SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY
DEPARTMENT OF STATISTICS
MASTER'S PROGRAM

Non-Parametric Estimation of Optimal Individualized Treatment Rules for Survival Data

by

Andre Ehrlich

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfillment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
September 3, 2024

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΤΕΧΝΟΛΟΓΙΑΣ & ΠΛΗΡΟΦΟΡΙΩΝ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

Μη Παραμετρική Εκτίμηση των Βέλτιστων Εξατομικευμένων Κανόνων Θεραπείας για Δεδομένα Επιβίωσης

του

Andre Ehrlich

Διατριβή
που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα, Ελλάδα
Ιούλιος 2024

Dedication

Acknowledgements

Abstract

Precision Medicine aims to deliver patient-centered care beyond one-size-fits-all treatment rules, by accounting for heterogeneity in treatment effect to determine patient subgroups. Individualization promises two benefits: (i) to extend the population mean survival (ii) to spare patients from aggressive treatment who are unlikely to benefit. The "Direct Value Search" approach frames the estimation of Individualized Treatment Rules as an Empirical Risk Minimization problem, allowing for the use of classification techniques which avoid strict assumptions of parametric regression models, and therefore minimize the possibility of misspecification.

This thesis reviews a series of developments in Outcome Weighted Learning (OWL) methods that handle right-censored survival outcomes. We evaluate the proposed efficiency-gain and distributional robustness of the Multistate Outcome Weighted Learning (MSOWL) method. MSOWL integrates Inverse-Probability-Censoring-Weighted individual stochastic benefit processes, including right-censored cases, at the cost of cubic computational complexity, which can be ameliorated via divide and conquer. We analyze a novel data set from an oncological clinical trial, which assesses the inclusion of immunotherapy to standard chemotherapy for Pancreatic Cancer.

Keywords: Precision Medicine, Survival Analysis, Causal Inference, Clinical Trials, Missing Data, Statistical Learning, Support Vector Machine

ΠΕΡΙΛΗΨΗ

Η Ιατρική Ακριβείας στοχεύει στην παροχή εξατομικευμένης φροντίδας πέρα από τους κανόνες θεραπείας «ένα μέγεθος για όλους», λαμβάνοντας υπόψη την ετερογένεια στην επίδραση της θεραπείας. Η εξατομίκευση της θεραπείας, δηλαδή ο εντοπισμός υποομάδων ασθενών για να λάβουν διακριτές θεραπευτικές επιλογές, μπορεί να ωφελήσει τους ασθενείς με δύο τρόπους: (ι) να επεκτείνει την επιβίωση πληθυσμών (ιι) να απαλλάξει τους ασθενείς από επιθετική θεραπεία που είναι απίθανο να ωφεληθούν.

Οι μέθοδοι Μάθησης Σταθμισμένης Έκβασης (Outcome Weighted Learning - OWL) επιλύουν άμεσα την εκτίμηση κανόνων θεραπείας χρησιμοποιώντας Ελαχιστοποίηση Εμπειρικού Κινδύνου, αποφεύγοντας έτσι αυστηρές υποθέσεις μοντέλου και ελαχιστοποιώντας την πιθανότητα εσφαλμένης προδιαγραφής μοντέλου. Εξετάζουμε τις κύριες ιδέες στη βιβλιογραφία της εκτίμησης βέλτιστων Εξατομικευμένων Κανόνων Θεραπείας, με έμφαση στις μεθόδους που χειρίζονται εκβάσεις επιβίωσης.

Η Πολυκαταστασιακή Μάθηση Σταθμισμένης Έκβασης (Multistate Outcome Weighted Learning - MSOWL) παράγει βέλτιστα αποδοτικές εκτιμήσεις ολοκληρώνοντας τις ατομικές στοχαστικές διαδικασίες, συμπεριλαμβανομένων των περιπτώσεων που διακόπτεται η παρατήρηση – αν και αυτή η στατιστική αποδοτικότητα έρχεται με το κόστος κυβικής υπολογιστικής πολυπλοκότητας, την οποία βελτιώνουμε μέσω διαίρεσης και κατάκτησης. Αξιολογούμε το κέρδος αποδοτικότητας και την κατανομική ευρωστία της μεθόδου MSOWL έναντι άλλων μεθόδων OWL & ITR.

Αξιολογούμε εκτιμητές ITR συμπεριλαμβανομένων των μεθόδων OWL, ενός εκτιμητή Παλινδρόμησης Έκβασης Cox, και των τυπικών κανόνων θεραπείας «ένα μέγεθος για όλους». Αξιολογούμε ένα νέο σύνολο δεδομένων από μια ογκολογική κλινική δοκιμή, η οποία εκτιμά την προσθήκη ανοσοθεραπείας στην τυπική χημειοθεραπεία είτε για Καρκίνο του Παγκρέατος.

Λέξεις-κλειδιά: Εξατομικευμένη Ιατρική, Κλινικές Δοκιμές, Ανάλυση Επιβίωσης, Αιτιακή Συμπερασματολογία, Ελλιπή Δεδομένα, Στατιστική Μάθηση, Μηχανή Διανυσμάτων Υποστήριξης

Contents

Nomenclature

Treatment Rule Acronyms

AIPTW	Augmented Inverse Probability Treatment Weighting
ATE	Average Treatment Effect
ATS	Adaptive Treatment Strategies
DTR	Dynamic Treatment Rule
EARL	Efficient Augmentation Relaxation Learning
IPCW	Inverse Probability of Censoring Weighting
IPTW	Inverse Probability of Treatment Weighting
IPW	Inverse Probability Weighting
ITE	Individual Treatment Effect
ITR	Individualised Treatment Rule
KM	Kaplan-Meier
OLS	Ordinary Least Squares
OWL	Outcome Weighted Learning
RCT	Randomised Controlled Trial
RMST	Restricted Mean Survival Time
SUTVA	Stable Unit Treatment Value Assumption
SVM	Support Vector Machine
ZOM	Zero-Order Model

Clinical Trial Acronyms

ECOG	Eastern Cooperative Oncology Group

HNSCC	Head & Neck Squamous Cell Cancer
RCT	Randomised Controlled Trial
RECIST	Response Evaluation Criteria in Solid Tumors
SPECTRUM	Study of Panitumumab Efficacy in Patients With Recurrent and/or Metastatic Head & Neck Cancer

Symbols

δ	Censoring Indicator
τ	Duration of Trial/Study
$\tilde{T}$	Observed Survival
A	Treatment Choice
C	Censoring Time
D	Treatment Rule:
N	Number of Patients
P	Number of Covariates per Patient
T	True Survival
X	Patient Covariates

List of Tables

List of Figures

List of Equations

List of Algorithms

Chapter 1

Introduction

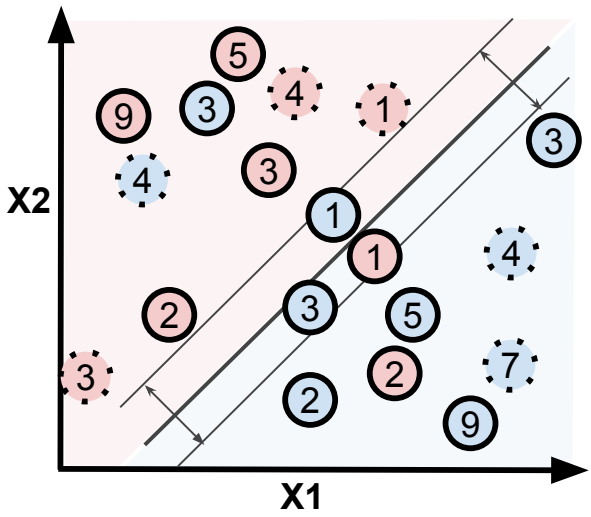


Figure 1.1: **Outcome Weighted Learning Diagram**
Weighted SVM treatment classifier for right-censored data.
Number: Missclassification cost, causally adjusted
Border: Solid=Uncensored, Dashed=Censored
Color: True Optimal Treatment

1.1 ITR Motivation

$$ITR = f : X \rightarrow A \tag{1.1}$$

An Individualized Treatment Rule (ITR) assigns treatment recommendation A_i patient i given their medical characteristics X. An optimal ITR maximizes the expectation of population health outcome $E[Y]$. ITR estimation is fundamental in Precision Medicine and patient-centered care that goes beyond the one-size-fits-all treatment approach to identify heteroge-

neous patient-treatment interactions, wherein patient *subgroups* receive distinct treatment recommendations. *Individualizing* treatment can benefit some patients by sparing them from aggressive treatment, given evidence that those patients are unlikely to benefit from such treatment Butler et al. (2018).

This thesis focuses on Outcome Weighted Learning (OWL) methods, which directly solve optimize treatment allocation problem using Empirical Risk Minimization (ERM), thereby avoiding strict model assumptions. We will introduce the main statistical ideas within the literature of Individualized Treatment Rules, with focus on the developments of OWL methods by Zhao et al. (2012), Zhao et al. (2015), and Bakoyannis (2023). We will compare the OWL models against a Cox Outcome Regression estimator, as well as the standard one-size-fits-all zero order models (ZOM) on real RCT data sets from cancer trials.

1.2 Problem Description

The aim of our statistical inquiry is to identify and evaluate Individualized Treatment Rule (ITR) estimators for *survival outcomes*. In randomized Controlled Trials (RCT), patient i with covariates $X_i \in \mathbb{R}^p, p \in \mathbb{N}_1$, will be administered a single treatment choice from binary (for simplicity) treatment options $A \in \{-1,1\}$. Health outcomes of the patient under the chosen treatment are denoted as $Y_i(A)$. We would like to determine the *optimal treatment* for each patient by comparing the *potential outcomes* under the different treatment options. The fundamental problem in *causal inference* is that we only ever observe a single treatment choice $A = a$ and its outcome $y_i(a)$, never the counterfactual outcome $y_i(A \neq a)$.

Survival analysis is concerned with time-to-event outcomes $Y_i \in \mathbb{R}^+$, which are subject to right-censoring, indicated by $\Delta_i = \mathbb{I}(Y_i \leq C_i)$, subject censoring time C_i, such that only the right-censored survival time $\tilde{Y}_i$ is observed for patient i. In a restricted survival analysis, we are not concerned with patient outcomes beyond the duration of the RCT, so the event indicator becomes $\Delta_i = \mathbb{I}(min(Y_i, \tau) \leq C_i)$.

From RCT data we observe the following variables:

i	Patient ID
$X_i^P, p \in \mathbb{N}_1$	Patient Covariates
$A_i \in \{-1,1\}$	Randomly Assigned Treatment
τ	End of Study
Y_i	Observed Survival Time
$\tilde{Y}_i$	True Survival Time
C_i	Observed Censoring Time
$\Delta_i = \mathbb{I}(Y_i \leq C_i)$	Event-Observed Status

Therefore, ITR estimation of survival data is characterized by two forms of missingness: (i) right-censored survival outcomes and (ii) counterfactual potential outcomes. We wish to infer patients outcomes under counterfactual treatment assignments.

$\hat{Y}_i$	Estimated Survival
$Y_i^*(A=0), Y_i^*(A=1)$	Potential Outcomes
$\mathbb{M}(X, A, Y, Y^*)$	ITR Model
$\hat{A}_i(X = x_i \vert \mathbb{M} = m)$	Optimal Treatment Assignment

1.3 Practical Considerations

We are interested to evaluate ITR estimation methodologies for right-censored survival outcomes. We are interested to solve this problem in a way that its statistical solution is usable in the real world. Therefore, we consider the following real-world constraints to inform our desiderata:

Right-Censoring plagues observations in survival analysis, prevalent in medical research and practice, due due to end-of-study or patient dropout. Merely ignoring censored cases potentially reduces precision, and can even introduce bias to the estimator if observations are not Missing At Random (MAR) $E[C,X] = E[C] * E[X]$

Counterfactual Outcomes Can we estimate how a patient would fair under alternative treatment? Potential outcomes are latent, hypothetical variable useful in causal inference to suppose the patient had alternatively taken the *other* treatment choice, estimation of Individual Treatment Effect would be trivial. (ITE) $\delta_i^* = y^*(A=a) - y^*(A \neq a)$.

Flexible Learning Can we discover an linear decision rule without posing untrue assumptions on the data? Approaches in statistical learning, classification, and machine learning require larger sample sizes for estimation and sometimes lack methods for statistical inference.

Flexible models accounting for non-linearity in finite samples which may not permit; comparison of Gaussian kernels to linear kernels shows no advantage in small samples.

Clinical Viability Decision-support assistance often requires an simple treatment rule (e.g. Linear), and quantification of uncertainty. Methods must be effective with small sample sizes.

Medical data have two overarching concerns for statistical modeling: (1) typical parametric model assumptions may not hold true, and (2) sample sizes are typically small, and with right-censoring, information is further reduced. Ideal estimators would not a priori assume structure on data, and be performant and economical with considering data from every patient under study, despite censoring.

We want to find optimal individualized treatment rules that account for counterfactuality, right-censorship, realistic small-sample sizes, and interpretable clinical use. Non-parametric and missing data techniques are therefore often used in towards this end.

1.4 Thesis Layout

Main

Chapter 1 Introduction to problem and research focus.
Chapter 2 Survival Analysis, Kaplan-Meier, Cox Model.
Chapter 3 Randomized Clinical Trials, Causal Inference, IPW, Outcome Regression
Chapter 4 Direct Value Search, Support Vector Machine, Outcome Weighted Learning.
Chapter 5 Precision Oncology case study with data from real clinical trials assessing Immunotherapy Panitumumab versus standard-of-care chemotherapy for cancers.
Chapter 6 Discussion of findings, and suggestions for future work.

Appendix

Appendix A Aggressive treatment and formalizing multi-objective trade-off.
Appendix B Sequential problems in Dynamic Treatment Regime (DTR) literature.
Appendix C Neural Networks for individualized treatment rule estimation.
Appendix D Basic Empirical Process Theory.
Appendix E Generating Survival Times from Cox and AFT Regression Models.
Appendix F Simulation Experiment description of Zhao et al. (2015) and Bakoyannis (2023).
Appendix G Annotated code from the MSOWL Value function and ITR estimation.
Appendix H Extra plots from related experiments from the literature for ease of reference.

Chapter 2

Survival Analysis

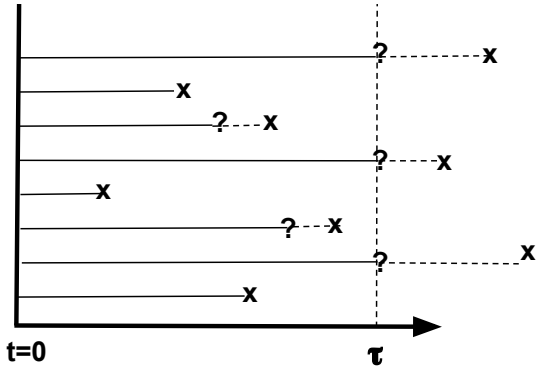


Figure 2.1: Right-Censored Survival Data.

Survival outcomes are ubiquitous outcome in medical research and care is continuous-time survival $T \in [0, +\infty]$, subject to right-censoring by either the end of study τ or by patient dropout C. Figure 2.1 generically visualizes the structure of the data.

