

Article

Graph-Regularized Survival Learning with Optimal Transport for Individual Employee Turnover Forecasting

Xia Wang, Qiuyun Guo and Wenting Wang *

Department of Economics and Management, Taiyuan Institute of Technology, Taiyuan 030008, China; wangx@tit.edu.cn (X.W.); guoqy@tit.edu.cn (Q.G.)

* Correspondence: wangwt@tit.edu.cn

Abstract

This study forecasts attrition or job-change risk for individual employee records; department- and city-level quantities are secondary aggregates of those predictions. We propose Graph-Regularized Survival Learning with Optimal Transport (GRSOT), which combines a survival-style risk objective, entropic optimal transport, and Laplacian regularization over a static employee-similarity graph. The general formulation accepts event times and censoring indicators, whereas the two public benchmarks used here are cross-sectional and provide neither longitudinal employment trajectories nor observed interaction networks. We therefore construct proxy time/event pairs from tenure-like variables and binary labels, and interpret the empirical results as survival-inspired risk forecasting rather than evidence from genuine longitudinal histories. The graph performs similarity-based risk smoothing and does not represent causal peer influence. The transport module targets covariate shift between controlled source and target partitions and assumes that the conditional proxy outcome mechanism remains stable. Relative to OT-Surv, GRSOT increases the C-index from 0.812 to 0.836 and reduces expected calibration error from 0.045 to 0.037 on IBM-HR; on HR-JC, the corresponding values improve from 0.781 to 0.803 and from 0.053 to 0.045. Training adds a sparse graph linear solve and Sinkhorn scaling, while deployment requires one encoder pass, local graph smoothing, and baseline-hazard evaluation. The resulting outputs support individual risk ranking, proxy-horizon calibration, and non-causal group-level decomposition for decision support.

Keywords: employee turnover forecasting; individual risk; survival learning; optimal transport; graph regularization; covariate shift; responsible human-resource analytics

MSC: 62R07



Academic Editor: Lei Qin

Received: 25 June 2026

Revised: 19 July 2026

Accepted: 28 July 2026

Published: 2 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Employee turnover affects staffing continuity, replacement costs, knowledge retention, and workforce planning. Behavioral research has linked voluntary turnover to job satisfaction, organizational commitment, search behavior, and workplace conditions [1–3]. Recent human-resource analytics studies use machine-learning models to identify complex associations in tabular employee data [4–6]. Most of these systems predict a binary outcome for each employee. The target of the present study is likewise individual-level risk; department, city, and enterprise summaries are computed only by aggregating individual predictions.

Time-to-event learning offers a useful way to distinguish early and late events and to represent incomplete follow-up. However, the public datasets used in many attrition

studies are cross-sectional. The IBM-HR benchmark is synthetic and records one row per employee, while HR-JC records job-change intent rather than a verified employment exit. Neither dataset contains entry and exit dates, repeated measurements, or a documented censoring process. The proposed method is therefore formulated generally for survival data, but its empirical evaluation uses proxy durations constructed from available tenure-like variables. This distinction limits the interpretation of the reported survival metrics and is stated explicitly throughout the paper.

A second challenge is covariate shift: the employee characteristics observed in a deployment cohort can differ from those in the training cohort. Optimal transport provides a principled mechanism for aligning empirical feature distributions across domains [7–9]. The present study applies this mechanism to controlled source and target partitions and assumes that the conditional proxy outcome mechanism remains stable. It does not claim to correct label shift, target-prevalence shift, or changes in the underlying turnover mechanism.

A third challenge is the use of relational structure. The public benchmarks do not contain communication links, reporting lines, project co-membership, or longitudinal social interactions. Accordingly, the graph in this study is a static employee-similarity graph constructed from observed attributes and group compatibility. Its Laplacian term smooths risk estimates across similar records; it is not a model of causal peer influence or dynamic social diffusion. This use of graph regularization follows the broader graph-learning principle that local structure can serve as a predictive prior [10–12].

To combine these ideas, we propose Graph-Regularized Survival Learning with Optimal Transport, abbreviated to GRSOT. The contributions are as follows. First, GRSOT provides an individual-level risk formulation that can use genuine event-time data when available and can be applied cautiously to proxy time/event pairs in cross-sectional benchmarks. Second, it specifies an entropic optimal-transport module for covariate shift under an explicit conditional-stability assumption. Third, it uses a static similarity graph as a non-causal regularizer and couples the smoothed risk with both the survival-style loss and transport cost. Fourth, it provides an alternating training procedure and a predictive decomposition that separates the feature score, graph adjustment, and source–target discrepancy.

The remainder of the paper is organized as follows. Section 2 reviews related studies. Section 3 defines the prediction target, proxy data construction, survival representation, optimal transport, and graph smoothness. Section 4 presents GRSOT and its optimization procedure. Section 5 reports the empirical evaluation. Section 6 discusses interpretation, limitations, and ethical use, and Section 7 concludes the paper.

2. Related Studies

2.1. Employee Turnover and Human-Resource Analytics

Turnover research has long treated leaving as a behavioral process shaped by individual attitudes, job conditions, and organizational context [1–3]. Machine-learning studies extend this literature by learning nonlinear associations from human-resource records. Fallucchi et al. compared common classifiers for employee attrition, Raza et al. evaluated several machine-learning approaches and feature effects, and Madanchian reviewed artificial-intelligence tools across recruitment and retention decisions [4–6]. These studies motivate predictive decision support, while the caution raised by Rudin remains important: high-stakes decisions benefit from interpretable models, transparent use constraints, and human oversight [13].

Most public attrition benchmarks support static classification because they provide one record and one outcome label per person. Such data are useful for comparing predictive models, but they do not by themselves establish an employee’s true event trajectory. The present work retains the individual-level forecasting goal and adds a survival-style

proxy representation, while explicitly separating that empirical construction from genuine longitudinal survival analysis.

2.2. Survival Prediction and Calibration

The Cox model introduced proportional-hazards regression, and random survival forests extended censored-data learning to nonlinear ensembles [14,15]. DeepSurv and DeepHit subsequently used neural networks for nonlinear risk and discrete-time event prediction [16,17]. These methods provide the basis for the survival component of GRSOT. Their standard interpretation presumes meaningful event times and censoring indicators, which is stronger than the information available in the two cross-sectional benchmarks used here.

Survival evaluation requires both ranking and probabilistic assessment. Haider et al. discuss individual survival distributions and appropriate evaluation, while expected calibration error summarizes the discrepancy between predicted probabilities and empirical event frequencies across probability bins [18,19]. In this paper, calibration is computed at a dataset-specific proxy horizon; it should not be read as censoring-adjusted longitudinal calibration.

2.3. Graph Regularization and Optimal Transport

Graph neural networks and graph attention methods show how neighborhood structure can regularize representation learning [10–12]. In the present setting, no observed social network is available, so graph structure is constructed from employee-feature similarity and group compatibility. The resulting Laplacian is a static smoothing prior rather than evidence of communication, influence, or risk propagation over time.

Entropically regularized optimal transport provides scalable distribution alignment, and optimal-transport domain adaptation transfers information between source and target samples through a soft coupling [7–9]. GRSOT uses this idea for controlled covariate shift. Its transport assumption is narrower than a general distribution-shift claim: the marginal distribution of observed attributes may change, but the conditional proxy outcome mechanism is assumed stable.

3. Preliminaries

3.1. Individual-Level Target and Observed Data

Let the enterprise contain n employee records. Record i has a feature vector $\mathbf{x}_i \in \mathbb{R}^p$, an observed duration $t_i > 0$, an event indicator $\delta_i \in \{0, 1\}$, and a group label g_i . The prediction target is the individual employee's risk and horizon-specific event probability. Group-level quantities are obtained only after prediction by averaging individual outputs; the model does not directly forecast a department-wide or company-wide turnover rate.

For genuine longitudinal data, t_i is the observed time from a defined entry point to exit or administrative censoring, and δ_i identifies whether exit was observed. The experiments in Section 5 use cross-sectional benchmarks that do not provide this information. There, t_i is a proxy derived from a tenure-like field and δ_i is the supplied attrition or job-change label. A non-event is treated as administratively censored only to construct the proxy objective. Thus, the methodological framework is a survival model, whereas the empirical benchmark is a survival-inspired reinterpretation of static records.

The dataset is denoted by $\mathcal{D} = \{(\mathbf{x}_i, t_i, \delta_i, g_i)\}_{i=1}^n$. Source and target subsets are written as $\mathcal{D}_s$ and $\mathcal{D}_t$, with sizes n_s and n_t . In the experiments, these subsets are controlled partitions of the training fold based on tenure or city-development-index covariates; they are not observed historical and future employment trajectories.

