is a central problem in statistics, with a wide range of applications in operations research and clinical trials. It seeks to answer questions of the form: "In how many days will a customer complete a purchase or subscription, or a patient experience disease progression or death?" At first glance the problem resembles classical regression. In practice, however, the collected data are typically *censored*: for some subjects we observe the event of interest (e.g. a customer who has subscribed), whereas for others we only know that the event has not yet occurred, without ruling out that it may occur in the future. Simply discarding these censored subjects not only wastes available data but also introduces bias into the resulting estimates (Kalbfleisch & Prentice, 2002).

Applications of survival analysis span many domains: finance (Stepanova & Thomas, 2002; Gepp & Kumar, 2008; Harrison & Ansell, 2002; Czado & Rudolph, 2002), where it is used to predict financial distress or cross-selling opportunities; environmental science (Kulkarni et al., 2013), e.g. for earthquake prediction; operations research (Maystre & Russo, 2022); churn prediction (Periáñez et al., 2016; Larivière & Van den Poel, 2004); and clinical trials (Prinja et al., 2010; Győrffy, 2021; Györffy et al., 2010). To address these problems, a variety of statistical models have been developed (George et al., 2014), the most widely used being the Cox proportional hazards (Cox-PH) model (Cox, 1972) and the accelerated failure time model (Wei, 1992).

In medicine, and in oncological clinical trials in particular, the Cox-PH model is widely used to estimate the survival of cancer patients across treatment arms (Unger et al., 2014; Katzman et al., 2018; Sako et al., 2026; Torres et al., 2021). Its main role is to estimate a per-patient risk function, which allows researchers to stratify patients into risk groups, detect imbalances between arms, and so on. Classically, the risk function is taken to be a linear model, which offers the advantage of

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

interpretable feature effects. Driven by the recent success of deep learning and the demand for more accurate survival predictions, however, there is now considerable interest in deep survival models (Katzman et al., 2018; Shu et al., 2021; Zhu et al., 2016; Fotso, 2018; Lee et al., 2018; Sako et al., 2026; Torres et al., 2021; Ranganath et al., 2016; Wiegrebe et al., 2024). Their inputs range from clinical and tabular data to transcriptomics, medical imaging, and multimodal combinations thereof. In the imaging case especially, the dimension of a CT or MRI volume can far exceed the number of patients enrolled in the trial ($d \gg n$). Training a deep survival model requires minimizing the *partial log-likelihood*, an objective built around a term known as the *risk set*: computing the gradient for a single data point involves $O(n)$ other data points. As a consequence, this objective does not mix well with standard stochastic gradient methods.

A line of work has sought to resolve this issue and train deep survival models with SGD (Kvamme et al., 2019; Tarkhan & Simon, 2020; Tarkhan et al., 2021; Shi & Ioannidis, 2025). In practice, however, these methods are either not applied to high-dimensional datasets or lack simplicity. This motivates the central question of our work:

Can we design deep survival training pipelines that are both scalable and simple?

We answer this question in the affirmative using the Bradley–Terry model, a simple model widely used in preference optimization. Training with a Bradley–Terry loss improves on standard methods such as DeepSurv (Katzman et al., 2018) and DeepHit Lee et al. (2018). Moreover, we exploit the rich toolbox available for the Bradley–Terry model—such as fine-grained (graded) labels—to improve the accuracy of the predicted survival horizon.

2 Prior Work

Over the past decade, deep learning has been widely applied to survival analysis, especially in medical settings (Wiegrebe et al., 2024; Chen, 2024; Ranganath et al., 2016). A seminal contribution is that of Katzman et al. (2018), who replaced the linear risk function of the Cox proportional hazards model—classically fit with second-order methods—by a neural network. Since then, a large body of work has adapted deep architectures to survival problems (Zhu et al., 2016; Fotso, 2018; Miscouridou et al., 2018; Torres et al., 2021), with inputs ranging from low-dimensional tabular (clinical) data (Katzman et al., 2018) to higher-dimensional 2D and 3D images (Shu et al., 2021; Sako et al., 2026; Torres et al., 2021).

Beyond the Cox family, a separate line of deep survival models deliberately drops the proportional-hazards assumption. The most prominent is the discrete-time, ranking-based DeepHit (Lee et al., 2018), which directly models the distribution of event times over a discretized horizon. Other approaches parametrize the event-time distribution explicitly, either through classical parametric families (e.g. Weibull or log-normal) or through flexible mixtures of such distributions (Nagpal et al., 2021; Zhong et al., 2021; Carroll, 2003; Royston, 2001), while more recent work captures the survival dynamics with neural ordinary differential equations (Tang et al., 2022). We nevertheless take the Cox partial likelihood as our primary point of comparison, both because of its predominance in practice and because it is the objective most directly affected by the minibatch obstacle we address.

Despite its wide use, the Cox model raises a difficulty in high-dimensional datasets, where full-batch gradient descent is infeasible. As the input dimension grows while the number of samples remains in the order of thousands, the risk set of each data point can no longer be approximated within a minibatch, which makes training on medical imaging datasets challenging (Sako et al., 2026; Torres et al., 2021). A number of works have therefore turned to stochastic gradient methods to overcome the risk-set obstacle (Kvamme et al., 2019; Tarkhan & Simon, 2020; Tarkhan et al., 2021; Shi & Ioannidis, 2025). Kvamme et al. (2019) propose `Cox-CC`, which approximates the risk set of each event by M sampled controls (typically $M = 1$) while still forming the loss over all n events. Tarkhan & Simon (2020), together with their follow-up, pursue a similar idea: they analyze simulated and moderate-dimensional ($28^2 = 784$) imaging data, split it into batches of size two, and approximate the partial log-likelihood with a risk set of size two. It is worth noting that in the special case of batch size two, `Cox-CC`, `BigSurv-SGD` (Tarkhan & Simon, 2020), `DeepSurv` (Katzman et al., 2018), and the Bradley–Terry objective all reduce to the same per-pair loss. They differ only in how the pairs are formed: `Cox-CC` pairs every event with M sampled controls and thus touches all n events at each update, which does not scale to large sample sizes, whereas the others draw a

minibatch uniformly at random and update only when it contains a comparable pair. More recently, Shi & Ioannidis (2025) introduce a spectral method that optimizes the Cox partial log-likelihood equation 1 independently of the batch size. Their method proceeds in outer iterations of the Alternating Direction Method of Multipliers (ADMM): each iteration expresses the per-patient risk as the stationary distribution of a Markov chain, fits the model to this risk by stochastic gradient descent, and then updates the dual variables. While this combination of Markov-chain and optimization tools is elegant, it is considerably more complex than ours—and than the rest of the deep survival analysis literature—and requires many more training epochs, so that its training time far exceeds that of standard methods (Katzman et al., 2018; Lee et al., 2018) and ours (see Section 5).

The Bradley–Terry–Cox connection is not itself new—Su & Zhou (2006) established it in the covariate-independent case and Nagpal (2025) revisit it—but both treat it as a relation between likelihoods rather than a training method. We turn it into one: the covariate-dependent pairwise law (Proposition 1) and its *unbiased* minibatch gradient (Proposition 2) make Bradley–Terry a drop-in replacement for the Cox partial likelihood that trains stably at the tiny batches forced by $d \gg n$ imaging, without the risk-set subsampling of `Cox-CC` or the extra machinery of the spectral method. Being a simple, well-studied model, it further opens a broad toolbox for survival analysis—the graded (fine-grained) labels we study being one example.

3 Preliminaries

3.1 Problem Setup & Notation

We study survival (time-to-event) prediction from high-dimensional covariates. Each instance is described by a covariate vector $\mathbf{X} \in \mathbb{R}^d$—for example the voxels of a 3D CT volume, possibly concatenated with clinical variables—and we are primarily interested in the regime in which the covariate dimension greatly exceeds the sample size, $d \gg n$.

Let $Y > 0$ denote the (latent) *time to the event of interest*, e.g. the time to death or disease progression, and let $C > 0$ denote the *censoring time*, i.e. the last time at which the instance is observed. Under right censoring we do not observe Y directly; instead we observe the pair

$$E := \min(Y, C), \qquad \Delta := \mathbf{1}\{Y \leq C\},$$

where E is the *observed time* and $\Delta \in \{0, 1\}$ is the *event indicator*: $\Delta = 1$ means the event was observed ($E = Y$), whereas $\Delta = 0$ means the observation was censored ($E = C$) and we only learn that $Y > C$.

A dataset of n instances is a collection of i.i.d. triplets

$$\mathcal{D} = \{(\mathbf{X}_i, E_i, \Delta_i)\}_{i=1}^n \overset{\text{iid}}{\sim} \mathcal{P},$$

drawn from an unknown distribution $\mathcal{P}$ over $\mathbb{R}^d \times \mathbb{R}_{>0} \times \{0, 1\}$, where $\mathbf{X}_i$, E_i, and Δ_i denote the covariates, observed time, and event indicator of instance i, respectively.

Throughout, we make the standard assumption of independent (non-informative) censoring.

Assumption 3.1 (Independent censoring)**.** *The event and censoring times are conditionally independent given the covariates,*

$$Y \perp\!\!\!\perp C \mid \mathbf{X}.$$

Assumption 3.1 states that, after conditioning on the observed covariates, the censoring mechanism carries no additional information about the event time: censoring may depend on, for example, the patient's current medical condition, but not on unobserved factors that also drive survival (Kalbfleisch & Prentice, 2002; Lillelund et al., 2025). This assumption makes $Y \mid \mathbf{X}$ identifiable from $\mathcal{P}$ despite censoring. For a self-contained, friendly introduction to survival analysis and its connections to deep learning, we refer the reader to the excellent surveys of Wiegrebe et al. (2024); Chen (2024).

3.2 Cox Proportional Hazards Model

Having described the censoring mechanism and the observed data, we now turn to modeling the latent event time Y. A central quantity in survival analysis is the *conditional hazard*, which measures

the instantaneous rate at which the event occurs at time t, given the covariates $\mathbf{X} = x$ and that the event has not occurred before t. A standard and well-established semi-parametric model for the conditional risk is the Cox proportional hazards (PH) model (Cox, 1972), which specifies the hazard as

$$\lambda(t \mid \mathbf{X} = x) = \lambda_0(t)\, \exp\big(r_\theta(x)\big),$$

where $\lambda_0(t)$ is an unspecified nonnegative *baseline hazard* and $r_\theta : \mathbb{R}^d \to \mathbb{R}$ is a *risk score*. As we mentioned, the model is semi-parametric: it leaves λ_0 fully nonparametric and only assumes that covariates act multiplicatively on the hazard, so the relative risk of any two subjects is constant in time. Classically $r_\theta(x) = x^\top \theta$ is linear (Cox, 1972; Therneau & Grambsch, 2000); `DeepSurv` (Katzman et al., 2018) and its successors instead parametrize the risk score by a neural network. Such neural Cox models have been adopted across data modalities—feed-forward networks on clinical and tabular covariates (Faraggi & Simon, 1995; Katzman et al., 2018; Kvamme et al., 2019), high-dimensional omics (Ching et al., 2018), and convolutional networks on medical images (Zhu et al., 2016; Torres et al., 2021; Sako et al., 2026; Huo et al., 2025)—and the Cox partial likelihood remains the most widely used training objective for deep survival models (Wiegrebe et al., 2024; Chen, 2024). Alternative families deliberately forgo the PH assumption; we take the Cox partial likelihood as our primary point of comparison precisely because of its wide usage.

