INTEGRATING MECHANISTIC MODELING WITH MACHINE LEARNING TO EVALUATE RADIOTHERAPY AND CHEMOTHERAPY OUTCOMES IN HEAD AND NECK CANCER

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

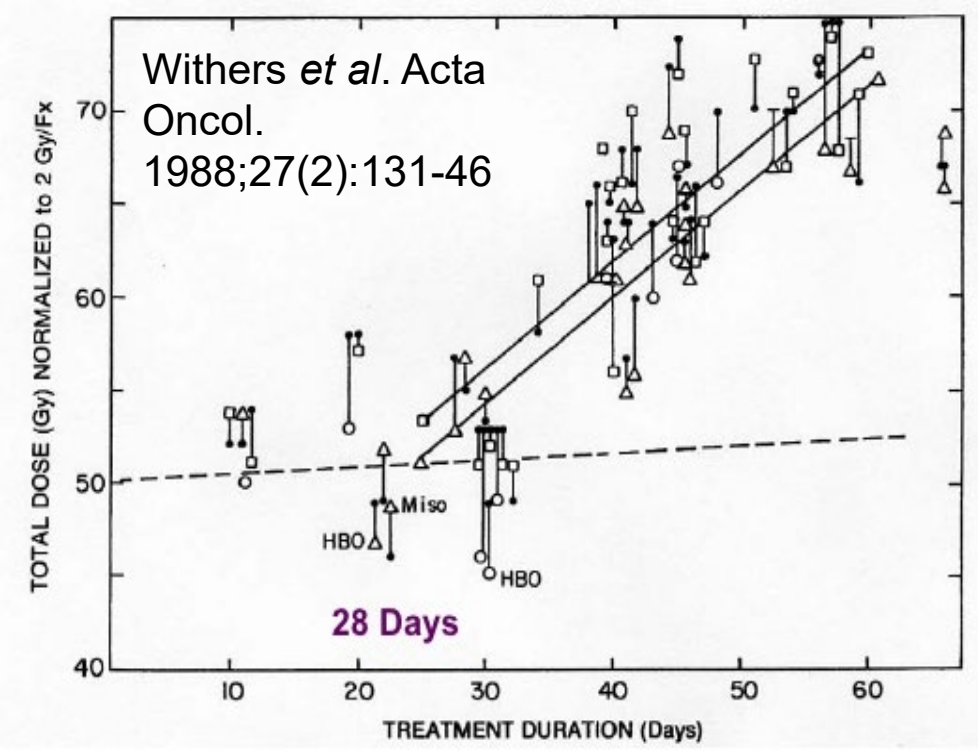
[1] Center for Radiological Research, Columbia University Irving Medical Center, USA
[2] Department of Computer Science, Brown University, USA
[3] Department of Mathematics, Columbia University, USA
[4] Department of Radiation Oncology, Princess Margaret Cancer Centre, Canada
[5] Department of Biostatistics,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

Integrating Mechanistic and ML Models

- Mathematical modeling in radiation biology/oncology has a long history (*e.g.* LQ model).
- Such models are based on diverse data sources: animal, *in vitro*, human clinical data.
- However, these models are simple and cannot include multiple relevant features: patient demographics, treatment and disease details, omics and imaging.
- In contrast, ML methods can integrate multiple features and modalities, generate accurate predictions, but are not as easy to interpret ("black box").
- **It makes sense to integrate these two approaches together to make more accurate and interpretable models:** concepts from simple models, like Biologically Effective Dose (BED), can enter into ML models as engineered features.
- This integration improves interpretability of ML models and can guide clinically actionable insights.
- By incorporating mechanistic elements, ML models can also benefit from a broader knowledge base, not limited to just the current dataset (especially if that dataset is small/limited).
- Here we present an example of integrating mechanistic and ML models on tabular clinical data, and our future plan is to extend this to multi-modal analysis incorporating image data.

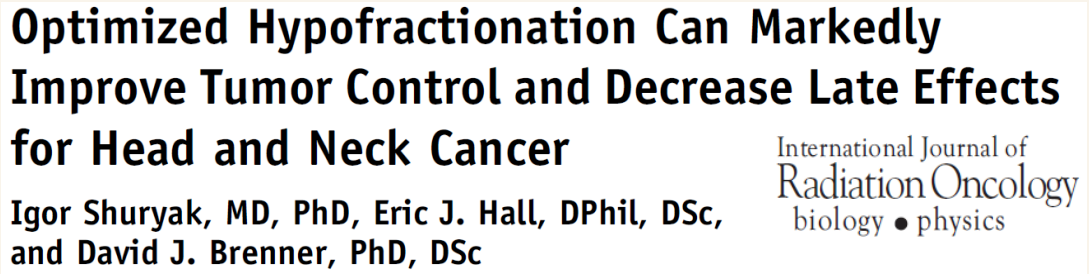
Example: Modeling Tumor Repopulation in HNSCC

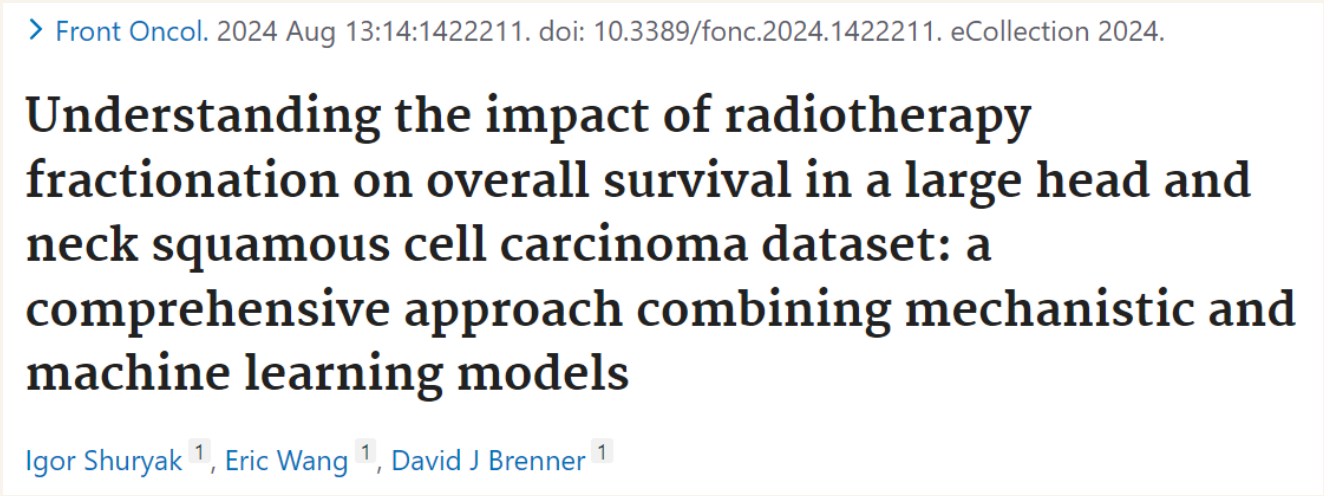
- Tumor repopulation is known to be a strong factor in HNSCC radiotherapy outcome.
- Shortening the radiotherapy helps to reduce the effect of repopulation because there is less time for tumor cells to proliferate, while gaps in treatment have the opposite effect.
- This was recognized a long time ago, leading to a model called the Withers "hockey stick". In this model, accelerated repopulation (AR) is assumed to start at a fixed time (T_k) after RT begins, and AR rate is assumed to be independent cell killing intensity.
- This can be called the Dose-Independent (DI) model.
- Our team was thinking how to improve this:
  - Since AR is likely a compensatory response to cell killing, the onset and rate of AR may depend on the "intensity" of cell killing during treatment.
- This is the rationale for a "Dose-Dependent (DD)" model.