τ	End of Study
$T \in [0, +\infty]$	Survival Time
$C \in [0, +\infty]$	Censoring Time
$\tilde{T} = min(T_i, \tau)$	Restricted Survival Time
$T^* = min(\tilde{T}_i, C_i)$	Observed Restricted Survival Time
$\Delta_i = \mathbb{I}(\tilde{T}_i \leq C_i)$	Event Status

Restricted Mean Survival Time (RMST) is a popular estimand within biomedical applications due to end-of-study censoring and few observable tail-probability. Royston and Parmar (2013) suggest that clinical comparison of the area under the curve may be more interpretable than interpreting hazard ratios and logrank test. Focusing on right-truncated survival endpoint

allows us to avoid difficult extrapolation of unrestricted survival beyond end-of-study. If a patient is alive at end of study $t = \tau$, their survival time is considered $T = \tau$ and their event status indicates an observed event. In this way, statistical methods are motivated to maximize, say, 5-year survival probability, rather than trying to maximize unrestricted survival time. Diagrams of the three possible scenarios of survival/censoring/end-of-study times are shown in Figure 2.2.

$$E_\tau[Y(t)] = \int_0^\tau Y(t)dt \tag{2.1}$$

Inverse Probability of Censoring Weighting Ignoring censored cases, and account for bias by inflating observed survivals by inverse censoring probability. It is an estimator commonly employed in the context of missing data with known population totals. Inverse Probability Weighting is described in more described in Sec. 3.4.

$$E[T] = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i T_i}{P(C_i|X_i)} \tag{2.2}$$

2.1 Survival Function

The Survival Function characterizes the fundamental quantity of interest, the probability that a patient will survive up to or past a time instant t.

$$S(t) = P(T \geq t) = 1 - F(T \leq t) \tag{2.3}$$

The Cumulative Distribution Function is a counting process accumulating instances of events over time.

$$F(t) = P(T \leq t) = \int_0^t f(u)\,du \tag{2.4}$$

The Probability Density Function computes the probability of an event occurring at time instant t; it can be decomposed into $f(t) = S(t) * h(t)$, the joint probability of survival $S(t)$ up until time instant t and then death at time instant t . This decomposition will be recognized in the likelihood function.

$$\begin{aligned} f(t) &= \lim_{\varepsilon \to 0} \frac{Pr(t \leq T < t + \varepsilon)}{\varepsilon} \\ &= P\big(h(t), S(t)\big) \\ &= P\big(h(t)|S(t)\big) * S(t) \\ &= \lim_{\varepsilon \to 0} \frac{Pr(t + \varepsilon|S(t))}{\varepsilon} * S(t) \\ &= h(t) * S(t) \end{aligned} \tag{2.5}$$

The Hazard Function specifies the instantaneous risk of death for a at time instant t conditioned on survival up until time t.

$$
\begin{aligned}
h(t) &= \lim_{\varepsilon \to 0} \frac{Pr\big(t < t + \varepsilon | T \geq t\big)}{\varepsilon} \qquad (2.6)\\
&= f(t)/S(t) \\
&= \frac{-d \log S(t)}{dt} \\
&=> \\
S(t) &= exp(-H(t)) \qquad (2.7)
\end{aligned}
$$

2.2 Kaplan-Meier Estimator Product-Limit Estimator

Estimate the survival function non-parametrically from observed data: at the point of time of each each event, calculate the instantaneous death rate, d_t/n_t. The probability of survival at each successive point in time is the product limit of one minus the instantaneous death rate, $P(T > t) = \prod\big(1 - \frac{d_i}{n_i}\big)$.

Censored cases are naturally included in n at time points before their censoring time, and ruled out after the occurrence of each event ruling out censored cases as time progresses (Eq. 2.8). The Kaplan-Meier estimator estimates the RMST, and does not extrapolate beyond the time-points of observed data. The Nelson-Aalen estimator similarly computes the cumulative hazard function $\sum \frac{d_i}{n_i}$ (Eq. 2.9).

$$\hat{S}(t) \approx \prod_{i:t_i \leq t} \Big(1 - \frac{d_i}{n_i}\Big) \qquad \text{Kaplan-Meier Estimator} \qquad (2.8)$$

$$\tilde{H}(t) = \sum_{j-1}^{i-1} \frac{d_i}{n_i},\ t_{j-1} \leq t < t_j \qquad \text{Nelson-Aalen Cum. Hazard Estimator} \qquad (2.9)$$

Satten and Datta (2001) elucidates several theoretical perspectives that help characterize the Kaplan-Meier Estimator, on top of the usual characterization of a Non-Parametric Maximum Likelihood Estimator (NPMLE) Kalbfleisch and Prentice (1978), also show the Kaplan-Meier as a coupling of Inverse Probability of Censoring Weighted (IPCW) estimators for both the Survival and Censoring distributions, Robins and Rotnitzky (1995), as well as imputation via mass redistribution to the right Efron (1977a), Expectation-Maximization Efron (1977b), self-consistency Efron (1977c).

2.3 Cox Proportional Hazards Regression

$$\hat{h}(t;z) = h_0(t)\psi(t) \quad \text{Cox Hazard Estimator} \tag{2.10}$$

$$h_0(t) \quad \text{Baseline Hazard}$$

$$\psi(t) = exp(z^T\beta),\ \beta \in \mathbb{R}^P \quad \text{Hazard Ratio Component}$$

$$\frac{\psi(t;z)}{\psi(t;z')} = exp((z-z')^T\beta) \quad \text{Proportional Hazard Ratio} \tag{2.11}$$

$$L_i(\beta) = \frac{\psi(Y_i|X_i)}{\sum_{j:Y_j \geq Y_i} \psi(Y_i|X_j)} \quad \text{Partial Likelihood} \tag{2.12}$$

$$\text{where } \psi_j = exp(X_j \cdot \beta)$$

Semi-parametric likelihood is a product of two factors, the data and by an infinitely-parameterized expression for the baseline hazard. (see Eq. 2.10) formulates It estimates non-parametrically the observed hazard rate at each occurrence of an event (as in the Kaplan-Meier estimator), counting the relative magnitude of risks at each observed event. The baseline hazard function $h_0(t)$ is left unspecified, treating it as an infinite-dimensional nuisance parameter. The baseline hazard may still be estimated parametrically, useful if extrapolation of the survival curve is of interest Demiris (2010). Gill (1984) describe an intuitive and mathematically rigorous proof of consistent MLE parameter estimates for the Cox Regression estimator using relatively simple theory from stochastic counting processes and martingales.

The assumption of proportional hazards of Cox's regression is the crucial assumption that the hazard rates between treatment groups is proportional, is important to understand. Misspecification can lead to grave outcomes in clinical decision-making. Proportional hazards regression has been widely successful and applicable over the last decades, but it is not without its faults. One such case: Immunotherapy typically has a delayed disease response with respect to standard chemotherapy Zhang and Klein (1995). In such circumstance of model misspecification due to non-proportional hazard, the inference from a Cox model does not hold.

Covariate-regression formula specification ought to remain relatively simple given practical consideration of small data samples. We must specify, for the purposes of ITR estimation, terms representing the *interaction* between patient-covariates X_P and treatment A, in order to uncover *heterogeneous treatment effects*.

2.4 Accelerated Failure Time Regression

The Accelerated Failure Time (AFT) posits the covariate-adjustment function $\psi(z;\beta) = e^{\beta^T z}$ as a scalar multiplier of T with respect to baseline T_0 at $z = 0$, such that $T = T_0\psi(z;\beta)$. The simplest form of AFT is the log-normal, involving log transformation of linear predictor with normally-distributed error term.

$$log(T) = X\beta + \varepsilon, \quad \varepsilon \sim N(0,1) \tag{2.13}$$

$$E(T|X) = exp(X\beta + \frac{\sigma^2}{2}) \tag{2.14}$$

$$S(T|X) = 1 - \Phi(\frac{log(T) - X\beta}{\sigma}), \quad \Phi \sim std.NormalCDF \tag{2.15}$$

Weibull Equivalence. There is a special case in which the AFT and Cox models are equivalent; that is in the case of Weibull-distributed data with $\psi_{PH}(z) = [\psi_{AL}(z)]^k$, where $\psi_{AL}(z) = exp(\beta_{AL}^T z)$, $\psi_{PH}(z) = exp(\beta_{PH}^T z)$, and $\beta_{PH} = \kappa\beta_{AL}$ (Cox and Oakes, 1984).

2.5 Classification Approach to Survival Analysis

A few articles which propose machine learning approaches for survival analysis; these may be worth another look when thinking of new computational approaches to survival problems. Zhong and Tibshirani (2019) pose survival analysis as a classification problem", explains the "stacking" method. Efron (1988) introduces Logistic Regression Survival Analysis Introduce the estimation of right-continuous survival curve via "partial" logistic regression on binomial survival outcome over time. The contribution induces structure in settings with few data, by using parametric method of logistic regression linear model. Their evaluation on real data is on a Head & Neck cancer RCT. Van Ness et al. (2023) "Heart Failure Prediction Through Explainable AI" shows modern machine learning approach to a similar problem as we are interested in

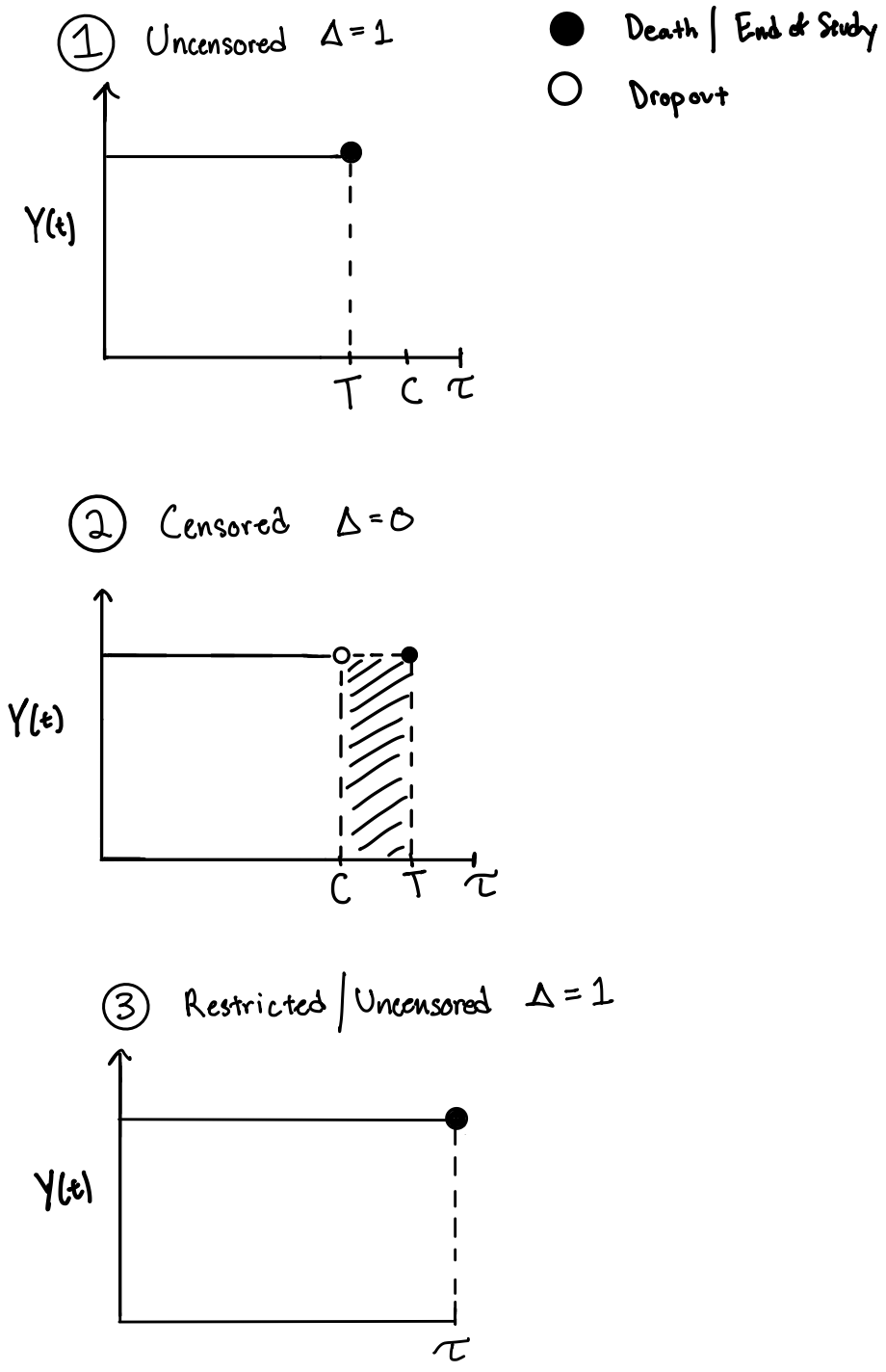
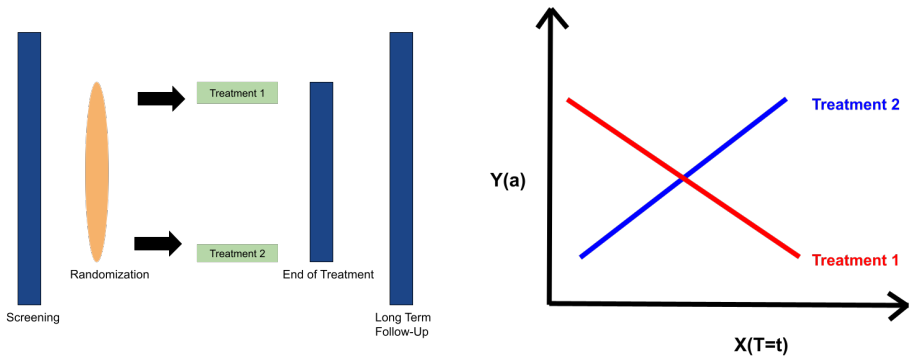


Figure 2.2: **Restricted Survival Analysis:** 3 Possible Scenarios per patient: (1) Uncensored (2) Censored (3) Restricted/Uncensored. The black dots indicate a death (or end-of-study) event, white dot indicates censoring time. The shaded region is the unobserved benefit due to censoring.

Chapter 3

Individualized Treatment Rules



(a) Binary Treatment Single-arm RCT (b) Prescriptive Treatment Effect

Figure 3.1: Discovering Individualized Treatment Rules from RCT

This chapter introduces ITR estimation, involving causal inference, outcome regression, and inverse probability weighting, notes summarize notions of single decision treatment from Chapters 2 and 3 of Dynamic Treatment Regimes (Chakaborty and Moodie, 2013).

3.1 Randomized Clinical Trials

Phase III Trials are confirmatory experiments administered on patients, either sick or at-risk to determine causal relationships between treatment and health outcomes (Fig. 3.1a). Participants are recruited, screened, and then randomized in part of a comparative statistical experiment, whose follow-up time often lasts 5 or 10 years. They are the most expensive component of modern medicinal development, with 40-60% success rate depending on the field. Law (2015)

Data are precious. [1] Relatively few patients (tens, hundreds, maybe thousands) are reasonable for administrative capability and cost. When considering survival data outcomes, right-censoring further reduces the number of completely observed data. Statistical methodology

[1] οὐκ ἐν τῷ μεγάλῳ τὸ εὖ κείμενον εἶναι, ἀλλὰ ἐν τῷ εὖ τὸ μέγα

must explicitly account for small sample sizes. Estimators that are asymptotically unbiased may show quite a bias in small samples.