The baseline hazard is eliminated by conditioning on the observed event times, giving Cox's *partial likelihood* (Cox, 1972; Therneau & Grambsch, 2000). Let $\mathcal{R}(t) := \{\, j : E_j \geq t \,\}$ be the *risk set* at time t (the subjects still event-free and under observation just before t), and write $r_i := r_\theta(\mathbf{X}_i)$. The negative log partial likelihood is

$$\mathcal{L}(\theta) \;:=\; -\frac{1}{n} \sum_{i:\, \Delta_i = 1} \left[r_i \;-\; \log \sum_{j \in \mathcal{R}(E_i)} \exp(r_j) \right], \tag{1}$$

where the outer sum runs over uncensored subjects and, for each such i, the inner term contrasts its risk with that of every subject still at risk at its event time E_i. Minimizing $\mathcal{L}$ by gradient-based optimization is the *de facto* training objective for deep Cox models (Katzman et al., 2018; Kvamme et al., 2019).

Unlike a standard supervised loss, $\mathcal{L}(\theta)$ is *not* a sum of independent per-subject terms: each summand couples subject i to its entire risk set through the log-sum-exp normalizer $\log \sum_{j \in \mathcal{R}(E_i)} \exp(r_j)$. Restricting the risk set to a minibatch $\mathcal{B}$—replacing $\mathcal{R}(E_i)$ by $\mathcal{R}(E_i) \cap \mathcal{B}$—therefore yields a *biased* estimate of both $\mathcal{L}$ and its gradient (Lemma 1 of Shi & Ioannidis (2025)): because $\log \sum \exp$ is nonlinear, the minibatch expectation does not commute with the logarithm, so $\mathbb{E}_{\mathcal{B}}\big[\nabla_\theta \mathcal{L}_{\mathcal{B}}(\theta)\big] \neq \nabla_\theta \mathcal{L}(\theta)$, where the stochastic loss $\mathcal{L}_{\mathcal{B}}$ is defined as

$$\mathcal{L}_{\mathcal{B}} := \; -\frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}:\, \Delta_i = 1} \left[r_i \;-\; \log \sum_{j \in \mathcal{R}(E_i) \cap \mathcal{B}} \exp(r_j) \right],$$

and when the batch does not contain any comparable pair, the loss is simply equal to zero.

This bias is well documented in past literature and motivates a range of specialized schemes: it is recently analyzed by Shi & Ioannidis (2025), mitigated by case–control subsampling of the risk set in Cox-CC/Cox-Time (Kvamme et al., 2019), and by strata-averaged reformulations in BigSurvSGD (Tarkhan & Simon, 2020; Tarkhan et al., 2021). The obstacle is worst precisely in our target regime: when $d \gg n$, memory forces small batches, so $\mathcal{R}(E_i) \cap \mathcal{B}$ is a small, high-variance fraction of the true risk set and the gradient bias is correspondingly large.

3.3 Objectives

Survival models are typically evaluated along two complementary axes. The first is *discrimination*, which measures the model's ability to rank patients according to their relative risk. The second is *absolute predictive accuracy*, which encompasses both *calibration* of the predicted survival probabilities and the accuracy of predictions on the event-time scale. Reporting both axes is essential: a model may rank patients correctly while systematically misestimating their survival probabilities or event times, and conversely, it may be accurate on average while failing to distinguish high-risk from low-risk patients. For comprehensive discussions of evaluation criteria and metrics for survival models, we refer the reader to Haider et al. (2020); Lillelund et al. (2025).

Throughout, let $\hat{S}(t \mid x) := \hat{\mathbb{P}}(Y > t \mid \mathbf{X} = x)$ be the model's predicted survival function and $\hat{G}(t)$ the Kaplan–Meier estimate (Kaplan & Meier, 1958; Kalbfleisch & Prentice, 2002) of the censoring survival $G(t) = \mathbb{P}(C > t)$.

Discrimination: Harrell's and Uno's concordance. A pair (i, j) is *comparable* if the earlier observed subject is uncensored, so that its true order is known. Harrell's concordance index (Harrell Jr et al., 1984) is the fraction of comparable pairs ranked correctly by the risk score,

$$C_{\mathrm{H}} = \frac{\sum_{i \neq j} \Delta_i \, \mathbf{1}\{E_i < E_j\} \, \mathbf{1}\{r_i > r_j\}}{\sum_{i \neq j} \Delta_i \, \mathbf{1}\{E_i < E_j\}}, \tag{2}$$

with ties in r counted as $\frac{1}{2}$. Because comparability itself depends on the censoring distribution, C_{H} is a biased estimate of the population concordance under heavy or covariate-dependent censoring. Uno's C-index (Uno et al., 2011), truncated at a horizon $\tau > 0$, corrects for this dependence by weighting each comparable pair by the inverse probability of observing that pair:

$$C_{\mathrm{Uno}}(\tau) = \frac{\displaystyle\sum_{i \neq j} \frac{\Delta_i}{G(E_i)^2} \mathbf{1}\{E_i < E_j\} \mathbf{1}\{E_i \leq \tau\} \mathbf{1}\{r_i > r_j\}}{\displaystyle\sum_{i \neq j} \frac{\Delta_i}{G(E_i)^2} \mathbf{1}\{E_i < E_j\} \mathbf{1}\{E_i \leq \tau\}}, \tag{3}$$

where $G(t) = \mathbb{P}(C > t)$ is the censoring survival function, which is replaced in practice by an estimate. This estimator is consistent for a concordance measure that does not depend on the censoring distribution.

Calibration and time-scale accuracy. From the risk score we recover a full survival curve via the Breslow estimator ($\hat{\Lambda}_0(t)$) of the baseline cumulative hazard ($\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$), $\hat{S}(t \mid x) = \exp\!\big(-\hat{\Lambda}_0(t)\, e^{\hat{r}(x)}\big)$, (Therneau & Grambsch, 2000). Two proper scoring rules summarize the quality of these probabilities over a horizon $[0, \tau]$. The *integrated Brier score* (Graf et al., 1999) is

$$\mathrm{IBS} = \frac{1}{\tau} \int_0^\tau \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\hat{S}(t \mid \mathbf{X}_i)^2 \, \mathbf{1}\{E_i \leq t,\, \Delta_i = 1\}}{\hat{G}(E_i)} + \frac{\big(1 - \hat{S}(t \mid \mathbf{X}_i)\big)^2 \, \mathbf{1}\{E_i > t\}}{\hat{G}(t)} \right] dt.$$

Finally, since ranking metrics are invariant to the time *scale*, we also report the root-mean-square error of the predicted restricted mean survival time $\widehat{\mathrm{RMST}}_i(\tau) = \int_0^\tau \hat{S}(t \mid \mathbf{X}_i)\, dt$ against the observed event time on uncensored subjects (Haider et al., 2020). In summary, C_{H} and C_{Uno} probe *ranking*, IBS probes *probabilistic calibration*, and the RMST error probes *time-scale accuracy*—a separation that matters for our method, whose graded-label variant improves the time scale while leaving the ranking essentially unchanged (see Section 5).

Stratification constitutes an additional metric that is typically assessed visually from survival curves. An effective survival model should be able to discriminate patients into groups exhibiting distinct risk levels (e.g., lower vs. higher risk). For further details, the reader is referred to Appendix E.

4 Methodology

Our method replaces the Cox partial likelihood, whose minibatch gradients are biased (Section 3.2), with a pairwise Bradley–Terry objective. We first make the connection between the two models precise (Section 4.1), and then derive the pairwise objective, show that it admits unbiased minibatch gradients, and give the resulting training algorithm together with its complexity.

4.1 Connection with the Bradley–Terry Model

The relationship between the Cox proportional hazards model and the Bradley–Terry model (Bradley & Terry, 1952) was established in the seminal contribution of Su & Zhou (2006) within the statistical literature, which established their equivalence in the covariate-independent setting. Even in the covariate-dependent setting, we can make a related observation: the probability that patient i experiences the event before patient j takes the form of a Bradley–Terry pairwise probability.

More formally, write $r_i^\star = r^\star(\mathbf{X}_i)$ for the true log-risk and $r_i = r_\theta(\mathbf{X}_i)$ for a candidate score, and define the sigmoid function as $\sigma(z) = (1 + e^{-z})^{-1}$.

Proposition 1 (Pairwise laws under proportional hazards)**.** *Suppose independent subjects have continuous event times with hazards $\lambda(t \mid x) = \lambda_0(t)e^{r^\star(x)}$, where $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$ is finite for finite t and $\Lambda_0(t) \to \infty$ as $t \to \infty$. Then*

$$\mathbb{P}(Y_i < Y_j \mid \mathbf{X}_i, \mathbf{X}_j) = \frac{e^{r_i^\star}}{e^{r_i^\star} + e^{r_j^\star}} = \sigma(r_i^\star - r_j^\star). \tag{4}$$

If, additionally, $Y_i \perp C_i \mid \mathbf{X}_i,\ \forall i \in [n]$ (Assumption 3.1), let

$$Q_{ij} = \{\Delta_i = 1, E_i < E_j\} \cup \{\Delta_j = 1, E_j < E_i\}$$

be the event that the pair is comparable. Whenever $\mathbb{P}(Q_{ij} \mid \mathbf{X}_i, \mathbf{X}_j) > 0$,

$$\mathbb{P}(Y_i < Y_j \mid \mathbf{X}_i, \mathbf{X}_j, Q_{ij}) = \sigma(r_i^\star - r_j^\star). \tag{5}$$

The proof is given in Appendix A. Proposition 1 shows that the latent pairwise ordering follows *exactly* a Bradley–Terry model (Bradley & Terry, 1952) with utilities $r_i^\star$. Applying the same argument to a set of subjects yields the Plackett–Luce probability for the complete failure order (Plackett, 1975; Luce, 2005), $\prod_k e^{r_{(k)}} / \sum_{\ell \geq k} e^{r_{(\ell)}}$. Evaluating these choice probabilities on the successive risk sets recovers the Cox partial likelihood of Section 3.2. Thus, Cox's partial likelihood has the form of a *listwise* Plackett–Luce likelihood, while Bradley–Terry describes its *pairwise marginal* (Su & Zhou, 2006; Nagpal, 2025).

However, the preceding proposition does not by itself establish that fitting a Bradley–Terry model and fitting a Cox model yield the same population parameters. We conjecture that they do, supported by our experimental observations in Section 5. A formal proof is beyond the scope of the present work, and we leave this question for future theoretical investigation.

For a dataset of n subjects, let

$$\mathcal{E} = \{(i, j) : E_i < E_j,\ \Delta_i = 1\}$$

be the set of comparable pairs. We minimize the pairwise logistic loss

$$\mathcal{L}_{\mathrm{BT}}(\theta) = \frac{1}{\binom{n}{2}} \sum_{(i,j)\in\mathcal{E}} \ell(r_i - r_j), \qquad \ell(z) = \log(1 + e^{-z}), \tag{6}$$

where $r_i = r_\theta(\mathbf{X}_i)$. The normalization uses the total number of possible subject pairs, with non-comparable pairs contributing zero.