Employees are represented as nodes in a static similarity graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$. The nonnegative weight A_{ij} is computed from observed feature similarity and available group compatibility. Because the public datasets contain no communication, reporting, project, or temporal interaction links, A_{ij} must not be interpreted as social influence. Let $\mathbf{D}$ be the degree matrix with $D_{ii} = \sum_j A_{ij}$ and let $\mathbf{L} = \mathbf{D} - \mathbf{A}$ be the graph Laplacian. The graph supplies a predictive smoothness prior.

3.2. Survival Representation of Employee Turnover

Let T_i denote the latent event time for employee i . The survival function is $S_i(t) = P(T_i > t \mid \mathbf{x}_i)$, and the hazard is modeled as

$$\lambda_i(t \mid \mathbf{x}_i) = \lambda_0(t) \exp\{r_\theta(\mathbf{x}_i)\},$$

where $\lambda_0(t)$ is the baseline hazard and $r_\theta(\mathbf{x}_i)$ is the individual log-risk. The cumulative hazard is $\Lambda_i(t \mid \mathbf{x}_i) = \Lambda_0(t) \exp\{r_\theta(\mathbf{x}_i)\}$ and $S_i(t) = \exp\{-\Lambda_i(t \mid \mathbf{x}_i)\}$.

With observed event times and valid censoring, the negative log-likelihood is

$$\mathcal{L}_{\text{surv}}(\theta, \lambda_0) = - \sum_{i=1}^n [\delta_i \log \lambda_i(t_i \mid \mathbf{x}_i) + \log S_i(t_i)].$$

This likelihood uses both events and right-censored observations. When it is applied to the proxy tuples in Section 5, it provides a structured risk objective, but its estimated time scale inherits the assumptions of the proxy construction and is not evidence of a measured employment trajectory. The probability within horizon τ is $1 - S_i(\tau)$.

3.3. Optimal Transport Under Covariate Shift and Graph Smoothness

Let $\mathbf{X}_s = \{\mathbf{x}_a^s\}_{a=1}^{n_s}$ and $\mathbf{X}_t = \{\mathbf{x}_b^t\}_{b=1}^{n_t}$ denote source and target samples, with empirical distributions $\mu_s = \sum_a u_a \delta_{\mathbf{x}_a^s}$ and $\mu_t = \sum_b v_b \delta_{\mathbf{x}_b^t}$. The transport module is intended for covariate shift: the marginal distribution of observed employee attributes may differ between source and target, while the conditional proxy outcome mechanism is assumed sufficiently stable for transfer. The method does not identify or correct label shift, target-prevalence shift, or survival-concept shift.

For a ground-cost matrix $\mathbf{C} \in \mathbb{R}^{n_s \times n_t}$, a coupling $\Pi \in \mathbb{R}_+^{n_s \times n_t}$ satisfies $\Pi \mathbf{1}_{n_t} = \mathbf{u}$ and $\Pi^\top \mathbf{1}_{n_s} = \mathbf{v}$. The entropic transport objective is

$$\mathcal{L}_{\text{ot}}(\Pi) = \langle \Pi, \mathbf{C} \rangle + \varepsilon \sum_{a=1}^{n_s} \sum_{b=1}^{n_t} \Pi_{ab} (\log \Pi_{ab} - 1),$$

where $\varepsilon > 0$ controls coupling smoothness.

Let $\mathbf{z} = (z_1, \dots, z_n)^\top$ denote graph-regularized individual risks. The Laplacian penalty is

$$\mathcal{R}_{\text{graph}}(\mathbf{z}) = \mathbf{z}^\top \mathbf{L} \mathbf{z} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n A_{ij} (z_i - z_j)^2.$$

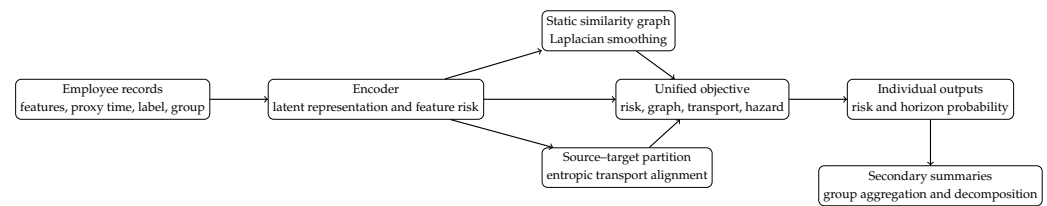
This term encourages similar records to have compatible risks while preserving individual variation. It is a static regularizer and does not imply that risk is transmitted between employees.

4. Methodology

4.1. Overview of GRSOT

GRSOT combines four operations: feature encoding, static graph smoothing, optimal-transport alignment, and survival-style probability estimation. The output is an individual

risk for each employee record. Group summaries are computed only after the individual probabilities have been produced. The block diagram below shows the information flow and separates the two structural inputs: the employee-similarity graph and the controlled source–target partition.



Model block diagram. GRSOT produces individual predictions; graph and transport terms regularize training, and group quantities are post-prediction summaries.

Stage one maps each employee vector $\mathbf{x}_i$ to a representation $\mathbf{h}_i = f_\theta(\mathbf{x}_i)$ and a feature-based risk $r_\theta(\mathbf{x}_i) = \mathbf{w}^\top \mathbf{h}_i + b$. Stage two uses the static similarity graph to smooth the risk while retaining the feature score. Stage three constructs a soft source–target coupling from representation distance, risk compatibility, and group compatibility. Stage four combines the proxy survival loss, transport loss, and graph penalty in one objective. The trained encoder, graph smoother, and baseline hazard then produce an individual probability, after which group summaries can be calculated.

Let $\mathbf{r}_\theta = (r_\theta(\mathbf{x}_1), \dots, r_\theta(\mathbf{x}_n))^\top$. The graph-regularized risk is the solution of

$$\mathbf{z}^* = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{r}_\theta\|_2^2 + \frac{\alpha}{2} \mathbf{z}^\top \mathbf{L} \mathbf{z} \right\},$$

which gives $\mathbf{z}^* = (\mathbf{I}_n + \alpha \mathbf{L})^{-1} \mathbf{r}_\theta$. The system is solved iteratively rather than by forming an inverse. This operation is called graph smoothing because the graph is static and feature-derived; it is not a dynamic or causal diffusion model.

The individual hazard uses z_i as the log-risk, $\lambda_i(t | \mathbf{x}_i, \mathcal{G}) = \lambda_0(t) \exp(z_i)$. For source employee a and target employee b , the transport cost is

$$C_{ab} = \|\mathbf{h}_a^s - \mathbf{h}_b^t\|_2^2 + \rho(z_a^s - z_b^t)^2 + \eta \mathbb{I}(g_a^s \neq g_b^t),$$

where ρ weights risk compatibility and η weights group compatibility. The coupling is a predictive alignment between controlled covariate partitions; it is not an employee-to-employee causal relation.

4.2. Unified Survival–Transport Objective

Training uses source and target subsets $\mathcal{D}_s = \{(\mathbf{x}_a^s, t_a^s, \delta_a^s, g_a^s)\}_{a=1}^{n_s}$ and $\mathcal{D}_t = \{(\mathbf{x}_b^t, t_b^t, \delta_b^t, g_b^t)\}_{b=1}^{n_t}$. In the benchmark experiments, both are constructed within the training fold by covariate partitions. The transport component assumes covariate shift and conditional stability of the proxy outcome mechanism. No claim is made that the method is robust to label shift or a changed conditional turnover process.

GRSOT uses a piecewise-constant baseline hazard. Let $0 = a_0 < a_1 < \dots < a_M$ partition the observation scale and let $\lambda_0(t) = \kappa_m$ for $t \in [a_{m-1}, a_m)$. The negative log-likelihood is

$$\mathcal{L}_{\text{surv}}(\theta, \mathbf{w}, b, \kappa) = - \sum_{i=1}^n [\delta_i \{\log \lambda_0(t_i) + z_i\} - \Lambda_0(t_i) \exp(z_i)].$$

With genuine longitudinal data, this is a censored survival likelihood. With the proxy tuples used here, it is a survival-style loss whose time interpretation is limited by

the cross-sectional construction. The horizon probability is $P_\theta(T_i \leq \tau \mid \mathbf{x}_i, \mathcal{G}) = 1 - \exp\{-\Lambda_0(\tau) \exp(z_i)\}$.