Applying DI and DD Models to HNSCC Data

- First, we fitted them to RT-only arms of older clinical trials (aggregated data).
- Then, we used them on a modern large dataset (RADCURE) on 2,651 patients with HNSCC with comprehensive radiotherapy, chemotherapy, clinical variables, and long-term cause of death data, from PMH in Toronto, Canada.
- In the RADCURE analysis we used a two-step approach that combines mechanistic modeling concepts with ML: Random Survival Forests (RSF) for an exploratory analysis followed by Causal Survival Forests (CSF) for a focused causal analysis.





BED for DI and DD Models

- BED consists of a cell killing term (LQ model) and a repopulation term.

DD model specifics:

- Killing term is the same as in DI model
- Repopulation term differs:
  - AR starts when −ln(cell surviving fraction) decreases below a constant −C
  - AR rate is proportional to the average fraction of cells killed per day

Simplification assumption:

- Assume $BED_{DD} = BE_{DI}$ for standard RT of 35 x 2 Gy over 7 weeks
- This helps calculate C, eliminating it as a free parameter and simplifying the model

Parameters and their meanings (plausible values):

$\alpha = 0.2\ Gy^{-1}$: Cell sensitivity to radiation (linear term)

$\alpha/\beta = 10\ Gy$: Ratio of L/Q cell sensitivity parameters

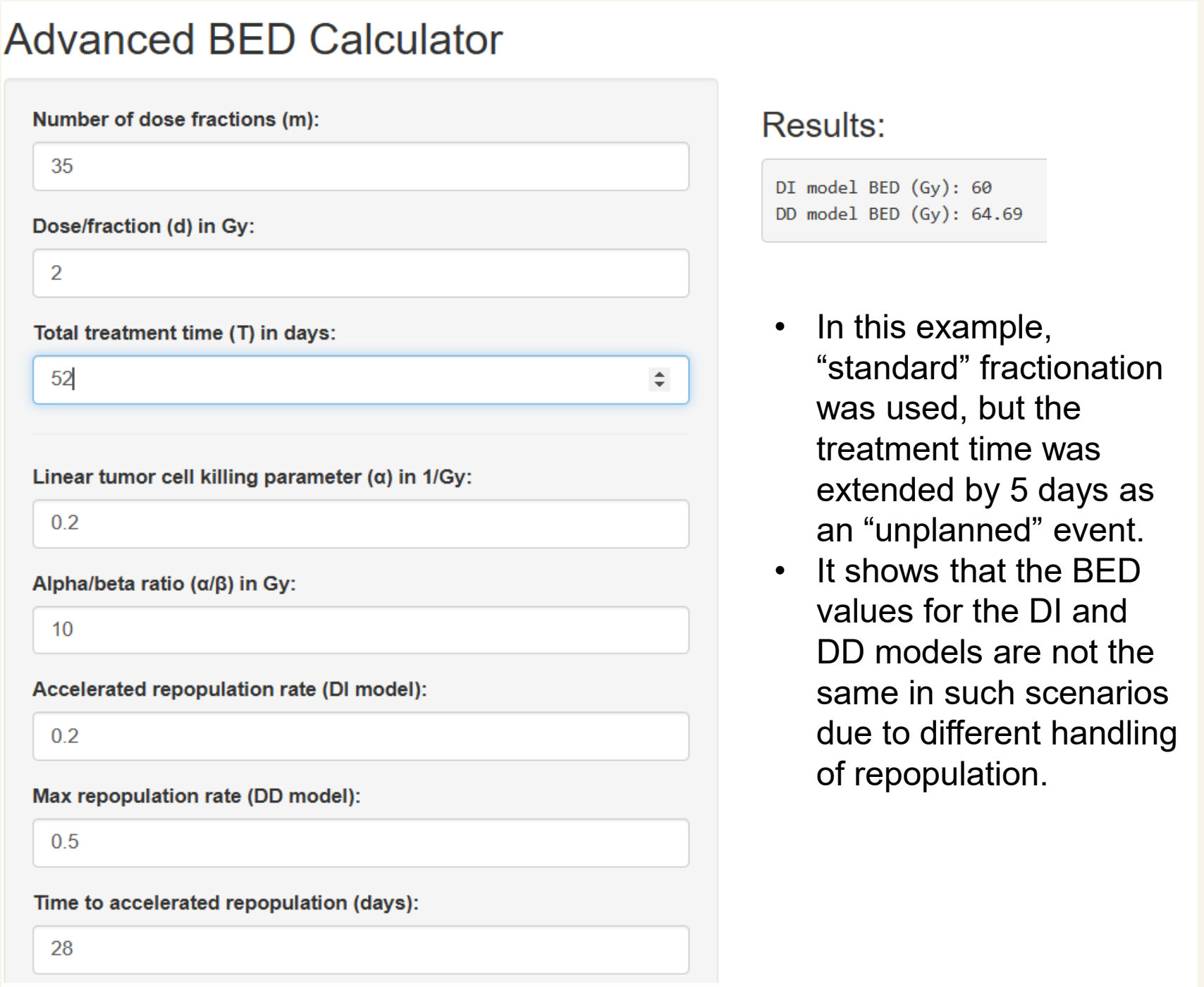
$\lambda_{DI} = 0.2\ days^{-1}$: Fixed AR rate for DI model

$\lambda_{DD} = 0.5\ days^{-1}$: Maximum AR rate for DD model

$T_k = 28\ days$: Fixed AR onset time for DI model

Simple Website Implementation of DI and DD BEDs

- We implemented these BED formulas in a simple website application using GitHub and Shiny R package: https://ishuryak.shinyapps.io/custom_bed_calculator/



- In this example, “standard” fractionation was used, but the treatment time was extended by 5 days as an “unplanned” event.
- It shows that the BED values for the DI and DD models are not the same in such scenarios due to different handling of repopulation.