At each update, we sample a batch of patients and compute the loss over all comparable pairs within that batch. The following proposition shows that this procedure yields an unbiased gradient of equation 6, whose proof comes from elementary computation.

Proposition 2 (Unbiased gradients from patient minibatches)**.** *Fix the observed dataset $\mathcal{D}$ and the current parameter θ. Assume $r_\theta(x)$ is a deterministic function of an individual subject's covariates and is differentiable in θ. Let $\mathcal{B}$ be sampled uniformly from all subsets of B subjects, where $2 \leq B \leq n$, conditional on $\mathcal{D}$ and θ. Define*

$$\mathcal{E}_\mathcal{B} = \{(i, j) \in \mathcal{E} : i, j \in \mathcal{B}\},$$

and the minibatch loss

$$\widehat{\mathcal{L}}_\mathcal{B}(\theta) = \frac{1}{\binom{B}{2}} \sum_{(i,j)\in\mathcal{E}_\mathcal{B}} \ell(r_i - r_j). \tag{7}$$

Then

$$\mathbb{E}_\mathcal{B}\left[\nabla_\theta \widehat{\mathcal{L}}_\mathcal{B}(\theta) \,\middle|\, \mathcal{D}, \theta\right] = \nabla_\theta \mathcal{L}_{\mathrm{BT}}(\theta). \tag{8}$$

The denominator in equation 7 remains $\binom{B}{2}$ regardless of how many pairs are comparable. Dividing instead by the random number of comparable pairs would generally invalidate the unbiased-gradient identity. A batch with no comparable pairs contributes zero loss and zero gradient.

Therefore, a natural algorithm is applying Stochastic Gradient Descent to minimize $-L_{\mathrm{BT}}(\theta)$. Algorithm 1 trains the model for T epochs, with $S = \lceil n/B \rceil$ updates per epoch. Each update samples B distinct patients uniformly from the full dataset, independently of earlier batch draws. Thus, an epoch specifies a budget of S updates rather than an exhaustive pass through the subjects. For completeness we add the pseudo-code of the above method at Appendix B.

Table 1: Semi-synthetic parameter recovery, $\|\theta_{\text{alg}} - \theta^\star\|_2 / \|\theta^\star\|_2$ (lower is better), under a known Cox proportional-hazards generative model (Bender et al., 2005), as a function of the minibatch size b (mean $\pm$ std over 5 seeds). Cox-standard is the full-batch Newton estimate and sets the recovery floor. Best per column (excluding the floor) in bold.

Method	b=2	b=4	b=8	b=16
GBSG (n=2232, d=7)				
Cox-standard (full batch, b-indep.)		0.266 ± 0.087		
Cox-minibatch	0.507 ± 0.068	$\mathbf{0.378} \pm 0.076$	$\mathbf{0.335} \pm 0.079$	$\mathbf{0.317} \pm 0.078$
Bradley–Terry	$\mathbf{0.459} \pm 0.073$	0.382 ± 0.082	0.360 ± 0.082	0.361 ± 0.079
Cox-CC	0.552 ± 0.193	0.451 ± 0.172	0.383 ± 0.149	0.331 ± 0.142
DBCD (n=295, d=4919)				
Cox-standard (full batch, b-indep.)		6.810 ± 0.649		
Cox-minibatch	$\mathbf{3.570} \pm 0.274$	4.130 ± 0.342	3.970 ± 0.319	3.487 ± 0.331
Bradley–Terry	4.497 ± 0.333	$\mathbf{3.758} \pm 0.396$	$\mathbf{2.624} \pm 0.256$	$\mathbf{1.809} \pm 0.174$
Cox-CC	39.453 ± 3.234	28.105 ± 2.567	19.604 ± 1.693	13.331 ± 1.197
CHOP (n=414, d=3833)				
Cox-standard (full batch, b-indep.)		4.643 ± 0.307		
Cox-minibatch	$\mathbf{1.727} \pm 0.063$	2.102 ± 0.060	2.079 ± 0.051	1.853 ± 0.052
Bradley–Terry	2.070 ± 0.089	$\mathbf{1.931} \pm 0.068$	$\mathbf{1.502} \pm 0.052$	$\mathbf{1.228} \pm 0.023$
Cox-CC	19.601 ± 1.177	13.902 ± 0.841	9.684 ± 0.618	6.690 ± 0.413

5 EXPERIMENTS

Before evaluating our method on standard metrics (see Section 3.3), we first examine experimentally the equivalence between the Bradley–Terry (BT) model and Cox proportional hazards, in terms of their population parameters. We show that when the Cox-PH assumption holds, fitting a BT model recovers the true parameter with small error. We compare the BT model against standard estimators: full-batch Cox (a second-order method), Cox-minibatch/DeepSurv (Katzman et al., 2018), and Cox-CC (Kvamme et al., 2019). To simulate the Cox proportional hazards model we follow the standard procedure of Bender et al. (2005). We report the relative ℓ_2 recovery error as a function of the minibatch size b in Table 1. Due to space limitations, the full methodology, the corresponding cosine-similarity table (Table 3), and the per-epoch convergence curves (Figure 1) are deferred to Section C.1. We use Bradley-Terry with uniform patient sampling (Algorithm 1) in all the following experiments.

For the parameter-recovery study we use one low-dimensional tabular cohort, GBSG (Schumacher et al., 1994) (as packaged by Kvamme et al. (2019) in `pycox` package), and two high-dimensional ($d \gg n$) gene-expression cohorts, DBCD (Van't Veer et al., 2002) and CHOP (Van Wieringen et al., 2009), which represent the regime where the minibatch obstacle is most severe. Across the rest of the paper we additionally use the tabular cohorts SUPPORT (Knaus et al., 1995) and METABRIC (Curtis et al., 2012), together with large full-resolution CT cohorts from The Cancer Imaging Archive (Clark et al., 2013): NSCLC-Radiomics (Aerts et al., 2014), NSCLC-Radiogenomics (Bakr et al., 2018), LungCT-Diagnosis (Grove et al., 2015), and TCGA-LUAD/LUSC (Albertina et al., 2016; Kirk et al., 2016).

The first row of each block at Table 1 is the full-batch Cox solution obtained with the Newton–Raphson method. This second-order estimator sets the recovery floor, but it is only feasible for small d and is therefore inapplicable to high-dimensional data such as medical imaging, where stochastic optimization is unavoidable.

Two regimes emerge. For low-dimensional data (GBSG), all stochastic methods recover the parameter well and are essentially indistinguishable from one another and, for $b \geq 4$, close to the second-order floor: the minibatch bias of Cox is mild and shrinks as b grows. In the high-dimensional $d \gg n$ regime (DBCD, CHOP), the picture changes. Here BT and Cox-minibatch recover the same *direc-*

tion — their cosine similarities to $\theta^\star$ agree to within noise (Table 3) — yet BT attains a substantially smaller relative ℓ_2 error, and the gap *widens* as the batch size increases (e.g. on DBCD at b=16, 1.81 for BT versus 3.49 for Cox-minibatch). The two objectives thus differ almost entirely in the *magnitude* they assign to θ: Cox-minibatch (DeepSurv) inflates the scale of the estimate, whereas BT keeps it calibrated. Cox-CC recovers the correct direction as well, but its magnitude diverges badly in this regime, leaving a relative ℓ_2 error an order of magnitude above BT. The per-epoch convergence curves in Figure 1 confirm that this separation holds along the whole optimization trajectory, not merely at a particular stopping point.

5.1 High-dimensional survival with medical imaging

We now evaluate our method in the high-dimensional regime, where each covariate vector is a volumetric CT scan. Every data point $(\mathbf{X}_i, E_i, \Delta_i)$ is a patient enrolled in an oncological study, with observed time $E_i = \min(Y_i, C_i)$ and censoring indicator $\Delta_i = \mathbf{1}\{Y_i \le C_i\}$. In general the covariates $\mathbf{X}$ may combine clinical variables, genomics, and imaging (CT, MRI); here we focus on the imaging modality, using cohorts of chest CT paired with an overall-survival endpoint. This is the setting that most sharply exposes the minibatch bias of the Cox partial likelihood: a single full-resolution CT occupies on the order of 10^7 voxels, so only a handful of volumes fit on a GPU and training is forced into the small-batch regime, precisely where the truncated risk-set denominator is most biased. As a concrete example, on a 40 GB GPU a full-resolution scan admits a per-device batch size of only about 2.

To obtain a moderate sample size we merge several publicly available lung-cancer trials from The Cancer Imaging Archive (Clark et al., 2013): NSCLC-Radiomics (Aerts et al., 2014) (421 patients), NSCLC-Radiogenomics (Bakr et al., 2018) (113), LungCT-Diagnosis (Grove et al., 2015) (53), TCGA-LUAD (Albertina et al., 2016) (21), and TCGA-LUSC (Kirk et al., 2016) (35), for a total of $n = 643$ patients with 463 observed deaths (28% censoring). Every scan is preprocessed identically: the patient table is removed with a TotalSegmentator (Wasserthal et al., 2023) body mask (voxels outside the body set to -1000 HU), and the volume is resampled to a common grid of $512 \times 512 \times 68 \approx 1.7 \times 10^7$ voxels, so that $d \gg n$ by four orders of magnitude. We split the merged dataset $70/15/15$ into train / validation / test ($450/96/97$ patients), stratified jointly by cohort *and* event, so that every fold matches on both trial composition and censoring rate, and all model selection (learning rate, early stopping) uses the validation concordance only. We train a 3D ResNet-18 (He et al., 2016) risk model with GroupNorm (batch-size-independent normalization, so that only the loss—not the normalization statistics—varies with batch size), AdamW, a cosine learning-rate schedule, and mixed precision, for up to 100 epochs. We sweep the global batch size over $\{4, 8\}$ and the learning rate over $\{3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}\}$. We observed little variation across learning rates; 10^{-3} gave the best results overall and is the setting we use for the cross-method comparison reported below.

All experiments are run on NVIDIA A100 GPUs with 40 GB of memory. Since a single full-resolution volume permits a per-GPU batch of only 2, we reach the effective batch sizes of 4 and 8 through data parallelism across multiple GPUs. We report $b \in \{4, 8\}$ and omit b=2 for two reasons. First, at a batch of two the Bradley–Terry and minibatched-Cox objectives coincide—a batch of two contributes the single term $-\log \sigma(r_i - r_j)$ under either loss—so b=2 cannot separate the methods. Second, b=2 maximizes the number of optimizer steps per epoch ($S = \lceil n/2 \rceil$), making multi-seed training substantially more expensive for no added comparative value. Our results are summarized in Table 2. Running times are summarized in Table 5. DeepSurv (Katzman et al., 2018), our method, and DeepHit (Lee et al., 2018) each complete within one hour of wall-clock time. The spectral method of Shi & Ioannidis (2025) is substantially slower: at every outer iteration it first recovers a ranking of the risk scores as the stationary distribution of a Markov chain—a comparatively cheap step—and then trains a neural network to regress onto those scores, so it performs many more forward and backward passes than the single-loss methods, pushing its runtime to roughly three hours. We also refer the reader to Appendix E, where we further evaluate the performance of these methods under stratification.