The entropic transport loss is

$$\mathcal{L}_{\text{ot}}(\theta, \mathbf{z}, \Pi) = \langle \Pi, \mathbf{C} \rangle + \varepsilon \sum_{a=1}^{n_s} \sum_{b=1}^{n_t} \Pi_{ab} (\log \Pi_{ab} - 1), \quad \Pi \in \mathcal{U}(\mathbf{u}, \mathbf{v}),$$

where $\mathcal{U}(\mathbf{u}, \mathbf{v})$ enforces the empirical source and target marginals. The complete objective is

$$\min_{\theta, \mathbf{w}, b, \kappa, \Pi} \mathcal{J} = \mathcal{L}_{\text{surv}} + \beta \mathcal{L}_{\text{ot}} + \gamma \mathbf{z}^\top \mathbf{L} \mathbf{z} + \omega \|\theta\|_2^2 + \psi \|\mathbf{w}\|_2^2,$$

subject to $\mathbf{z} = (\mathbf{I}_n + \alpha \mathbf{L})^{-1} \mathbf{r}_\theta$, $\Pi \in \mathcal{U}(\mathbf{u}, \mathbf{v})$, and positive baseline hazards. The shared risk z_i enters the hazard, transport cost, and graph penalty, which couples the three modules.

4.3. Alternating Optimization

The constrained coupling, graph-smoothed risk, baseline hazards, and encoder parameters are updated alternately. Given the current representation, the graph system produces $\mathbf{z}$. The cost matrix is then rebuilt and the transport coupling is updated by Sinkhorn scaling. Let $K_{ab} = \exp(-C_{ab}/\varepsilon)$; the coupling has the form $\Pi = \text{diag}(\mathbf{a}) \mathbf{K} \text{diag}(\mathbf{c})$, with $\mathbf{a} \leftarrow \mathbf{u} \oslash (\mathbf{K} \mathbf{c})$ and $\mathbf{c} \leftarrow \mathbf{v} \oslash (\mathbf{K}^\top \mathbf{a})$.

For interval exposure $\Delta_{im} = \max\{0, \min(t_i, a_m) - a_{m-1}\}$, the baseline update is

$$\kappa_m = \frac{\sum_{i=1}^n \delta_i \mathbb{I}(t_i \in [a_{m-1}, a_m])}{\sum_{i=1}^n \exp(z_i) \Delta_{im} + \xi}.$$

The remaining parameters are updated by gradient optimization of $\mathcal{J}$.

The computational costs are stated before Algorithm 1. With q conjugate-gradient iterations, sparse graph smoothing costs $O(q(|\mathcal{E}| + n))$. Full Sinkhorn scaling costs $O(K_{\text{sk}} n_s n_t)$ and can be reduced to $O(K_{\text{sk}} B_s B_t)$ with source and target mini-batches. Baseline-hazard estimation costs $O(nM)$. The graph solve and transport step therefore dominate training, whereas ordinary neural baselines omit these two operations.

Training stability is improved by rebuilding the transport cost after the risk update instead of fixing a coupling from an untrained representation. This alternating schedule lets the representation, graph-smoothed risk, and coupling evolve together while preserving the explicit covariate-shift assumption.

Algorithm 1 Alternating optimization for GRSOT training

Require: Employee subsets $\mathcal{D}_s$ and $\mathcal{D}_t$, graph Laplacian $\mathbf{L}$, weights $\mathbf{u}$ and $\mathbf{v}$, interval partition $\{a_m\}_{m=0}^M$, and model coefficients.

Ensure: Encoder and risk parameters, baseline hazards, and transport coupling.

- 1: Initialize the encoder, risk head, and positive baseline hazards.
 - 2: **for** each training epoch **do**
 - 3: Compute latent representations and feature-based individual risks.
 - 4: Solve $(\mathbf{I}_n + \alpha \mathbf{L}) \mathbf{z} = \mathbf{r}_\theta$ for graph-smoothed risks.
 - 5: Construct the transport cost from representations, risks, and group compatibility.
 - 6: Update Π by Sinkhorn scaling.
 - 7: Update each baseline hazard from risk-weighted exposure.
 - 8: Update the encoder and risk head by minimizing $\mathcal{J}$.
 - 9: **end for**
 - 10: Return the learned parameters and coupling.
-

4.4. Individual Prediction and Predictive Decomposition

For employee i , GRSOT produces

$$\hat{p}_i(\tau) = 1 - \exp\{-\hat{\Lambda}_0(\tau) \exp(z_i)\}.$$

This is an individual probability. A department or city value is a later average of individual probabilities and is not a separately trained aggregate forecast.

For a new employee with attributes $\mathbf{x}_{n+1}$, a similarity vector $\mathbf{a}_{n+1}$ connects the record to candidate neighbors. The local extension is

$$z_{n+1} = \frac{r_\theta(\mathbf{x}_{n+1}) + \alpha \sum_{j=1}^n a_{n+1,j} z_j}{1 + \alpha \sum_{j=1}^n a_{n+1,j}}.$$

The calculation is a weighted graph adjustment based on static similarity, not a causal peer effect.

The predictive decomposition contains the feature score $r_\theta(\mathbf{x}_i)$, the graph adjustment $\Delta_i^{\text{graph}} = z_i - r_\theta(\mathbf{x}_i)$, and the target-to-source discrepancy

$$\Delta_b^{\text{ot}} = z_b^t - \frac{\sum_{a=1}^{n_s} \Pi_{ab} z_a^s}{\sum_{a=1}^{n_s} \Pi_{ab} + \xi}.$$

These quantities explain how the predictor combines a record's features, static neighborhood, and matched source records. They do not identify causal effects of management, compensation, or co-worker behavior.

Deployment cost is stated before Algorithm 2. A prediction requires one forward pass through the encoder, similarity evaluation against a restricted neighbor set, one local weighted average, and baseline-hazard evaluation. Thus, the additional complexity is concentrated in training; the per-employee deployment path is lightweight and can use an indexed neighbor set.

Algorithm 2 Individual prediction and decomposition

Require: A target record $(\mathbf{x}_{n+1}, g_{n+1})$, trained parameters, historical risks and representations, and horizon τ .

Ensure: Individual probability and predictive-decomposition tuple.

- 1: Compute the target representation and feature-based risk.
 - 2: Construct similarities to candidate neighbors.
 - 3: Compute the locally smoothed risk z_{n+1} .
 - 4: Compute $\hat{p}_{n+1}(\tau)$.
 - 5: Compute the graph adjustment and transport discrepancy when a coupling is available.
 - 6: Return the individual probability and decomposition.
-

For group c , the reported probability is $\hat{P}_c(\tau) = |\mathcal{V}_c|^{-1} \sum_{i \in \mathcal{V}_c} \hat{p}_i(\tau)$, with analogous averages for the graph adjustment and transport discrepancy. These group summaries support descriptive workforce analysis but should not be treated as causal diagnoses.

5. Experiments

5.1. Experimental Settings

The empirical study compares methods on two public cross-sectional human-resource datasets. Each row is evaluated as an individual record. Table 1 reports the observed labels, proxy durations, graph groups, and data limitations.

Neither dataset contains longitudinal employment trajectories, entry and exit dates, repeated employee measurements, or a documented censoring mechanism. IBM-HR supplies a synthetic attrition label and YearsAtCompany; HR-JC supplies job-change intent and a categorical last-new-job field. For the shared objective, these variables are

converted into pseudo-time/event tuples. IBM non-attrition records and HR-JC records with target zero are treated as administratively censored only within this proxy construction. Consequently, the empirical study evaluates survival-inspired ranking and proxy-horizon probabilities, not validated longitudinal survival functions.

Table 1. Dataset characteristics and proxy time/event construction.

Dataset	Records	Observed Label	Proxy Duration	Graph Group	Data Character	Event Rate
IBM-HR ^a	1470	Attrition = Yes	max(YearsAtCompany, 0.5)	Department	Synthetic cross-sectional benchmark; no trajectories	237/1470 (16.12%)
HR-JC ^b	19,158	target = 1 (job-change intent)	mapped last_new_job	City	Cross-sectional candidate survey; no verified exit time	4777/19,158 (24.93%)

^a <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset> (accessed on 20 May 2026); ^b <https://www.kaggle.com/datasets/arashnic/hr-analytics-job-change-of-data-scientists> (accessed on 20 May 2026).

Categorical variables were one-hot encoded, numerical variables were standardized using training-fold statistics, and missing categories were retained as Unknown. For HR-JC, last_new_job values one through four were mapped to their corresponding year values, greater than four was mapped to five, and never was mapped to the midpoint of the experience bin clipped to one through six. Table 2 gives the common protocol.

Table 2. Experimental protocol.