Randomization is a strong scientific and experimental method to avoid hidden confounding (Eq. 3.2). Administering a *randomized, controlled experiment* requires much greater resources than analyzing *observational data* aggregated post-hoc. Scientific hypotheses and trial procedures must be designed and enumerated before trial. For precision oncology, a typical primary question is, *"which treatment yields a better mean overall survival?"*.

Individualized Treatment refines treatment policy beyond one-size-fits-all. We wish to identify *which*, if at any, patients respond better according to different treatment, i.e. There exists a *heterogeneous prescriptive treatment effect* (Fig. 3.1b). If there are indeed heterogeneous prescriptive treatment effects determinable from the data, then it may be possible to determine an *Individualized Treatment Rule* with which to aid clinical decision-making for future patients.

3.2 Causal Inference

The Potential Outcomes framework was introduced by Jerzy Neyman (in his 1923 master's thesis) concerning randomized experiments. Neyman and Rubin (1978) introduced a framework to evaluate treatment efficacy with valid causal inference. The fundamental problem of causal inference is that we can only ever observe the outcome of one treatment for a given patient, $Y_i(A = a)$. We must instead estimate the potential counterfactual outcomes, $\{Y_i^*(A = -1), Y_i^*(A = 1)\}$, in order to make valid statements about causal relations.

The Individual Treatment Effect (ITE):

$$\delta_i^* = Y_i^*(1) - Y_i^*(0)$$

The Average Treatment Effect (ATE):

$$\delta^* = E[Y^*(1) - Y^*(0)]$$

Three main assumptions are necessary for the causal inference framework:

A1. Stable Unit Treatment Value Assumption (SUTVA) :

I. Independence between patients of treatment and health outcomes, i.e patients do not influence other patients' treatment allocation. An example of a violating case is that of infectious disease which can be transmitted between patients under study.

II. Consistency: the estimated *potential* outcome of a treatment d on patient i, $\hat{Y}_i(d)$, is equal to the true *observed* outcome when patient indeed received treatment d, $Y_i(d)$. That is, the potential outcomes are aligned with the observed outcomes.

$$Y_i = Y_i^*(1)A_i + Y_i^*(0)(1 - A_i), i = 1, ..., n \tag{3.1}$$

A2. Strong Ignorability Assumption :

There do not exist unmeasured confounders. All covariates under physician consideration are available to the analyst. RCTs are considered the gold-standard of experimental studies, due to randomization of key variables; Observational studies typically involve selection bias of treatment that violates the Strong Ignorability assumption ().

$$\{Y^*(1), Y^*(0)\} \perp\!\!\!\perp A, X \tag{3.2}$$

A3. Positivity Assumption : Each patient has a possibility of receiving each treatment.

$$P(A = a|X = x) > 0, a = 0, 1 \forall x \in X \text{ s.t. } P(X = x) > 0 \tag{3.3}$$

3.3 Outcome Regression

A standard ITR estimation approach posits a parametric conditional mean outcome model $E[Y|A,X]$ as the data generating process. See details in Chapter 3.5.1 Kosorok and Moodie (2015). Treatment prescription for a patient is then the treatment that yields the maximum plug-in estimate of all possible treatment (Eq. 3.5).

Linear regression models offer well-developed statistical machinery such as coefficient interpretation, confidence intervals, regularization, and random effects. They by nature impose linear form on the treatment rule to be estimated, which is clinically interpretable per se.

$$E[Y|A,X] =$$
$$Q(x,a;\beta) = \beta_1 + \beta_2 a + \beta_3^T x + \beta_4^T xa + ||\lambda I|| \tag{3.4}$$
$$V_Q(d^*) = E\Big\{\max_{a \in \mathbb{A}} Q(X,a;\beta)\Big\} \tag{3.5}$$

Q-Learning in Reinforcement Learning literature is concerned with optimal control for an agent navigating an action-reward state space. The so-called Quality function is the expected reward for an agent occupying a particular state X and taking action a (Eq. 3.5). Backwards inductive procedure operates such that each successive treatment allocation checkpoint is modeled by an outcome regression model that includes also the treatment choice and health outcome of the previous stage. The connection is useful for Dynamic Treatment Regime estimation, wherein sequential treatments choices are analyses. Sequential analysis is briefly introduced in Appendix B.3.

Model Misspecification is a serious concern for both regression models. Clinical errors can be grave. Cox's Proportional Hazard is the standard covariate-adjusted regression model for survival analysis, and assumes proportional hazard rates between groups. In medical context

very likely to be misspecified, and so it is questionable if this will be truly useful in clinical decision-making Lunceford and Davidian (2004).

For reference, a few models proposed to minimize misspecification of outcome regression ITR models:

Regularization Murphy (2003) details regularized Q-learning for ITR estimation, in which we aim to select only a few critical variables for clinical decision-making.

A-Learning narrows our focus to only the interaction terms that distinguish value between treatments, leaving aside main effects irrelevant to distinguish policy decisions (Eq. 3.6)

$$A(x,a;\beta) = \beta_3^T x + \beta_4^T xa \tag{3.6}$$

Prescriptive Lasso Along those lines, Lu et al. (2013) attends the Lasso operator to select *prescriptive* variables rather than predictive ones

Random Survival Forest proposed by Cho et al. (2022) utilize also tree-based methods, and shows promise for use in ITR analysis.

Bayesian Additive Regression Trees (BART) Li and Moodie (2023) introduces a boosted ensemble decision trees. Decision trees are recursive sample splitting capable of capturing non-linear relationships in data, and not inhibited by multicollinearity, and other strict regression impositions. Two types of regularization are then involved, tree-depth-pruning for robustness and boosting / ensemble learning for optimality. Integrating this machine learning approach for causal inference requires specific procedure to procure valid estimates. Honvoh et al. (2022) Introduce jackknife estimator for ITR with Inverse Probability of Censoring Weighted (IPCW) adjustment. They compare against the CoxPH Regression approach, and a BART model.

3.4 Inverse Probability Weighting

The missing counterfactual treatment outcomes Neyman and Rubin (1978) causal framework can be thought of as missing data, and therefore relevant statistical techniques for missing data problems can be employed. A simple and ubiquitous missing data technique is the Inverse Probability Weighting (IPW) estimator, introduced by Horvitz and Thompson (1952) to estimate population averages in cases of known sampling probabilities for clinical trials. It has since found use in a plethora of applications and statistical methodologies.

The key idea is to overcome bias in estimates due to missing data by inflating the observed values Y_i by the sampling probabilities π^-1 of each like units X, so as to reconstruct a "data-complete" pseudo-population with which we can derive consistent estimates of a population average.

$$\hat{E}[X] = n^{-1} \sum_{i=1}^{n} \frac{x_i}{P(X = x_i)} \tag{3.7}$$

IPW is a widely used statistical technique that can yield consistent estimates for functions of known population totals when there is missing data. For our purposes, we IPW estimators used in at least two places: (i) correcting bias from right-censored survival observations, detailed in Sec. 2.2, and (ii) for estimating counterfactual potential outcome, which we will detail now..

$$w_i = \pi(A_i|X_i)^{-1} \tag{3.8}$$

$$E[Y] = n^{-1}\sum_{i=1}^{n} w_i Y_i \tag{3.9}$$

We can estimate unbiased marginal mean counterfactual outcomes $Y^*(A)$ for population X and so naturally the Average Treatment Effect (ATE) is the expected difference in potential outcomes, $\hat{\Delta}^*_{Y,A}$.

Essentially, weight observations by their propensity score of treatment, $\pi\big(X_i, A_i\big) = \mathbb{P}(A_i = a|X_i = x)$, is used to weight observed patient outcomes. $\hat{\pi}$ is typically estimated via Logistic Regression in cases when it is not known, such as observational studies. Although, even in experimental studies it is done as a calibration maneuvre.

A note on notation:

- Y^* the asterisk denotes a *potential* outcome.
- $\Delta^*_{Y,A}$ the asterisk denotes that the difference is between *potential* outcomes.
- d^* the asterisk denotes an *optimal* rule.

$$\Delta^*_{Y,A} = E\big[Y^*(A=1) - Y^*(A=0)\big] \tag{3.10}$$

$$= n^{-1}\sum_{i=1}^{n}\left(\frac{A_i Y_i}{\pi(X_i)]} - \frac{(1-A_i)Y_i}{1-\pi(X_i)}\right) \tag{3.11}$$

Consistency. IPW is a non-parametric, unbiased M-estimator, with known approximately normal large sample distribution. The Law of Total Expectation provides the theoretical proof of IPW consistency, invoking a double expectation over a conditional expectation. Conditional probability is often used to estimate quantities given incomplete data when joint probability density function is well-defined and expectations integrable.

Law of Total Expectation:

$$
\begin{aligned}
E[X] &= E[E[X|Y]] \\
E[X] &= \int xPr[X=x]dx \\
E[X|Y=y] &= \int xPr[X=x|Y=y]dx \\
E[E[X|Y]] &= \int \Big(\int xPr[X=x|Y=y]dx \Big) Pr[Y=y]dy \\
&= \int \int xPr[X=x,Y=y]dxdy \\
&= \int x \Big(\int Pr[X=x,Y=y]dy \Big) dx \\
&= \int xPr[X=x]dx \\
&= E[X]
\end{aligned}
$$

Change of Probability Measure. The IPW technique estimates the expectation of a random variable with missing data by employing a change of probability measure, under assumption of absolutely continuous P_d w.r.t P_π. That means that trajectories from policy d must have a positive probability under exploration policy (i.e. Treatment propensity) π. Chakaborty and Moodie (2013) note that the change of probability measure is similar to that of importance sampling in Monte Carlo simulation.

The Radon-Nikodym derivative function induces a change of probability measure $\frac{d\mathbb{Q}}{d\mathbb{P}}$ by relating two probability measures $\mathbb{P}$ and $\mathbb{Q}$ for any measurable set A, if they are continuous with respect to each other,

$$
Q(A) = \int_A \frac{dQ}{dP} dP \tag{3.12}
$$

For IPW estimation, we define the Radon-Nikodym derivative $\frac{dP_d}{dP_\pi}$ between P_d and P_π the probability measures of the treatment rule and propensity, respectively, as:

$$
\frac{dP_d}{dP_\pi} = \frac{I(a = D(x)}{P(A=a)} \tag{3.13}
$$

The change of probability measure for the expected population survival $E[R]$ under regime $\mathscr{D}$: The indicator $I(A = D(X))$ in Eq. 4.2 selects instances X_i in the data where the observed treatment, A_i matches the ITR's prescribed treatment, $D(X)$. The denominator in the bracketed expression is the expected value of treatment.

$$
\mathbb{E}^{\mathscr{D}}(R) = \int R dP^{\mathscr{D}} = \int R \frac{dP^{\mathscr{D}}}{dP} dP = \mathbb{E}\left[\frac{\mathbb{I}(A = \mathscr{D}(X))}{A\pi + (1-A)/2} R \right] \tag{3.14}
$$

Drawback. "The IPTW estimator is highly variable due to the presence of the non-smooth indicator functions inside the weights." Chakaborty and Moodie (2013). There are two known drawbacks when extending to the multi-stage context: (1) when T grows large, many of the terms will be zero (2) .... (Kosorok, 2019).

$$W_a(Y,xX,A,\pi,Q) = \frac{YI(A=a)}{\pi(a;X)} - \frac{I(A=a)-\pi(a;X)}{\pi(a;X)}Q(X,a) \tag{3.15}$$

Augmented Inverse Probability Weighting (AIPTW) estimator introduces additional information from non-selected treatment cases using an inverted outcome model, $E[Y|X,A]$. It is theoretically robust against outcome model misspecification in the sense that if *either* the model for propensity score, $\hat{\pi}\big(A|X\big)$, is correctly specified, or the outcome model is correctly specified, then the estimator is guaranteed to be consistent (Bang and Robins, 2005). "Doubly Robust" methods have been introduced in survival literature, for which complex estimators must account for multiple sources of information such as survival, censoring, propensity, etc... Tsiatis and Davidian (2004), ITR estimation Zhao et al. (2015) introduce a Doubly-Robust Outcome Weighted Learning ITR estimator (Section 4.5).

Chapter 4

Direct Value Search

The chapter progresses in the following sequence: The Value Function is discussed in Section 4.1. The classification perspective is introduced in Section 4.2. Support Vector Machine (SVM), a statistical learning theory approach to classification and regression, introduced in Section 4.3. Outcome Weighted Learning (OWL) methodology is introduced in Section 4.4 and constitute our primary interest. Supporting notions from Empirical Process Theory are given in Appendix D. We focus on the developments within OWL geared towards survival data, Zhao et al. (2015) and Bakoyannis (2023) in Sections 4.5 and 4.6, respectively.

Direct Value Search estimators discover an optimal treatment rule D^* that maximizes the expected-value $E_D(Y)$ on a defined function-space $\mathscr{F}_D$ by means of *Empirical Risk Minimization* (ERM). Direct Value Search approach is concerned with arriving at an optimal rule within specified class of functions, without imposing restrictive assumptions on the data.

$$D^* \in \arg\max_{\mathscr{D}} E^D(Y) \tag{4.1}$$

Optimal linear rules. Linear functions are interpretable for clinical decision support, and amenable to statistical inference; non-linear functions are not. We wish to learn the optimal linear treatment rule, which maximizes the expected population survival outcome over the class of linear functions. Glivenko-Cantelli Theorem guarantees weak converge for linear functionals under regularity conditions Bickel and Doksum (1977).

Convergence. For Linear Regression models, convergence to optimum is not guaranteed if the model is misspecified. Empirical Risk Minimization (ERM) classification methodologies discover the optimal linear rule without imposing such harsh assumptions on the data generating process. The flexibility of non-parametric methods comes at a price of slower convergence and the need for larger sample sizes. When restricted to the linear class of functions, ERM classifiers have guaranteed rate of weak convergence of $n^{-1/2}$ (Tsybakov, 2004).