On discrimination the objectives are statistically indistinguishable on this cohort: Harrell's and Uno's C differ by less than one standard deviation across the 8 splits (Table 2), in line with recent cautions that the concordance index is often too coarse to separate survival models (Lillelund et al., 2025). The differences surface on the other axes. At b=4 Bradley–Terry attains the best

Table 2: Held-out **test** metrics on the public multi-cohort CT dataset (n=643) at fixed learning rate 10^{-3}, using the same 3D CNN backbone for all losses: Cox partial likelihood (DeepSurv (Katzman et al., 2018)), Bradley–Terry, DeepHit (Lee et al., 2018), and the spectral method of Shi & Ioannidis (2025), at global batch 4 and 8. Mean ± 1 std over 8 cohort-and-event-stratified $70/15/15$ splits. Uno's C is truncated at $\tau = 25/50/75\%$ of the maximum patient survival horizon. C/AUC higher is better; RMST-RMSE (in days) lower is better; best overall per column in bold.

Loss	b	Harrell C ↑	Uno$_{25\%}$ ↑	Uno$_{50\%}$ ↑	Uno$_{75\%}$ ↑	IPCW-AUC↑	RMST-RMSE(d)↓
Cox (DeepSurv)	4	0.614 ±.028	0.617 ±.029	0.615 ±.026	0.607 ±.026	0.639 ±.039	959 ±55
BT	4	**0.629** ±.028	**0.634** ±.027	**0.629** ±.025	**0.623** ±.032	**0.661** ±.038	964 ±133
DeepHit	4	0.589 ±.020	0.588 ±.024	0.592 ±.018	0.586 ±.019	0.604 ±.026	961 ±95
Spectral	4	0.595 ±.043	0.602 ±.041	0.595 ±.045	0.594 ±.039	0.622 ±.053	1167 ±302
Cox (DeepSurv)	8	0.597 ±.016	0.601 ±.017	0.597 ±.015	0.590 ±.019	0.619 ±.025	977 ±71
BT	8	0.604 ±.033	0.608 ±.036	0.604 ±.030	0.599 ±.031	0.628 ±.050	**932** ±62
DeepHit	8	0.617 ±.024	0.621 ±.022	0.617 ±.023	0.611 ±.025	0.641 ±.033	1018 ±164
Spectral	8	0.585 ±.041	0.592 ±.044	0.583 ±.038	0.580 ±.039	0.610 ±.060	1073 ±355

IPCW-AUC (0.661 vs. 0.639 for Cox), and it yields the strongest risk stratification: splitting the held-out cohort by predicted risk, Bradley–Terry separates the Kaplan–Meier curves more sharply than every baseline (two-group hazard ratio 2.06 vs. 1.92 for Cox, 1.79 for DeepHit and 1.75 for Spectral; three-group high-versus-low 3.12 vs. 2.54 for Cox), which is precisely the property used to define clinical risk groups (Appendix E).

5.2 Graded preference labels

Because the Bradley–Terry objective is a pairwise preference model, it accepts graded feedback for free: a pair in which the two patients die far apart is a stronger preference than one in which they die close together. For every comparable pair (i, j) with $E_i < E_j$ and $\Delta_i = 1$ we form the event-time gap $g_{ij} = E_j - E_i$ and turn it into a training signal in three ways, all sharing the same linear risk $r = \theta^\top x$ so they are directly comparable: (i) hard binary BT, $\ell = \mathrm{softplus}(-z)$ with $z = r_i - r_j$; (ii) soft, a graded target $y_{ij} = \frac{1}{2} + \frac{1}{2}\widehat{F}(g_{ij})$ (empirical gap-CDF on the training set) trained with cross-entropy $\mathrm{BCE}(\sigma(z), y)$ — the margin heuristic; and (iii) ordinal, the principled cumulative-link (ordered-logit) model of Afsharrad et al. (2026): bin g_{ij} into K ordered levels and fit $P(\text{level} \le k \mid z) = \sigma(c_k - z)$ with learnable thresholds $c_0 < \cdots < c_{K-2}$. We add a linear Cox proportional-hazards model as an external anchor. We treat these graded targets purely as heuristics: the gap g_{ij} is computed from observed times and we apply no censoring correction. The ordinal levels use calendar cutpoints at $\{1, 2, 3\}$ years (four ordered levels), the finest resolution that yields non-degenerate bins on these cohorts. We evaluate on METABRIC (n=1904, d=9), SUPPORT (n=8873, d=14) and GBSG (n=2232, d=7) with $70/15/15$ stratified splits standardised on the training fold, 20 seeds, model selection by validation Harrell C, and up to 2×10^5 subsampled comparable pairs per fit. We report Harrell C, Uno C, IPCW-AUC, the integrated Brier score (IBS), and the RMST-RMSE (root-mean-square error of the predicted restricted mean survival time against observed event times), averaged across seeds. In summary, the graded soft target yields a substantial reduction in the RMST-RMSE — on the breast-cancer cohorts an improvement on the order of 1–3 months of survival-time accuracy — at negligible ranking cost, whereas the principled ordinal model does not transfer to this setting. Owing to space constraints, we defer the full results table and analysis to Appendix D.

6 Conclusion and Limitations

We showed that a Bradley–Terry loss is a drop-in surrogate for the Cox partial likelihood: it gives unbiased minibatch gradients in the $d \gg n$ small-batch regime, matches Cox on discrimination while stratifying risk more sharply, and inherits the graded-label toolbox of preference learning. Its main limitations are the reliance on proportional hazards, graded-label results shown only on linear models, and a single moderate imaging cohort (n=643); multimodal and longitudinal survival are natural next steps.

AI USE STATEMENT

We used generative AI tools in two capacities: (i) polishing the writing of this manuscript (grammar, clarity, and phrasing), and (ii) assisting with software development, including writing, refactoring, and debugging the code used for our experiments and analyses. We did not use generative AI tools for the conceptual contributions of this work: the derivation of our method, and the design and selection of the experiments were carried out by the authors. All AI-assisted text was reviewed and edited by the authors, and all AI-assisted code was read, tested, and verified for correctness by the authors before being used to produce any reported result. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI. This disclosure follows the ICLR 2027 AI Policy for Authors.

REPRODUCIBILITY STATEMENT

We provide all details needed to reproduce our results, together with anonymized code, at `https://anonymous.4open.science/r/deepSurvivalAnalysis-D58F/README.md`.

REFERENCES

Odd O Aalen, Ørnulf Borgan, and Håkon K Gjessing. Survival and event history analysis: a process point of view. Springer, 2008.

Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications, 5(1):4006, 2014.

Amirhossein Afsharrad, Ruida Zhou, Luca Viano, Sanjay Lall, and Mohammad Ghavamzadeh. Beyond binary preferences: A principled framework for reward modeling with ordinal feedback. arXiv preprint arXiv:2603.02232, 2026.

Brad Albertina, Mark Watson, Chandra Holback, Rose Jarosz, Shanah Kirk, Yueh Lee, Kimberly Rieger-Christ, and John Lemmerman. The cancer genome atlas lung adenocarcinoma collection (tcga-luad). The Cancer Imaging Arch, 2016.

Shaimaa Bakr, Olivier Gevaert, Sebastian Echegaray, Kelsey Ayers, Mu Zhou, Majid Shafiq, Hong Zheng, Jalen Anthony Benson, Weiruo Zhang, Ann NC Leung, et al. A radiogenomic dataset of non-small cell lung cancer. Scientific data, 5(1):180202, 2018.

Ralf Bender, Thomas Augustin, and Maria Blettner. Generating survival times to simulate cox proportional hazards models. Statistics in medicine, 24(11):1713–1723, 2005.

Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. Biometrika, 39(3/4):324–345, 1952.

M Sc Kevin J Carroll. On the use and utility of the weibull model in the analysis of survival data. Controlled clinical trials, 24(6):682–701, 2003.

George H Chen. An introduction to deep survival analysis models for predicting time-to-event outcomes. Foundations and Trends in Machine Learning, 17(6):921–1100, 2024.

Travers Ching, Xun Zhu, and Lana X Garmire. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. PLoS computational biology, 14(4): e1006076, 2018.

Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of digital imaging, 26(6): 1045–1057, 2013.

David R Cox. Regression models and life-tables. Journal of the royal statistical society: Series B (methodological), 34(2):187–202, 1972.

Christina Curtis, Sohrab P Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M Rueda, Mark J Dunning, Doug Speed, Andy G Lynch, Shamith Samarajiwa, Yinyin Yuan, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, 486(7403): 346–352, 2012.

Claudia Czado and Florian Rudolph. Application of survival analysis methods to long-term care insurance. Insurance: Mathematics and Economics, 31(3):395–413, 2002.

David Faraggi and Richard Simon. A neural network model for survival data. Statistics in medicine, 14(1):73–82, 1995.

Stephane Fotso. Deep neural networks for survival analysis based on a multi-task framework. arXiv preprint arXiv:1801.05512, 2018.

Brandon George, Samantha Seals, and Inmaculada Aban. Survival analysis and regression models. Journal of nuclear cardiology, 21(4):686–694, 2014.

Adrian Gepp and Kuldeep Kumar. The role of survival analysis in financial distress prediction. International research journal of finance and economics, 16(16):13–34, 2008.

Erika Graf, Claudia Schmoor, Willi Sauerbrei, and Martin Schumacher. Assessment and comparison of prognostic classification schemes for survival data. Statistics in medicine, 18(17-18):2529–2545, 1999.

Olya Grove, Anders E Berglund, Matthew B Schabath, Hugo JWL Aerts, Andre Dekker, Hua Wang, Emmanuel Rios Velazquez, Philippe Lambin, Yuhua Gu, Yoganand Balagurunathan, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PloS one, 10(3):e0118261, 2015.

Balázs Győrffy. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. 2021.

Balazs Győrffy, Andras Lanczky, Aron C Eklund, Carsten Denkert, Jan Budczies, Qiyuan Li, and Zoltan Szallasi. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast cancer research and treatment, 123(3):725–731, 2010.

Humza Haider, Bret Hoehn, Sarah Davis, and Russell Greiner. Effective ways to build and evaluate individual survival distributions. Journal of Machine Learning Research, 21(85):1–63, 2020.

Frank E Harrell Jr, Kerry L Lee, Robert M Califf, David B Pryor, and Robert A Rosati. Regression modelling strategies for improved prognostic prediction. Statistics in medicine, 3(2):143–152, 1984.

Tina Harrison and Jake Ansell. Customer retention in the insurance industry: Using survival analysis to predict cross-selling opportunities. Journal of Financial Services Marketing, 6(3):229–239, 2002.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Zepeng Frazier Huo, Jason Fries, Alejandro Lozano, Jeya Maria Jose Valanarasu, Ethan Steinberg, Louis Blankemeier, Akshay Chaudhari, Curtis Langlotz, and Nigam Shah. Time-to-event pretraining for 3d medical imaging. In International Conference on Learning Representations, volume 2025, pp. 84072–84108, 2025.

John D Kalbfleisch and Ross L Prentice. The statistical analysis of failure time data. John Wiley & Sons, 2002.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282):457–481, 1958.

Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology, 18(1):24, 2018.

