CAST: Causal Modeling of Time-Varying Treatment Effects on Head and Neck Cancer

Everest Yang[1,2]*, Ria Vasishtha[1,3], Luqman K. Dad[4], Lisa A. Kachnic[4], Andrew Hope[5], Eric Wang[1], Xiao Wu[6], Yading Yuan[4,7], David J. Brenner[1], Igor Shuryak[1]

[1] Center for Radiological Research, Columbia University Irving Medical Center, USA
[2] Department of Computer Science, Brown University, USA
[3] Department of Applied Mathematics, Columbia University, USA
[4] Department of Radiation Oncology, Columbia University Irving Medical Center, USA
[5] Department of Radiation Oncology, Princess Margaret Cancer Centre, Canada
[6] Department of Biostatistics, Columbia University Irving Medical Center, USA
[7] Data Science Institute, Columbia University, USA

Abstract

Causal machine learning (CML) is quickly gaining recognition in medical research because it offers better strategies to estimate treatment effects from complex real-world data, helping guide treatment optimization. Causal survival forests (CSF) are a powerful CML method for estimating heterogeneous treatment effects in survival outcomes, which are essential for informed healthcare decision-making. However, CSF only estimates effects at a fixed horizon, rather than at multiple time points. We introduce Causal Analysis for Survival Trajectories (CAST), a novel extension of CSF that models treatment effects as continuous parametric and non-parametric effect trajectories over time. Applied to the RADCURE dataset [1] of 2,651 head and neck cancer patients, CAST reveals how the effects of chemotherapy (added to radiotherapy) evolve over time at the population and individual levels. By capturing the temporal dynamics of treatment response, CAST can help clinicians determine when and for which patient subgroups treatment benefits are maximized.

1 Introduction

Methodological gap: Traditional statistical and machine learning models are based on correlations, making them unsuitable to answer the causal questions critical to clinical research [2, 3]. Causal machine learning (CML) addresses this by modeling causal effects to estimate individualized or subgroup-specific treatment responses [4]. Causal survival forests (CSF) extend CML to survival outcomes and flexibly estimate heterogeneous treatment effects - critical for medical applications - but estimate effects only at a fixed time, missing how effects evolve over time [5, 6, 7].

Proposed approach: We present CAST (Causal Analysis for Survival Trajectories), which extends causal survival forests to model treatment effects continuously over time. Combining parametric and non-parametric methods, CAST models the full trajectory of treatment response (Figure 1), which is key to understanding the complex and time-varying effects of cancer therapy [8, 9, 10]. Figure 1 includes imaging for context, though only clinical and treatment data were modeled here. Because all patients received radiotherapy, the treatment being modelled is chemotherapy yes/no only.

Clinical motivation: We apply CAST to head and neck squamous cell carcinoma (HNSCC), a cancer with increased incidence [11] and changing demographics due to HPV. HPV-related tumors, more

*Corresponding Author: Everest Yang (everest_yang@brown.edu)

Accepted to the NeurIPS 2025 Workshop on CauScien: Uncovering Causality in Science.

common in younger patients, differ in radiosensitivity and prognosis [12, 13, 14], highlighting the need for individualized treatment. Treatment typically combines chemotherapy and radiation with responses that vary between patients and manifest gradually [15, 16, 17, 18, 19].

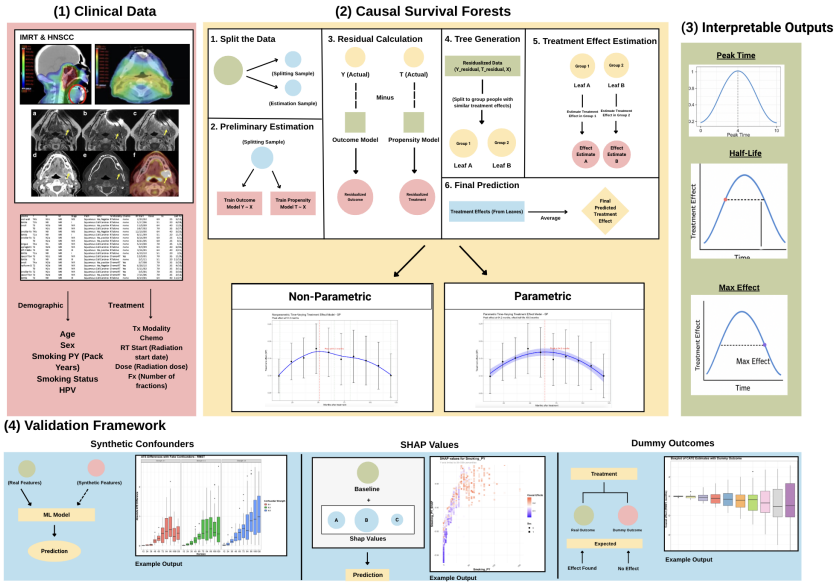


Figure 1: Overview of the CAST framework

2 Related Work

Recent CML methods estimate individual treatment effects (ITE) from time-to-event data [20, 21], but typically do so at discrete time points. This is particularly limiting in oncology, where treatment effects evolve through different biological phases driven by tumor response and toxicity [22, 23]. The response to treatment in HNSCC is highly heterogeneous, influenced by factors such as HPV status, gender, and stage of the disease [24, 25, 26]. Traditional survival models (e.g. Cox proportional hazards) assume constant effects over time, while more flexible models like RSF or DeepSurv improve risk prediction, but do not isolate causal effects [27, 28, 29]. CAST provides a unified framework to capture the full response trajectory.

3 Methodology

Causal Machine Learning Framework: Our approach uses a CML framework to isolate treatment effects beyond traditional correlational methods. CML allows us to understand the causal impact of interventions [30] with the goal of dissecting how treatments shape survival outcomes across patient subgroups. Given the observational, non-randomized nature of our clinical data, we rely on the following assumptions:

- **Unconfoundedness**: Treatment assignment is independent of potential outcomes conditional on observed covariates (also called ignorability or no unmeasured confounding)
- **Positivity (Overlap)**: Every subject has a non-zero probability of receiving each treatment
- **Consistency**: A subject's observed outcome under their received treatment equals their potential outcome for that treatment
- **Non-interference**: The treatment of one subject does not affect the outcome of another subject.

To address selection bias in observational data, we modelled propensity scores using elastic net logistic regression: $\hat{e}(X) = P(W = 1|X)$ with hyperparameters optimized through 10-fold cross-validation. Patients with extreme propensity scores (outside $[0.10, 0.90]$) were trimmed to ensure overlap between treatment groups. See Appendix C.1 for balance diagnostics.

3.1 CAST: Causal Analysis for Survival Trajectories

The theoretical foundation of CAST rests on modeling the effect trajectory as a function of time. Our target estimand is the conditional average treatment effect (CATE) at time t, given covariates X:

$$\tau(x,t) = \mathbb{E}[Y(1,t) - Y(0,t) \mid X = x] \tag{1}$$

where $Y(w,t)$ represents the potential outcome at time t under treatment w, and x denotes the covariates of an individual. We consider two types of time-varying estimands: the difference in restricted mean survival time (RMST) and the difference in survival probability (SP) between treatment groups. Unlike prior methods that estimate effects at fixed time points, CAST models treatment effects as smooth functions of time. We used a smoothing spline to estimate the continuous effect trajectory, and a quadratic fit to derive interpretable metrics.

3.1.1 Parametric Modeling Component

Our parametric modeling component employs a quadratic function: $\tau(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ to capture the rise and fall of treatment effects. The parameters are estimated using weighted least squares:

$$\min_{\beta_0,\beta_1,\beta_2} \sum_t w(t)(\hat{\tau}(t) - (\beta_0 + \beta_1 t + \beta_2 t^2))^2 \tag{2}$$

where $w(t) = 1/\sigma^2(t)$ are weights based on the variance of the effect estimates at each time point. This approach yields clinically interpretable parameters, including the peak effect time ($t_{\text{peak}} = -\beta_1/2\beta_2$), the maximum effect magnitude ($\tau(t_{\text{peak}})$), and the treatment effect half-life, defined as the time it takes for the effect to diminish by 50% from its peak.