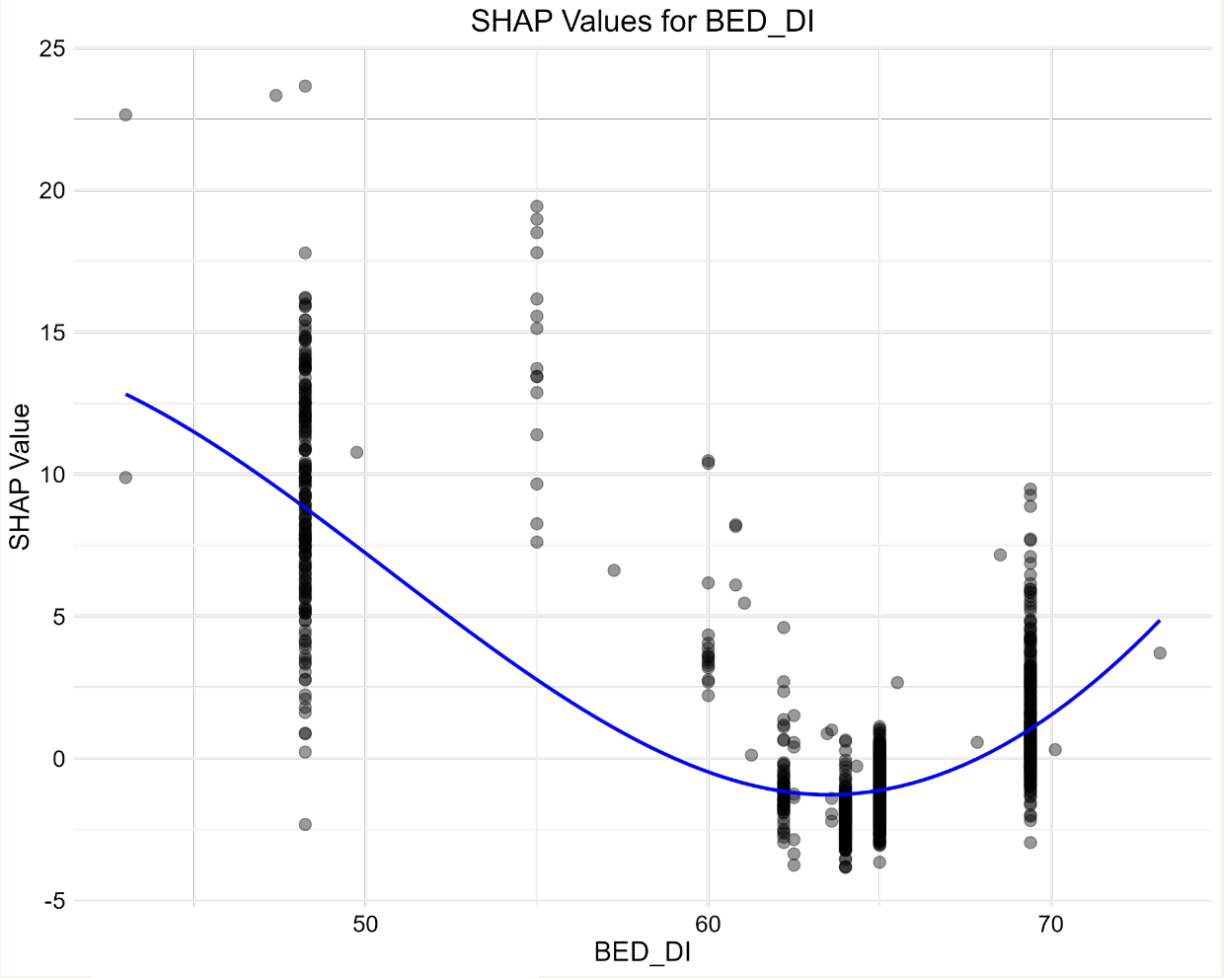
Behaviors on RADCURE data

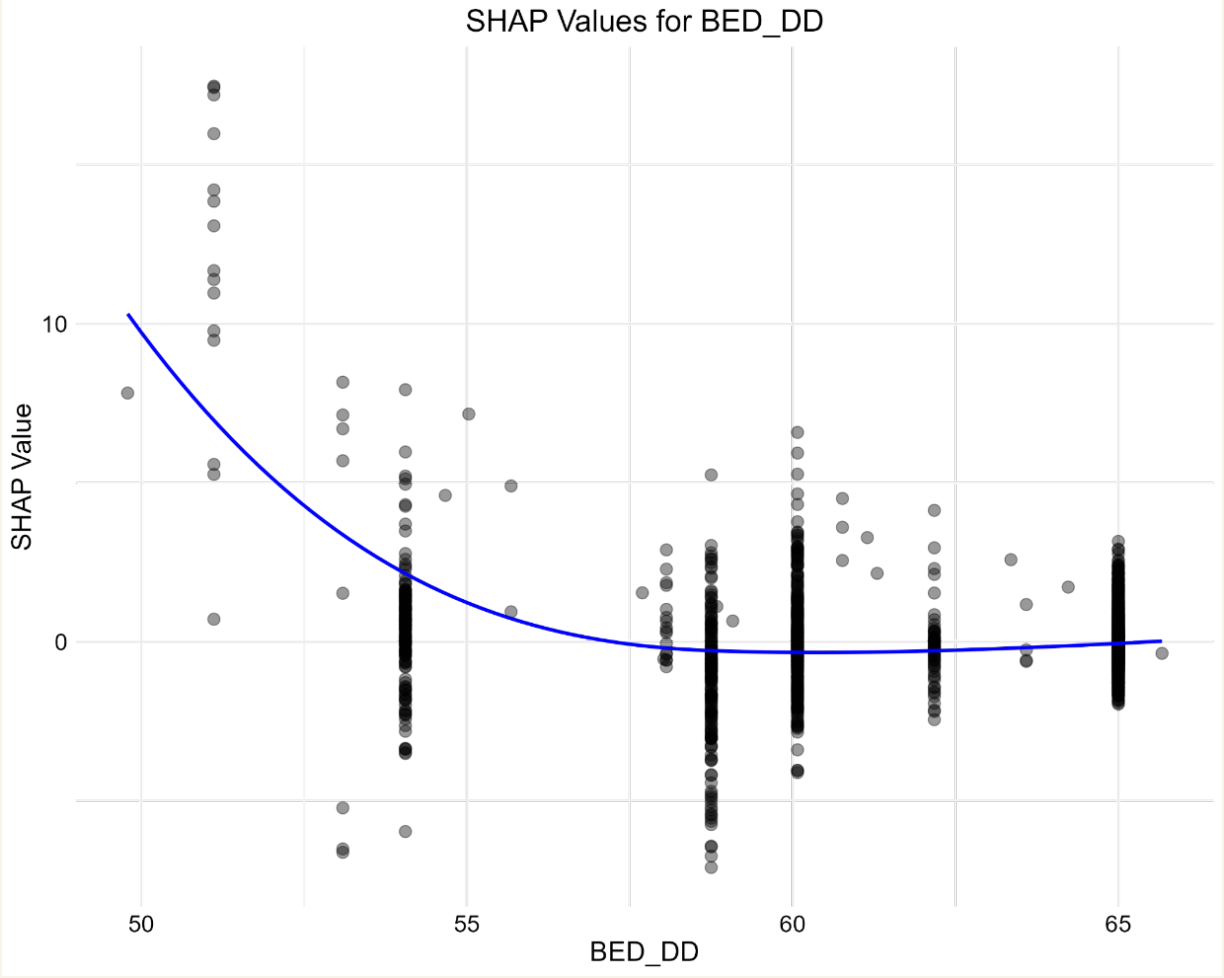
- In a simple Cox regression model, BED_{DD} (but not BED_{DI}) was a significant predictor of overall survival.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
BED_DI	0.009276	1.009320	0.005812	1.596	0.11046	
BED_DD	-0.037980	0.962732	0.013246	-2.867	0.00414	**
Age	0.030824	1.031304	0.003337	9.236	< 2e-16	***
Sex	0.054472	1.055983	0.078374	0.695	0.48704	
Smoking_PY	0.009133	1.009175	0.001200	7.612	2.69e-14	***
Stage_numeric	0.544222	1.723268	0.041860	13.001	< 2e-16	***
HPV_Positive	-1.137305	0.320682	0.097512	-11.663	< 2e-16	***
HPV_Unknown	-0.342279	0.710150	0.078626	-4.353	1.34e-05	***
Chemo	-0.528934	0.589233	0.086121	-6.142	8.16e-10	***
RT_year	-0.047312	0.953790	0.011288	-4.191	2.77e-05	***