S Kirk, Y Lee, P Kumar, J Filippini, B Albertina, M Watson, K Rieger-Christ, and J Lemmerman. The cancer genome atlas lung squamous cell carcinoma collection (tcga-lusc)(version 4)[data set]. The Cancer Imaging Archive, 15, 2016.

William A Knaus, Frank E Harrell, Joanne Lynn, Lee Goldman, Russell S Phillips, Alfred F Connors, Neal V Dawson, William J Fulkerson, Robert M Califf, Norman Desbiens, et al. The support prognostic model: Objective estimates of survival for seriously ill hospitalized adults. Annals of internal medicine, 122(3):191–203, 1995.

Ram Kulkarni, Ivan Wong, Judith Zachariasen, Chris Goldfinger, and Martin Lawrence. Statistical analyses of great earthquake recurrence along the cascadia subduction zone. Bulletin of the Seismological Society of America, 103(6):3205–3221, 2013.

Håvard Kvamme, Ørnulf Borgan, and Ida Scheel. Time-to-event prediction with neural networks and cox regression. Journal of Machine Learning Research, 20(129):1–30, 2019.

Bart Larivière and Dirk Van den Poel. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. Expert Systems with Applications, 27(2):277–285, 2004.

Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.

Christian Marius Lillelund, Shi-ang Qi, Russell Greiner, and Christian Fischer Pedersen. Stop chasing the c-index: This is how we should evaluate our survival models. arXiv preprint arXiv:2506.02075, 2025.

R Duncan Luce. Individual choice behavior: A theoretical analysis. Courier Corporation, 2005.

Lucas Maystre and Daniel Russo. Temporally-consistent survival analysis. Advances in Neural Information Processing Systems, 35:10671–10683, 2022.

Xenia Miscouridou, Adler Perotte, Noémie Elhadad, and Rajesh Ranganath. Deep survival analysis: Nonparametrics and missingness. In Machine Learning for Healthcare Conference, pp. 244–256. PMLR, 2018.

Chirag Nagpal. Preference models assume proportional hazards of utilities. arXiv preprint arXiv:2508.13189, 2025.

Chirag Nagpal, Xinyu Li, and Artur Dubrawski. Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. IEEE Journal of Biomedical and Health Informatics, 25(8):3163–3175, 2021.

África Periáñez, Alain Saas, Anna Guitart, and Colin Magne. Churn prediction in mobile social games: Towards a complete assessment using survival ensembles. In 2016 IEEE international conference on data science and advanced analytics (DSAA), pp. 564–573. IEEE, 2016.

Robin L Plackett. The analysis of permutations. Journal of the Royal Statistical Society Series C: Applied Statistics, 24(2):193–202, 1975.

Shankar Prinja, Nidhi Gupta, and Ramesh Verma. Censoring in clinical trials: review of survival analysis techniques. Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine, 35(2):217, 2010.

Rajesh Ranganath, Adler Perotte, Noémie Elhadad, and David Blei. Deep survival analysis. In Machine Learning for Healthcare Conference, pp. 101–114. PMLR, 2016.

P Royston. The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. Statistica Neerlandica, 55(1):89–104, 2001.

Chiharu Sako, Brenda F Kurland, Taly G Schmidt, Jing He, Marius De Groot, Arpan A Patel, Dwight H Owen, Arya Amini, Brendan D Curti, Ronan J Kelly, et al. Deep-learning serial ct prediction of survival in immunotherapy-treated non–small cell lung cancer. JAMA Network Open, 9(1):e2555759, 2026.

Martin Schumacher, Guenter Bastert, Henryk Bojar, Karl Hübner, Manfred Olschewski, Willi Sauerbrei, Cornelia Schmoor, Christa Beyerle, Rolf L Neumann, and Heinz F Rauschecker. Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. german breast cancer study group. Journal of Clinical Oncology, 12(10): 2086–2093, 1994.

Chengzhi Shi and Stratis Ioannidis. Spectral survival analysis. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2, pp. 2538–2549, 2025.

Michelle Shu, Richard Strong Bowen, Charles Herrmann, Gengmo Qi, Michele Santacatterina, and Ramin Zabih. Deep survival analysis with longitudinal x-rays for covid-19. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4046–4055, 2021.

Maria Stepanova and Lyn Thomas. Survival analysis methods for personal loan data. Operations Research, 50(2):277–289, 2002.

Yuhua Su and Mai Zhou. On a connection between the bradley–terry model and the cox proportional hazards model. Statistics & probability letters, 76(7):698–702, 2006.

Weijing Tang, Jiaqi Ma, Qiaozhu Mei, and Ji Zhu. Soden: A scalable continuous-time survival model through ordinary differential equation networks. Journal of Machine Learning Research, 23(34):1–29, 2022.

Aliasghar Tarkhan and Noah Simon. Bigsurvsgd: Big survival data analysis via stochastic gradient descent. arXiv preprint arXiv:2003.00116, 2020.

Aliasghar Tarkhan, Noah Simon, Thomas Bengtsson, Kien Nguyen, and Jian Dai. Survival prediction using deep learning. In Survival Prediction-Algorithms, Challenges and Applications, pp. 207–214. PMLR, 2021.

Terry M Therneau and Patricia M Grambsch. The cox model. In Modeling survival data: extending the Cox model, pp. 39–77. Springer, 2000.

Felipe Soares Torres, Shazia Akbar, Srinivas Raman, Kazuhiro Yasufuku, Carola Schmidt, Ahmed Hosny, Felix Baldauf-Lenschen, and Natasha B Leighl. End-to-end non–small-cell lung cancer prognostication using deep learning applied to pretreatment computed tomography. JCO Clinical Cancer Informatics, 5:1141–1150, 2021.

Joseph M Unger, William E Barlow, Diane P Martin, Scott D Ramsey, Michael LeBlanc, Ruth Etzioni, and Dawn L Hershman. Comparison of survival outcomes among cancer patients treated in and out of clinical trials. Journal of the National Cancer Institute, 106(3):dju002, 2014.

Hajime Uno, Tianxi Cai, Michael J Pencina, Ralph B D'Agostino, and Lee-Jen Wei. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in medicine, 30(10):1105–1117, 2011.

Wessel N Van Wieringen, David Kun, Regina Hampel, and Anne-Laure Boulesteix. Survival prediction using gene expression data: a review and comparison. Computational statistics & data analysis, 53(5):1590–1603, 2009.

Laura J Van't Veer, Hongyue Dai, Marc J Van De Vijver, Yudong D He, Augustinus AM Hart, Mao Mao, Hans L Peterse, Karin Van Der Kooy, Matthew J Marton, Anke T Witteveen, et al. Gene expression profiling predicts clinical outcome of breast cancer. nature, 415(6871):530–536, 2002.

Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence, 5(5):e230024, 2023.

Lee-Jen Wei. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. Statistics in medicine, 11(14-15):1871–1879, 1992.

Simon Wiegrebe, Philipp Kopper, Raphael Sonabend, Bernd Bischl, and Andreas Bender. Deep learning for survival analysis: a review. Artificial Intelligence Review, 57(3):65, 2024.

Qixian Zhong, Jonas W Mueller, and Jane-Ling Wang. Deep extended hazard models for survival analysis. Advances in Neural Information Processing Systems, 34:15111–15124, 2021.

Xinliang Zhu, Jiawen Yao, and Junzhou Huang. Deep convolutional neural network for survival analysis with pathological images. In 2016 IEEE international conference on bioinformatics and biomedicine (BIBM), pp. 544–547. IEEE, 2016.

APPENDIX

A DEFINITIONS AND PROOFS FOR THE BRADLEY–TERRY CONNECTION

A.1 DATA, HAZARDS, AND NOTATION

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be the covariate space and let P_X denote the marginal law of the covariate vector $\mathbf{X}$. The parameter space is Θ, and $r_\theta : \mathcal{X} \to \mathbb{R}$ is a measurable log-risk function. In the linear model, $\Theta \subseteq \mathbb{R}^d$ and $r_\theta(x) = x^\top \theta$.

Assumption A.1 (Sampling, censoring, and model specification)**.** *The latent triples $(\mathbf{X}_i, Y_i, C_i)$ are independent and identically distributed. The nonnegative event time Y_i is conditionally absolutely continuous given $\mathbf{X}_i$, and the censoring time C_i takes values in $[0, \infty]$. We assume $Y_i \perp C_i \mid \mathbf{X}_i$ and observe $Z_i = (\mathbf{X}_i, E_i, \Delta_i)$, where $E_i = \min(Y_i, C_i)$ and $\Delta_i = \mathbf{1}\{Y_i \le C_i\}$. The conditional event hazard is*

$$\lambda(t \mid x) = \lambda_0(t) e^{r^\star(x)}, \qquad r^\star = r_{\theta^\star} \text{ for some } \theta^\star \in \Theta.$$

Here $\lambda_0 : [0, \infty) \to [0, \infty)$ is a measurable, locally integrable baseline hazard, its cumulative hazard is $\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$, and $\Lambda_0(t) \to \infty$ as $t \to \infty$. Scores are finite and $\mathbb{E}|r_\theta(\mathbf{X})| < \infty$ for every $\theta \in \Theta$.

The integrability condition ensures the expectations appearing below are finite. Censoring may depend on covariates and may include a deterministic administrative cutoff. Ties involving a failure have probability zero under these assumptions; ties between censoring times cause no difficulty.

For fixed covariates x_i, x_j, define, for $k \in \{i, j\}$,

$$F_k(t) = \mathbb{P}(Y_k \le t \mid \mathbf{X}_k = x_k), \qquad f_k(t) = \frac{d}{dt} F_k(t),$$
$$S_k(t) = 1 - F_k(t) = \mathbb{P}(Y_k > t \mid \mathbf{X}_k = x_k), \qquad \lambda_k(t) = \frac{f_k(t)}{S_k(t)},$$
$$r_k^\star = r^\star(x_k), \qquad a_k = e^{r_k^\star} > 0.$$

Thus F_k, f_k, S_k, and λ_k are the conditional event-time CDF, density, survival function, and hazard, respectively. Densities and hazard identities are understood almost everywhere in time. Since

$$\frac{d}{dt} \log S_k(t) = -\lambda_k(t) = -a_k \lambda_0(t), \qquad S_k(0) = 1,$$

integration and differentiation give

$$S_k(t) = e^{-a_k \Lambda_0(t)}, \qquad F_k(t) = 1 - e^{-a_k \Lambda_0(t)}, \qquad f_k(t) = a_k \lambda_0(t) e^{-a_k \Lambda_0(t)}. \tag{9}$$

A.2 Proof of the latent pairwise law

We first prove equation 4. Conditional on $Y_i = t$ and the covariates, the event $Y_i < Y_j$ is equivalent to $Y_j > t$. Conditional independence therefore yields

$$\begin{aligned}\mathbb{P}(Y_i < Y_j \mid \mathbf{X}_i = x_i, \mathbf{X}_j = x_j) &= \int_0^\infty \mathbb{P}(Y_j > t \mid \mathbf{X}_j = x_j) f_i(t)\, dt \\ &= \int_0^\infty [1 - F_j(t)] f_i(t)\, dt. \end{aligned} \tag{10}$$