These parameters directly quantify key clinical aspects of the treatment response: when the maximum benefit occurs, how large that benefit is, and how quickly it diminishes—information critical for clinical decision-making that traditional methods cannot provide. See Appendix C.3 for fitted coefficients and summary statistics from the parametric model.

3.1.2 Non-parametric Modeling Component

Our non-parametric component employs cross-validated smoothing splines:

$$\tau(t) = g(t), \quad \text{where} \quad g = \arg\min_f \left\{ \sum_t w(t)\left(\hat{\tau}(t) - f(t)\right)^2 + \lambda \int f''(t)^2\, dt \right\} \tag{3}$$

where λ is selected via cross-validation. This approach adapts to the data without imposing a predetermined functional form, revealing subtle inflection points in the effect trajectory that correspond to biological phase transitions in the treatment response.

We calculate the first and second derivatives of the fitted spline to identify key features of the treatment effect trajectory: local maxima and minima where $g'(t) = 0$, acceleration and deceleration phases based on sign changes in $g''(t)$, and inflection points where $g''(t) = 0$. The non-parametric model complements the parametric fit by capturing complex, less predictable patterns, especially during later follow-up periods, when biological processes like accelerated repopulation and late toxicities may cause deviations from the smooth quadratic trend.

Theoretical Guarantees: See Appendix A for theorem statements establishing the consistency of CAST parametric and nonparametric estimators, along with full algorithmic implementations.

4 Experiments

Data and Preprocessing: We used 2,651 patients from the publicly available RADCURE dataset, focusing on head and neck squamous cell carcinoma. We filtered incomplete records, standardized variables, and computed radiotherapy dose features using established radiobiological models [31].

The assignment of treatment (yes / no chemotherapy) was modeled by elastic net logistic regression, with overlap trimming to reduce confounding. The cohort was split into 75/25 for training / testing. See Appendix B for full hyperparameter settings, preprocessing and propensity modeling details, and sensitivity analyses.

Implementation & Heterogeneity Analysis: We used causal survival forests with Nelson-Aalen estimators to address right-censoring [32, 33], training models at 12–120 month horizons with 5,000 trees and doubly robust estimation using standardized covariates and propensity scores. Uncertainty was quantified via the infinitesimal jackknife. To analyze heterogeneity, we computed approximate SHAP values (1,000 Monte Carlo samples, $\epsilon = 0.01$), normalized to reflect deviations from the mean. HPV status and smoking history emerged as key predictors of chemotherapy response.

Validation Methods: We validated treatment effects using robustness checks, including dummy outcome tests (20 permutations), adding irrelevant covariates correlated with treatment but not outcome, negative control tests with random treatment assignments, and random noise injection. All refutation tests yielded null or stable effects. Additional visualizations are provided in Appendix C.4.

5 Results

CAST reveals a time-varying chemotherapy benefit that rises post-treatment, peaks between 48–60 months, and declines thereafter (Figure 2). This pattern is consistent across both the parametric and non-parametric models, suggesting a peak in benefit between 50 and 65 months with gradual tapering due to recurrence, long-term toxicity, or competing risks. On the testing set, chemotherapy increased survival probability by $15.2 \pm 6.0\%$ at 36 months and $15.0 \pm 6.7\%$ at 60 months, with RMST gains of 3.6 ± 1.4 and 7.1 ± 2.6 months, respectively.

Individualized effect distributions: Individual treatment effects varied across patients. While most experienced moderate benefit, CAST identified a long right tail of high responders and a subset with near-zero or negative effects.

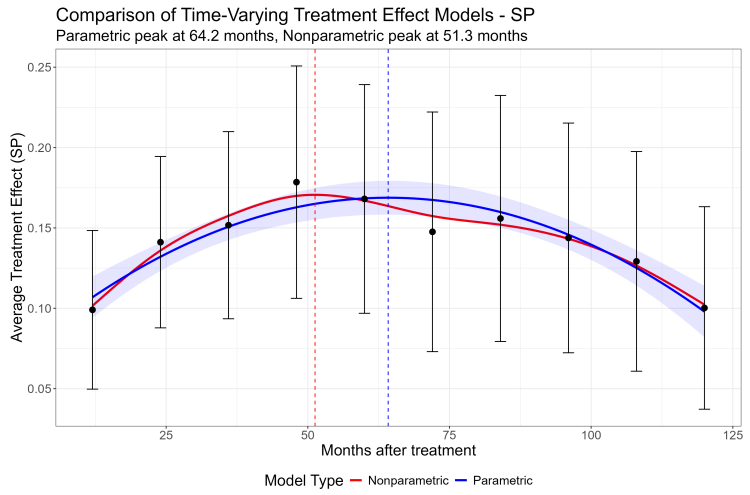


Figure 2: Comparison of time-varying treatment effect models using CAST. The red curve shows the parametric estimate with 95% CIs; the blue curve shows the non-parametric spline. Black dots denote average treatment effects $\pm$ standard errors on the survival probability scale.

Subgroup variation: Correlation matrices and SHAP analyses identified smoking pack-years as the strongest negative predictor of chemotherapy benefit, with HPV positivity and older age showing modest positive associations. Further SHAP visualizations are provided in Appendix C.2.

6 Conclusion

CAST reveals clinically significant chemotherapy patterns: strong short- to mid-term benefits that decline over time, likely due to recurrence or delayed toxicity [34]. These findings support adaptive

monitoring and strategies to maintain therapeutic effects. CAST also exposes treatment heterogeneity, with HPV status and smoking as key modifiers, and a modest age inflection near 50–60 years. Rather than replacing existing survival tools, CAST complements them by modeling continuous-time effects and capturing patient-level variation. Although currently limited to binary treatments, the framework can extend to dose, multiarm, and longitudinal designs. Beyond cancer, CAST applies to settings with evolving treatment effects—such as infectious disease interventions—to pinpoint critical windows, tailor care, and adapt strategies.

References

[1] M. L. Welch, S. Kim, A. Hope, S. H. Huang, Z. Lu, J. Marsilla, M. Kazmierski, K. Rey-McIntyre, T. Patel, B. O'Sullivan, J. Waldron, J. Kwan, J. Su, L. Soltan Ghoraie, H. B. Chan, K. Yip, M. Giuliani, Neck Site Group Princess Margaret Head, S. Bratman, and T. Tadic. Computed tomography images from large head and neck cohort (radcure) (version 4). *The Cancer Imaging Archive*, 2023. doi: 10.7937/J47W-NM11.

[2] M. Hung, J. Bounsanga, and M. W. Voss. Interpretation of correlations in clinical research. *Postgraduate Medicine*, 129(8):902–906, November 2017. doi: 10.1080/00325481.2017.1383820.

[3] H. A. Miot. Correlation analysis in clinical and experimental studies. *Jornal Vascular Brasileiro*, 17(4):275–279, December 2018. doi: 10.1590/1677-5449.174118.

[4] K. Shiba and K. Inoue. Harnessing causal forests for epidemiologic research: key considerations. *American Journal of Epidemiology*, 193(6):813–818, June 2024. doi: 10.1093/aje/kwae003.

[5] A. Venkatasubramaniam, B. A. Mateen, B. M. Shields, A. T. Hattersley, A. G. Jones, S. J. Vollmer, and J. M. Dennis. Comparison of causal forest and regression-based approaches to evaluate treatment effect heterogeneity: an application for type 2 diabetes precision medicine. *BMC Medical Informatics and Decision Making*, 23(1):110, June 2023. doi: 10.1186/s12911-023-02207-2.

[6] G. Solana-Lavalle, M. D. Cusimano, T. Steeves, R. Rosas-Romero, and P. N. Tyrrell. Causal forest machine learning analysis of parkinson's disease in resting-state functional magnetic resonance imaging. *Tomography*, 10(6):894–911, June 2024. doi: 10.3390/tomography10060068.

[7] Yifan Cui, Michael R. Kosorok, Erik Sverdrup, Stefan Wager, and Ruoqing Zhu. Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *Journal of the Royal Statistical Society: Series B*, 85(2):380–403, 2023. doi: 10.1093/jrsssb/qkac020.

[8] Z. Huang, N. A. Mayr, M. Gao, S. S. Lo, J. Z. Wang, G. Jia, and W. T. C. Yuh. The onset time of tumor repopulation for cervical cancer: first evidence from clinical data. *International Journal of Radiation Oncology*Biology*Physics*, 84(2):478–484, October 2012. doi: 10.1016/j.ijrobp.2011.12.037.