Component	IBM-HR Setting	HR-JC Setting
Train/validation/test split	Stratified split with 60%/20%/20% proportions over five random seeds	Stratified split with 60%/20%/20% proportions over five random seeds
Source–target construction	$\mathcal{D}_s$ and $\mathcal{D}_t$ formed by lower and upper tenure-cohort partitions within the training fold	$\mathcal{D}_s$ and $\mathcal{D}_t$ formed by lower and upper city-development-index partitions within the training fold
Graph construction	Mutual k -nearest-neighbor graph with $k = 10$ and additional department compatibility weight	Mutual k -nearest-neighbor graph with $k = 15$ and additional city compatibility weight
Encoder f_θ	Two hidden layers with dimensions 64 and 32, ReLU activation, dropout 0.20	Two hidden layers with dimensions 128 and 64, ReLU activation, dropout 0.20
Survival intervals	$M = 6$ piecewise-constant baseline-hazard intervals over observed t_i	$M = 6$ piecewise-constant baseline-hazard intervals over observed t_i
Main GRSOT coefficients	$\alpha = 0.25, \beta = 0.10, \gamma = 0.05, \rho = 0.50, \eta = 0.20, \varepsilon = 0.05$	$\alpha = 0.20, \beta = 0.15, \gamma = 0.05, \rho = 0.50, \eta = 0.15, \varepsilon = 0.05$
Optimization	Adam optimizer, learning rate 10^{-3} , weight decay 10^{-4} , batch size 128	Adam optimizer, learning rate 10^{-3} , weight decay 10^{-4} , batch size 512
Evaluation metrics	C-index, time-dependent AUC, integrated Brier score, expected calibration error	C-index, time-dependent AUC, integrated Brier score, expected calibration error

The source and target subsets in Table 2 are controlled covariate partitions inside the training fold. They test changes in the marginal distribution of tenure or city-development-index-related attributes. The experiments assume that the conditional proxy outcome mechanism remains stable and do not test label shift, target-prevalence shift, or survival-concept shift. The graph is a static mutual-nearest-neighbor graph built from processed attributes with department or city compatibility; it contains no observed communication or reporting links.

The workflow has eight stages. First, one cross-sectional row is retained per employee or candidate. Second, the supplied label and tenure-like field form a proxy event–time pair. Third, each seed creates stratified training, validation, and test folds. Fourth, encoding, scaling, and missing-value handling are fitted on the training fold only. Fifth, the training fold is divided into controlled source and target covariate partitions. Sixth, a static similarity graph is

built without using test records. Seventh, GRSOT is trained by alternating graph smoothing, Sinkhorn alignment, baseline-hazard estimation, and gradient updates. Eighth, the held-out fold is scored inductively and evaluated by ranking and proxy-calibration metrics.

Expected calibration error (ECE) is estimated at the dataset-specific proxy horizon using ten equal-width probability bins. Let I_b be the test records in bin b , $\bar{p}_b$ their mean predicted probability, and $\bar{y}_b$ their mean binary event indicator. Then

$$\text{ECE} = \sum_{b=1}^{10} \frac{|I_b|}{n} |\bar{y}_b - \bar{p}_b|.$$

Empty bins are omitted. ECE is computed separately on each test split and summarized across five seeds. In this cross-sectional setting, it measures calibration of the supplied label at the proxy horizon; it is not an inverse-probability-of-censoring-weighted estimate of longitudinal calibration [19].

Figure 1 shows the class imbalance that motivates ranking and calibration metrics rather than raw accuracy.

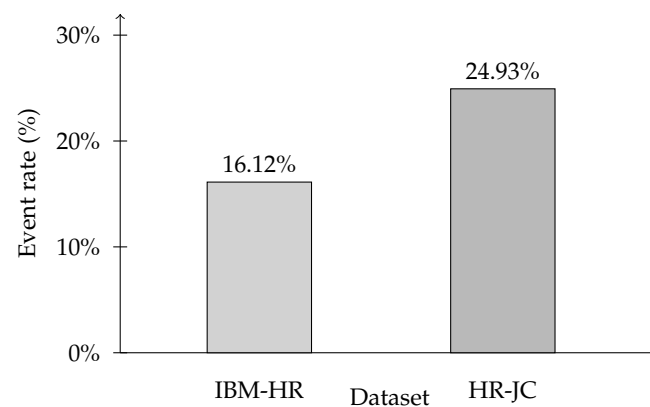


Figure 1. Event rates in the two datasets.

IBM-HR has an event rate of 16.12%, and HR-JC has a job-change-intent rate of 24.93%. The reported C-index and AUC assess ordering under the proxy construction, while integrated Brier score and ECE assess probability error.

5.2. Overall Predictive Performance

The baselines include static classifiers, survival models applied to the same proxy tuples, a graph-smoothed Cox variant, and an optimal-transport survival variant. Table 3 reports IBM-HR results.

Table 3. Overall performance on IBM-HR.

Method	C-Index ↑	AUC ↑	IBS ↓	ECE ↓
Logistic Regression	0.742 ± 0.018	0.768 ± 0.021	0.141 ± 0.006	0.073 ± 0.009
Random Forest	0.758 ± 0.015	0.781 ± 0.018	0.136 ± 0.005	0.068 ± 0.007
XGBoost	0.771 ± 0.014	0.792 ± 0.016	0.132 ± 0.004	0.061 ± 0.006
CoxPH	0.765 ± 0.017	0.786 ± 0.019	0.134 ± 0.005	0.064 ± 0.007
Random Survival Forest	0.779 ± 0.013	0.801 ± 0.015	0.129 ± 0.004	0.058 ± 0.006
DeepSurv	0.788 ± 0.012	0.812 ± 0.014	0.126 ± 0.004	0.054 ± 0.006
DeepHit	0.795 ± 0.011	0.817 ± 0.013	0.124 ± 0.004	0.052 ± 0.005
Graph-Cox	0.807 ± 0.010	0.829 ± 0.012	0.119 ± 0.003	0.047 ± 0.005
OT-Surv	0.812 ± 0.010	0.835 ± 0.011	0.117 ± 0.003	0.045 ± 0.004
GRSOT	0.836 ± 0.009	0.859 ± 0.010	0.109 ± 0.003	0.037 ± 0.004

On IBM-HR, GRSOT records a C-index of 0.836 and ECE of 0.037. Relative to OT-Surv, the absolute C-index gain is 0.024 and the ECE reduction is 0.008. Relative to XGBoost, the corresponding differences are 0.065 and 0.024. These values support improved ranking and smaller proxy-calibration error; they do not convert the cross-sectional benchmark into a longitudinal survival study. Table 4 reports HR-JC results.

On HR-JC, GRSOT obtains a C-index of 0.803 and ECE of 0.045, compared with 0.781 and 0.053 for OT-Surv. The added graph term therefore contributes beyond transport alignment on this controlled covariate partition. The model also has greater training complexity than a single neural survival model because it solves a sparse graph system and performs Sinkhorn updates, as quantified by the complexity expressions before Algorithm 1.

Figure 2 visualizes the IBM-HR ranking results.