Substituting the CDF and density from equation 9,

$$\begin{aligned}\mathbb{P}(Y_i < Y_j \mid \mathbf{X}_i = x_i, \mathbf{X}_j = x_j) &= \int_0^\infty e^{-a_j \Lambda_0(t)} \left[a_i \lambda_0(t) e^{-a_i \Lambda_0(t)}\right] dt \\ &= a_i \int_0^\infty \lambda_0(t) e^{-(a_i + a_j)\Lambda_0(t)}\, dt \\ &= a_i \int_0^\infty e^{-(a_i + a_j)u}\, du \\ &= \frac{a_i}{a_i + a_j} = \frac{e^{r_i^\star}}{e^{r_i^\star} + e^{r_j^\star}} = \sigma(r_i^\star - r_j^\star).\end{aligned}$$

The third line uses $u = \Lambda_0(t)$, $du = \lambda_0(t)\, dt$, $\Lambda_0(0) = 0$, and $\Lambda_0(\infty) = \infty$. This change of variables remains valid if the cumulative hazard has flat intervals, since those intervals contribute zero to the integral. For the reverse ordering,

$$\mathbb{P}(Y_i > Y_j \mid \mathbf{X}_i = x_i, \mathbf{X}_j = x_j) = \sigma(r_j^\star - r_i^\star).$$

Remark 1 (Why the cumulative-hazard assumption is needed). If $\Lambda_0(\infty) = L < \infty$, the event-time law is defective, with positive probability of $Y_k = \infty$. The calculation instead gives

$$\mathbb{P}(Y_i < Y_j \mid x_i, x_j) = \frac{a_i}{a_i + a_j}\big(1 - e^{-(a_i + a_j)L}\big).$$

Consequently the unconditional latent law cannot be asserted for an arbitrary baseline without an assumption excluding infinite event times.

A.3 Proof of the comparable-pair law under censoring

Define the conditional censoring survival function $G_k(t) = \mathbb{P}(C_k > t \mid \mathbf{X}_k = x_k)$, and define

$$A_{ij} = \{\Delta_i = 1, E_i < E_j\}, \qquad A_{ji} = \{\Delta_j = 1, E_j < E_i\}, \qquad Q_{ij} = A_{ij} \cup A_{ji}.$$

The events A_{ij} and A_{ji} are disjoint. Apart from null tie events, A_{ij} occurs precisely when $Y_i = t$, $C_i > t$, $Y_j > t$, and $C_j > t$ for some t. Independence across subjects and $Y_k \perp C_k \mid \mathbf{X}_k$ therefore imply

$$\begin{aligned}\mathbb{P}(A_{ij} \mid x_i, x_j) &= \int_0^\infty f_i(t) S_j(t) G_i(t) G_j(t)\, dt \\ &= a_i K(x_i, x_j),\end{aligned}$$

where

$$K(x_i, x_j) = \int_0^\infty \lambda_0(t) e^{-(a_i + a_j)\Lambda_0(t)} G_i(t) G_j(t)\, dt. \tag{11}$$

Interchanging i and j gives $\mathbb{P}(A_{ji} \mid x_i, x_j) = a_j K(x_i, x_j)$. Hence, writing

$$q(x_i, x_j) = \mathbb{P}(Q_{ij} \mid x_i, x_j) = (a_i + a_j) K(x_i, x_j), \tag{12}$$

we obtain, whenever $q(x_i, x_j) > 0$,

$$\mathbb{P}(Y_i < Y_j \mid x_i, x_j, Q_{ij}) = \frac{\mathbb{P}(A_{ij} \mid x_i, x_j)}{q(x_i, x_j)} = \frac{a_i}{a_i + a_j} = \sigma(r_i^\star - r_j^\star).$$

This proves equation 5. The censoring distributions affect which covariate pairs are observed, but cancel from the conditional ordering probability. No censoring weights are needed for this correctly specified conditional pairwise model. This conclusion specifically uses comparable pairs; restricting to pairs with both events observed is a different selection rule and is not justified by this calculation.

Algorithm 1 Bradley–Terry survival training with patient minibatches

Require: Dataset $\mathcal{D}$ with n subjects; score function r_θ; epochs T; patient-batch size $2 \le B \le n$; learning rate $\eta > 0$

Initialize θ; set $S \leftarrow \lceil n/B \rceil$
for $t = 1, \dots, T$ **do**
  for $s = 1, \dots, S$ **do**
    Sample a fresh batch $\mathcal{B}$ of B distinct patients uniformly from $\mathcal{D}$
    Compute $r_i = r_\theta(\mathbf{X}_i)$ for each $i \in \mathcal{B}$
    $\mathcal{E}_\mathcal{B} \leftarrow \{(i,j) \in \mathcal{B} \times \mathcal{B} : E_i < E_j,\ \Delta_i = 1\}$
    $\widehat{\mathcal{L}} \leftarrow \binom{B}{2}^{-1} \sum_{(i,j)\in\mathcal{E}_\mathcal{B}} \log(1 + \exp(r_j - r_i))$
    $\theta \leftarrow \theta - \eta \nabla_\theta \widehat{\mathcal{L}}$
  end for
end for
return θ

B Bradley-Terry SGD

Each iteration of Algorithm 1 requires a network forward and backward pass on B patients. Constructing the pairwise comparison mask and computing the pairwise losses and their derivatives with respect to the scores require $\mathcal{O}(B^2)$ additional time. A direct matrix implementation uses $\mathcal{O}(B^2)$ additional memory, beyond the network's parameters and minibatch activations.

Each epoch consists of $S = \lceil n/B \rceil$ iterations. It therefore processes $BS = \mathcal{O}(n)$ subject instances through the network and performs $\mathcal{O}(SB^2) = \mathcal{O}(nB)$ pairwise work. The additional memory for pairwise computations remains $\mathcal{O}(B^2)$.

C Additional Experimental Details

C.1 Semi-synthetic parameter recovery

We generate data from a known Cox proportional-hazards model, following the standard procedure of Bender et al. (2005). For each cohort we hold out 15% of the samples and fit a ridge-regularized Cox model on them to obtain a ground-truth parameter $\theta^\star$; this fixes a proportional-hazards ground truth that lives in the feature space of the real covariates. On the remaining 85% we regenerate event times from $\theta^\star$: for covariate x_i we draw $U_i \sim \mathrm{Unif}(0,1)$ and set $T_i = -\log(U_i)\,\exp(-x_i^\top \theta^\star)/\hat{h}_0$, where the constant baseline hazard $\hat{h}_0 = (\#\text{events})/\sum_i t_i \exp(x_i^\top \theta^\star)$ is the exponential-model maximum-likelihood estimate, and we apply independent exponential censoring. Each estimator is then trained on the regenerated data and scored by how well it recovers $\theta^\star$, through the relative ℓ_2 error $\|\hat{\theta} - \theta^\star\|_2 / \|\theta^\star\|_2$ and the cosine similarity $\langle \hat{\theta}, \theta^\star \rangle / (\|\hat{\theta}\|_2 \, \|\theta^\star\|_2)$.

The second-order baseline (Cox-standard) is the full-batch Newton–Raphson solution from `lifelines`. All stochastic estimators are trained for 100 epochs with AdamW and a small ridge penalty, and we report the Polyak-averaged iterate; numbers are mean $\pm$ std over 5 random seeds. The relative-ℓ_2 results are given in the main text (Table 1) and the matching cosine similarities in Table 3. The per-epoch convergence curves of Figure 1 instead use *plain* SGD, so as to display the raw optimization trajectory without the confound of an adaptive optimizer; there we plot the Polyak-averaged iterate, which is the estimator we report.

D A closer look at graded preference labels (Section 5.2)

Ranking is unchanged; survival-time calibration improves. All pairwise variants match Cox on discrimination: Harrell/Uno C and IPCW-AUC move by less than 0.01 between Cox, hard BT and soft BT (Table 4). The graded *soft* target leaves ranking essentially intact while significantly reducing the RMST-RMSE on all three cohorts: the model's restricted mean survival time becomes a better estimate of the realised survival time. Concretely, the error drops by 2.9 months on METABRIC ($89.1 \to 86.2$, ≈ 90 days, 3.3%), by 1.4 months on GBSG ($27.8 \to 26.4$, ≈ 43 days, 5.1%), and

Table 3: Semi-synthetic parameter recovery, cosine similarity $\langle \theta_{\mathrm{alg}}, \theta^\star \rangle / (\|\theta_{\mathrm{alg}}\|_2 \|\theta^\star\|_2)$ (higher is better), under the same generative model as Table 1, as a function of the minibatch size b (mean $\pm$ std over 5 seeds). Cox-standard is the full-batch Newton estimate. Best per column (excluding the floor) in bold. All learning-based methods recover essentially the same *direction*: the cosine values agree to within noise even where the relative ℓ_2 errors of Table 1 differ by an order of magnitude (e.g. Cox-CC), confirming that the separation observed there is one of *magnitude*, not direction.

Method	b=2	b=4	b=8	b=16
GBSG (n=2232, d=7)				
Cox-standard (full batch, b-indep.)		0.969 ± 0.018		
Cox-minibatch	0.927 ± 0.035	0.944 ± 0.026	$\mathbf{0.952} \pm 0.025$	$\mathbf{0.957} \pm 0.024$
Bradley–Terry	$\mathbf{0.941} \pm 0.031$	$\mathbf{0.948} \pm 0.028$	0.951 ± 0.026	0.952 ± 0.024
Cox-CC	0.892 ± 0.073	0.919 ± 0.051	0.938 ± 0.034	0.951 ± 0.028
DBCD (n=295, d=4919)				
Cox-standard (full batch, b-indep.)		0.555 ± 0.011		
Cox-minibatch	0.518 ± 0.022	0.541 ± 0.022	$\mathbf{0.553} \pm 0.019$	0.555 ± 0.017
Bradley–Terry	$\mathbf{0.546} \pm 0.021$	$\mathbf{0.543} \pm 0.022$	0.552 ± 0.023	$\mathbf{0.558} \pm 0.021$
Cox-CC	0.509 ± 0.024	0.515 ± 0.027	0.513 ± 0.020	0.515 ± 0.020
CHOP (n=414, d=3833)				
Cox-standard (full batch, b-indep.)		0.332 ± 0.019		
Cox-minibatch	0.268 ± 0.012	0.291 ± 0.013	$\mathbf{0.308} \pm 0.016$	$\mathbf{0.318} \pm 0.018$
Bradley–Terry	0.294 ± 0.014	0.296 ± 0.014	0.301 ± 0.014	0.307 ± 0.015
Cox-CC	$\mathbf{0.300} \pm 0.015$	$\mathbf{0.302} \pm 0.016$	0.304 ± 0.017	0.303 ± 0.016

Table 4: Graded preference labels on three survival cohorts (mean $\pm$ std over 20 seeds). C/AUC higher is better; IBS/RMST-RMSE lower is better. RMST-RMSE is in each cohort's native time unit (months for METABRIC and GBSG, days for SUPPORT). Best per column in bold. The soft graded target consistently improves RMST-RMSE at negligible ranking cost, whereas the principled ordinal model degrades ranking; Cox tracks hard BT.