[9] C. Petersen and F. Würschmidt. Late toxicity of radiotherapy: a problem or a challenge for the radiation oncologist? *Breast Care (Basel)*, 6(5):369–374, October 2011. doi: 10.1159/000334220.

[10] I. Shuryak, E. J. Hall, and D. J. Brenner. Dose dependence of accelerated repopulation in head and neck cancer: Supporting evidence and clinical implications. *Radiotherapy and Oncology*, 127(1):20–26, April 2018. doi: 10.1016/j.radonc.2018.02.015.

[11] D. E. Johnson, B. Burtness, C. R. Leemans, V. W. Y. Lui, J. E. Bauman, and J. R. Grandis. Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers*, 6(92):1–22, November 2020. doi: 10.1038/s41572-020-00224-3.

[12] M. E. Sabatini and S. Chiocca. Human papillomavirus as a driver of head and neck cancers. *British Journal of Cancer*, 122(3):306–314, February 2020. doi: 10.1038/s41416-019-0602-7.

[13] D. C. Beachler and G. D'Souza. Nuances in the changing epidemiology of head and neck cancer. *Oncology (Williston Park)*, 24(10):924–926, September 2010. PMID: 21138173.

[14] G. M. P. van Kempen, R. J. Baatenburg de Jong, and R. J. H. Borra. Hpv and head and neck cancers: Towards early diagnosis and prevention. *Oral Oncology*, 128:105214, September 2022. doi: 10.1016/j.oraloncology.2022.105214.

[15] Janet Tu. How long does it take chemotherapy to shrink tumors? *Cancerwise, MD Anderson Cancer Center*, 2024. https://www.mdanderson.org/cancerwise/how-long-does-it-take-chemotherapy-to-shrink-tumors.h00-159696756.html.

[16] UCSF Health. Coping with chemotherapy. *Patient Education, UCSF Health*, 2025. https://www.ucsfhealth.org/education/coping-with-chemotherapy.

[17] S. R. Rathod, S. Gupta, S. Ghosh-Laskar, V. Murthy, A. Budrukkar, J. Agarwal, and K. Kannan. Quality-of-life (qol) outcomes in patients with head and neck squamous cell carcinoma treated with intensity-modulated radiation therapy (imrt) compared to three-dimensional conformal radiotherapy (3d-crt): Evidence from a prospective randomized study. *Oral Oncology*, 49(6): 634–640, June 2013. doi: 10.1016/j.oraloncology.2013.02.013.

[18] A. Viganò, F. De Felice, N. A. Iacovelli, D. Alterio, R. Ingargiola, A. Casbarra, N. Facchinetti, O. Oneta, A. Bacigalupo, E. Tornari, S. Ursino, F. Paiar, O. Caspiani, A. Di Rito, D. Musio, P. Bossi, P. Steca, B. A. Jereczek-Fossa, L. Caso, N. Palena, A. Greco, and E. Orlandi. Quality of life changes over time and predictors in a large head and neck patients' cohort: secondary analysis from an italian multi-center longitudinal, prospective, observational study—a study of the italian association of radiotherapy and clinical oncology (airo) head and neck working group. *Supportive Care in Cancer*, 31(4):220, March 2023. doi: 10.1007/s00520-023-07661-2.

[19] R. Yang, A. C. Freeman-Cook, H. C. Kurnik, and D. C. Kirouac. Dissecting variability in responses to cancer chemotherapy through systems pharmacology. *Clinical Pharmacology & Therapeutics*, 88(1):34–38, July 2010. doi: 10.1038/clpt.2010.96.

[20] Y. Zhang, N. Kreif, V. S. Gc, and A. Manca. Machine learning methods to estimate individualized treatment effects for use in health technology assessment. *Medical Decision Making*, 44 (7):756–769, October 2024. doi: 10.1177/0272989X241263356.

[21] V. Chernozhukov, C. Hansen, N. Kallus, M. Spindler, and V. Syrgkanis. Applied causal inference powered by ml and ai. *arXiv preprint*, arXiv:2403.02467, March 2024. doi: 10.48550/arXiv.2403.02467.

[22] W. J. Allard and L. W. M. M. Terstappen. Ccr 20th anniversary commentary: Paving the way for circulating tumor cells. *Clinical Cancer Research*, 21(13):2883–2885, July 2015. doi: 10.1158/1078-0432.CCR-14-2559.

[23] J. A. Langendijk, P. Doornaert, I. M. Verdonck de Leeuw, C. R. Leemans, N. K. Aaronson, and B. J. Slotman. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *Journal of Clinical Oncology*, 26(22): 3770–3776, August 2008. doi: 10.1200/JCO.2007.14.6647.

[24] Y. Sun, Z. Wang, S. Qiu, and R. Wang. Therapeutic strategies of different hpv status in head and neck squamous cell carcinoma. *International Journal of Biological Sciences*, 17(4):1104–1118, March 2021. doi: 10.7150/ijbs.58077.

[25] K. K. Ang, J. Harris, R. Wheeler, R. Weber, D. I. Rosenthal, P. M. Nguyen-Tan, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *New England Journal of Medicine*, 363(1):24–35, July 2010. doi: 10.1056/NEJMoa0912217.

[26] Y. Wu, Y. Wang, J. Liu, Y. Wang, Y. Li, Y. Hu, H. Qiu, Z. Liang, Y. Wei, and H. Zhong. Hpv-positive status is a favorable prognostic factor in non-nasopharyngeal head and neck squamous cell carcinoma patients: a population-based study. *Frontiers in Oncology*, 11:765, October 2021. doi: 10.3389/fonc.2021.765.

[27] N. Jiang, Y. Wu, and C. Li. Limitations of using cox proportional hazards model in cardiovascular research. *Cardiovascular Diabetology*, 23(219), June 2024. doi: 10.1186/s12933-024-02302-2.

[28] L. Xu, S. Jiang, T. Li, and Y. Xu. Limitations of the cox proportional hazards model and alternative approaches in metachronous recurrence research. *Gastric Cancer*, 27(6):1348–1349, November 2024. doi: 10.1007/s10120-024-01554-x.

[29] S. Saha. Survival analysis with bayesian additive regression trees and its application. https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations/5158/, 2017. Northern Illinois University Thesis.

[30] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):C1–C68, January 2018. doi: 10.1093/ectj/uty017.

[31] I. Shuryak, E. Wang, and D. J. Brenner. Understanding the impact of radiotherapy fractionation on overall survival in a large head and neck squamous cell carcinoma dataset: A comprehensive approach combining mechanistic and machine learning models. *Frontiers in Oncology*, 14: 1422211, August 2024. doi: 10.3389/fonc.2024.1422211.

[32] S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, July 2018. doi: 10.1080/01621459.2017.1319839.

[33] S. Athey, J. Tibshirani, and S. Wager. Generalized random forests. *Annals of Statistics*, 47(2): 1148–1178, April 2019. doi: 10.1214/18-AOS1709.

[34] I. Shuryak, E. J. Hall, and D. J. Brenner. Optimized hypofractionation can markedly improve tumor control and decrease late effects for head and neck cancer. *International Journal of Radiation Oncology, Biology, Physics*, 104(2):272–278, June 2019. doi: 10.1016/j.ijrobp.2019.02.025.

Ethics Statement

Existing at the intersection of machine learning (ML), healthcare, and causal inference, our work inevitably raises ethical considerations. By bringing ML methods to oncology research, we strive to advance personalized medicine and treatment strategies. However, our estimates are based on observational data and may be biased by unmeasured confounding. While the dataset includes a comprehensive description of variables including age, sex, smoking history, and HPV status, it omits race, ethnicity, and socioeconomic status data. These factors are key to understanding structural barriers to healthcare that could possibly affect outcomes. This risks amplifying existing biases in the data. ML models in oncology must be used cautiously and should not replace clinical judgment, but rather act as a supplement. Our findings require further clinical validation before integration into decision-making workflows.

A Theoretical Justification of CAST

We provide formal justification for the consistency and identifiability of the time-varying treatment effect estimator $\hat{\tau}(t)$ used in the CAST framework.