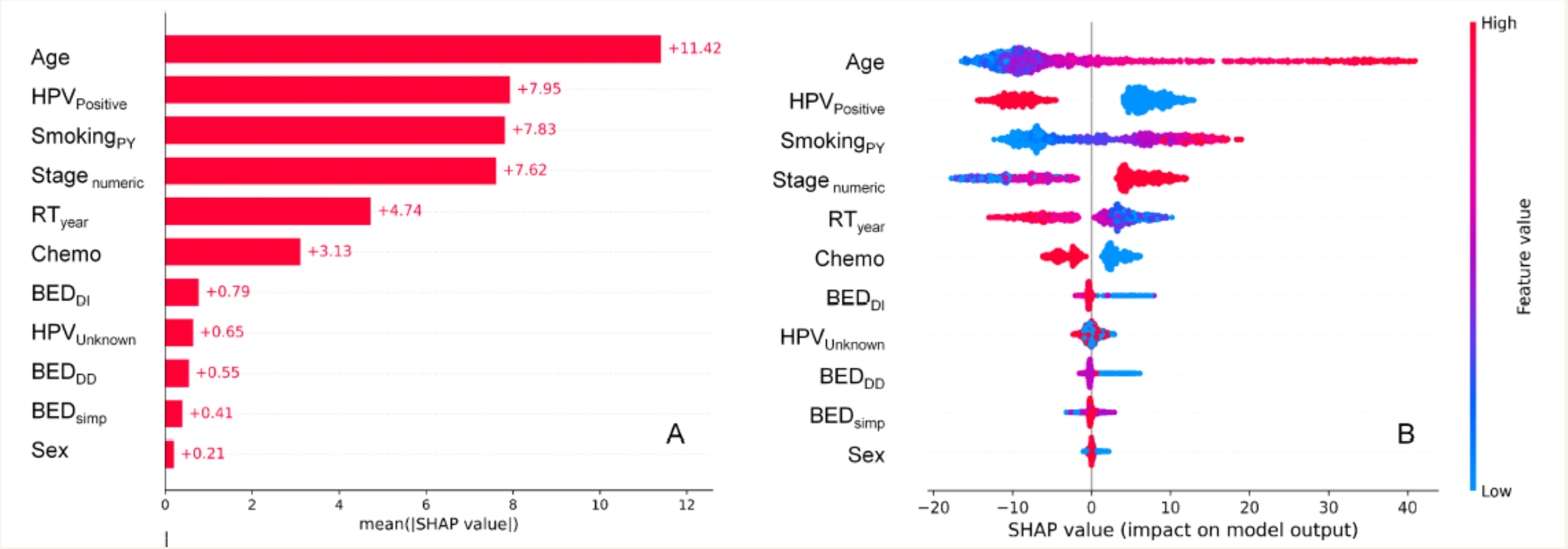
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- In a more complex (but flexible) random survival forest (RSF) model on the same data, both BED_{DI} and BED_{DD} were associated with reduced mortality.
- BED_{DD} had a more monotonic effect.





SHAP values for the RSF model for many features



- Of course, these are predictive (not causal) models.
- The patterns can be affected by confounding.
- Next step – causal ML models.

Causal Machine Learning (CML)

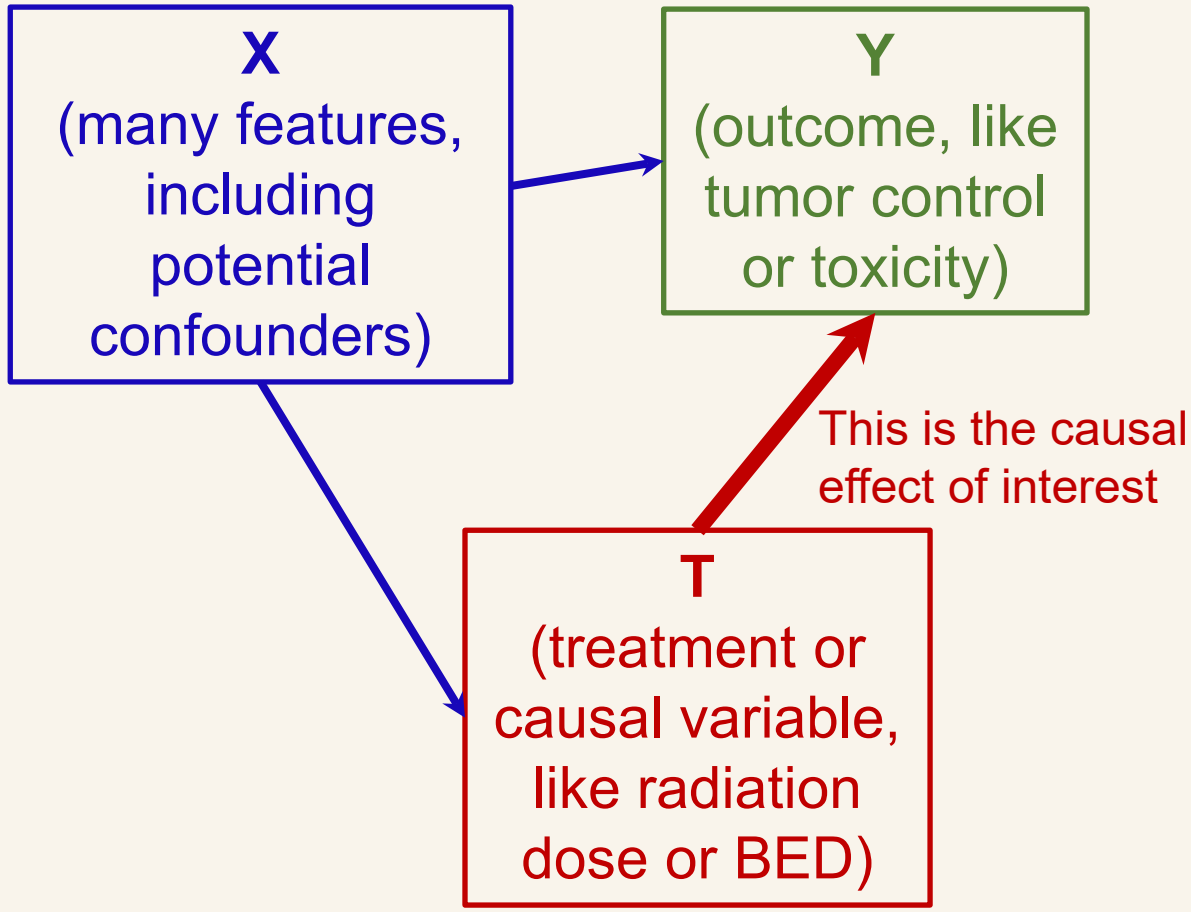
- Using ML to study *causal* effects seems counterintuitive, since ML is commonly used for *predictive* tasks which operate with correlations/associations – and *correlation and causation are conceptually different*
- However, causal ML techniques exist, and this field is evolving rapidly because exploring causality is scientifically important
- Key advantages of causal ML (CML):
  - For personalized medicine, it is important to study heterogeneous treatment effects (*i.e.* how do the treatment effects vary by disease details, mutations, patient demographics) to identify which patients / subgroups benefit most / least from a particular treatment
  - Unlike correlations, causal effects can better translate to other data sets where the data distributions and correlation structures can be different
  - Observational clinical data (which are much more widely available than RCTs) can potentially provide causal insights about treatment effects using CML – not as reliable as RCTs of course, but more reliable than predictive modeling techniques
  - CML can also be useful for clinical trials data to identify subgroups whose response to the treatment may differ strongly (even in sign) from the average population's

- In predictive tasks, there are two types of variables: inputs (X, features, predictors) and outputs (Y, outcomes, targets).
- In contrast, in the causal framework, the cause (often called treatment, T) is conceptually distinct from other features (X).
- So, causal tasks involve three types of variables: inputs (X), interventions (T, which represent the treatment / causal variable), and outputs (Y).
- In CML methods described below, we operate under specific assumptions derived from domain knowledge:
  - X can cause T, both X and T together can cause Y, and importantly, T does not cause X.

➢ Main assumptions:

➢ No unmeasured confounding/ignorability = all variables that influence both the treatment and the outcome are observed and accounted for.

➢ Overlap/common support/positivity = every individual has a positive probability of receiving each treatment level.



Our main objective is to quantify the causal effect of T on Y. Meanwhile, the effects of X on T and X on Y can be treated as "nuisance parameters".