Method	**Harrell** $C \uparrow$	**Uno** $C \uparrow$	**IPCW-AUC**$\uparrow$	**IBS**$\downarrow$	**RMST-RMSE**$\downarrow$
METABRIC (RMST-RMSE in months)					
Cox	$0.630 \pm .017$	$\mathbf{0.637} \pm .016$	$0.656 \pm .022$	$\mathbf{0.192} \pm .006$	90.64 ± 5.27
Hard BT	$0.631 \pm .023$	$0.624 \pm .029$	$0.671 \pm .029$	$0.193 \pm .007$	89.15 ± 4.39
Soft BT	$\mathbf{0.632} \pm .024$	$0.623 \pm .027$	$\mathbf{0.674} \pm .031$	$0.194 \pm .006$	$\mathbf{86.21} \pm 4.37$
Ordinal	$0.626 \pm .023$	$0.620 \pm .029$	$0.665 \pm .030$	$0.195 \pm .007$	89.34 ± 5.72
SUPPORT (RMST-RMSE in days)					
Cox	$0.573 \pm .007$	$0.576 \pm .007$	$0.573 \pm .012$	$\mathbf{0.232} \pm .002$	573.6 ± 16.9
Hard BT	$\mathbf{0.577} \pm .008$	$0.577 \pm .008$	$\mathbf{0.597} \pm .012$	$\mathbf{0.232} \pm .002$	569.0 ± 17.1
Soft BT	$0.575 \pm .007$	$\mathbf{0.578} \pm .007$	$0.582 \pm .012$	$0.233 \pm .001$	$\mathbf{561.0} \pm 15.6$
Ordinal	$0.552 \pm .009$	$0.558 \pm .009$	$0.534 \pm .016$	$0.236 \pm .002$	574.9 ± 16.8
GBSG (RMST-RMSE in months)					
Cox	$0.666 \pm .020$	$0.657 \pm .020$	$0.728 \pm .025$	$\mathbf{0.185} \pm .006$	$\mathbf{26.34} \pm 0.92$
Hard BT	$\mathbf{0.671} \pm .018$	$\mathbf{0.661} \pm .019$	$\mathbf{0.734} \pm .023$	$0.186 \pm .009$	27.83 ± 1.35
Soft BT	$0.669 \pm .019$	$0.659 \pm .019$	$0.732 \pm .025$	$0.189 \pm .005$	26.43 ± 0.80
Ordinal	$0.662 \pm .018$	$0.652 \pm .018$	$0.724 \pm .024$	$0.193 \pm .005$	26.84 ± 0.84

by 8 days on SUPPORT ($569 \to 561$, 1.4%). In other words, without changing who is predicted to be at higher risk, graded labels sharpen the survival-time calibration: they recover roughly three months of survival-time accuracy on the breast-cancer cohorts and about a week on the critical-care cohort. This is consistent with the mechanism that hard BT, by driving perfectly separable pairs to

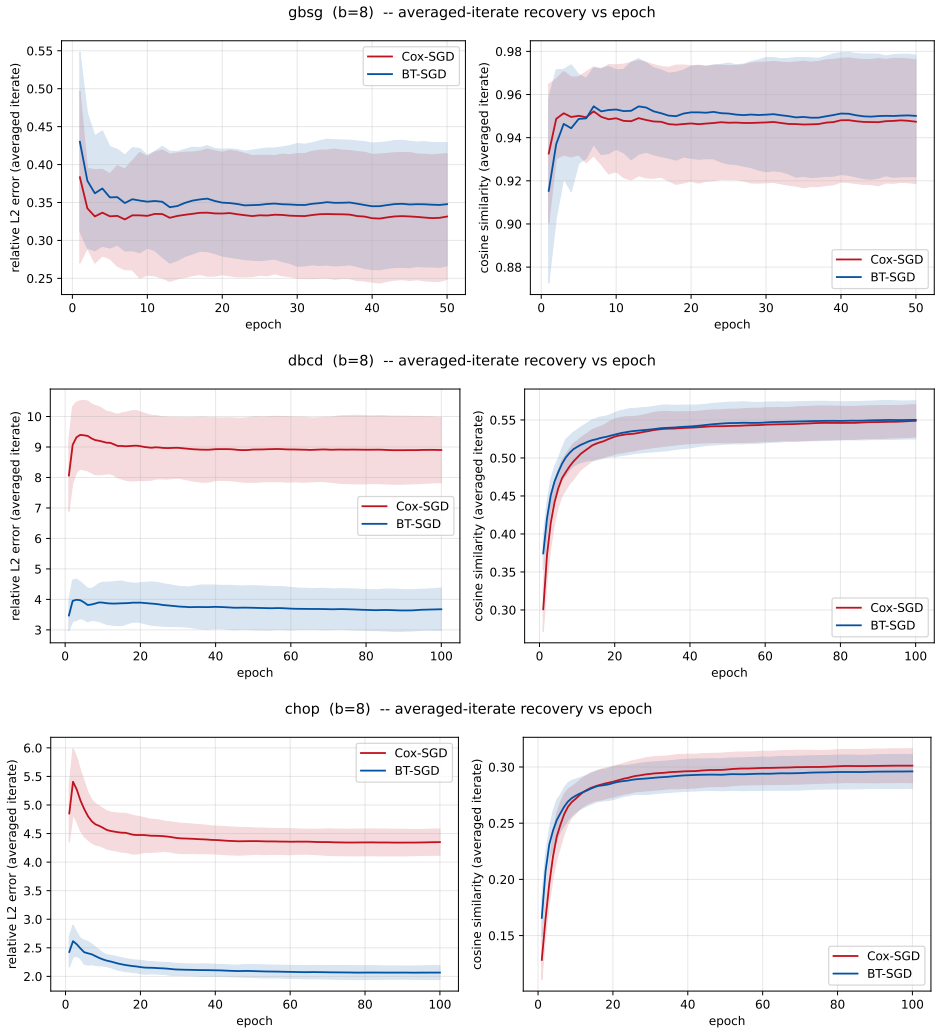


Figure 1: Per-epoch parameter recovery (Polyak-averaged iterate) of Cox-SGD vs BT-SGD at batch size b=8, on GBSG (top), DBCD (middle), and CHOP (bottom). Left column: relative ℓ_2 error; right column: cosine similarity to $\theta^\star$. Curves are mean $\pm$ std over 5 seeds. On the low-dimensional GBSG cohort the two objectives coincide; on the high-dimensional $d \gg n$ cohorts they track the same cosine throughout training while BT converges to a markedly lower relative ℓ_2 error, confirming that Cox-SGD's disadvantage is a persistent magnitude effect rather than an artefact of a particular stopping epoch.

extreme scores, over-disperses the risk distribution and thus mis-calibrates the survival-time scale that RMST integrates over.

Ordinal regression does not transfer. The principled ordinal (cumulative-link) model of Afsharrad et al. (2026), which is the state of the art for graded feedback in reward modelling, fails here: it yields no RMST improvement over hard BT and significantly degrades ranking, and the degradation grows with the informativeness of the grading — on SUPPORT, whose year-level bins are the most evenly populated, IPCW-AUC collapses from 0.597 to 0.534 (near chance). The reason is structural: the cumulative-link loss forces the score gap $z = r_i - r_j$ to match the magnitude of the time gap, but under proportional hazards z encodes the log hazard ratio, which is only monotone—not cardinal—in the survival-time gap. Demanding cardinal agreement therefore suppresses z on close-but-clearly-ordered pairs and sacrifices discrimination, exactly where the grading is richest.

Table 5: Wall-clock training time on the public multi-cohort CT dataset (NVIDIA A100, 40 GB), at global batch 4 and 8. The single-loss methods finish within an hour, whereas the spectral method's inner regression step makes it roughly an order of magnitude slower. Best per column in bold.

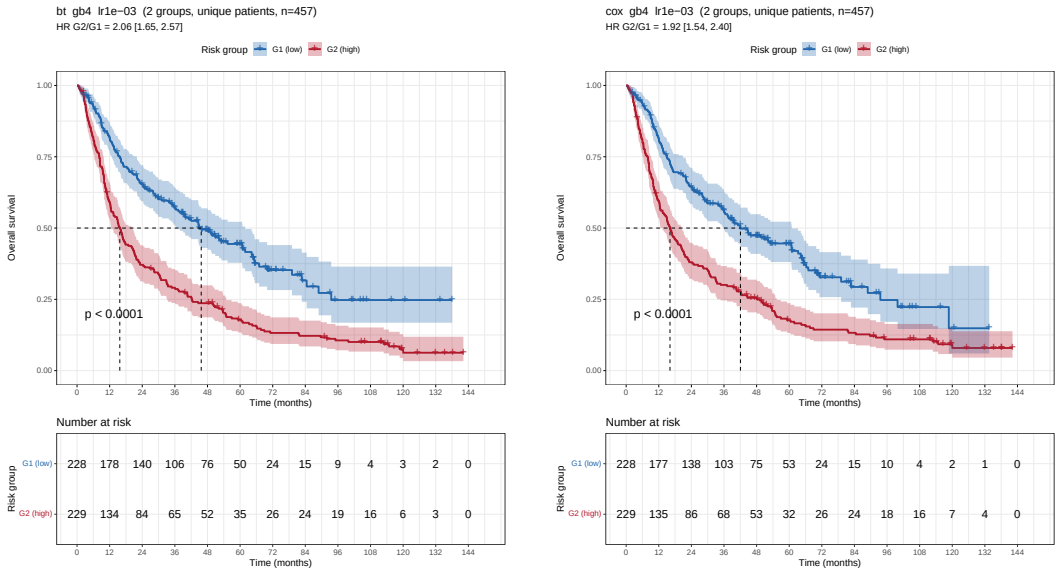
Method	b=4	b=8
Cox (DeepSurv)	**46** min	51 min
BT	51 min	**38** min
DeepHit	53 min	48 min
Spectral	$\sim$3 h	$\sim$3 h

The practical lesson is that survival benefits from graded preference labels, but through a simple soft target rather than ordinal regression.