A.1 Problem Setting

Let $\mathcal{D} = \{(X_i, W_i, T_i, \delta_i)\}_{i=1}^n$ be a dataset of n i.i.d. samples where: - $X_i \in \mathbb{R}^p$ is a vector of observed covariates, - $W_i \in \{0, 1\}$ is a binary treatment indicator, - T_i is the observed event or censoring time, - $\delta_i \in \{0, 1\}$ is the event indicator (1 if the event occurred, 0 if censored).

Let $Y(w, t)$ denote the potential outcome (e.g., survival status at time t) under treatment $w \in \{0, 1\}$.

We define the time-varying Conditional Average Treatment Effect (CATE) as:

$$\tau(x, t) := \mathbb{E}[Y(1, t) - Y(0, t) \mid X = x].$$

CAST estimates $\tau(x, t)$ using a doubly-robust causal survival forest followed by a spline or quadratic fit across time.

A.2 Assumptions

We adopt standard causal inference and survival analysis assumptions:

(A1) **Unconfoundedness:** $(Y(0, t), Y(1, t)) \perp W \mid X$ for all t.

(A2) **Positivity:** $0 < P(W = 1 \mid X) < 1$ almost surely.

(A3) **Consistency:** $Y = Y(W, t)$ if W is received.

(A4) **Non-informative Censoring:** $C \perp (Y(0, t), Y(1, t)) \mid X, W$ for censoring time C.

(A5) **Consistency of Forest Estimators:** The causal survival forests used yield consistent estimates of conditional survival functions $S_w(t \mid X)$.

A.3 Theorem: Pointwise Consistency of $\hat{\tau}(t)$

[Pointwise Consistency] Under assumptions (A1)–(A5), for each fixed t:

$$\hat{\tau}(t) := \mathbb{E}_X[\hat{S}_1(t \mid X) - \hat{S}_0(t \mid X)] \xrightarrow{p} \tau(t) := \mathbb{E}_X[S_1(t \mid X) - S_0(t \mid X)]$$

as $n \to \infty$, where $\hat{S}_w(t \mid X)$ is the estimated conditional survival function under treatment w from causal survival forests.

This follows from: 1. Consistency of $\hat{S}_w(t \mid X)$ (A5), 2. The continuous mapping theorem, since subtraction and expectation are continuous, 3. Trimming enforces overlap (A2), ensuring bounded inverse propensity weights.

A.4 Identifiability of $\tau(t)$ from Observational Data

[Identifiability] Under assumptions (A1)–(A4), the marginal time-varying treatment effect

$$\tau(t) := \mathbb{E}_X[\mathbb{E}[Y \mid W = 1, X, T \geq t] - \mathbb{E}[Y \mid W = 0, X, T \geq t]]$$

is identified from observational data using inverse probability weighting or doubly-robust estimation.

Under unconfoundedness and non-informative censoring, we can consistently estimate the conditional means $\mathbb{E}[Y(w, t) \mid X]$ from observed data. The difference in conditional expectations across treatment groups yields an identifiable estimator of $\tau(t)$.

A.5 Estimability of Peak Effect Time in CAST-Parametric

Let the parametric effect trajectory be:

$$\tau(t) = \beta_0 + \beta_1 t + \beta_2 t^2,$$

and suppose $\hat{\beta}_1, \hat{\beta}_2$ are estimated using weighted least squares.

[Consistency of Estimated Peak Time] If $\hat{\beta}_1 \xrightarrow{p} \beta_1$, $\hat{\beta}_2 \xrightarrow{p} \beta_2$ with $\beta_2 < 0$, then the estimated peak time

$$\hat{t}^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

is a consistent estimator of the true peak $t^* = -\frac{\beta_1}{2\beta_2}$.

This follows from Slutsky's theorem. Since both $\hat{\beta}_1$ and $\hat{\beta}_2$ converge in probability to non-zero limits, and the mapping $f(a, b) = -a/(2b)$ is continuous for $b \neq 0$, it follows that:

$$\hat{t}^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2} \xrightarrow{p} -\frac{\beta_1}{2\beta_2} = t^*.$$

A.6 Algorithmic Implementation of CAST

Algorithm 1 CAST-PARAMETRIC

Input: Horizons $\mathcal{H}$, ATEs $\{\hat{\tau}_h\}$, SEs $\{\hat{\sigma}_h\}$
Output: Temporal function $\hat{\tau}(t)$, peak time t^*, half-life λ
$\mathcal{W} \leftarrow \{w_h = 1/\hat{\sigma}_h^2\}$ ▷ Inverse-variance weights
$\hat{\tau}(t) \leftarrow$ FITQUADRATICMODEL$(\mathcal{H}, \hat{\tau}, \mathcal{W})$
$\beta_1, \beta_2 \leftarrow$ coefficients from fit
if $\beta_2 \neq 0$ **then**
  $t^* \leftarrow -\beta_1/(2\beta_2)$ ▷ Time of peak effect
  $\lambda \leftarrow$ SOLVE$(\hat{\tau}(t^* + \lambda) = \hat{\tau}(t^*)/2)$
else
  $t^*, \lambda \leftarrow$ NA ▷ Degenerate case
end if
return $\hat{\tau}(t), t^*, \lambda$

CAST-Parametric: This algorithm models treatment effects over time using a weighted quadratic fit to the estimated ATEs across discrete horizons. Inverse-variance weighting emphasizes more confident estimates. The peak effect time is derived analytically, while the half-life is computed by numerically solving for the point where the curve falls to half its maximum. This approach yields interpretable summaries of treatment dynamics, aligning with radiobiological phenomena such as delayed benefit and diminishing returns.

Algorithm 2 CAST-NONPARAMETRIC

Input: Horizons $\mathcal{H}$, ATEs $\{\hat{\tau}_h\}$, SEs $\{\hat{\sigma}_h\}$
Output: Spline $\hat{\tau}(t)$, peak t^*, inflections $\{t_i\}$
$\mathcal{W} \leftarrow \{w_h = 1/\hat{\sigma}_h^2\}$
$\hat{\tau}(t) \leftarrow$ FITSPLINE$(\mathcal{H}, \hat{\tau}, \mathcal{W})$
$D_1(t), D_2(t) \leftarrow$ first and second derivatives of $\hat{\tau}(t)$
$t^* \leftarrow$ ARGMAX$(\hat{\tau}(t))$ ▷ Peak effect
$\{t_i\} \leftarrow$ ZEROCROSSINGS$(D_2(t))$ ▷ Inflection points
if t^* not in $[\min(\mathcal{H}), \max(\mathcal{H})]$ **then**
  $t^* \leftarrow$ NA
end if
return $\hat{\tau}(t), t^*, \{t_i\}$

CAST-Nonparametric: This algorithm fits a smoothing spline to the estimated treatment effects across time using inverse-variance weights. It computes the first and second derivatives of the spline to identify key dynamics: the peak effect time via the curve's global maximum and biological phase transitions via inflection points. This method captures delayed and non-monotonic effect trajectories often missed by parametric models, reflecting immune response, tissue adaptation, or timing heterogeneity.

B Expanded Dataset Subsection

Overview

Our analysis uses the RADCURE dataset from The Cancer Imaging Archive (TCIA), the largest to our knowledge publicly accessible head and neck cancer imaging dataset. The data spans from 2005 to 2017 and includes computed tomography (CT) images for 3,346 patients, from which we selected a subset of 2,651 patients after filtering for only HNSCC cases. These images are linked to clinical, demographic, and treatment metadata. Following standardized clinical imaging protocols, the RADCURE project includes CT images, pictured alongside manually-reviewed contours differentiating between the planning tumor volume (PTV) and the organs at risk (OARs). All patients in this dataset received radiotherapy, and some received chemotherapy.

The clinical data accounts for patient demographics, including age, gender, and HPV status. It also details tumor staging using the 7th edition TNM system to describe the cancer, in addition to treatment information. While the dataset primarily focuses on oropharyngeal cancer, it also covers laryngeal, nasopharyngeal, and hypopharyngeal cancers.