4.1 Value Function

The Value Function calculates the expected population health outcome $E^D[Y(d)]$ of a given a treatment rule d on a population X. This can be completely determined by observed data. We can estimate the value similarly to how we estimate the ITR themselves, with techniques such as Outcome Regression or IPW.

$$V(D) = E_D(Y) = \int Y dP_D = \int Y \frac{dP_D}{dP} dP = E\Big[\frac{I(A = D(X))}{A\pi + (1-A)/2} Y\Big] \tag{4.2}$$

Empirical Value. To compute the value of any given ITR f on RCT data X, the procedure simply selects participants whose randomized treatment A_i coincides with the ITR's recommended treatment $f(X_i)$ by the evaluating the sign of their product, $sign(A_i f(X_i)) > 0$ and average their rewards.

$$\hat{V}(\hat{f}) = E[R_i sign(A_i * f(X_i)) > 0] \tag{4.3}$$

Censoring. In cases of censoring, we must actually estimate the population expectation. Zhao et al. (2015) propose the Inverse Censoring Outcome (ICO) methodology, which employs IPW with respect to the censoring distribution in order to correct for bias from MAR data. We introduce properly in Section 4.5.

$$\theta_0 = \vee_{j=1}^n \mu_{0,j} = max(\mu_0, ..., \mu_n) \tag{4.4}$$

Confidence Sets around comparitive statistics such as the value difference of estimated optimal ITR to the true optimal ITR are described and distinguished in Laber and Qian (2018). Asymptotic Normal approximations of the sampling distribution of the estimator is the limiting distribution of componentwise maximum of vector of means according to M-Estimation theory (Chakaborty and Moodie, 2013). Although we are performing inference not on the parameters of a model, but rather the expected value induced under the estimated treatment rule, the convergence of the value estimates follows similarly to that of the convergence of parameter estimates as the sample size grows.

$$n^{-1/2}(\hat{\theta} - \theta) \rightsquigarrow \vee_{j=1}^n \mu_j \tag{4.5}$$

4.2 Classification Perspective

Zhang et al. (2012) cast Individualized Treatment Rule estimation as classification of treatment, disentangling two important aspects of the ITR estimation problem: the treatment contrast

function estimator and the optimization procedure. First, induce a functional form $\mathscr{F}_D$ on the treatment rule, and then secondly optimize to find optimal rule $\tilde{d}^*_D$ within class $\mathscr{F}_D$. A key idea: Set the loss function as the inverse of the value function, so that we maximise the value of the treatment function by minimising the empirical risk of the classifier. Another key idea: Weighting of missclassification cost by the difference in potential outcome. This allows us to use of essentially any Empirical Risk Minimisation classifier (Eq. 4.6), according to respective problem and classifier constraints.

$$D^* \in \arg\min_{\mathscr{D}} \mathscr{R}(D) \tag{4.6}$$

The value of rule $V(d) = E^D[Y(d)]$ is the expected population health outcome under the rule. The discrepancy in value between optimal and proposed treatments is equivalent to the discrepancy in optimal and empirical risk:

$$V(D^*) - V(D) = R(f) - R(f^*)$$

where the weighted misclassification error rate is:

$$R(f) = P(\hat{A} \neq A^*) = n^{-1} \sum_{i=1}^{n} \hat{W}_i$$

To illutrate the flexibility of model choice within the classification framework, consider that Zhang et al. (2012) specify an ITR classifier using AIPWE estimator of contrast function, and CART for classification. They compared this estimator against a logistic outcome regression in simulation study.

Asymptotically linear estimators. Asymptotically linear estimators, for example those which can be formulated as population averages, enjoy uniform consistency and weak convergance convergence. It is important to understand for the purposes of OWL estimation, that choosing a linear kernel for the SVM will specify a linear treatment rule. According to Empricial Process Theory, Linear rules are P-Donsker and enjoy favorable $n^{-1/2}$ rate of convergance. The limiting distribution of estimators of linear functions $\mathscr{F}$, $\mathbb{G}_n f$ converges in distribution to a normal r.v. $N(0, \mathbb{E}(f - \mathbb{E}f)^2)$. The Glivenko-Cantelli theorem provides guaruntees of uniform convergece, stronger than pointwise convergence. The Donsker class of functions are Glivenko-Cantelli functions which converge weakly. A very brief definition of empirical processes is given in Appendix D.

Finite Sample Performance. Of course, in RCT data we are dealing with small sample sizes and it is not clear when exactly the asymptotics "kick-in" so to speak, especially given the No Free Lunch theorem, but the efficient rate of convergence for linear functions is favorable. Empirical performance is assessed in Chapter 5 .

4.3 Support Vector Machine

The Support Vector Machine is a theoretically principled statistical learning methodology, and empirically useful even in small data regimes. SVMs can accomplish complex classification and regression involving uncertainty.

A discriminatory hyperplane between two convex classes of mostly separable data, within a soft-margin of error, and is shown to yield more robust class separation than some preceeding classification methods (such as LDA).

$$\Big[\frac{1}{n}\sum_{i=1}^{n}\phi\Big(y_i(w^T x-b)\Big)\Big]+\lambda|w|^2 \tag{4.7}$$

Primal problem guarunteed global optimum, solvable by langrange multipliers a_i. a_i are the dual coefficients.

$$\min_{\beta,\beta_0,\xi}\frac{1}{2}||\beta||^2+C\sum_{i=1}^{n}\xi_i \tag{4.8}$$

$$s.t.\quad \xi_i\geq 0, i=1,...,n \tag{4.9}$$

$$y_i\big(x_i^T\beta+\beta_0\big)\geq 1-\xi_i, i=1,...,n \tag{4.10}$$

Dual problem Langrange function is differnetiable w.r.t.β and $beta_0$. and subject to KKT Stationarity Conditions (stationarity, complementary slackness, and primal and dual feasibility).

$$L(\beta,\beta_0,\alpha)=\frac{1}{2}\beta^T\beta-\sum_{i=1}^{n}\alpha_i(y_i(\beta^T x_i+\beta_0)-1$$

$$\nabla_\beta L(\beta,\beta_0,\alpha)=0 \quad\Rightarrow\quad \beta=\sum_{i=1}^{n}\alpha_i y_i x_i$$

$$\nabla_{\beta_0} L(\beta,\beta_0,\alpha)=0 \quad\Rightarrow\quad \sum_{i=1}^{n}\alpha_i y_i=0$$

Soft-Margin relaxation of hyperplane the ξ slack variable allows discovery of an optimal decision boundary when the two target classes are not linearly seperable in product space. So-called support points are instances a_i of the data that lie on the margin of the separating hyperplane.

Covariance Structure for SVMs can be flexibly defined on Reproducing Hilbert Function Spaces (RHFS), which are geometric dot-product spaces with defined covariance structure between data with a formal distance metric. While there exist many rich covariance models, our analysis is concerned with simple kernels.

$$Euclidean: \qquad K=X^T*X \tag{4.11}$$

$$Gaussian: \qquad K(xi,xj)=-\gamma\|x_i-x_j\|^2 \tag{4.12}$$

Non-Linear Kernel Overfits in Small Samples (Bakoyannis, 2023) shows through simulation that the linear kernel shows better performance than more flexible Gaussian kernel, with variance set to 2 or 5, likely due to overfitting on small training sample. As n increases, we expect to eventually reap a modest benefit in terms of ITR-value from non-linear kernel.

Prediction is then a linear combination of support vectors and the covariance between the training data and the new data (Eq. 4.13).

$$f(x) = \beta_0 + \sum_{i=1}^{N} \Big(a_i k(x, x_i) \Big) \tag{4.13}$$

Penalisation governed parameter $\lambda_n = C\frac{1}{\sqrt{(n)}}$, balances bias-variance. It's specification must be according to the asymptotic convergence of the Central Limit Theorem $\sqrt{n}(\hat{\theta} - \theta)$. The λ penalty term should vanish asymptotically, such that the difference between the empirical sample average and the population expectation approaches zero.

$$\begin{aligned}
\sqrt{(n)}(\mathbb{P}_n - P)\sqrt{\lambda} L_{f_n, \Lambda_0, \pi_9} = \mathbb{O}_{\text{ı}}(1) \qquad & \lambda \to 0 \text{ as } n \to \infty \\
\mathbb{P}_n f := \frac{1}{n} \sum_{i=1}^{n} f(D_i) \qquad & \text{Empirical Sample Average} \\
Pf := \int_D f dP = E\big(f(D)\big) \qquad & \text{Population Expectation}
\end{aligned}$$

Tuning. The learning rate parameter can be tuned according to grid search using the k-fold cross validation, with jackknife estimator of predictive accuracy. Although this is time-consuming depending on the number of possible parameter values chosen for comparison. λ ought to vanish as $n \to \infty$, so we can choose $\lambda = n^{-1/2}$ which satisfies regularity conditions, but may not yield an optimal model.

An ill-specified learning rate parameter λ_n can lead to suboptimal or even non-sensical estimates, because the loss function is over or under penalised, the model is not fitted properly, and estimates may be inconsistent. The variability of the estimates can grow significantly.

In monte carlo simulation experiment, we generate many, many (say B=1000) random samples from various combinations of conditions {sample size, censoring rate, and survival distribution}. Tuning for the learning rate parameter should be done for each randomly generated data set. It is not sufficient to tune the parameter for distinct combination of conditions. Depending on the sample size, complexity of the model, and procedure of tuning, this can add non-trivial computational cost to simulation experiments.

Zhao et al. (2015) tune the learning rate parameter according to a grid search (although the values were not specified), from which they choose the predictively-optimal λ^* as the λ with

the greatest sum of values across its 5-fold cross validation.

$$\lambda^* := \arg\max_{\lambda} \frac{1}{k} \sum_{j=1}^{k} V_j(X^{(-j)}) \tag{4.14}$$

As k grows its the portion of data to which the model is fitted grows, and the portion of data on which to evaluate the fitted model shrinks. limit is n, so it becomes the LOOCV or jackknife estimator. Standard practice employs $k = 5$ or $k = 10$, as do Zhao et al. (2015).

$$\lim_{n\to\infty} \mathbb{P}\Big(d(\tilde{f}_n, f^*) > \varepsilon\Big) \to 0, \forall \varepsilon > 0 \tag{4.15}$$

Convergence requirements for non-parametric estimators depend on empirical process and optimization theory. Eq. 4.15 denotes the limit as sample size n approaches inf the learned rule converges *almost surely* in probability to the true optimal rule. The so-called equicontinuity assumption states that at least one continuous covariate must be involved in order to ensure that the empirical process can converge to the tight mean gaussian limiting distribution. Tsybakov (2004) details the convergence properties of classification methods. Larger the dimensionality of covariate-space the slower the convergence of the estimator, which in large-P cases we confront the *curse of dimensionality*. Zhao et al. (2012) section 3 details theoretical notions for the SVM's rate of convergance, which depend partially on the separability of the given data set and the number of support points lying along the separating margin.

4.4 Outcome Weighted Learning

The focus of this thesis is the performance of Outcome Weighted Learning (OWL) estimators. They are a Direct Value Search methodology within the classification perspective, that employs empirical risk minimization to learn an optimal treatment rule. We will show that OWL methods enjoy several advantageous properties that make them a competitive methodology for Individual Treatment Rule estimation. We will pay especially close attention to Bakoyannis (2023)'s MSOWL estimator which introduces an ITR estimator but also a maximally efficient estimator of the ITR value function for right-censored data, which is useful in evaluating any type of ITR from any type of methodology consistently and efficiently.

$$\mathscr{D}^* \in \arg\max_{\mathscr{D}} \mathbb{E}\left[\frac{\mathbb{I}(A = \mathscr{D}(X))}{A\pi + (1-A)/2} R\right]. \tag{4.16}$$

The Efficient Augmentation and Relaxation Learning (EARL) framework introduces surrogate loss for the binary classification loss. This transforms the loss function from a spikey, non-convex, non-smooth risk function to a smooth, well-behaved loss function. Such smooth

loss functions are conducive to statistical, which will learn, or iteratively reweight parameters until the specified surrogate loss function's empirical risk is minimised according to the procedure of their associated classification algorithm (Zhao et al., 2019).

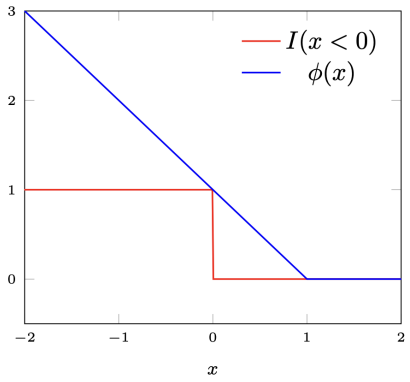


Figure 4.1: SVM Hinge Loss
$\phi(x) = max(0, 1-x)$

(Zhao et al., 2012) introduces Outcome Weighted Learning methodology, using a soft-margin SVM classifier which employs the hinge surrogate loss binary classification error. This method is consistent real-valued and uncensored outcomes $Y \in \mathbb{R}^P$. The hinge loss and SVM classifier are just one of many possible configuration. In this chatper, we will discuss the advantages of OWL (and the underlying SVM classifier) in terms of convergence and finite-sample property.

$$\mathscr{R} = E\left[\frac{I(A = D(X))}{A\pi + (1-A)/2}\phi(Af(x))\right] \tag{4.17}$$

In the case of Zhao et al. (2012), outcome weighting merely assigns the observed health outcome as the misclassification risk, such that patients covariates similar to patients with high outcomes will tend to be assigned treatment that they were actually given in the study. The outcome weighting mechanism is then evolved in sophisticated ways by Zhao et al. (2015) to provide consistent estimates in the case of right-censored data, and Bakoyannis (2023) to more effficiently utilise more information from censored cases and incorporate patient preferences in a multistate model.