Double Debiased Machine Learning (DML)

➢ DML (V. Chernozhukov *et al*., https://arxiv.org/abs/1608.00060) involves the following steps:

- ❖ Model the **treatment** (T) based on the covariates (X), using any ML method. This is a "de-confounding" operation.
- ❖ Model the **outcome** (Y) based on the covariates (X), using any ML method, but ignore the treatment. This is for "de-biasing/de-noising".
- ❖ Build a third model to **relate the residuals from the first two models to each other** – this relationship is interpreted as the causal effect. Also include X in this model.
- ❖ In place of unknown true treatment effects, this method involves using differences between observed and ML-predicted values for "nuisance functions": dependences of T and Y on X. This is used to achieve "orthogonality", reducing sensitivity to nuisance functions.

- DML is doubly robust: it can provide reliable causal effect estimates if either the treatment our outcome model (but not necessarily both) is correctly specified.
- In practice, the nuisance function models need to be reasonably accurate but not perfect, and can be generated using any ML method, provided cross-fitting is used.

Causal Forest (CF)

Causal forest grf 2.4.0

https://grf-labs.github.io/grf/reference/causal_forest.html

Athey, Susan, Julie Tibshirani, and Stefan Wager. "Generalized Random Forests". Annals of Statistics, 47(2), 2019.

1. Data Splitting

- Honestly split data into two parts:
  - Splitting sample: Determine tree structure
  - Estimation sample: Calculate treatment effects

2. Preliminary Estimation

- Estimate separate machine learning models:
  - Conditional mean of outcome: $\hat{m}(X) = \mathbb{E}[Y \mid X]$
  - Propensity score: $\hat{e}(X) = \mathbb{P}(T = 1 \mid X)$

3. Residual Calculation

- Compute residuals:
  - Residualized outcome: $Y_{\text{residual}} = Y - \hat{m}(X)$
  - Residualized treatment: $T_{\text{residual}} = T - \hat{e}(X)$

4. Tree Generation

- Create trees by:
  - Splitting features to maximize treatment effect heterogeneity
  - Partitioning feature space into leaves
  - Estimating a constant treatment effect within each leaf:
    - Calculated as the coefficient from regressing Y_residual on T_residual within the leaf
    - This single coefficient represents the treatment effect for all observations in the leaf

5. Treatment Effect Estimation

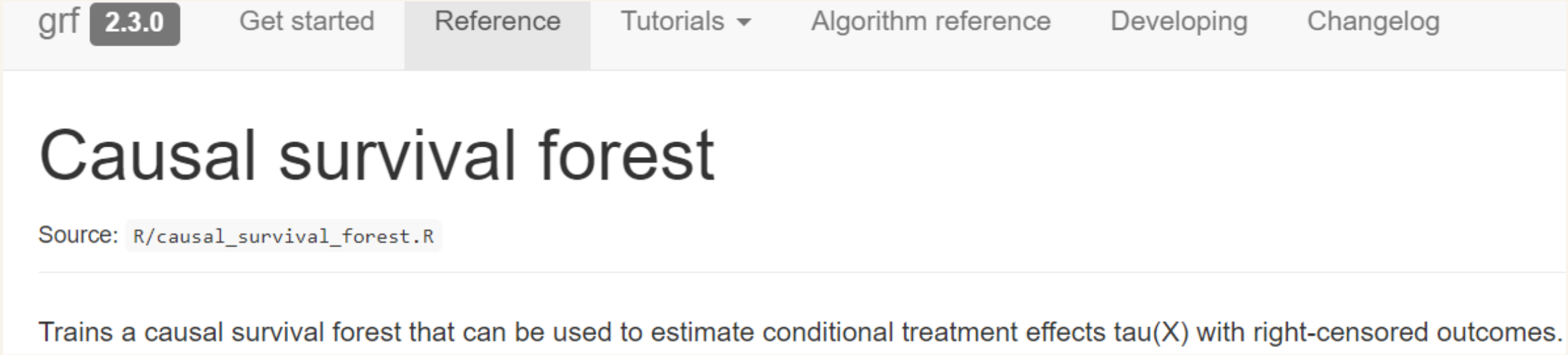
- Apply determined tree structure to estimation data
- Assign weights $\alpha(x)$ based on:
  - Fraction of trees where training point falls in same leaf as test point
  - This fraction represents proximity/similarity in feature space

6. Final Prediction

- Compute weighted average of treatment effects

Using CML on Survival Data Like RADCURE

- Here the outcome is overall survival, so we used the causal survival forest (CSF) – a CF variant which uses censoring-robust estimating equations.
- CSF incorporates models for the treatment propensity score, for the censoring probability, and for the survival time.
- It assumes that the censoring process is random, conditional on the covariates.



Estimands: Survival probability (SP) or Restricted Mean Survival Time (RMST). SP = "vertical" difference in survival probabilities between treated and untreated groups at a specific time (horizon). RMST = "integral" of the SP difference from time zero up to the horizon time.

Causal analyses on RADCURE suggested that high BED_{DD} or BED_{DI} increased patient restricted mean survival time (RMST) by 0.5-1.0 years and increased survival probability (SP) by 5-15% several years after treatment.