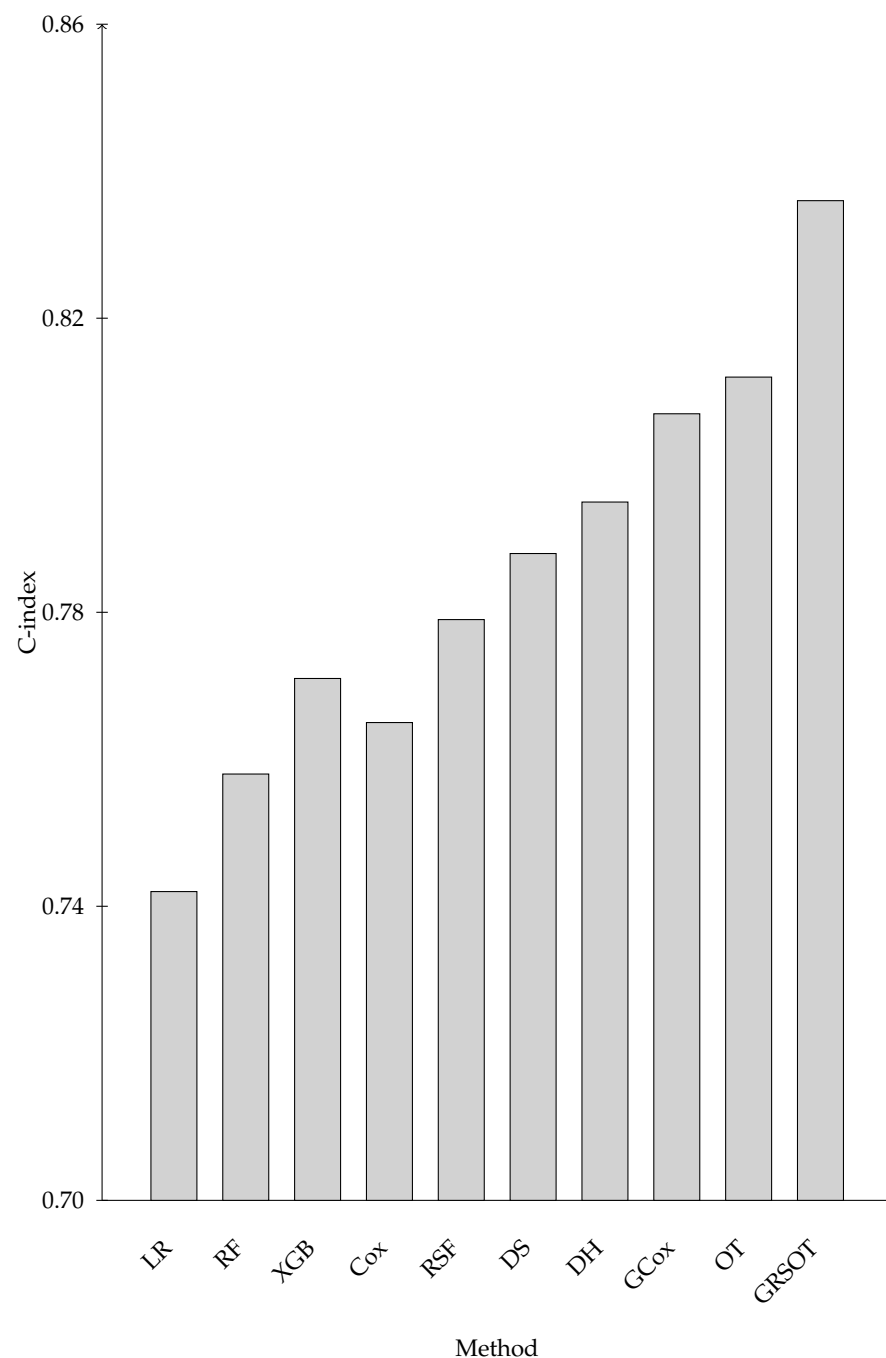


Figure 2. IBM-HR C-index by method.

Table 4. Overall performance on HR-JC.

Method	C-Index $\uparrow$	AUC $\uparrow$	IBS $\downarrow$	ECE $\downarrow$
Logistic Regression	0.716 ± 0.006	0.729 ± 0.007	0.177 ± 0.003	0.081 ± 0.004
Random Forest	0.731 ± 0.005	0.744 ± 0.006	0.171 ± 0.003	0.076 ± 0.004
XGBoost	0.748 ± 0.005	0.762 ± 0.006	0.166 ± 0.002	0.069 ± 0.003
CoxPH	0.727 ± 0.006	0.739 ± 0.006	0.173 ± 0.003	0.078 ± 0.004
Random Survival Forest	0.746 ± 0.005	0.759 ± 0.006	0.167 ± 0.002	0.071 ± 0.003
DeepSurv	0.759 ± 0.004	0.773 ± 0.005	0.161 ± 0.002	0.064 ± 0.003
DeepHit	0.764 ± 0.004	0.778 ± 0.005	0.159 ± 0.002	0.062 ± 0.003
Graph-Cox	0.774 ± 0.004	0.789 ± 0.005	0.154 ± 0.002	0.056 ± 0.003
OT-Surv	0.781 ± 0.004	0.795 ± 0.004	0.151 ± 0.002	0.053 ± 0.003
GRSOT	0.803 ± 0.003	0.818 ± 0.004	0.145 ± 0.002	0.045 ± 0.002

The ordering in Figure 2 matches Table 3: the graph and transport baselines improve on feature-only models, and their coupled form has the highest C-index. Figure 3 presents the HR-JC integrated Brier score with a conventional zero baseline and explicit axis labels.

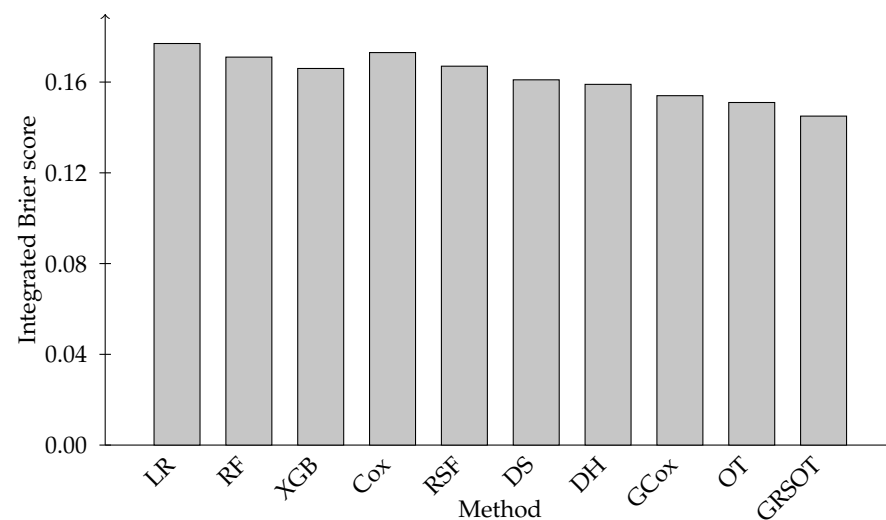
**Figure 3.** HR-JC integrated Brier score by method.

Figure 3 shows the lowest integrated Brier score for GRSOT, 0.145, compared with 0.151 for OT-Surv and 0.154 for Graph-Cox.

5.3. Ablation Analysis

The ablations remove the proxy survival loss, graph regularization, optimal transport, risk compatibility in the transport cost, or group compatibility. All variants use the same splits and encoder. Table 5 reports IBM-HR results.

Table 5. IBM-HR ablation results.

Variant	C-Index $\uparrow$	AUC $\uparrow$	IBS $\downarrow$	ECE $\downarrow$
GRSOT-CIs	0.799 ± 0.012	0.821 ± 0.013	0.123 ± 0.004	0.053 ± 0.005
GRSOT-w/o-Graph	0.814 ± 0.010	0.837 ± 0.011	0.117 ± 0.003	0.046 ± 0.004
GRSOT-w/o-OT	0.818 ± 0.010	0.841 ± 0.011	0.115 ± 0.003	0.044 ± 0.004
GRSOT-w/o-RiskOT	0.824 ± 0.009	0.848 ± 0.010	0.113 ± 0.003	0.042 ± 0.004
GRSOT-w/o-GroupOT	0.827 ± 0.009	0.851 ± 0.010	0.112 ± 0.003	0.041 ± 0.004
GRSOT	0.836 ± 0.009	0.859 ± 0.010	0.109 ± 0.003	0.037 ± 0.004

On IBM-HR, replacing the survival-style objective with binary cross-entropy lowers the C-index from 0.836 to 0.799. Removing graph regularization lowers it by 0.022, and

removing optimal transport lowers it by 0.018. These comparisons isolate the contribution of each component under the same proxy construction. Table 6 reports HR-JC results.

Table 6. HR-JC ablation results.

Variant	C-Index $\uparrow$	AUC $\uparrow$	IBS $\downarrow$	ECE $\downarrow$
GRSOT-Cl _s	0.761 ± 0.005	0.775 ± 0.005	0.160 ± 0.002	0.064 ± 0.003
GRSOT-w/o-Graph	0.783 ± 0.004	0.798 ± 0.004	0.151 ± 0.002	0.053 ± 0.003
GRSOT-w/o-OT	0.778 ± 0.004	0.792 ± 0.004	0.153 ± 0.002	0.056 ± 0.003
GRSOT-w/o-RiskOT	0.790 ± 0.004	0.805 ± 0.004	0.149 ± 0.002	0.050 ± 0.003
GRSOT-w/o-GroupOT	0.793 ± 0.004	0.808 ± 0.004	0.148 ± 0.002	0.049 ± 0.003
GRSOT	0.803 ± 0.003	0.818 ± 0.004	0.145 ± 0.002	0.045 ± 0.002

On HR-JC, removing optimal transport raises ECE from 0.045 to 0.056 and lowers the C-index from 0.803 to 0.778. Removing graph regularization yields an ECE of 0.053 and a C-index of 0.783. Figure 4 visualizes the IBM-HR C-index ablation.