The classification perspective allows variety of specification of the contrast function, $\hat{C}(X_i, A_i, Y_i)$. The IPTW-estimator of the so-called contrast function in Eq. 4.20. Zhang et al. (2012) show that the OWL estimator of Zhao et al. (2012) loss function is equivalent to an IPTW Constrast function estimator:

$$W_i = Y^*(X_i, A = +1) - Y^*(X_i, A = -1) \tag{4.18}$$

$$W_{AIPTW} = \Big| \frac{A_i}{\pi(X, \hat{y})} Y_i - \frac{1 - A_i}{1 - \pi(X, \hat{y})} \Big| \tag{4.19}$$

$$= \frac{Y_i}{A_i \pi(X_i, y) + (1 - A_i)(1 - \pi(X_i, y)} I\{A_i \neq g(X_i)\} \tag{4.20}$$

The SVM procedure then solves the empirical risk minimization of $\min \mathscr{R}$ with Tikhonov (l_2) regularisation with surrogate hinge loss (Eq. 4.21). If limited to linear class of functions $\mathscr{F}$, i.e. P-Donsker, given the property of stochastic equicontinuity, the SVM classifier is guaranteed to converge to the optimal linear rule at rate of $n^{-1/2}$.

$$\hat{f}(x) \arg\min_{\mathscr{H}^k} \sum_{i=1}^{n} W_i \frac{\phi\{A_i f(X_i)\}}{\pi(A_i; X_i)} + \lambda_n ||f||^2 \tag{4.21}$$

The treatment recommendation for a given patient $\hat{D}(x_i)$ is derived from the sign of the learned optimal function $\hat{f}$ as follows:

$$\hat{D}(x) = sign\{\hat{f}(x)\} \tag{4.22}$$

4.5 Inverse Censoring Outcome Methodology

Zhao et al. (2015) introduced the Inverse Censoring Outcome (ICO), extending the OWL ITR estimator (Eq. 4.20) to handle right-censored data. This approach assumes non-informative censoring, only considers information from non-censored cases, and selects data via the event status indicator δ in the numerator of the expression in Eq. 4.23, with bias-correction adjustment that inflates bserved survival times by inverse probability of censoring weighting. The censoring distribution can be estimated with a Nelson-Aalen estimator.

$$W_i = \frac{\Delta_i Y_i}{\hat{S}_C(Y_i | A_i, X_i)} \tag{4.23}$$

Then obtain $\hat{f}(x)$ by minimizing $\sum_{i=1}^{n} W_i \frac{\phi\{A_i f(X_i)\}}{\pi(A_i; X_i)} + \lambda_n ||f||^2$ according to SVM quadratic programming. The Zhao et al. (2015) ICO classifier's loss function (Eq. 4.24. Notice inverse probability weighting is invoked on two terms: The individual benefit is weighted by the inverse probability of censoring $\hat{S_C}$ and the inverse probability of treatment $\pi(A|X)$.

$$L = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i Y_i}{\hat{S}_c(Y_i|A_i, X_i)} \frac{IA_i \neq D(X_i)}{\pi(A_i; X_i)} \tag{4.24}$$

Doubly Robust Model is also introduced by Zhao et al. (2015). The Doubly Robust estimation is consistent given the correct specification of *either* the Conditional Outcome $E[H_j(\phi_j)]$ *or* the Censoring distribution $P(C|X,A)$.

$$W_i = \frac{\Delta Y}{S_C(Y|A,X)} - \int E_{\tilde{T}}(T|T > t, A, X)\{\frac{dN_C(t)}{S_C(t|A,X)} + I(Y \geq t)\frac{dS_C(t|A,X)}{S_C(t|A,X)^2}\} \tag{4.25}$$

where $E_T(T|T > t, A, X)$ is the Conditional Mean Residual Life-Time.

$$\hat{E}_{\tilde{T}}(T|T > t, A, X) = \frac{\tau \hat{ST}(\tau|A,X)}{\hat{S}_{\tilde{T}}(T|A,X)} - \int_t^{\tau} u d\hat{ST}(u|A,X) \tag{4.26}$$

This requires us to: Estimate Survival Distribution $\hat{S_T}(T|A,X)$ and Censoring Distribution C $\hat{S}_C(t|A,X)$ with models such as Nelson-Aalen or Cox Regression estimators. or any of the survival models discussed in Section 2,

The ICO and DR estimators completely disregard censored individuals from the loss function, according to the event status indicator δ in the numerator (Eq. 4.23 and 4.25). If the data are Missing Completely At Random (MCAR), point estimates will be consistent but with higher variance due to smaller effective sample size – i.e. not optimally efficient. In Section 4.6, we will see how the MSOWL methodology integrates over each individual benefit process up until time of censoring, therefore incorporating information from censored cases. This means that in cases of censoring, the MSOWL will be more efficient, deriving a smaller variance due to the increased effective sample size. We are interested to see empirically how much information from censored cases improve value estimates in an applied analysis of clinical trial data. See Section F.2.1 for the plots of simulation results from the Zhao et al. (2015).

4.6 Multistate Outcome Weighted Learning

Multi**S**tate **O**utcome-**W**eighted **L**earning (MSOWL) estimator (Eq. 4.33) , introduced by Bakoyannis (2023), incorporates patient-preferences of disease progression; Importantly, the MSOWL ITR and value estimators are theoretically more efficient than previously introduced OWL methods.[1] The non-parametric estimation approach for the state transition does not impose any additional assumption on the state-transition, not even the common Markovian. Note, Datta and Satten (2001) shows the consistency of the Nelson-Aalen estimator for multistate processes even when they are not Markovian.

[1]The R code is available at `github.com/gbakoyannis/msowl`.

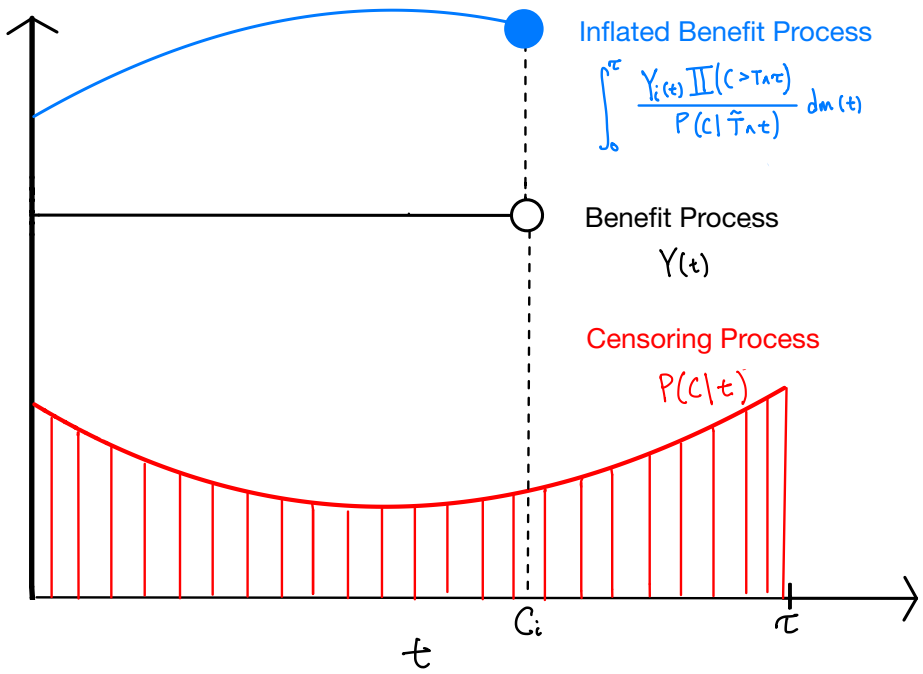


Figure 4.2: MSOWL IPCW-Inflated Stochastic Integration.
Processes: **Observed Benefit**, **Inflated Benefit**, **Censoring**

Multistate Methodology. Refining analysis to account for disease progression helps to deliver patient-center care, a goal of precision medicine. Bakoyannis (2023) introduces a multi-state disease progression model into the OWL formulation; patient-ellicited Quality-of-Life preferences are incorporated as state-weights within the summation of the multistate benefit process. State of Disease is weighted by Patient Preference, e.g. $w = (0.5, 1, 0)'$, time in state 1 is reduced by 50%. Therefore, patients may choose to optimize only their progression-free, others may value the time in progression-free disease state say twice as much as the time spent in progressive disease state.

$$\text{States of Disease:} \qquad S = \{1, \ldots, S\} \tag{4.27}$$

$$\text{Disease as function of time:} \qquad \tilde{X}(t) = (\mathbb{I}\{X(t) = 1\}, \ldots, \mathbb{I}\{X(t) = S\}) \tag{4.28}$$

$$\text{Time in } j\text{th state for patient } i\text{:} \qquad D(X_i, j) = \int_0^\tau \mathbb{I}\{X_i(t) = j\} dm(t) \tag{4.29}$$

$$\text{Benefit Processes:} \qquad \{Y_w(t) dm(t), w \in \mathbb{W}\} \tag{4.30}$$

$$\text{Individual Benefit Process:} \qquad \{Y_{i,w}(t) = w'(I\{X_i(t) = 1, \ldots, I\{X_i(t) = S\})'\} \tag{4.31}$$

where the integrator function $m(t) = t$ induces a Lebesque measure on the Borel σ-algebra $\mathscr{B}([0,1])$.

Value Function $\mathscr{V}(d)$:

$$\mathscr{V}(d) = E\left(\left[\int_0^\tau \frac{Y_{w_{\text{pref}}}(t) I(C \geq T \wedge t)}{exp(-\Lambda_0(\tilde{T} \wedge t)} dm(t)) \right] w_{d,\pi} \right) \tag{4.32}$$

Multistate Outcome Weighted Learning value function estimator $\hat{\mathscr{V}}(d)$:

$$\hat{\mathscr{V}}(d) = \frac{1}{n}\sum_{i=1}^{n}\left(\left[\int_0^{\tau}\frac{Y_w(t)I(C \geq (T_i \wedge t))}{exp(-\hat{\Lambda}_0((\tilde{T} \wedge t))}dm(t)\right]\frac{I(A_i = d(Z_i))}{A_i\hat{\pi} + (1-A_i)/2}\right) \tag{4.33}$$

Multistate Change of Probability Measure:

$$V(d) = E_d Y = \int Y\left(\frac{dP_d}{dP_\pi})dP_\pi\right) = E\left\{\int_0^{\tau} Y_w^*(t;d)dm(t)\right\} = \int w_{d,\pi} Y dP_\pi$$

where $w_{j,\pi} = \prod_{j=1}^{K}\frac{\mathbb{I}[A_j = d_j(H_j)]}{\pi_j(A_j|H_j)}$ is a Radon-Nikodym derivative between P_d and P_π; integrator $dm(t) = t$ and induces Lebesque measure on Borel σ-algebra $\mathscr{B}([0,\tau])$; $\int_0^{\tau} I\{X(t) = j\}dm(t)$ is time spent in jth state during study.

Censoring-Efficiency. A key advantage of the Bakoyannis (2023) MSOWL methodology with respect to its predecessor: For any right-censored data the MSOWL method will be more efficient in the estimation of an optimal ITR than the Zhao OWL methodologies. MSOWL accomplishes such increased efficiency by integrating the stochastic benefit process $Y_w(t)$: $t \in [0,\tau]$ of each individual, until their individual time of censoring $I(C \geq T \wedge t)$, where $a \wedge b = min(a,b)$, Inverse-Probability-of-Censoring Weighting the benefit process $Y_w(t)$ per state by the patient's QoL preference w_i and censoring distribution $\Lambda(\tilde{T} \wedge t)$. Information is then averaged across patients. The patient preference weighting $w_i \in [0,1]$ is a state-occupation-time multiplier, allowing patients to value their time in various states (initial disease state, tumour response) differently in the value equation.

Stochastic Integration of Right-Censored Benefit Process. The MSOWL reward function computes a per-patient IPCW-inflated benefit process, which is described within the bracketed expression of Eq.4.33 and isolated in Eq.4.34. For each patient, MSOWL integrates from $t = 0$ to $t = \tau$, computing the area under the curve between discrete observed events, and inflating each area by the instantaneous probability of censoring. A visualisation of an IPCW-inflated benefit process is shown in Figure 4.2. The relevant MSOWL code for computing the reward is profiled in Appendix G.

$$\int_0^{\tau}\frac{Y_w(t)I(C \geq (T_i \wedge t))}{exp(-\hat{\Lambda}_0((\tilde{T} \wedge t))}dm(t) \tag{4.34}$$

Influence Function. ψ governs the variance of the estimator. The sources of variation stem from the variation of the survival process f_T, the censoring process f_C, the treatment propensity f_π, and the latent counterfactual treatment outcomes f_{T^*}. Bakoyannis (2023) Theorem 3 implies that, for any measurable decision function f , $\varepsilon_{n,w}(f) = o_p(1)$, and if $\mathscr{F}$ is the space of linear functions, then $\sup_{f\in\mathscr{F}}|\varepsilon_{n,w}(f)| = o_p(1)$ and the class of influence functions

$\{\psi_w(f) : f \in \mathscr{F}\}$ is P-Donsker. The error term $\varepsilon_{n,w}(f)$ is asymptotically negligible (converges to 0 in probability), such that the estimator's deviation from the true value is constituted by the sum of the influence functions, which is asymptotically normal.

$$\mathbb{G}_{n,w}(f) := \sqrt{n}\left(\hat{V}n, w(\text{sgn}(f)) - V_w(\text{sgn}(f))\right) \tag{4.35}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{i,w}(f) + \varepsilon_{n,w}(f), \quad w \in \mathscr{W} \tag{4.36}$$

Confidence intervals derive via the following, in which the $c_\alpha = 1 - \alpha$ percentile can be estimated by simulation procedure.

$$\hat{\mathscr{V}}(sgn(\hat{f}_{n,w})) \pm c_\alpha \frac{\sigma(\hat{f}_{n,w})}{\sqrt{n}}, \tag{4.37}$$

Jackknife estimation works similarly to Leave-One-Out-Cross-Validation (LOOCV), by fitting $n-1$ models, each round j leaving one datum out X_j such that we fit a model on $X^{(-j)}$ and evaluate, and assessing each model's predictive accuracy on the single left-out datum of the round $\hat{\mathscr{V}}^{jk}(X_j)$. This procedure allows us to estimate the variability of the model with respect to a finite data sample.

$$\hat{\mathscr{V}}^{jk}_{n,w}(\hat{d}_{n,w}) = \frac{1}{n} \sum_{i=1}^{n} \left(\int_0^t Y_{i,w}(t) I(C_i \geq T_i \wedge t) \frac{dm(t)}{\exp\left\{-\hat{\Lambda}_n(T_i \wedge t)\right\}} \times \frac{I(A_i = \hat{d}^{(-i)}_{n,w}(Z_i))}{A_i \hat{\pi}_n + (1 - A_i)/2} \right) \tag{4.38}$$

While better than merely evaluated the model fit on the full training data, because it "leaves out" a datum each round, the jackknife procedure is still expected to overestimate the true value of the fitted model on out-of-sample data, because the data were still collected together and may share many underlying proporties that are not universal. In order to evaluate personalised medicine models, proper attention must be paid to the distance and transport of probability distributions. Factors such as geographical location and ethnicity of the participants are often incompletely covered by oncological trials; In Section 5 we will discuss individualized treatment estimation from RCT data and briefly expound on such demographic concerns.

Computational Complexity. The cost of MSOWL value integrating information from censored cases, is cubic computational complexity $\mathscr{O}(n^3)$. The expression of interest within Eq. 4.33 performs more computation than ICO's simple use of the Nelson Aalen estimator. Pair-wise comparison of all events is compared at each event-time t in the study. Which can produce impractically long computation times for even modest sample sizes (see Figure 4.3).