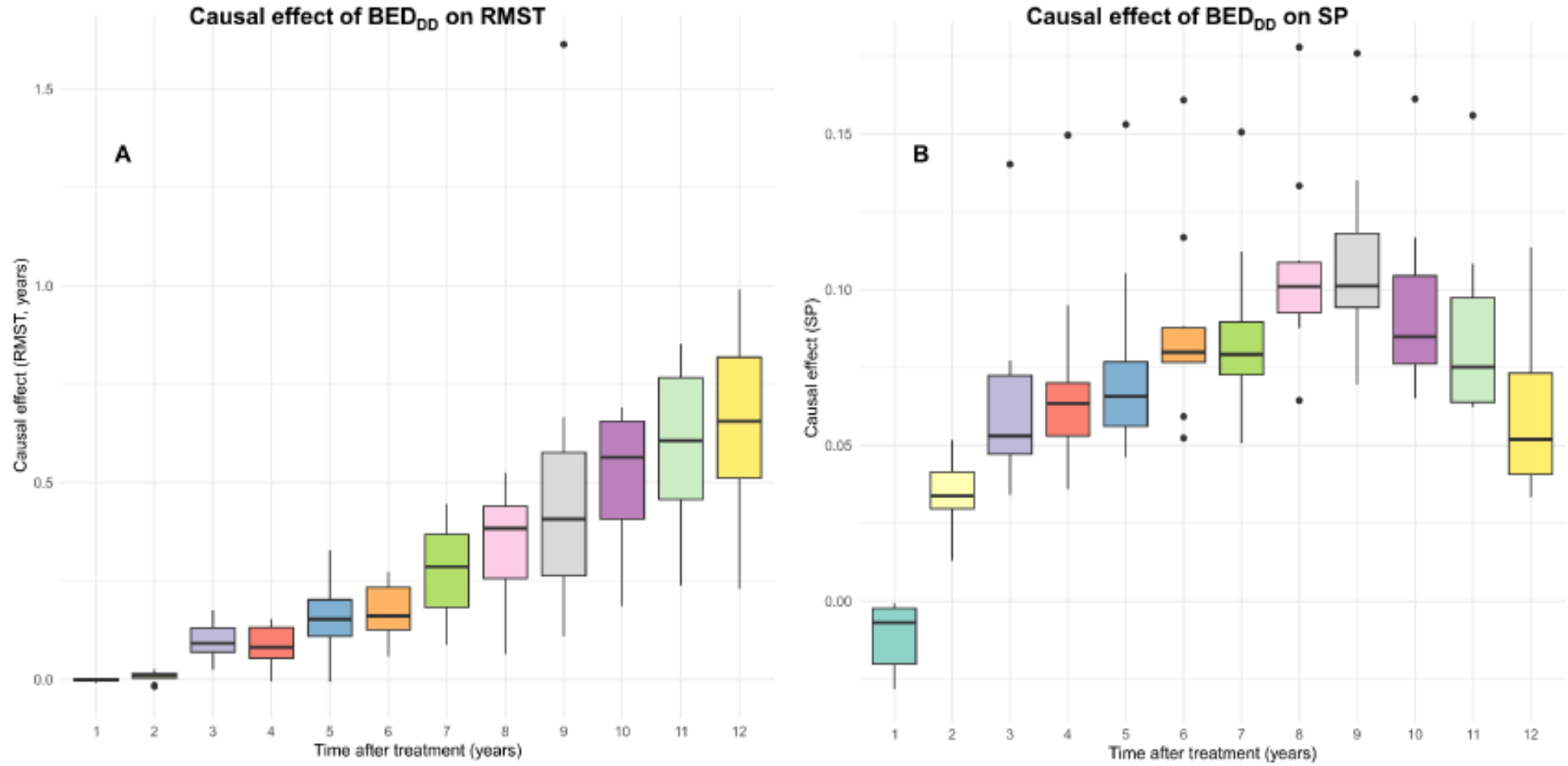
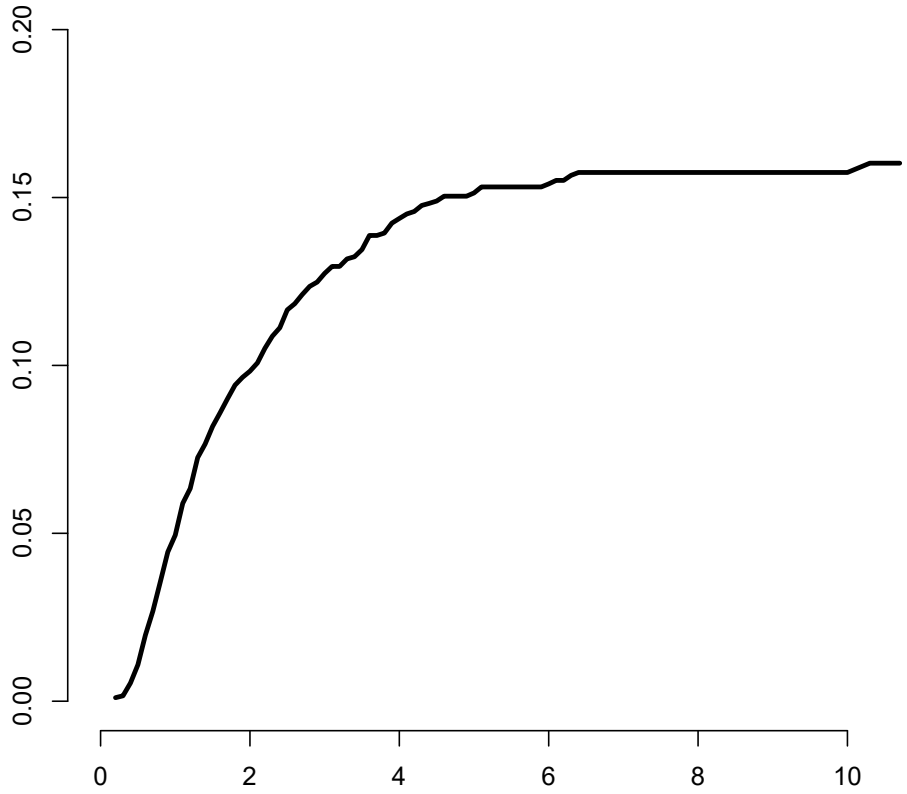


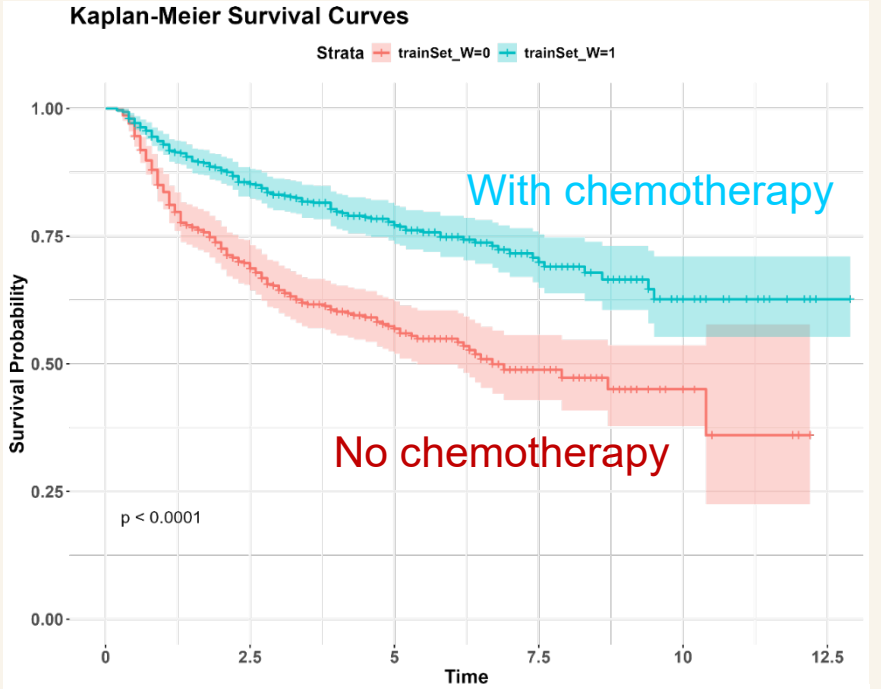
FIGURE 6

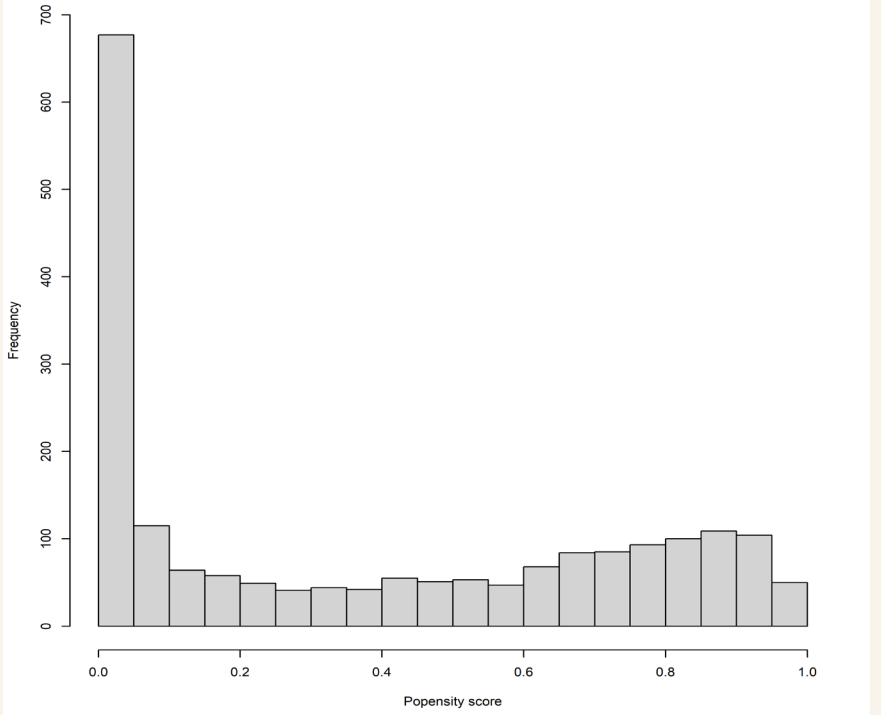
Causal effect estimates of BED_{DD} from the CSF analyses. The boxplots show the distributions of restricted mean survival time (RMST) **(A)** or survival probability (SP) **(B)** causal effects over 10-fold cross validation folds on the training data.