Data Preprocessing

In the preprocessing stage, we filtered out incomplete patient profiles to ensure the dataset included relevant variables and appropriately represented potential confounders. We standardized all continuous variables to have zero mean and unit variance to ensure comparability and optimize model performance. The dataset comprehensively describes treatment details—dose/fraction, number of fractions, and total days of radiotherapy—which we used to calculate Biologically Effective Dose (BED) values. We implemented both dose-independent (DI) and dose-dependent (DD) BED models to capture the biological effects of radiation therapy, using established radiobiological parameters ($\alpha = 0.2$ Gy^{-1}, $\alpha/\beta = 10$ Gy, accelerated repopulation rates and onset times). This allowed us to quantify the effective radiation dose accounting for different fractionation schedules. We employed a stratified data partitioning strategy, creating training (75%) and testing (25%) sets while maintaining consistent event rates across partitions. Both subsets contained similar proportions of survival events, allowing for unbiased evaluation of treatment effects.

Table 1 summarizes the estimated average treatment effects across time for both restricted mean survival time (RMST) and survival probability (SP) metrics. These values were computed using causal survival forests on held-out test data. We observe that the estimated effects generally increase with longer follow-up, particularly under the RMST metric, reflecting the accumulating benefit of treatment over time. Standard errors are included to reflect model uncertainty at each horizon.

Table 1: Summary statistics of the real dataset

Statistic	**Control Group**	**Treated Group**
Event Rate (%)	79.8	
Treatment Rate (%)	44.9	
Median Survival (months)	17.0	24.0
12-month Survival (%)	70.3	90.1
24-month Survival (%)	20.2	45.5
36-month Survival (%)	1.9	7.3
48-month Survival (%)	0.0	0.1
Age (mean)	60.42	59.23
TNM Stage (mean)	1.73	3.46
HPV Positivity Rate	0.68	0.51
Sex (Male = 1)	0.48	0.49

Computing Resources: All experiments were conducted with a 13th Gen Intel Core i7-1355U CPU, 16GB RAM, and integrated Intel Iris Xe Graphics. No discrete GPU or cloud resources were used, though such resources would significantly reduce runtime for large-scale extensions of this work.

C Additional Results

In this section, we present additional results that extend and validate the findings reported in the main paper. These include visualizations of treatment effect heterogeneity across time, a summary of average treatment effects, and robustness checks to support the reliability of our causal estimates.

C.1 Summary Table of Average Treatment Effects

Table 2 summarizes the estimated average treatment effects across time horizons using both RMST and survival probability metrics. These values were computed using causal survival forests on the held-out test set. The treatment effects tend to increase over time under both metrics, with RMST showing a steeper upward trend reflecting cumulative benefit. Standard errors are included for each estimate. The early rise in both SP and RMST suggests initial treatment efficacy, while the plateauing in later months reflects diminishing returns, possibly due to recurrence or late toxicity. The RMST gains—peaking at over 16 months—highlight how cumulative survival benefit continues to accrue even as survival probability differences taper off. These patterns support the biological intuition that treatment effects rise quickly post-intervention and then gradually attenuate.

Table 2: Estimated average treatment effects (ATE) across time using RMST and survival probability (SP). SE represent standard errors

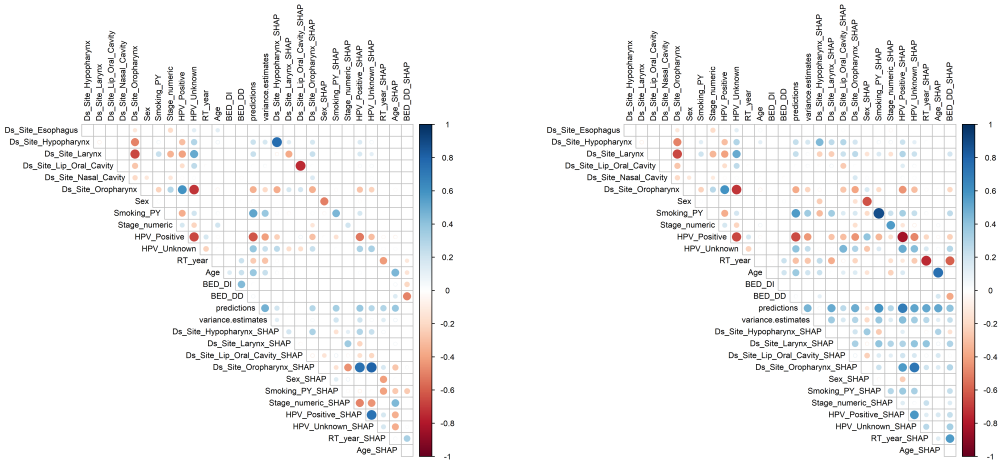
Months	ATE (SP)	SE (SP)	ATE (RMST)	SE (RMST)
12	0.099	0.049	0.44	0.26
24	0.141	0.053	1.88	0.80
36	0.152	0.058	3.58	1.46
48	0.178	0.072	5.80	2.31
60	0.168	0.071	7.39	2.73
72	0.148	0.075	8.38	3.52
84	0.156	0.077	11.08	4.76
96	0.143	0.071	13.89	5.90
108	0.129	0.068	14.76	6.16
120	0.100	0.063	16.11	6.92

These summary statistics also inform the CAST modeling strategies described in Section 3.3. The steady increase followed by tapering motivates the use of both quadratic and spline-based approaches to flexibly capture the full temporal arc of treatment efficacy.

C.2 SHAP-Based Interpretability Analysis

While SHAP provides valuable insights into feature influence, the estimates generated here using the fastshap R package are approximate and may be noisy, particularly in the context of survival analysis. We calculated approximate SHAP values because an exact SHAP explainer does not yet exist for the causal survival forest model.

To identify drivers of treatment heterogeneity, we computed Pearson and Spearman correlation matrices between clinical covariates, SHAP values, and estimated treatment effects (Figure 3a,b). Pearson captures linear relationships, while Spearman reflects monotonic trends, offering complementary views of variable influence. Smoking pack-years showed the strongest and most consistent negative correlation across both matrices, reinforcing its role in reducing chemotherapy benefit. HPV-negative status also correlated with higher chemotherapy benefit, while HPV positivity was associated with reduced benefit. Older age showed a modest positive correlation with treatment effects.

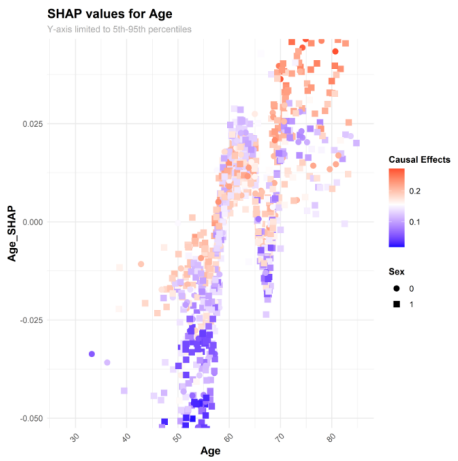


(a) Pearson Correlation

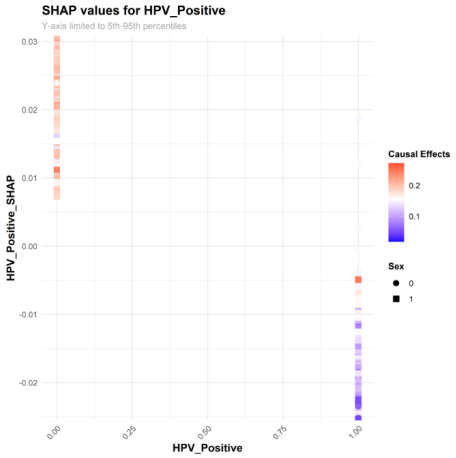
(b) Spearman Correlation

Figure 3: Correlation matrices between covariates, SHAP values, and treatment effects

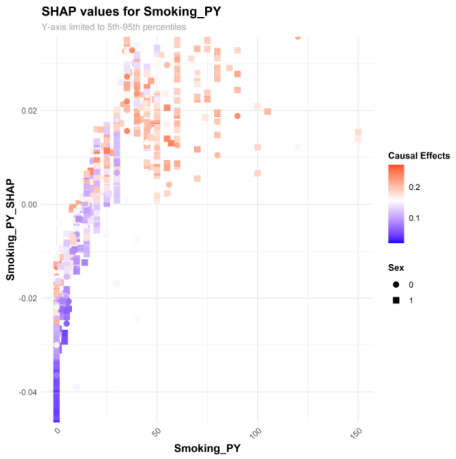
Figures 4(a–c) show SHAP plots for the three most influential variables—age, HPV status, and smoking pack-years—highlighting clear heterogeneity in treatment benefit across subgroups. Additional SHAP plots for other covariates—such as tumor site, treatment timing, dose metrics, and TNM stage—are also provided below. These variables had smaller contributions to the model, but are shown for completeness and transparency.