Divide & Conquer. Fortunately, we can employ a simple procedure to reduce the computational complexity time to be quasi-linear. Divide & Conquer approximation randomly batches the data reduces the total number of pair-wise patient comparisons required. Statistically, the

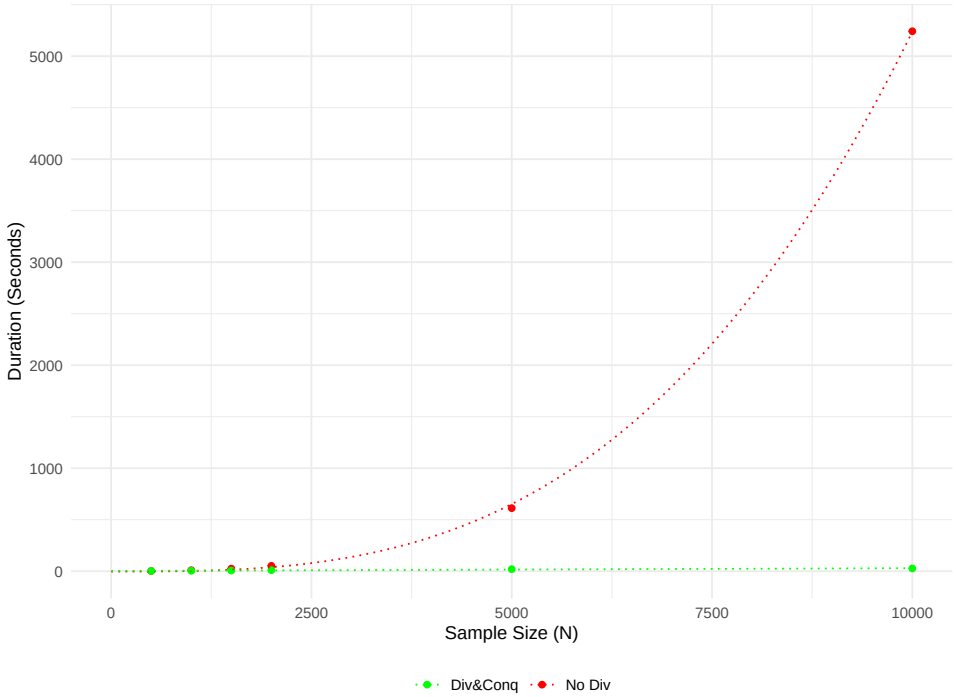


Figure 4.3: MSOWL Value Estimator: Sample Size x Processing Time

estimate will be consistent according to the law of Total expectation. Chen et al. (2021) review Divide & Conquer methods, general approaches towards computing big data and complex procedures. for several types of statistical estimators. Dividing into batches can lead to massive speed up in processing time for data analysis, which is necesary for the MSOWL Value Function is involves cubic computational complexity.

$$\hat{E}[\hat{\mathscr{V}}] = \frac{1}{k}\sum_{i=1}^{k} \hat{\mathscr{V}}(X_k) \tag{4.39}$$

The Divide & Conquer procedure is as follows: We desire to estimate the value for a a large population. Divide the population X into k batches, compute the individual weights $(W_{k1}, ..., W_{k2}) = W_{ki}$ for each patient i in each batch X_k, and then recombine the weights into the full W by taking the mean of all batches' value statistic, to render an unbiased estimate of population mean. Each patient is therefore assigned a weighting that is calibrated to the random-batch to which they belong, but in aggregrate the recombined value will be consistent.

Figure 4.3 shows the processing time for increasing sample sizes with various Value Computation approaches – (1) naive full computation, (2) Pre-Calculate Divide & Conquer, (3) Simple Divide & Conquer. A single evaluation of the value function zeta_t function on $N = 10,000$ dataset is measured to take about 1.5hrs on M2 Pro processor. If $N = 10,000$ is divided into k batches of $n = 200$, the Divide & Conquer approach to compute the value takes approximately 30 seconds.

An additional amortisation of value computation with respect to the number of ITRs under

comparison, that we may be able to reap. The time intensive aspect of the MSOWL Value estimation is the computation of the outcome weights, which does not involve ITR assigned treatment term.

Chapter 5

Application to Data

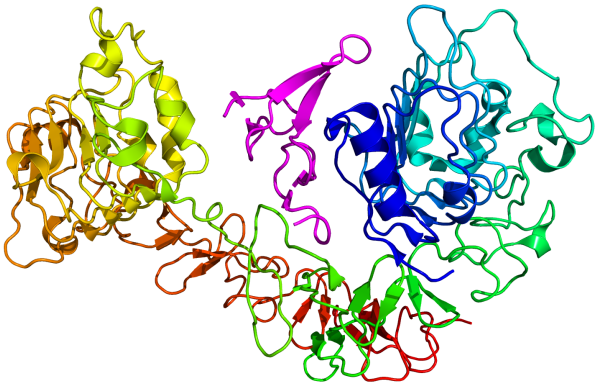


Figure 5.1: Epidermal Growth Factor Recepter (EGFR) Boghog (2008)

We perform a case study in Precision Oncology, assessing Individualized Treatment Rule estimators on real data from a Randomized Clinical Trial, the PRIME trial (Douillard et al., 2010). This analysis follows the application of Bakoyannis (2023) MSOWL estimator on the SPECTRUM trial to compare against standard one-size-fits-all approaches and the pre-existing survival-adjusted OWL methodologies of Zhao et al. (2015).

Trial	**N**	**P**	**Censoring**
SPECTRUM	512	4	10%
PRIME	800	4	10 %

Table 5.1: RCT Data Sets

Panitumumab immunotherapy has been shown to have promising results for several types of cancer, as it targets the Epidermal Growth Factor Recepter (EGFR), which regulates cellular proliferation among other things Ciardiello and Tortora (2008). The discovery of which was awarded a Nobel Prize in 1986. EGFR is a transmembrane molecular pathway that resides in the epidermus (Figure 5.1).

Let us overview a few key biological mechanisms at play in precision oncology and the statistical model therein:

Aggressive Adverse Events are more prevalent with Immunotherapy treatment than chemotherapy. Non-fatal adverse events include 90% chance of skin toxicity, typically apparent within two weeks, and 15 % of which were severe. There is also a 10% chances for fatal stroke or heart attack. Panitumumab targeting the EGFR which regulates cell production in the epidermus, and the severe skin toxicities were associated with improved progression free survival and overall survival (Peeters et al., 2009). Individualized models might aim formalize predictive relationship of patient adverse events from aggressive treatment, although the measurement and data fidelity of adverse events in RCTs is not of high quality (Brown and Green, 2015).

Multistate Analysis may be beneficial approach towards patient-centered care. The benefit of some immunotherapies is known to specifically increase the progression-free survival (PFS), not so much the overall survival (OS). The difference may be important for patients when considering Quality-of-Life outcomes among treatment options. Maximising PFS may be more important to some patients' Quality-of-Life preference (Douillard et al., 2010). Delgado and Guddati (2021) provides a primer on important endpoints in clinical oncology, such as Tumor Response, Disease Progression, Death, End-of-Study, Dropout.

Estimand	**Start**	**End**
Time to Response	Randomization	Tumor Response
Progression-Free Survival	Randomization	Disease Progression
Overall Survival	Randomizaion	Death/End-of-Study
Duration of Response	Tumor Response	Disease Progression

Table 5.2: Clinical Endpoints in Oncology

Proportional Hazards assumption often finds itself violated due to the delayed response of the treatment with respect to standard-of-care chemotherapy. Ristl et al. (2021) explores the dynamics of delayed treatment response, such as when administering immunotherapy for patients with KRAS gene; this case we will analyse ourselves in Section 5.2.

In praxis, we would like our ITR estimators to formally account for such patient Quality-of-Life, such as minimising aggressive adverse events, or minimising time in progressive state, and to be flexible to handle data which do not satisfy the proportional hazard rates between groups.

5.1 SPECTRUM Trial

The Study of Panitumumab Efficacy in Patients With Recurrent and/or Metastatic Head & Neck Cancer (SPECTRUM) [1] was a multi-center, two-arm Randomized Controlled Trial to assess

[1] Data available with permission from ProjectBioSphere.com

the effectiveness of the addition of Panitumumab in treatment for Head & Neck Cancer with a focus on Progression-Free-Survival (PFS) of sustained tumour response (Vermorken et al., 2013). 520 Patients were assigned randomly two groups gven the following four variables which are clinically known to be predictive of survival outcomes: age, primary tumour site, patient history prior treatment for squamous-cell carcinoma, and baseline patient ECOG status (ambulatory vs. active).

The primary aim of the study was the assessment of immunotherapy inclusion to standard chemotherapy for patients suffering from Recurrent or Metastatic Squamous-Cell Carcinoma of the Head and Neck (SCCHN) on progression-free survival (PFS). The primary finding was that Panitumumab treatment yielded a positive survival difference of almost 2 months, of borderline significance (Vermorken et al., 2013).

Misiukiewicz and Posner (2013) explains the lack of differentiation between the two treatments under study as due in part to confounding by second-line treatments; Clinical solutions for such disease typically require sequential and combinatoric consideration of several first and second-line treatments due to complex disease mechanisms. Also important to note that efficacy varied by high-level geographic regions, potentially due to underlying genetic differences. HPV history is most important biomarker, but unfortunately, 40% of HPV History are missing in the SPECTRUM data set.

Bakoyannis (2023) MSOWL SVM-optimized linear rule outperforms one-size-fits-all rules, and both slightly increased value and variance with respect to ICO and DR methods in the case of censoring.

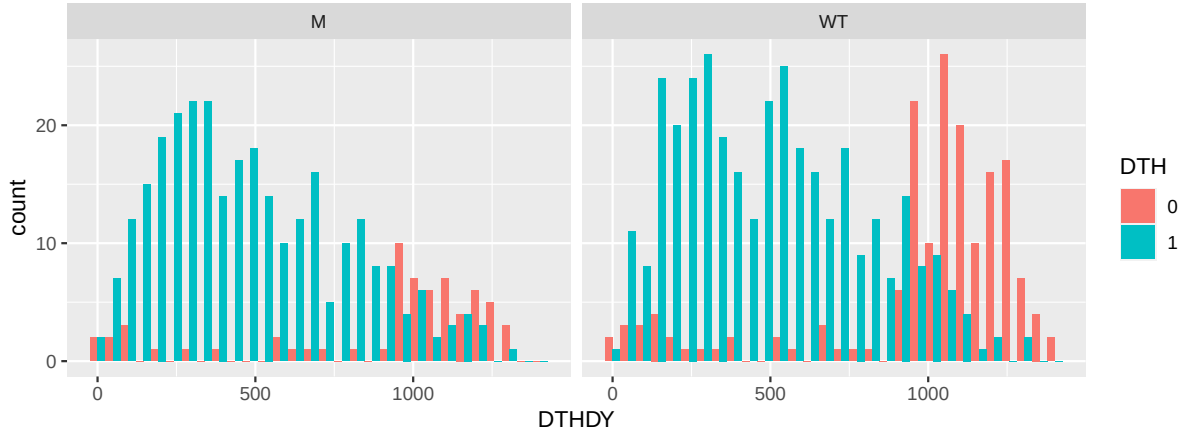
5.2 PRIME Trial

We are interested to assess Individualized Treatment Rules on a novel data set that involves survival outcomes with right-censoring. The PRIME Trial (Douillard et al., 2010) assesses Panitumumab (immunotherapy) vs FOLFOX (standard care, chemotherapy) primarily on Progression-Free Survival, with focus on wild-type vs mutant KRAS gene Douillard et al. (2010). More than 1,000 patients were randomized on (1) Geographic Region (Western Europe, Canada and Australia VS Rest of World) (2) ECOG Status. Two groups by stratified to KRAS exon 2 (c12/13) biomarker: 352 Mutant (M), 514 Wild-type (WT), 69 N/A. The publically available dataset has anonymized the patient data and only includes a random subset containing 80% of the patients.

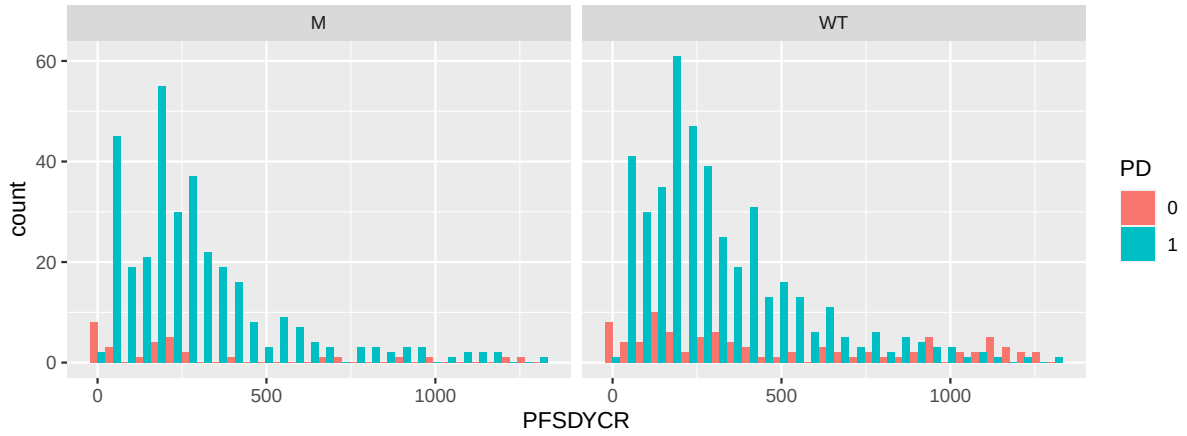
5.2.1 Exploratory Analysis

The KRAS biomarker is known to be a significant prescriptive variable which may differentiate health outcomes across patients. Figure 5.2 shows histograms of survival times, color indicates censoring status, and separated by Overall Survival and Progression-Free Survival. As expected, there are more censored cases in the Overall Survival data; Censoring may follow

a kind of bathtub-shaped distribution, in which a few dropouts occur at the begining of study, and then things settle down until the end of the survival distribution in which patients come very near to death and are typically subjected to higher rates of censoring, which may potentially be Missing Not At Random (MNAR) censoring rate is approximately 10%.



(a) Overall Survival (OS)



(b) Progression Free Survival (PFS)

Figure 5.2: PRIME: Histogram of OS & PFS Times by KRAS type

Delayed Reponse. Note the divergence of the curves between the immunotherapy (solid) and chemotherapy (dashed) groups during the middle portion of the timeline. Patients with Wild-Type KRAS receiving immunotherapy slightly outperform those receiving standard chemotherapy. Patients with Mutant KRAS biomarker receiving immunotherapy underperform significantly during the middle segment (months 30 to 60), before reconverging towards the end. This may potentially lead to a violation of the proportional hazards assumption, but this will be confirmed upon model checking of the Schoenfeld residuals of the Cox regression model in Sec. 5.2.2.

Heterogeneity. For Progression-Free Survival, the Wild-Type KRAS group receiving immunotherapy perform best. Chemotherapy curves for both KRAS groups are very similar, with a slight divergence appearing between months 10 to 16. Lastly, the Mutant KRAS group re-

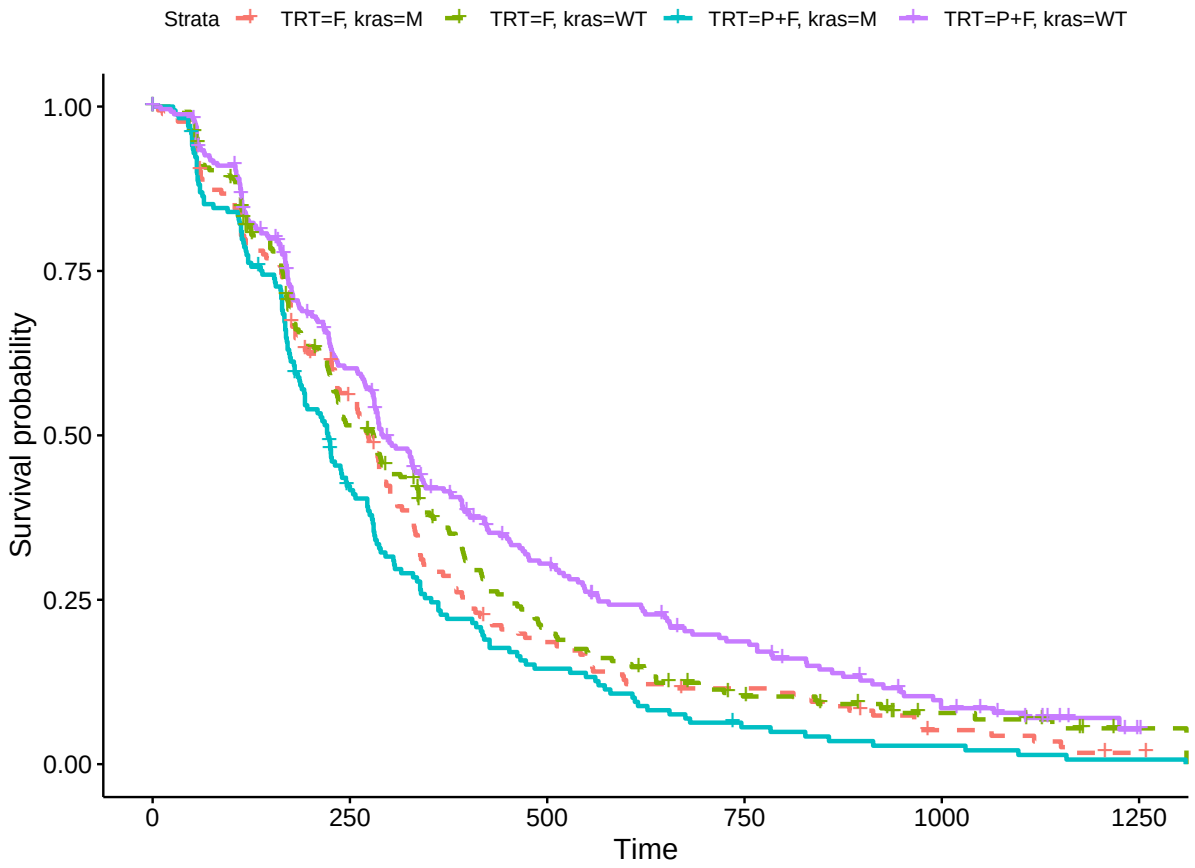


Figure 5.3: **PRIME: Progression Free Survival (PFS)**
Survival curves grouped by treatment and KRAS type.
Solid = Immunotherapy, Dashed = Chemotherapy.

ceiving Immunotherapy which suffer a the most drastic drop in survival probability after 6 months.