0.00
0.05
0.10
0.15
0.20
0
2
4
6
8
10

Causal ML Analysis for Chemotherapy

- About 37% of patients received chemotherapy.
- Its effect is evident even in a “naïve” univariate KM curve comparison.
- We used CSF to estimate the chemotherapy effect more rigorously in a causal framework, considering other variables.
- Elastic net regression was used for propensity score prediction.
- Patients with propensity scores <0.1 or >0.9 (those very unlikely or very likely to get chemotherapy) were dropped from analysis to generate stable causal effect estimates.
- The propensity scores varied by tumor site.
- They tended to increase with tumor Stage and radiotherapy BED, and to decrease with Age.

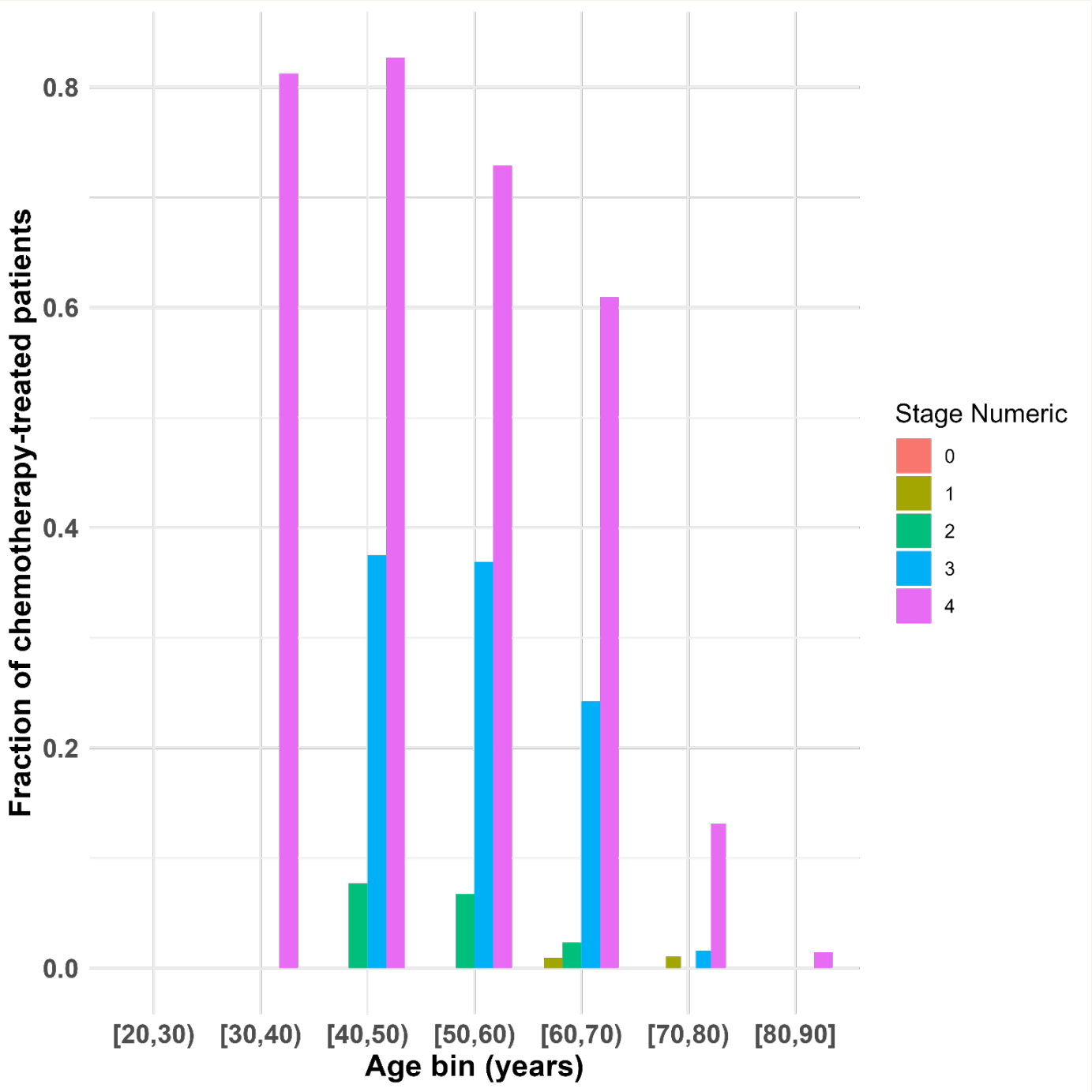




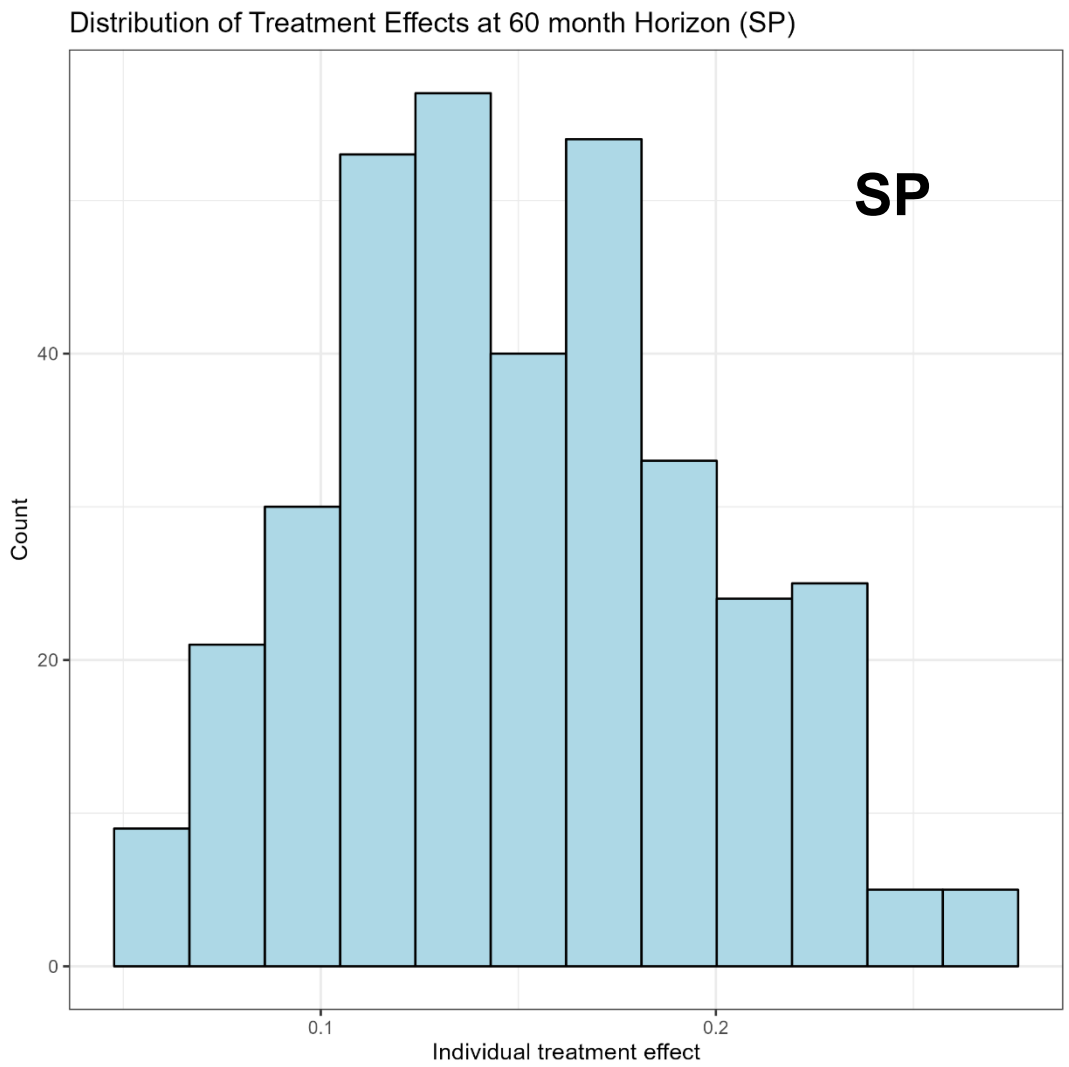
- Summary statistics for some variables of interest