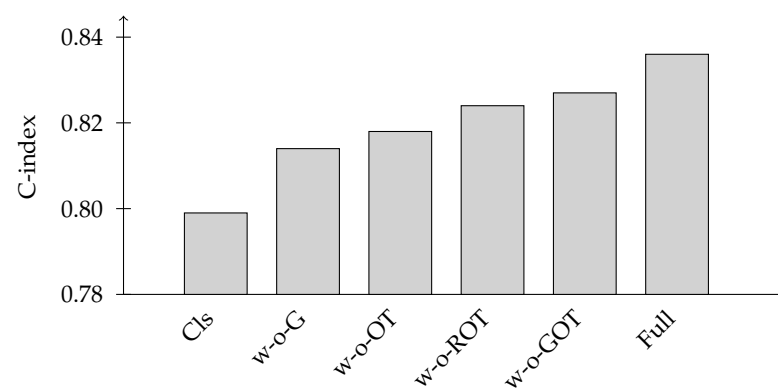


Figure 4. IBM-HR ablation C-index.

Figure 4 confirms that the full model has the highest ranking score. Figure 5 shows the HR-JC calibration ablation with a zero-based vertical axis.

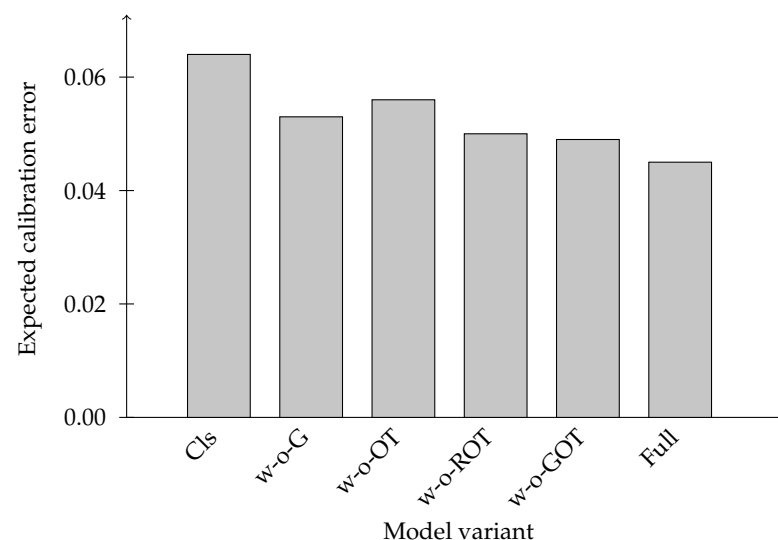


Figure 5. HR-JC ablation ECE.

At a probability-bin level, overconfidence means that the mean predicted probability exceeds the observed event frequency, or $\bar{p}_b > \bar{y}_b$. ECE aggregates the absolute gaps and therefore does not establish that every bin is overconfident. The quantitative conclusion supported here is narrower: the full model reduces calibration mismatch. Its HR-JC ECE is

0.045, compared with 0.056 without optimal transport, an absolute reduction of 0.011 and a relative reduction of 19.6%.

5.4. Robustness Under Controlled Covariate Shift

The shift study changes the separation between source and target covariate partitions. IBM-HR uses increasingly separated tenure partitions, and HR-JC uses increasingly separated city-development-index partitions. This design changes marginal covariate distributions; it does not test label shift, a changed event prevalence mechanism, or survival-concept shift. Table 7 reports IBM-HR results.

Table 7. IBM-HR performance under controlled covariate shift.

Shift Level	Method	C-Index $\uparrow$	AUC $\uparrow$	IBS $\downarrow$	ECE $\downarrow$
None	DeepHit	0.801 $\pm$ 0.011	0.823 $\pm$ 0.012	0.122 $\pm$ 0.004	0.050 $\pm$ 0.005
None	Graph-Cox	0.812 $\pm$ 0.010	0.834 $\pm$ 0.011	0.117 $\pm$ 0.003	0.046 $\pm$ 0.004
None	OT-Surv	0.817 $\pm$ 0.010	0.840 $\pm$ 0.011	0.115 $\pm$ 0.003	0.043 $\pm$ 0.004
None	GRSOT	0.840 $\pm$ 0.009	0.862 $\pm$ 0.010	0.108 $\pm$ 0.003	0.036 $\pm$ 0.004
Mild	DeepHit	0.792 $\pm$ 0.012	0.815 $\pm$ 0.013	0.125 $\pm$ 0.004	0.053 $\pm$ 0.005
Mild	Graph-Cox	0.804 $\pm$ 0.011	0.826 $\pm$ 0.012	0.121 $\pm$ 0.004	0.049 $\pm$ 0.004
Mild	OT-Surv	0.810 $\pm$ 0.010	0.832 $\pm$ 0.011	0.118 $\pm$ 0.003	0.046 $\pm$ 0.004
Mild	GRSOT	0.833 $\pm$ 0.009	0.856 $\pm$ 0.010	0.110 $\pm$ 0.003	0.038 $\pm$ 0.004
Moderate	DeepHit	0.779 $\pm$ 0.013	0.802 $\pm$ 0.014	0.130 $\pm$ 0.004	0.058 $\pm$ 0.005
Moderate	Graph-Cox	0.790 $\pm$ 0.012	0.813 $\pm$ 0.013	0.126 $\pm$ 0.004	0.054 $\pm$ 0.005
Moderate	OT-Surv	0.801 $\pm$ 0.011	0.824 $\pm$ 0.012	0.121 $\pm$ 0.003	0.049 $\pm$ 0.004
Moderate	GRSOT	0.824 $\pm$ 0.010	0.847 $\pm$ 0.011	0.113 $\pm$ 0.003	0.041 $\pm$ 0.004
Severe	DeepHit	0.758 $\pm$ 0.015	0.781 $\pm$ 0.016	0.139 $\pm$ 0.005	0.067 $\pm$ 0.006
Severe	Graph-Cox	0.769 $\pm$ 0.014	0.792 $\pm$ 0.015	0.134 $\pm$ 0.005	0.062 $\pm$ 0.006
Severe	OT-Surv	0.786 $\pm$ 0.012	0.809 $\pm$ 0.013	0.126 $\pm$ 0.004	0.054 $\pm$ 0.005
Severe	GRSOT	0.813 $\pm$ 0.011	0.836 $\pm$ 0.012	0.117 $\pm$ 0.003	0.044 $\pm$ 0.004

From no shift to severe shift, the GRSOT C-index decreases from 0.840 to 0.813, a loss of 0.027. The corresponding losses are 0.031 for OT-Surv and 0.043 for both DeepHit and Graph-Cox. Table 8 reports HR-JC results.

Table 8. HR-JC performance under controlled covariate shift.

Shift Level	Method	C-Index $\uparrow$	AUC $\uparrow$	IBS $\downarrow$	ECE $\downarrow$
None	DeepHit	0.769 $\pm$ 0.004	0.783 $\pm$ 0.005	0.157 $\pm$ 0.002	0.060 $\pm$ 0.003
None	Graph-Cox	0.780 $\pm$ 0.004	0.795 $\pm$ 0.005	0.152 $\pm$ 0.002	0.054 $\pm$ 0.003
None	OT-Surv	0.787 $\pm$ 0.004	0.801 $\pm$ 0.004	0.149 $\pm$ 0.002	0.051 $\pm$ 0.003
None	GRSOT	0.807 $\pm$ 0.003	0.822 $\pm$ 0.004	0.143 $\pm$ 0.002	0.043 $\pm$ 0.002
Mild	DeepHit	0.761 $\pm$ 0.004	0.775 $\pm$ 0.005	0.160 $\pm$ 0.002	0.063 $\pm$ 0.003
Mild	Graph-Cox	0.772 $\pm$ 0.004	0.787 $\pm$ 0.005	0.155 $\pm$ 0.002	0.057 $\pm$ 0.003
Mild	OT-Surv	0.781 $\pm$ 0.004	0.796 $\pm$ 0.004	0.151 $\pm$ 0.002	0.053 $\pm$ 0.003
Mild	GRSOT	0.802 $\pm$ 0.003	0.817 $\pm$ 0.004	0.145 $\pm$ 0.002	0.045 $\pm$ 0.002
Moderate	DeepHit	0.747 $\pm$ 0.005	0.762 $\pm$ 0.006	0.165 $\pm$ 0.002	0.070 $\pm$ 0.003
Moderate	Graph-Cox	0.757 $\pm$ 0.005	0.772 $\pm$ 0.005	0.161 $\pm$ 0.002	0.064 $\pm$ 0.003
Moderate	OT-Surv	0.771 $\pm$ 0.004	0.786 $\pm$ 0.005	0.155 $\pm$ 0.002	0.058 $\pm$ 0.003
Moderate	GRSOT	0.793 $\pm$ 0.004	0.808 $\pm$ 0.004	0.148 $\pm$ 0.002	0.049 $\pm$ 0.002
Severe	DeepHit	0.728 $\pm$ 0.006	0.743 $\pm$ 0.007	0.173 $\pm$ 0.003	0.081 $\pm$ 0.004
Severe	Graph-Cox	0.739 $\pm$ 0.006	0.754 $\pm$ 0.006	0.168 $\pm$ 0.003	0.074 $\pm$ 0.004
Severe	OT-Surv	0.758 $\pm$ 0.005	0.773 $\pm$ 0.006	0.160 $\pm$ 0.002	0.066 $\pm$ 0.003
Severe	GRSOT	0.783 $\pm$ 0.004	0.798 $\pm$ 0.005	0.152 $\pm$ 0.002	0.053 $\pm$ 0.003

Under severe HR-JC shift, GRSOT records a C-index of 0.783 and ECE of 0.053. OT-Surv records 0.758 and 0.066, Graph-Cox records 0.739 and 0.074, and DeepHit records

0.728 and 0.081. Thus, the reported evidence supports smaller calibration gaps under the tested covariate shift. Figure 6 shows the IBM-HR C-index trend.