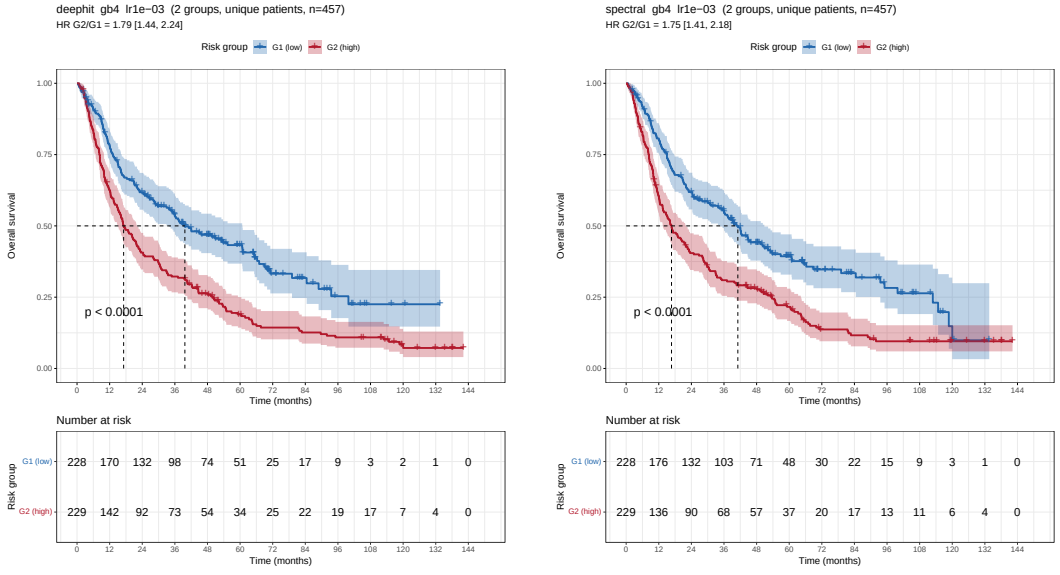
E STRATIFICATION

As described in the main text, another objective that practitioners care about is *stratification*. It is usually assessed visually by splitting the cohort into distinct risk groups and plotting their Kaplan–Meier survival curves (Aalen et al., 2008; Sako et al., 2026; Torres et al., 2021). For instance, having trained a model and obtained a risk function $\hat{r}(\cdot) : \mathcal{X} \to \mathbb{R}$, we order the patients $\{\mathbf{X}_i\}_{i=1}^n$ by their risk score and split them into the lower and upper halves, then plot the survival curve of each half. A good risk score should stratify the cohort well: the survival curve of the low-risk group should lie clearly above that of the high-risk group—equivalently, the low-risk group should have a higher probability of survival and, in particular, a longer median survival time. Broadly, the larger the gap between the two curves, the better the stratification. The same property can also be quantified numerically by the *hazard ratio* between the two groups, which compares their instantaneous event rates; the further it is from one, the stronger the separation. As we mentioned before, we ran experiments for 8 random seeds (different train-validation-test splits), and therefore if a patient appears more than once in aggregated test set, we averaged their scores.

Visually, for two risk groups, all four methods stratify the cohort well, recovering statistically separated survival curves over most of the time horizon (Figure 2). The Bradley–Terry and Cox models separate the two groups earlier, already from the initial stages of the time horizon, with the Bradley–Terry model achieving the best stratification overall.

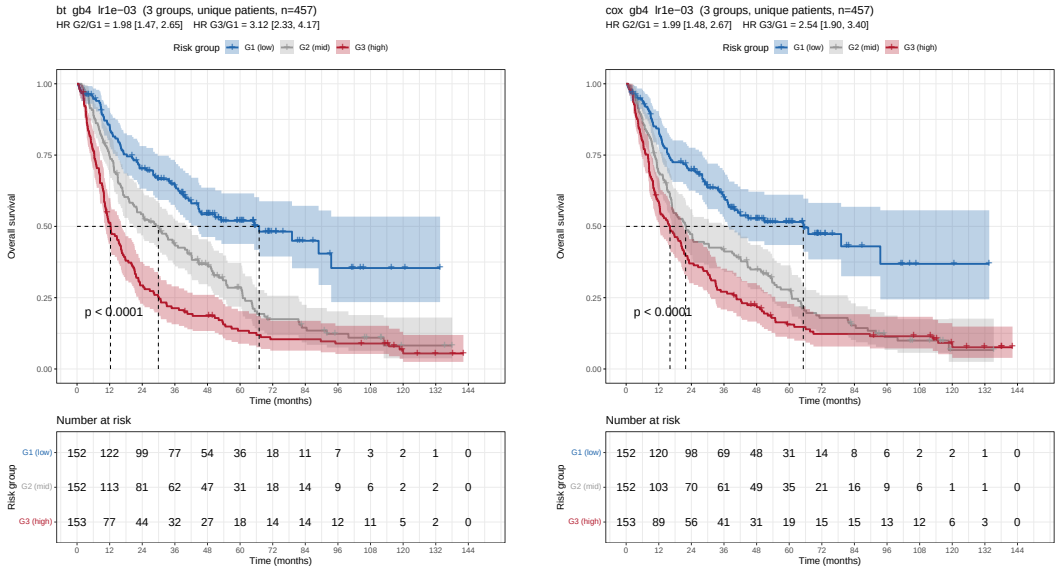


(a) Bradley–Terry (b) Cox partial likelihood

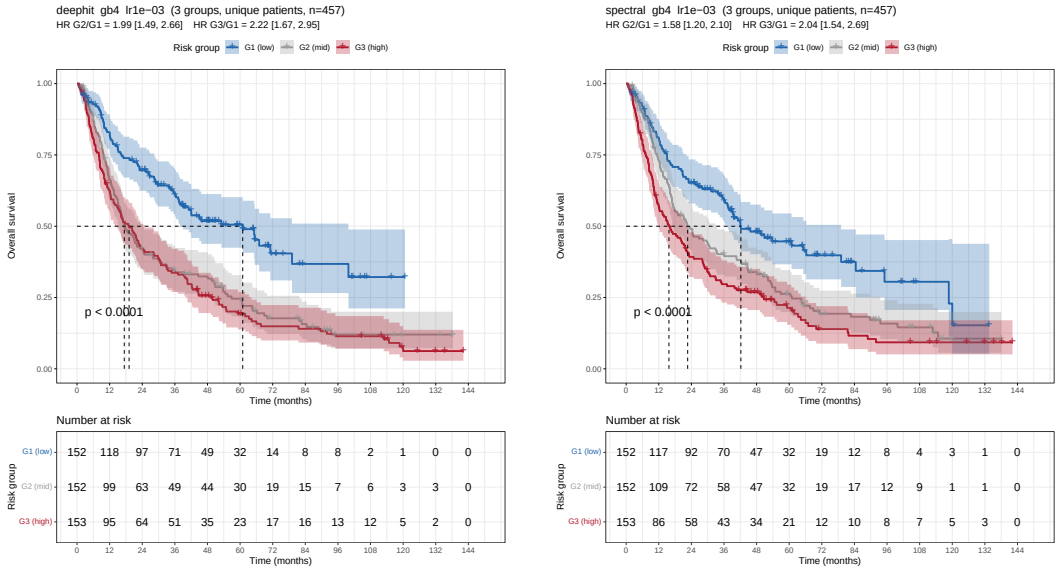


(c) DeepHit (d) Spectral

Figure 2: Kaplan–Meier survival curves for the low- and high-risk halves of the public multi-cohort CT test set (risk-group size 2), using the risk scores obtained from (a) the Bradley–Terry loss, (b) the Cox partial likelihood, (c) DeepHit, and (d) the spectral method of Shi & Ioannidis (2025) (all trained with global batch size 4 and learning rate 10^{-3}). A larger gap between the two curves indicates stronger stratification.

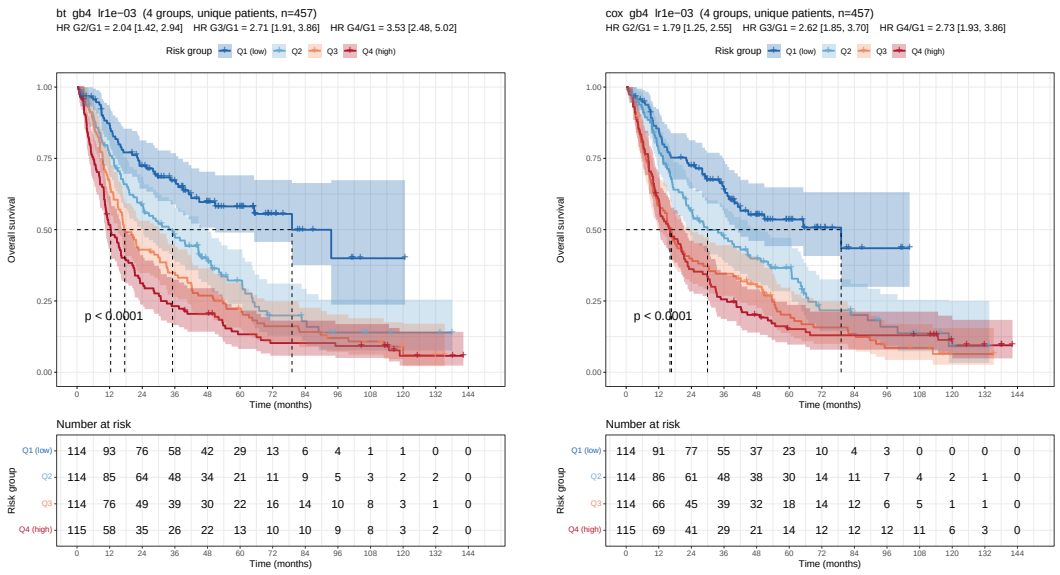


(a) Bradley–Terry (b) Cox partial likelihood

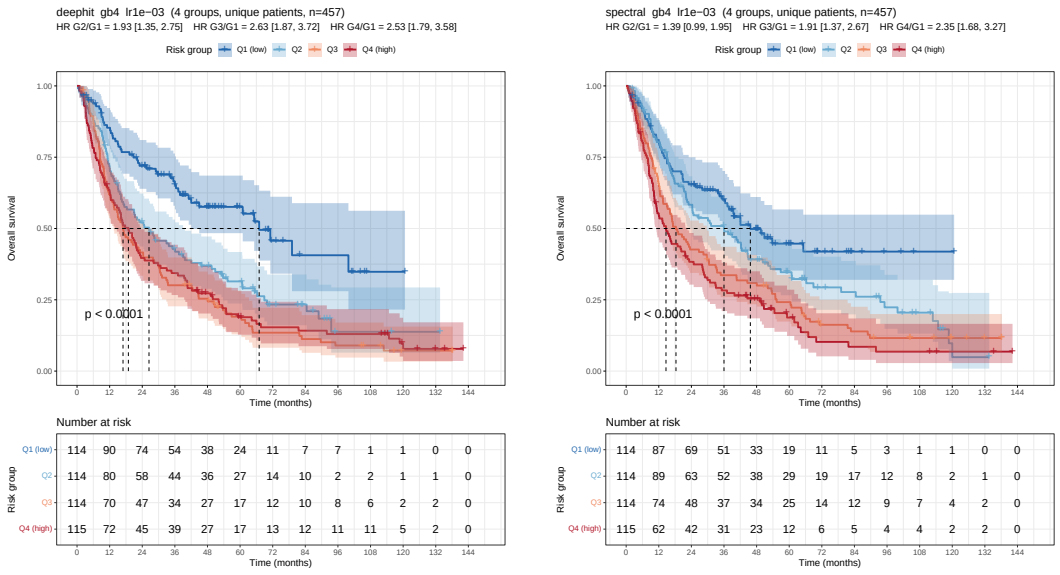


(c) DeepHit (d) Spectral

Figure 3: While all methods stratify the cohort well into two risk groups, DeepHit fails to separate three risk groups. In contrast, the Bradley–Terry method cleanly distinguishes low-, moderate-, and high-risk patients up to 60 months—a valuable property that can help detect imbalances between treatment arms in a clinical trial.



(a) Bradley–Terry (b) Cox partial likelihood



(c) DeepHit (d) Spectral

Figure 4: None of the four methods cleanly separates four risk groups over a substantial portion of the time horizon. Nevertheless, Cox and DeepHit visibly collapse some of the groups together, whereas the Bradley–Terry and Spectral order the group median survival times correctly. Overall, the Bradley–Terry method remains the most robust across all risk groups.