5.2.2 ITR Analysis

Let us now compare ITR estimators on the PRIME dataset for Progression-Free Survival. The experimental treatment (immunotherapy + chemotherapy) is encoded as $A = +1$, the standard of care is encoded as $A = -1$. The decision problem is posed such that from the estimated treatment effect function $\hat{f}$ we can derive the treatment rule as $\hat{d} = sgn(\hat{f})$.

Model Fitting. We tune MSOWL and ICO models independently according to an exhaustive grid-search involving cross-validation schema. Both k-fold and jackknife can be employed to assess the variability of the estimator given a finite sample of data. K-Fold Cross Validation where $k = 5$ or 10 is standard practice, and reduces computational time with respect to jackknife by 80-160 times for the data set of $n = 800$. See Figure H.3 for cross-fold validation results per lambda.

Value Comparison. Table 5.3 showcases the value estimated under three Individualized

Estimator	$\hat{V}$	SE	ll	ul
ICO	11.10	0.40	10.20	11.90
MSOWL	12.20	0.6	11.00	13.40
COX	12.40	0.50	11.40	13.40
+1	11.40	0.40	10.60	12.30
-1	11.10	0.40	10.20	11.90

Table 5.3: PRIME PFS: ITR Value Comparison (Months)

Treatment Rule estimators and the two homogoenous one-size-fits-all rules. The MSOWL estimator yields an estimated value of approximately 12.2 (+/- 0.6) months, outpeforming the one-size-fits-all treatment rules which yield estimated restricted mean survival time. 11.4 (+/- 0.4) and 11.1 (+/- 0.4) months. Therefore by the criterion of value alone, the MSOWL individualized rule only slights outperforms, by about one standard error, immunotherapy for all (the better overall one-size-fits-all treatment). The Cox Proportional Hazard Model actually shows a good fit here, according to both the Schoenfeld residual plots which do not betray any discernable pattern (Fig. **??**), and the overall value of the Cox-ITR 12.4 (+/- 0.5) months.

Estimated Optimal Individualized Treatment Rules.

MSOWL Estimated Optimal Individualized Treatment Rule:

$$\hat{f}_{MSOWL} = 0.02 - 3.5e^{-8}X_{\text{AGE}} + 0.73X_{\text{SEX}} + 0.21X_{\text{ECOG}} - 1.06X_{\text{KRAS}} \tag{5.1}$$

ICO Estimated Optimal Individualized Treatment Rule:

$$\hat{f}_{ICO} = -0.3 - 1.3e^{-7}X_{\text{AGE}} + 0.13X_{\text{SEX}} + 0.03X_{\text{ECOG}} - 0.19X_{\text{KRAS}} \tag{5.2}$$

COX Estimated Optimal Individualized Treatment Rule:

$$\begin{aligned}\hat{f}_{COX} = \frac{h(t)}{h_0(t)} &= 0.97X_{\text{AGE}} + 1.00X_{\text{SEX}} + 0.76X_{\text{ECOG}} - 1.07X_{\text{KRAS}} \\ &+ X_A\Big(1.07X_{\text{AGE}} + 0.82X_{\text{SEX}} + 0.80X_{\text{ECOG}} + 1.42X_{\text{KRAS}}\Big)\end{aligned} \tag{5.3}$$

Model Interpretation. While slightly underperforming the well-fitted COX-ITR model, MSOWL modestly outperforms the fitted ICO model; this is to be expected given there are about 10% of cases censored. MSOWL finds an optimal rule in which the patient characteristics of SEX, ECOG, and KRAS have coefficients of +0.7, +0.2, and -1 months respectively. Such a linear decision rule can be used by clinicians to reason about patient treatment by understanding scalar multiplicative effects of patients' particular characteristics. The coefficient for AGE is very small as to not affect the function at all. AGE is the only continuous variable of the four patient biomarkers; at least one continuous variable is needed for the equicontinuity condition, perhaps another continuous biomarker can be used instead of AGE, to yield even stronger benefit from individualisation.

Treatment Aggression. From a multi-objective optimization perspective, ITR not only matches or slightly outperforms the homogenous rule in terms of expected restricted survival of the population, but also it allows some patients to not partake in the more aggressive (and often, but not always, more effective) treatment may match the expected value of one-size-fits-all rules, but prescribing either treatment depending on the subgroup, such that the individualisation property produce a treatment rule that is not dominated by homogeneous treatment rules. Individualising spares some patients from the aggressive immunotherapy treatment, thus improving Quality-of-Life for those patients.

Multistate Analysis. The MSOWL estimator allows us to go further and refine our analysis of a patients disease trajectory through a multistate model (e.g.: Initial, Response, Progression, Death). Further analysis of the PRIME clinical trial formally accounting for multistate trajectory is of interest, but due to time constraints is left for further work.

Chapter 6

Discussion

This thesis reviews the recent literature in Individual Treatment Rule (ITR) estimation for right-censored outcomes, oriented around a dialogue between model-based, so-called parametric and non-parametric theory. Parametric estimators include linear regression or Cox Proportional Hazards regression, and specifically in the Precision Medicine literature technique such as G-Estimation (Sec. B.2), Outcome Regression Q-learning (Sec.3.3), etc. Non-parametric estimators include Kaplan-Meier Hazard Function Estimator (Sec. 2.2), Inverse Probability Weighting (Sec. 3.4), Support Vector Machine classification (Sec. 4.3).

The prevalence of small sample sizes in RCTs, and typical right-censoring in survival data, motivates data-efficiency in statistical estimation. Meanwhile, we also wish to employ non-parametric estimators wherever possible, so as to impose as few restrictions as possible on the data. Such methodologies are of interest due to their universal nature, which can be generically applied to a plethora of data sets with differing assumptions.

Our main focus was on Outcome Weighted Learning (OWL) estimators, which aim to directly estimate treatment rules within a predefined class of functions via weighted classification methods. A data set from a real oncological clinical trial was analysed, with ITR models under comparison including: ZOM (one-size-fits-all), Cox Regression, and Outcome Weighted Learning (ICO and MSOWL).

MSOWL uses more information from censored cases than ICO, which consequently reduces variability and yields higher overall expected survival outcomes (Sec. 4.6). MSOWL integrates over the stochastic benefit process per patient, inflating each segment between observed events by the inverse probability of weighting, and then take an expectation with respect to the popoulation. This procedure has computational complexity on the order of a cubic polynomial; such computational bottle-neck can be addressed for moderate-to-large sample sizes of $N > 2000$ using a Divide & Conquer approach (Sec. 4.6).

In Section 5 we perform ITR analysis on a novel data set from a real clinical trial, which evaluated immunotherapy Panitumumab with respect to standard care chemotherapy. We see MSOWL outperform the ICO estimator, and just slightly underperform the COX-ITR which happens to be well-fitted to the data set, i.e. the proportional hazards assumption holds in this case. Further work, to assess more clinical trial data to determine cases in which the

crucial assumptions for regression do not hold, a supposedly common occurance in clinical trials assessing immunotherapy vs chemotherapy.

A broad issue underlying problems of our interest is missing data or censoring. In survival data, right-censoring is standard. In precision medicine, missing data might be in the form of important patient health information, biomarkers which can greatly determine optimal treatment strategy and expected health outcomes. For example, HPV is a known biomarker for Head & Neck cancer outcomes, but in the multi-center SPECTRUM trial, HPV data was missing for 40% of patients. Missing data estimators are typically built on probability theory of conditional expectation.

The Law of Total Expectation, $E[Y] = E[E[Y|X]]$, appeared in two different aspects of the MSOWL estimation, allowing us to decompose the expectation of a variable of interest in order to estimate an expectation despite missing data. In Section 3.4, we introduced the IPW method which relies on the conditional expectation of the outcome with respect to treatment. In Section 4.6, we discuss the Divide & Conquer approach which relies computes the expectation of the conditional expectation of the data on assignments to small batches.

The term *individualization* is a bit fraught in the sense that statistical estimation and decision-making optimization are done with respect to the expectation of overall population health outcomes. Individualisation really refers to the tailoring of treatment rules to *subgroups* of patient population. optimization and Inference in ITR estimation is performed at population level, with respect to expectation of population health outcomes. More generally, the nature of statitical prediction is based on population-level gauruntees. In truth, *heterogeneity* of treatment interaction with respect to patient covariates is the statistical aim of Precision Medicine.

The OWL framework views Individualized Treatment Rule estimation as decision-making under uncertainty, it structures the problem as that of Empirical Risk Minimization. An interesting possibility begins to open regarding how exactly we prioritise patients. Under the initial OWL proposal, patients are weighted according to their overall observed health outcomes, and thus patients with larger survival times contribute more to the estimation of population survival. This configuration implicitly follows a *Utilitarian* framework, in which aggregate population outcome is the criterion, and no individual-level mandates are imposed. The SVM will knowingly misallocate treatment, especially when we desire to limit the complexity of the learned decision function, for example to be linear. Would it ever be necessary to impose some individual-level constraints? Ethical consideraitions of regarding undue harm may be necessary. Further work may focus on applying individualized treatment to various real data sets, assessing which cases avail to individualisation benefit.

Appendices

Appendix A

Aggressive Treatment

We might find a heterogenous treatment rule that is of a similar estimated value as an aggressive one-size-fits-all rule, , then we might say there is an individualisation benefit from merely sparing a minority of patients from a needlessly aggressive treatment. Butler et al. (2018) discuss ways to formally incorporate trade-offs of treatment aggression, patient preferences, and overall survival outcomes. A common approach is to introduce a trade-off variable ψ that is determined by preference ellicitation.

Adverse Event data may be incorporated into survival analysis, or perhaps a multistate model. MSOWL methodology only considers side-effects insofar as they affect survival. Non-fatal side-effects can be severe, painful, and require second-line treatment, but may not be detected by the variables mentioned in analysis so far. There is no formal inclusion of adverse events from treatment in Eq. 4.33, and it is difficult to imagine a criterion which would reflect immunotherapy's aggression in a principled and consistent way. Can we formally account for side effects and/or treatment aggression disparities into our model? Accounting for side-effects may be an important facet to achieving patient preference's of Quality-of-Life-adjusted ? outcomes.

Ethics come into play when we try to calibrate patient-preferences and avoidance of side effects. Patient can provide a state-preference-weighting, but how calibration of these into a survival model can still be difficult. Expert clinical input can provide notions of the exchange rate between life-year and duration or severity of adverse events. Let say there are two patient state-preference weightings of $w_A = (1, 0.5, 0)$ vs $w_B = (1, 0, 0)$ according to the state-weighting schema of MSOWL. In both cases, patients want to maximise their time in the initial disease state, i.e. progression-free survival. Then patient A values their time in state of tumor response half as much, and patient B does not assign any value to this state.

Causality of adverse events is also difficult, due to the sequential and sometimes delayed timing and impact of immunotherapy. Positive and negative effects ought to be incorporated into causal experiment. In some cases, adverse events predicate tumor response with some certainty. We need to identify whether it is in the best interest of the patient to promote or inhibit the adverse-event of the patient. We wouldn't want to encourage patients to avoid unpleasant side-effects at the risk of losing precious life-years. Only for patients who really wouldn't

receive the benefit of immunotherapy we'd like to spare them the negative effects, as well.

Data Quality of adverse event records in RCTs is notoriously poor. It may prove difficult to infer any strong conclusions. Preference-illicitation of personal quality-of-life value-judgements is a clinically employed procedure.

Three ideas further analysis of treatment aggession:

1. Minamally Aggressive Optimal Rule Find the least aggressive optimal treatment rule that is within the confidence interval of the optimal rule. Or perhaps within the SVM margin of error. What is the optimal rule that assigns the fewest aggressive treatment? What is the variability in patient treatment allocation within the class of optimal or near-optimal rules? How much do the treatment assignments differ as we move around that space?

$$f_{min-agg} = \arg\min_{f^*} \sum_{i=1}^{n} I(sgn(\hat{f}(x)) > 0) \tag{A.1}$$

$$\text{where } f^* \in \arg\max_{f} E[Y|f(x)] \tag{A.2}$$

2. Aggression Penalty Can we introduce a penalty term for aggressive treatment that biases the misclassification error towards non-aggressive treatment? This term would then be arbitrarily set, tuned by hand by what possible preference ellicitation? Would it be useful to map the change in treatment rule value as we increase the aggression penalty ? A.3.

$$W_i = \frac{\Delta_i Y_i}{S_C(Y_i|A_i, X_i)} + \xi \mathbb{I}(A_i = +1) \tag{A.3}$$

3. Multi-state Inclusion of Adverse Events proposed by Bakoyannis. Count the time that patients endure adverse events on the same multistate process as the survival analysis. This would bring variables in line to incorporate nicely under the state weighting according to patient-preference. We would then extend the three states: **S1: Initial Disease, S2: Tumor Recession, S3: Progression/Death** to include either a single state for adverse events: **S4: Adverse Event** or possibly to capture the second-order effects of adverse effects on tumor response, we can define more complex state definitions such as: **S5: Adverse Event during Disease Progression** and **S6: Adverse Event during Tumor Response**. Although it remains to be seen how much information we can expect to gain from this finer graduation of patient disease states.