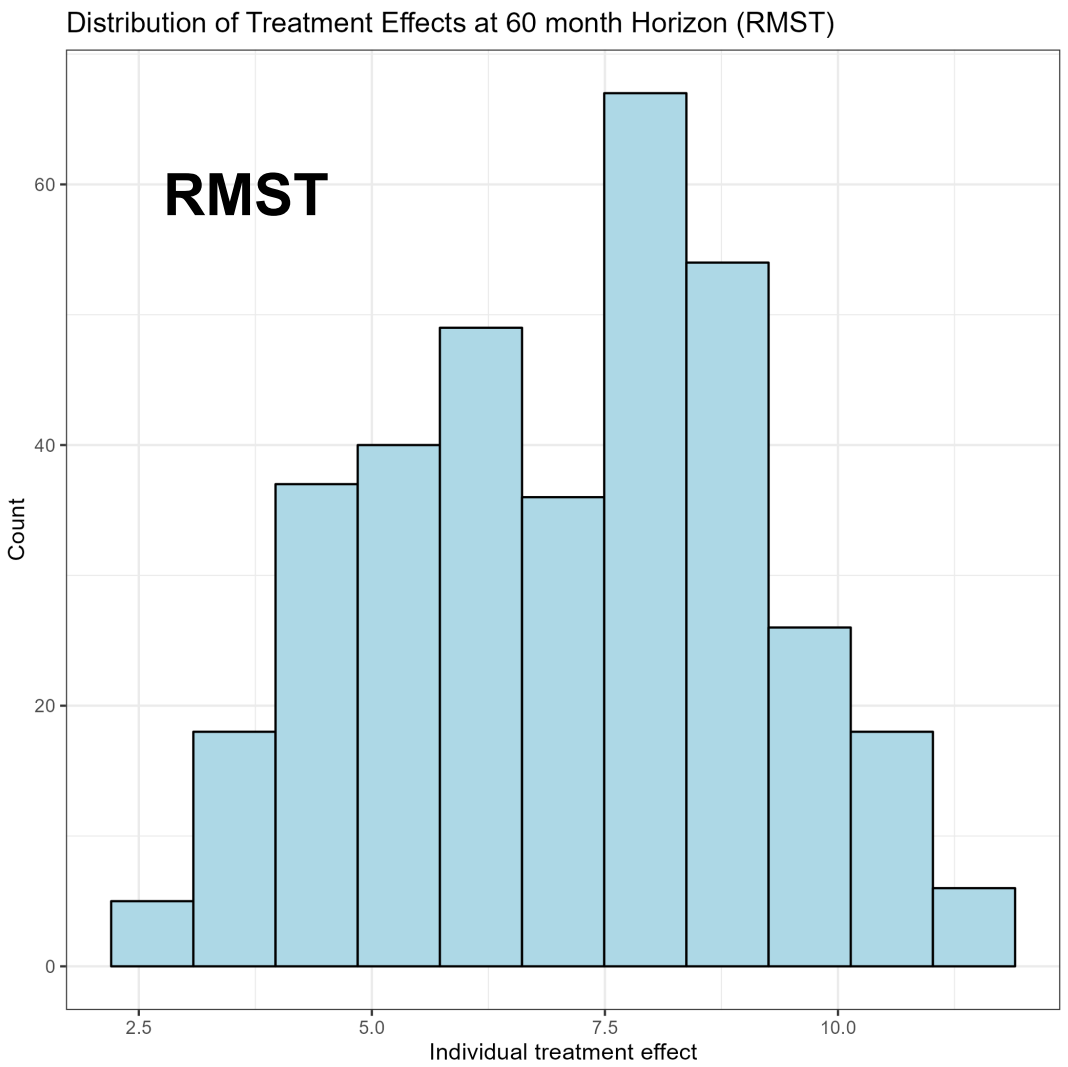
	Age	Sex (0=F, 1=M)	Stage	Chemo (0=NO, 1=YES)
Min	22.30	0	0	0
Max	90.00	1	4	1
Mean	63.49	0.81	3.21	0.37
Median	63.10	1	4	0
25%	55.90	1	2	0
75%	70.80	1	4	1

- Distribution of chemotherapy by age and stage →



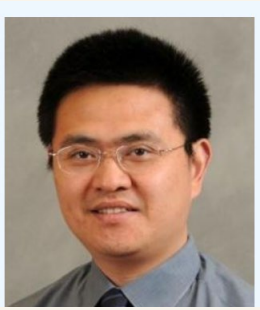
CSF results: chemotherapy increased survival probability by 15.2 ± 6.0% at 3 years and 15.0 ± 6.7% at 5 years on the testing set.
RMST improved by 3.6 ± 1.4 months at 3 years and 7.1 ± 2.6 months at 5 years.
Considerable heterogeneity between patients is seen in the histograms.





Future Directions: Incorporating Image Data

- The RADCURE data set also contains CT images and radiomics features extracted from them.
- Combining the tabular clinical data and image features into a more powerful analysis seems very promising for our team: several members have image analysis expertise.



Yading Yuan

Columbia University Irving Medical Center
Herbert and Florence Associate Professor of Radiation Oncology (Physics) (in the Data Science Institute) at the Columbia University Medical Center

DSI Member

yy3332@cumc.columbia.edu

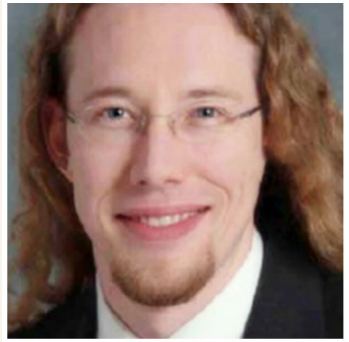
ORCESTRA for Radiomics

Explore multimodal Radiomics Datasets (Radiomic sets)

Xiao Wu, PhD
Assistant Professor of Biostatistics

xw2892@cumc.columbia.edu

Associate Professor
Andrew Hope
MD, FRCPC



Conclusions

- We combined simple mechanistic mathematical modeling concepts with predictive and causal ML methods to investigate the effects of radiotherapy and chemotherapy on HNSCC patient survival in the large RADCURE data set.
- High BED_{DD} or BED_{DI}, and chemotherapy, significantly increased RMST and survival probability several years after treatment. The magnitudes of these causal effects varied substantially between patients.
- These findings are in line with current knowledge, but chemotherapy effect estimates are larger than in published meta-analyses, possibly due to the tendency for younger/healthier patients to receive chemotherapy more frequently, other population differences, incomplete fulfillment of causal modeling assumptions, and evolving treatment protocols.
- This study presents an example of implementing the concept of incorporating mechanistic modeling insights into ML analyses of cancer treatment data.
- We think that this type of approach has a lot of potential for enhancing knowledge about treatment effects from non-randomized clinical data to complement RCT analyses, generate new hypotheses, and support personalized medicine.

Thank you very much for your interest!

My email is: is144@cumc.columbia.edu