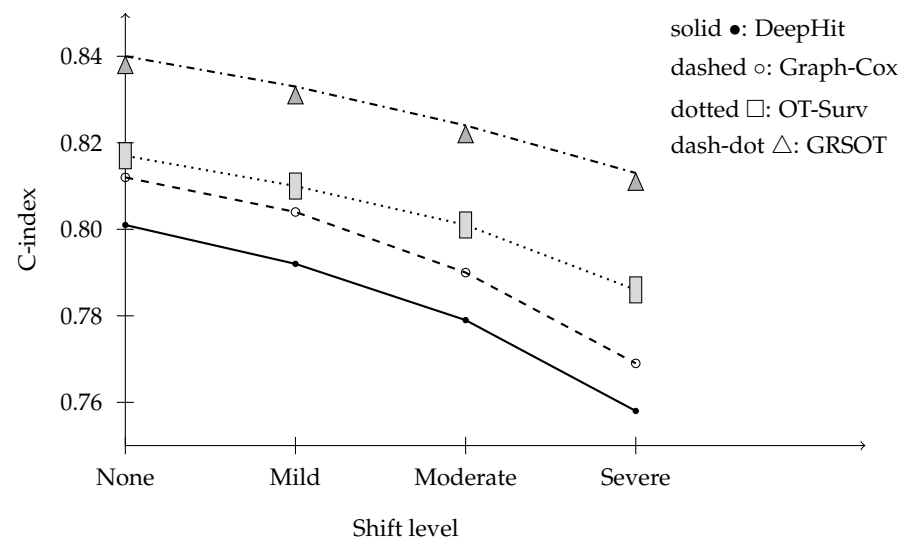


Figure 6. IBM-HR C-index under controlled covariate shift.

Figure 6 shows that GRSOT remains above the other methods at each shift level. Figure 7 shows the HR-JC ECE trend.

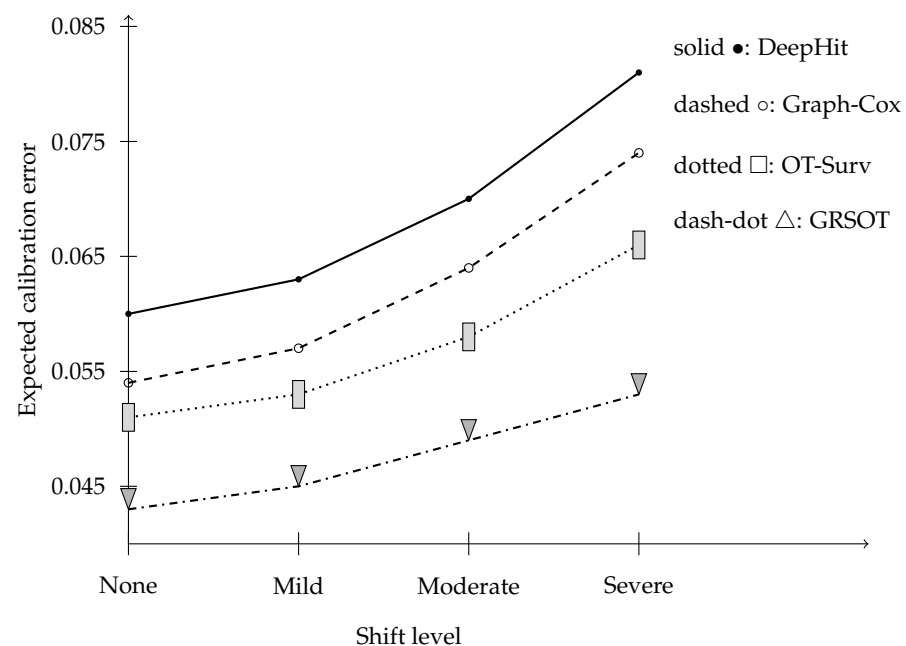


Figure 7. HR-JC ECE under controlled covariate shift.

Tables 7 and 8, together with Figures 6 and 7, demonstrate that GRSOT degrades more slowly than the compared methods under the controlled covariate partitions. This statement is limited to the tested covariate-shift setting and does not imply robustness to a changed conditional outcome mechanism.

5.5. Sensitivity and Interpretability Analysis

Table 9 reports C-index sensitivity to the graph smoothing coefficient α and transport coefficient β .

Table 9. C-index sensitivity to graph and transport coefficients.

Dataset	α	$\beta = 0.00$	$\beta = 0.05$	$\beta = 0.10$
IBM-HR	0.00	0.801 ± 0.011	0.808 ± 0.010	0.812 ± 0.010
IBM-HR	0.10	0.814 ± 0.010	0.824 ± 0.010	0.830 ± 0.009
IBM-HR	0.25	0.818 ± 0.010	0.829 ± 0.009	0.836 ± 0.009
IBM-HR	0.50	0.811 ± 0.011	0.821 ± 0.010	0.827 ± 0.010
HR-JC	0.00	0.765 ± 0.005	0.773 ± 0.004	0.778 ± 0.004
HR-JC	0.10	0.774 ± 0.004	0.786 ± 0.004	0.795 ± 0.004
HR-JC	0.20	0.778 ± 0.004	0.792 ± 0.004	0.799 ± 0.003
HR-JC	0.30	0.775 ± 0.004	0.789 ± 0.004	0.803 ± 0.003

Moderate graph smoothing improves ranking, whereas larger values reduce performance through over-smoothing. Increasing the transport weight improves performance within the tested range, with a larger effect on HR-JC. Table 10 reports descriptive group averages of the individual predictions and decomposition terms.

Table 10. Group-level predictive decomposition.

Dataset	Group c	$\hat{P}_c(\tau)$	$\bar{r}_c$	Graph Adjustment	OT Discrepancy
IBM-HR	Sales	0.284	0.312	0.041	0.028
IBM-HR	Human Resources	0.263	0.286	0.036	0.019
IBM-HR	Research and Development	0.181	0.147	−0.022	0.011
HR-JC	City group A	0.337	0.418	0.052	0.046
HR-JC	City group B	0.286	0.301	0.018	0.024
HR-JC	City group C	0.214	0.196	−0.031	0.009

The values in Table 10 are predictive summaries, not causal effects. For example, the positive graph adjustment for Sales indicates that similarity-based smoothing raises its average risk relative to the feature-only score; it does not show that co-workers cause turnover. Figure 8 visualizes the sensitivity to α .

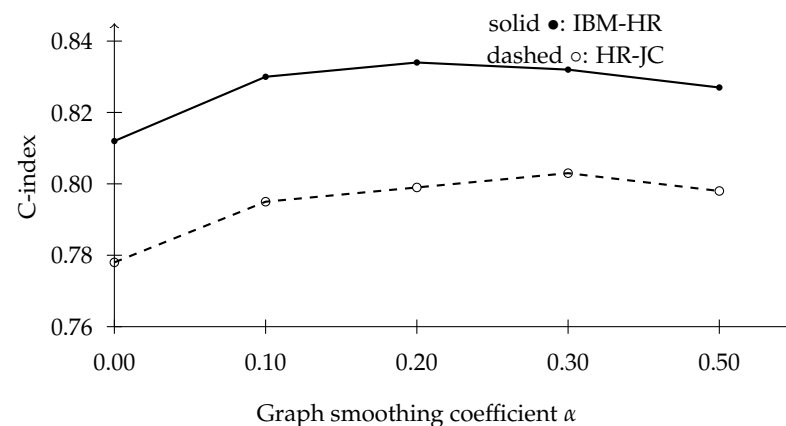
**Figure 8.** C-index sensitivity to graph smoothing.