(a) Age

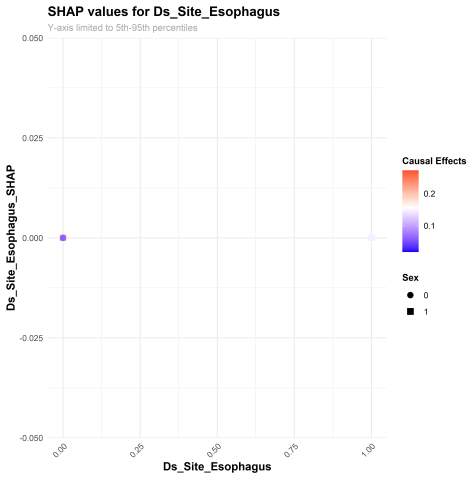


(b) HPV Status

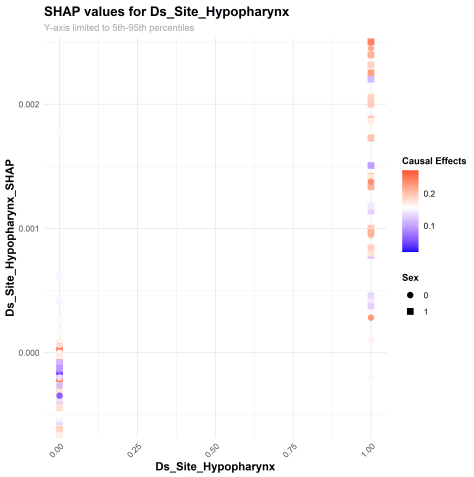


(c) Smoking Pack-Years

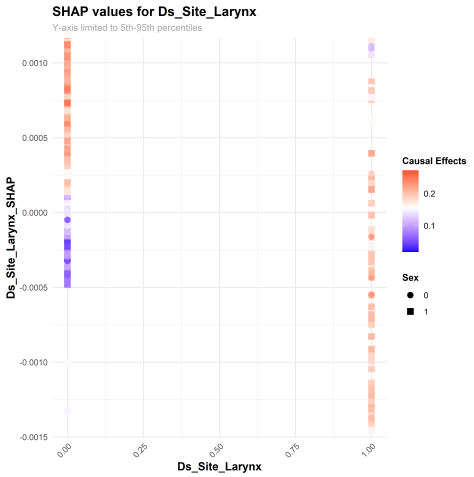
Figure 4: SHAP analysis of covariates driving treatment effect heterogeneity. (a) Older age is linked to greater chemotherapy benefit. (b) HPV-negative patients consistently show higher contributions. (c) Smoking Pack-Years is one of the features of strongest association treatment effect heterogeneity.



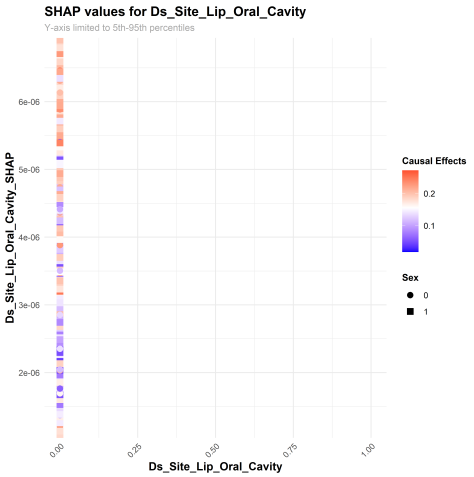
(a) Esophagus



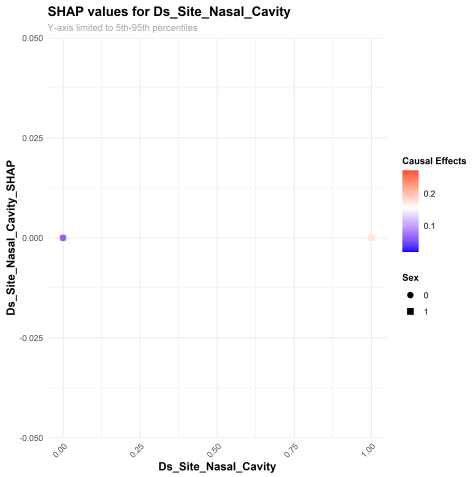
(b) Hypopharynx



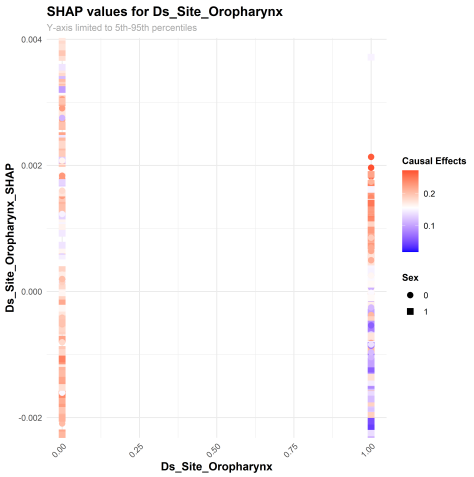
(c) Larynx



(d) Lip/Oral Cavity

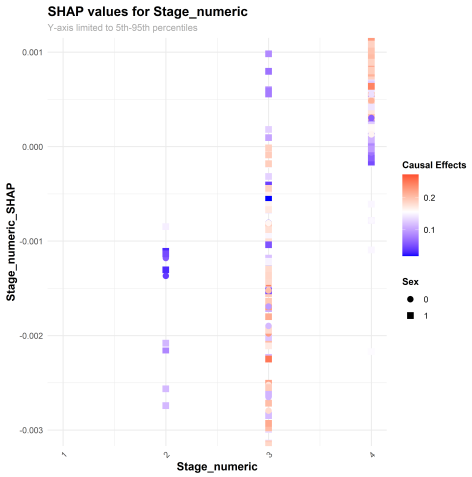


(e) Nasal Cavity

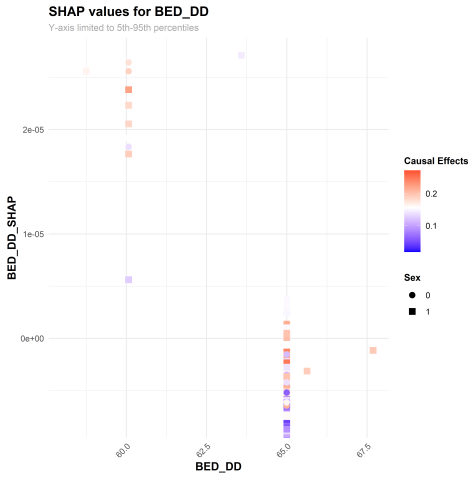


(f) Oropharynx

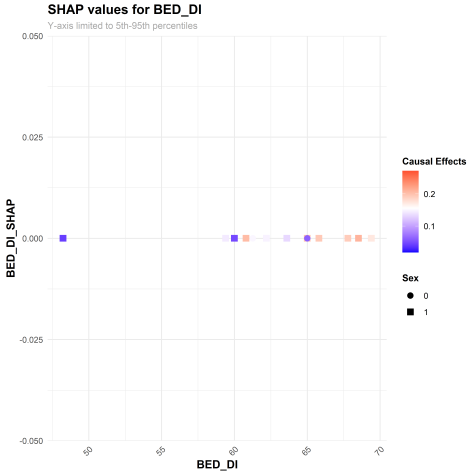
Figure 5: SHAP values for primary tumor site. These anatomical subgroups exhibited low or diffuse contributions to treatment effect heterogeneity, though subtle site-specific trends may still hold clinical value.



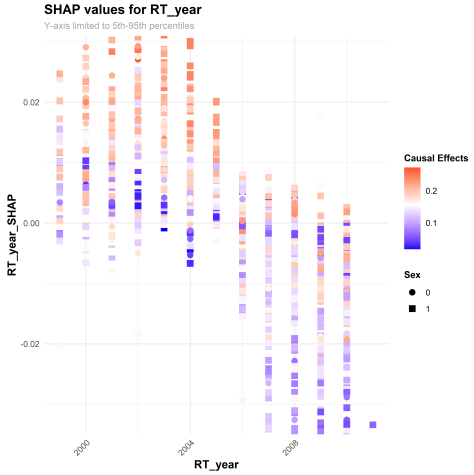
(a) TNM Stage



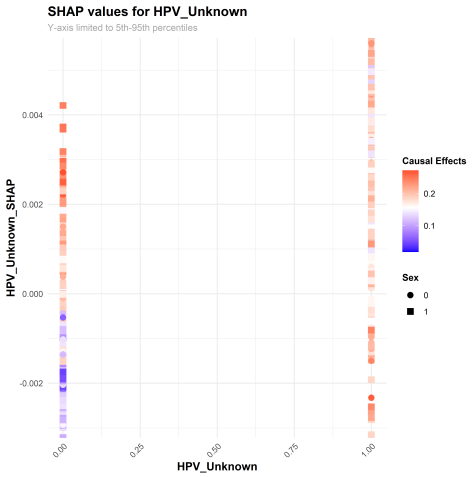
(b) BED (Dose-Dependent)



(c) BED (Dose-Independent)



(d) Year of RT



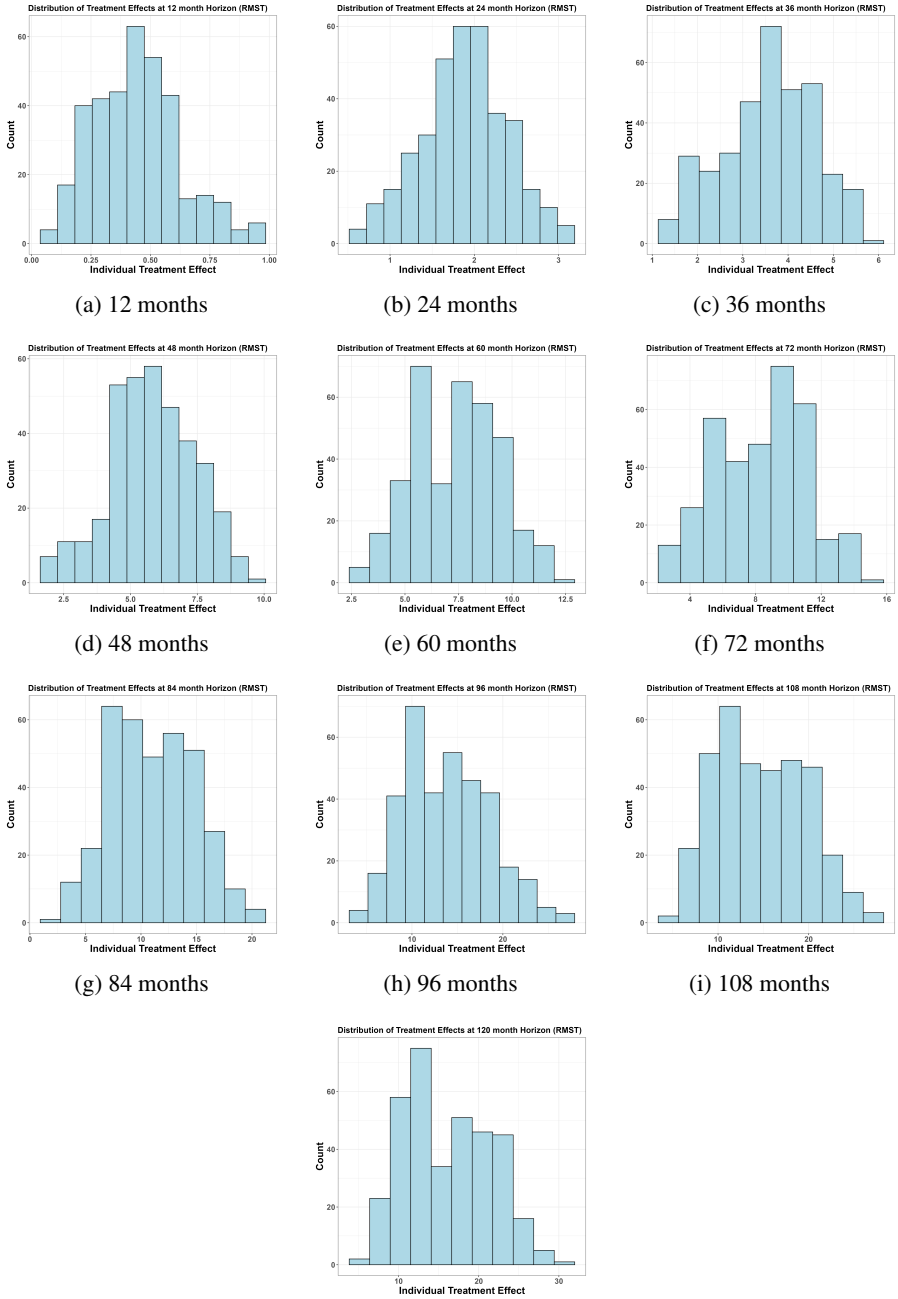
(e) HPV Unknown

Figure 6: SHAP values for additional covariates, including TNM stage, treatment year, and dose-related metrics. These features showed limited or context-specific contributions to treatment effect heterogeneity.

C.3 Distributions of Individualized Treatment Effects

We visualize the estimated treatment effect distributions for both RMST and survival probability (SP) at intervals ranging from 12 to 120 months. Figures 4 and 5 show individual-level causal effects derived from the causal survival forest at each time horizon.

RMST Treatment Effect Distributions



(a) 12 months (b) 24 months (c) 36 months

(d) 48 months (e) 60 months (f) 72 months

(g) 84 months (h) 96 months (i) 108 months

(j) 120 months

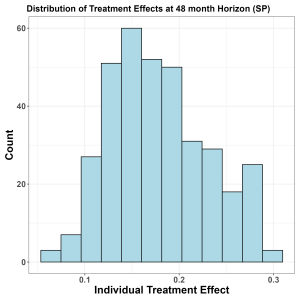
Figure 7: Distributions of estimated RMST-based treatment effects over time. Each panel shows the individual-level causal effect at a specific horizon as learned by the causal survival forest.

Survival Probability Treatment Effect Distributions

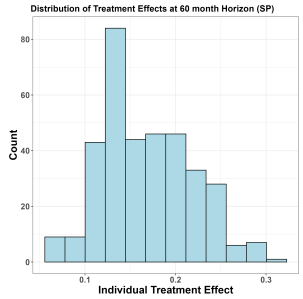
(a) 12 months

(b) 24 months

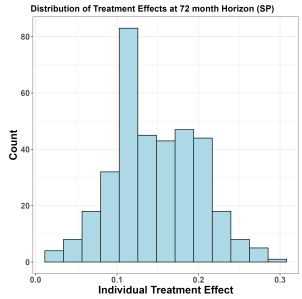
(c) 36 months



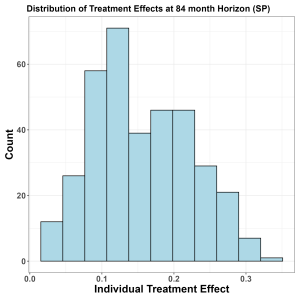
(d) 48 months



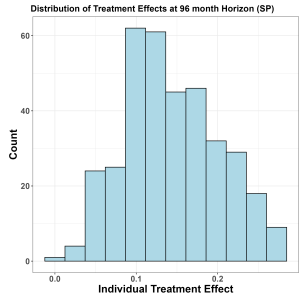
(e) 60 months



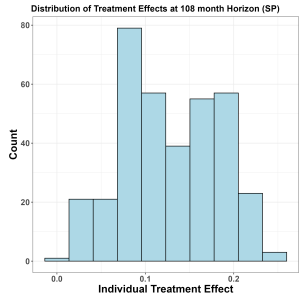
(f) 72 months



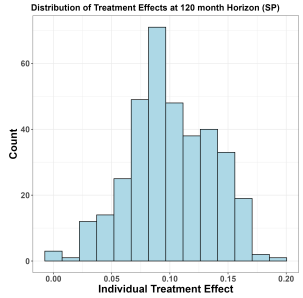
(g) 84 months



(h) 96 months



(i) 108 months



(j) 120 months

Figure 8: Distributions of estimated survival-probability-based treatment effects over time. Each panel shows the individual-level causal effect at a specific horizon as estimated by the causal survival forest.