Figure 8 confirms the non-monotonic effect of graph smoothing on both datasets. Figure 9 visualizes the two structural decomposition terms.

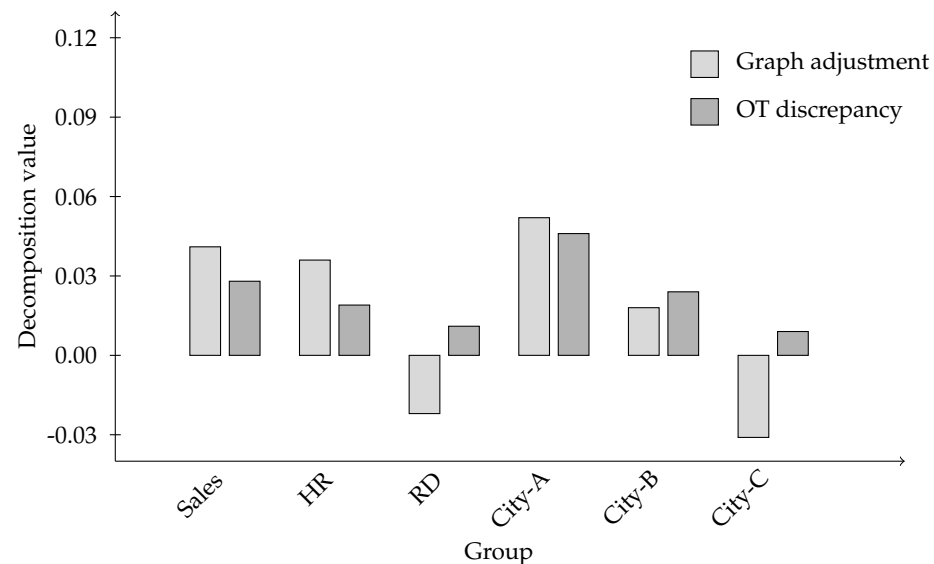


Figure 9. Group-level graph adjustment and transport discrepancy.

The experimental evidence is summarized by Tables 1–10 and Figures 1–9. The overall results show higher ranking scores and smaller proxy-calibration errors for GRSOT; the ablations attribute gains to the survival-style, graph, and transport components; the controlled-shift results show slower degradation under marginal covariate change; and the sensitivity analysis shows that excessive graph smoothing can be harmful. These findings are interpreted subject to the cross-sectional proxy construction described in Table 1.

6. Discussion

The results align with three strands of prior work. First, nonlinear machine-learning models are effective for employee attrition records [4,5]. Second, survival models provide a useful structure for risk ranking and horizon-specific probabilities when event times and censoring are available [14–17]. Third, graph regularization and optimal transport can supply complementary structural priors under local similarity and covariate shift [8,10,12]. The present findings extend these ideas in one objective, but the benchmark data support only a survival-inspired, similarity-graph interpretation.

The main limitations follow directly from the data. IBM-HR is synthetic and cross-sectional, and HR-JC measures job-change intent rather than verified turnover. Neither dataset contains employee trajectories, explicit entry and exit dates, repeated measurements, or a documented censoring process. The proxy durations therefore impose modeling assumptions that cannot be validated from the public records. The employee graph is constructed from static feature similarity and group labels rather than observed relations, so graph adjustments are neither causal peer effects nor dynamic risk diffusion. The transport evaluation covers controlled covariate shift and assumes conditional stability; it does not establish robustness to label shift or a changed turnover mechanism. Finally, the public benchmarks do not support a complete fairness audit across legally or organizationally relevant groups.

Individual turnover-risk scores create ethical risks if they are used to deny promotion, training, compensation, or continued employment. The model is intended only for aggregate planning and supportive, voluntary retention programs with human review. Appropriate governance includes data minimization, access controls, documented purpose limitation, subgroup performance and calibration audits, employee notice where required, a process for contesting consequential decisions, and a prohibition on punitive action based

solely on a predicted risk. The decomposition is predictive and should not be presented as a causal explanation of an employee's intentions.

7. Conclusions

GRSOT combines individual survival-style risk learning, static graph regularization, and optimal-transport alignment for controlled covariate shift. On the two public benchmarks, Tables 3 and 4 show improved ranking and smaller proxy-calibration error, while Tables 5 and 6 identify contributions from the three components. Tables 7 and 8, together with Figures 6 and 7, show slower degradation under the tested covariate partitions. These conclusions remain limited by the cross-sectional proxy durations and static similarity graphs.

Future research should evaluate the method on longitudinal workforce records with defined entry, exit, and censoring processes; replace proxy similarity with validated organizational relations when ethically and legally available; test label and concept shift separately; and conduct intersectional fairness and utility studies before deployment. Human-resource governance bodies should require documented model purpose, independent validation, ongoing calibration and disparity audits, and human appeal procedures. Labor and data-protection authorities can further clarify limits on individual risk profiling and prohibit solely automated or punitive employment decisions based on predicted turnover propensity.

Author Contributions: Conceptualization, X.W., Q.G. and W.W.; Methodology, X.W., Q.G. and W.W.; Software, X.W., Q.G. and W.W.; Validation, X.W., Q.G. and W.W.; Formal analysis, X.W., Q.G. and W.W.; Investigation, X.W., Q.G. and W.W.; Resources, X.W., Q.G. and W.W.; Data curation, X.W., Q.G. and W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mobley, W.H. Intermediate Linkages in the Relationship Between Job Satisfaction and Employee Turnover. *J. Appl. Psychol.* **1977**, *62*, 237–240.
2. Cotton, J.L.; Tuttle, J.M. Employee Turnover: A Meta-Analysis and Review with Implications for Research. *Acad. Manag. Rev.* **1986**, *11*, 55–70.
3. Rubenstein, A.L.; Eberly, M.B.; Lee, T.W.; Mitchell, T.R. Surveying the Forest: A Meta-Analysis, Moderator Investigation, and Future-Oriented Discussion of the Antecedents of Voluntary Employee Turnover. *Pers. Psychol.* **2018**, *71*, 23–65.
4. Fallucchi, F.; Coladangelo, M.; Giuliano, R.; De Luca, E.W. Predicting Employee Attrition Using Machine Learning Techniques. *Computers* **2020**, *9*, 86. <https://doi.org/10.3390/computers9040086>.
5. Raza, A.; Munir, K.; Almutairi, M.; Younas, F.; Fareed, M.M.S. Predicting Employee Attrition Using Machine Learning Approaches. *Appl. Sci.* **2022**, *12*, 6424.
6. Madanchian, M. From Recruitment to Retention: AI Tools for Human Resource Decision-Making. *Appl. Sci.* **2024**, *14*, 11750. <https://doi.org/10.3390/app142411750>.
7. Cuturi, M. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems 26*; Curran Associates, Inc.: Red Hook, NY, USA, 2013; pp. 2292–2300.
8. Courty, N.; Flamary, R.; Tuia, D.; Rakotomamonjy, A. Optimal Transport for Domain Adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1853–1865.
9. Peyré, G.; Cuturi, M. Computational Optimal Transport: With Applications to Data Science. *Found. Trends Mach. Learn.* **2019**, *11*, 355–607. <https://doi.org/10.1561/22000000073>.
10. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the International Conference on Learning Representations*, Toulon, France, 24–26 April 2017.

11. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
12. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Networks Learn. Syst.* **2021**, *32*, 4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>.
13. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215.
14. Cox, D.R. Regression Models and Life-Tables. *J. R. Stat. Soc. Ser. B* **1972**, *34*, 187–220.
15. Ishwaran, H.; Kogalur, U.B.; Blackstone, E.H.; Lauer, M.S. Random Survival Forests. *Ann. Appl. Stat.* **2008**, *2*, 841–860.
16. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y. DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network. *BMC Med Res. Methodol.* **2018**, *18*, 24.
17. Lee, C.; Zame, W.R.; Yoon, J.; van der Schaar, M. DeepHit: A Deep Learning Approach to Survival Analysis with Competing Risks. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
18. Haider, H.; Hoehn, B.; Davis, S.; Greiner, R. Effective Ways to Build and Evaluate Individual Survival Distributions. *J. Mach. Learn. Res.* **2020**, *21*, 1–63.
19. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On Calibration of Modern Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*; Precup, D., Teh, Y.W., Eds.; PMLR (Proceedings of Machine Learning Research): Cambridge, MA, USA, 2017; Volume 70, pp. 1321–1330.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.