C.4 Dummy Outcome Refutation Tests

To assess whether CAST detects spurious treatment effects in the absence of a true signal, we performed dummy outcome tests. For each time horizon, we randomly shuffled treatment assignments and outcome times across 20 repetitions to simulate a null setting. If the model was overfitting or improperly attributing causal structure, it would produce non-zero treatment effect estimates even under randomization. As shown in the boxplots below, the estimated treatment effects for both RMST and survival probability are centered around zero, especially at relatively short times (≤ 60 months), when the number of patients still at risk was large. This confirms that CAST does not learn artifacts from the data and is robust to randomization of causal structure.

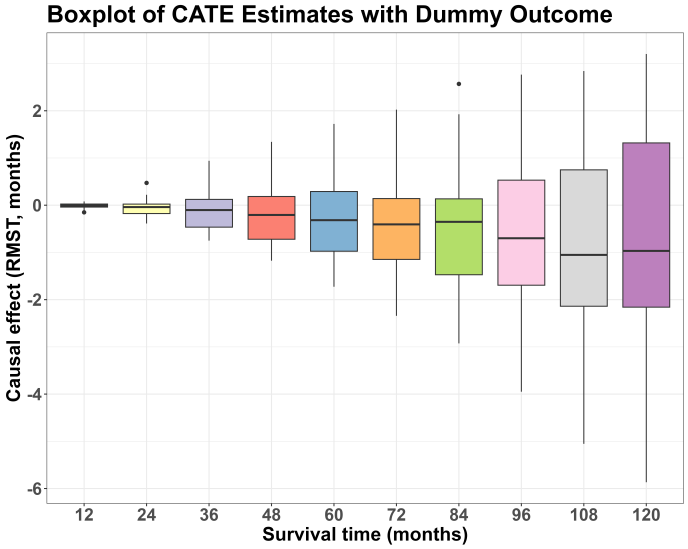


Figure 9: Dummy outcome test for RMST-based ATE estimates. Across 20 shuffles per horizon, treatment effects are centered near zero, consistent with the null.

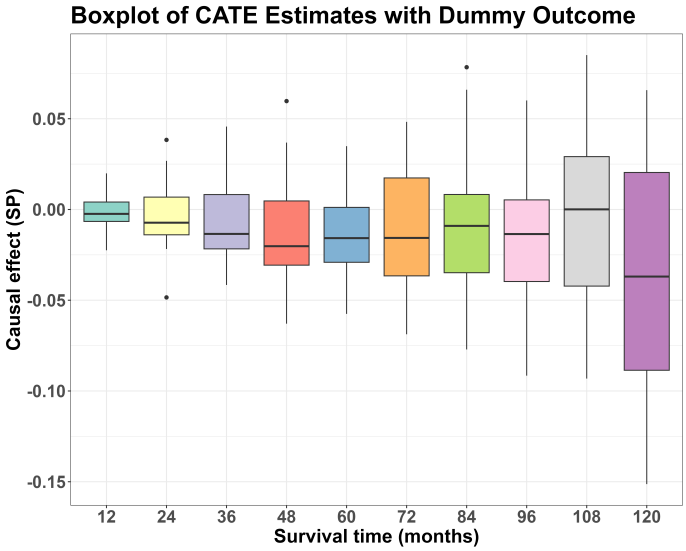


Figure 10: Dummy outcome test for survival probability-based ATE estimates. The model correctly reports no significant treatment effects under randomized labels.

To assess the robustness of CAST estimates to unobserved confounding, we performed a sensitivity analysis by injecting synthetic covariates with varying correlation to treatment assignment ($r = 0.1$, 0.3, 0.5). We then measured the resulting shifts in ATE estimates across time horizons for both RMST and survival probability outcomes.

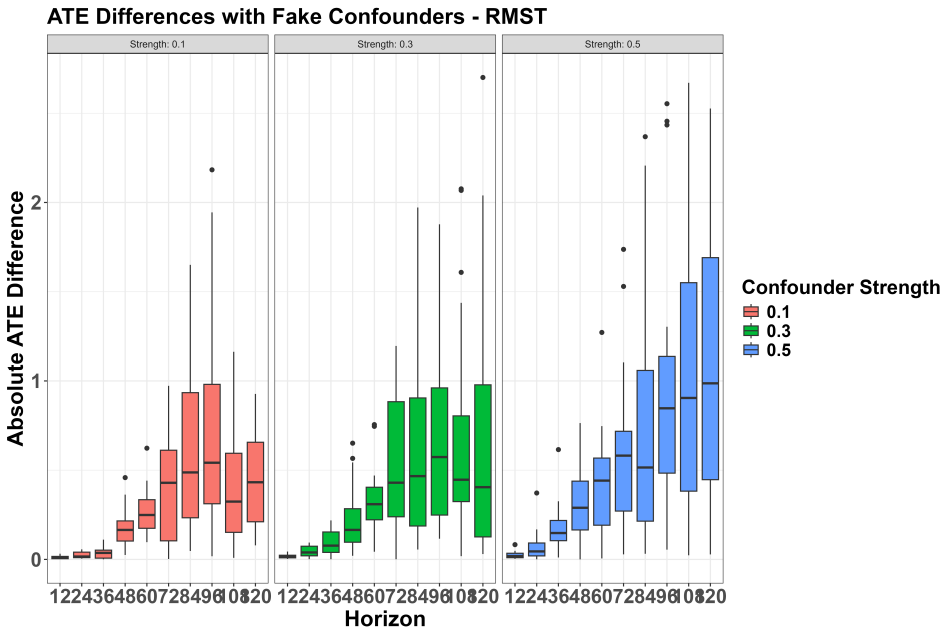


Figure 11: Absolute ATE differences in RMST under varying confounder strengths ($r = 0.1$, 0.3, 0.5). Estimates are stable under weak strengths but diverge at longer horizons and higher strengths.

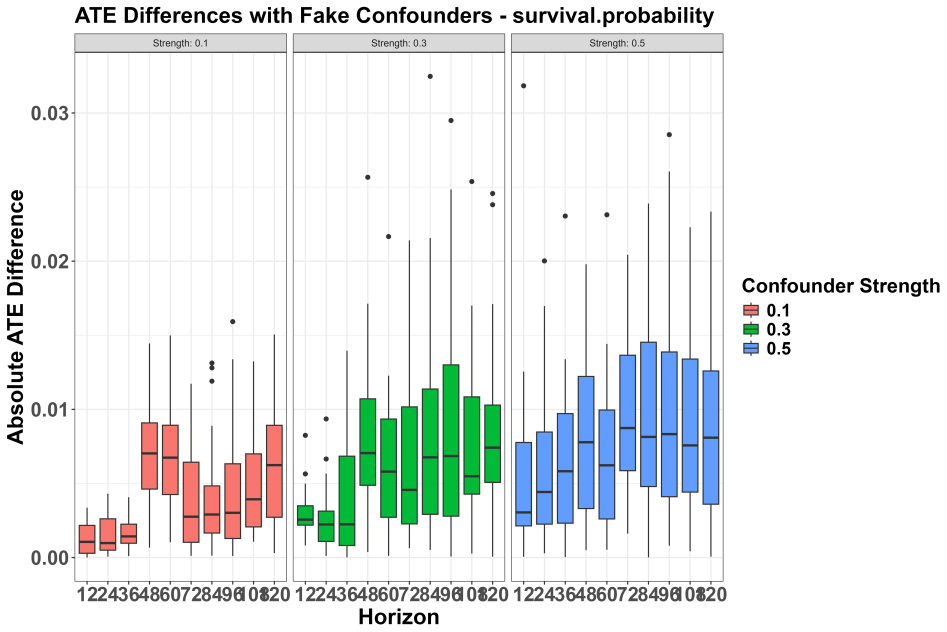


Figure 12: Absolute ATE differences in SP under varying confounder strengths ($r = 0.1$, 0.3, 0.5). CAST estimates remain stable under weak strengths, with modest shifts at stronger levels and longer horizons.