Appendix B

Dynamic Treatment Regimes

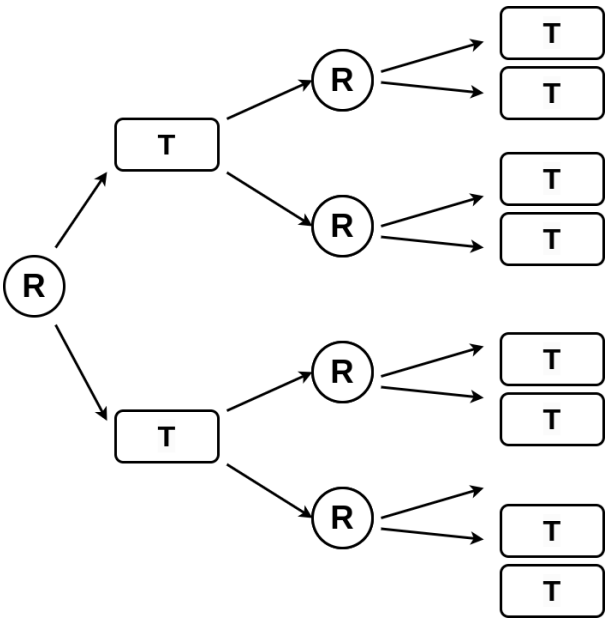


Figure B.1: SMART 2-Stages each with 2-Arms. Randomized experiment to measure the effect of sequentially prescribed treatments, with the aim to discover prescriptive effects. Legend: (R) Randomization, (T) Treatment.

Dynamic Treatment Regime extend ITR analysis to a sequential treatment-assignment context, where a DTR is a function that prescribes a sequence of treatment allocations given patient-specific, time-varying health status.

Sequential, Multiple Assignment, Randomized Trials (SMART) are clinical trials designed to answer comparative questions about sequential treatment decisions. Test scientific hypotheses that investigate questions such as *what is the best first-round treatment for alcohol-dependence recovery when marginalising over all second-round treatment choices?*.

Responsive treatment policies, swifter trial efficacy by allowing unfortunate patients to switch treatment according to their preference.

Multi-Stage Decision-Making Chronic-disease treatment involves long-term clinical follow-up to monitor patient health status and adjusting or changing treatments accordingly. Statistical

and computational models must take into the complexity of consideration time-varying covariates and sequential treatment effects with prescriptive, moderating, or treatments to future treatments. Treatment-to-Treatment Interaction Effects measure the increase in final health outcome when comparing patients who recieved various sequences of treatments over time. Intuitively, treatment efficacy can differ if paired sequentially with other treatments. Variability in the number and timing of patient follow-ups introduces non-trivially complexity.

An example of this phenomenom is exemplified in the ExTENd study for alcohol-dependant individuals (Chakaborty and Moodie, 2013), where Cognitive Behavioral Therapy (CBT) can improve the efficacy of Telephonic Monitoring at subsequent stages.

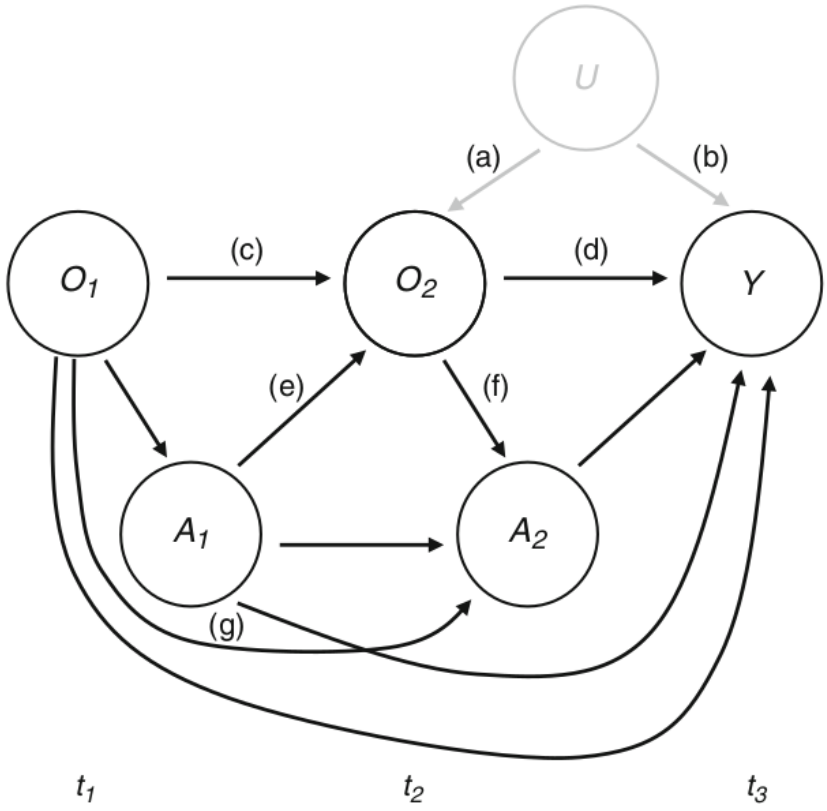


Figure B.2: Simple Causal Dependency Graph Two-Stage Trial (Chakaborty and Moodie, 2013)

B.1 SMART

Real Data Sets. Prevalent examples of SMART in the literature are within Precision Psychiatry and Substance Use Disorder, where the outcome variables are clinician specified quantifications of depressivity or binge-consumption.

For survival outcome data, publically available data are not easy to find. From this paper: https://www.nature.com/articles/s41416-022-02110-z , citations 10-24 are supposedly on SMART oncological trials with results This trial for example: https://clinicaltrials.gov/study/NCT00634725.

SMART. Artman et al. (2018) introduced an efficient methodology for the design of *Sequential, Multiple Assignment, Randomized Trials* (SMART) in order to statistically determine the optimal via Multiple Comparison with the Best (MCB) amongst the standardised mean outcomes of the embedded DTRs (EDTR).

Even modest studies require intentful calculations to properly size and power a SMART towards particular primary and secondary research questions. A research question may resemble the following:

What is the best first-round treatment marginalised over all second treatments?

What is the best three-round treatment regime to optimize QoL-Adjusted Progression-Free Survival according to patient preferences?.

Further information may be sought, to test a wider range of hypotheses, such that we need to consider the necessary sample-sizes, treatment options, and study design to make meaningful comparisons.

$$\mathfrak{B} := \{EDTR_i : \hat{\theta}_i \geq \max_{j \neq i}[\hat{\theta}_j - c_{i,\alpha}\sigma_{i,j}]\}$$

where $c_{i,a}$ depends on the covariance matrix Σ, and α type-1 error rate for excluding the best EDTR from $\mathfrak{B}$.

Correlation Structure. Special care given to the correlation structure amongst the DTRs. *"The correlation structure between EDTR outcomes arises, in part, due to overlapping interventions in distinct DTRs and because patients' treatment histories may be consistent with more than a single DTR."* (Artman et al., 2018).

Missing Data. Special care in trial design as to how to handle non-responders. Careful: The term "non-responder" is often used in the literature for subtly different purposes. Let us specify two near-by ideas in DTR literature: (1) non-response can refer to cases where patients do not show positive health response to the administered treatment. (2) Non-responders to a survey / study can also refer to patients or participants who do not participate, or stop participating at some point in the study. Sometimes patients may drop-out *because* they had no response to treatment, and illness worsens or treatment is not worth the trouble. Patients can drop-out either at first-stage randomization or at follow-up stages, as well as cases where patient visits were skipped. The more flexible nature of SMART may improve study participation as there is alternative potential treatment for patients that may otherwise decide to not continue their partitipcation **citation needed**. Conservative estimates of power and sample size are encouraged to leave room for undesirable trial implementation circumstancesm such as loss to follow up.

B.2 G-Estimation

Moodie et al. (2007) describes the methodologies of Murphy and Robins and shows them to be mathematically equivalent formulations of the same problem, we will refer to both as

G-Estimation. These models are continuations of research in the field of Casual Inference (Robins, 1999). They properly account for the time-dependant confounding and mediation of covariates and treatments, which otherwise lead to biased estimation in the more standard methods of Multivariate Regression and Cox Hazard Model.

Structured Nested Mean Models (SNMM). Expected difference between counterfactual response on specific regime t_{j+1} onward and another specific regime from t_j conditioned on history. The comparison is focussed on deviation of a prescribed treatment from optimal (optimal blip-to-reference function) or deviation from prescribed to standard treatment (blip-to-zero). The structured nature of the model allows estimation of optimal treatment after t_j given treatment at t_j, for any time-point j arbitrary covariate and treatment history up until t_j.

Estimate the optimal DTR by finding the parameters ψ of the optimal blip-to-zero (aka regret) function via g-estimation.

$$U(\psi, s) = \sum_{j=1}^{2} H_j(\psi)\{S_j(A_j) - E[S_j(A_j)|H_j]\}$$

with probability of being treated (i.e. propensity score) $p_j(a_j|h_j;a)$

B.3 Q-Learning

The typical formulation of Q-Learning in the Precision Medicine context models directly the conditional mean of the value function via a series of regression estimators that are solved via backwards induction.

Reinforcement Learning (RL) aims to solve sequential, multi-state decision problems. A simple representation of an RL problem is that an agent traverses a discrete state-space in which each state is assigned a value. RL seeks to analyze the value of various decision-making policies, and typically must balance exploration and exploitation of policy choices. An agent assesses the value of various choices given their current state, and then takes an action, and then recieves some reward, and repeats this process until an end-state is reached. The core theoretical idea of this framework is that an agent must think sequentially, and learn the *quality* of taking certain actions at certain states, this is known as "Q-learning".

In the context of Precision Medicine, we are dealing with finite samples of ppopulation under any given treatment regime, and human data that it is not easy (and in some cases not ethical) to experiment with all possibilities. We may not observe all hypothetical scenaria from the data, and therefore do not know the full probability distribution of the learning environment (Chakaborty and Moodie, 2013).

Extending MSOWL to sequential treatment allocation: If time-points are fixed, things might be ok. But gathering such data, or expecting such stability during model appliction is unreasonable. **Feasibility:** *"Q-learning converges to the optimum action-values with probability 1 so long as all actions are repeatedly sampled in all states and the action-values are represented*

discretely." (Watkins and Dayan, 1989). we need to consider the exclusion of any possible covariate patterns from consideration. so any datum should have non-zero probability of getting either treatment. This is the same as the Positivity Assumption in Causal inference. **Curse of Dimensionality:** RL encounters difficulty in high-dimensional problems (Chakaborty and Moodie, 2013).

Appendix C

Neural Networks

Neural networks offer the possibilitty of uncovering of flexible functional forms that uncover hidden structure and non-linear relationships in complex data. A brief literature review of causal and survival deep learning methods is followed by the introduction of a proposed "deep causal survival" model, and finally a rough-sketch proposal of an Outcome Weighted Learning neural network.

Non-Linearity. What gains in population health outcome might be possible when allowing non-linear treatment effect form to be posited in the ITR estimation procedure. Are they worth the loss of interpretability and inference?

Pattern Recognition approaches like those of support vector machines, random forests, and neural networks, have become popular and for some purposes state-of-the-art techniques. The exponential growth of computation since the second-half of the 21st century allows previously infeasible computation. A major tradeoff between non-parametric and parametric methods, is that of estimators which place fewer assumptions of structure in the model require more data to realise the limit of the estimators capability.

Classical Non-Parametric methods are likely more appropriate for small-sample consistency in our RCT-focussed data setting; Support Vector Machine, K-Nearest Neighbors, Non-Linear Regression, and Trees, for example. These methods typically require more data than parametric methods (as less structure is superimposed by the parametrisation of the model), and do not extrapolate well. But in the right context, they avoid model mispecification, and are of interest.

Despite laudatory optimism for deep learning, serious concerns of mass training data requirement, model complexity and overfitting, make the models irrelevant for many purposes. It is not straight-forward to adapt notions from literature in Image Recognition to Randomized Control Trial data. If we are indeed dealing with tabular data, Boosting and random forests methods have been shown to outpeform neural networks on benchmark, tabular datasets Borisov et al. (2022). Boosting runs the risk overfitting and lack interperetability and structure in expense for their higher predictive scores. Neural networks offer a rich modality of instilling structure in statistical estimation.

Shallow Networks may be able to balance flexibility, interperatability, and data-economy.

Liestøl et al. (1994) show neural networks to match the state-of-art methods in survival analysis. Ba and Caruana (2014) "Do Deep Nets Really Need to be Deep?" show shallow networks matching performance of some sota deep nets. Suggests that shallow model architecture designs have room for improvement. Brigato and Iocchi (2020) Interrogate Deep Learning with small data sample-sizes, reporting competitive results on typical image classification dataset (MNIST) with shallow neural networks, with respect to SoTA models with shallow neural networks, with respect to SoTA models.

Causal Learning structure can be imposed into neural networks, and lots of literature is interested in learning heterogenous treatment effects , particularly from *observational data*.

C.1 Architecture

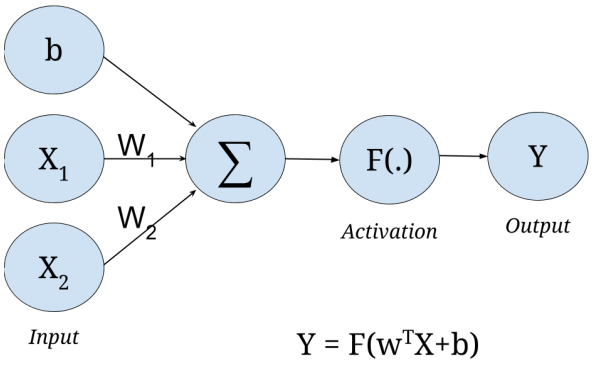


Figure C.1: Neural Regression.

Neurons, the atomic unit of the neural network, have the following components: activation function, weighted data input edges and bias term (Fig. C.1). The activiation function has several possible options, two of which are the Identity function, creating a Linear Regression, and the Logit, creating a Logistic Regression.

Neural Networks are a directed graphs of neurons that allow a phenomenon called "feature representation" to take place; theoretically, an arbitrarily complex function or features can be learned from the data. Weights initially random to widely explore the gradient topography during training and then iteratively updated via stochastic gradient descent (state-of-the-art is ADAM), where prediction error is back-propagated through the neural network. Hyperparameters include learning rate, batch size, training dropout, variable layer depth and width.

Dropout for Predictive Uncertainty as Bayesian Approximation was introduced by Gal and Ghahramani (2016). They employ a form of bayesian approximation by implementing a randomized output-layer dropout, and then repeatedly drawing predictions from the network, thereby exploring a variety of "nearby" networks in Kullback-Leibler space. Fong and Holmes (2019) show the equivalence of marginal likelihood and cumulative leave-p-out-validation.

Posing the bayesian update as a similar model within distance on posterior likelihood with respect to the data. If we consider a robust statistical model to be one that is not surprised by outliers, then it may be advantageous to construct a sphere of nearby models in KL-space. This opens up interesting potential for statistical inference and uncertainty quantification, whereby nearby networks can be thought of as models models within this aforementioned sphere in KL-space. In small networks, though, dropout on the output layer during prediction may not be very feasible because of the lack of breadth per layer / output layer.

C.2 Casual Survival Prediction

Jeong and Jia (2022) is a pre-print proposal of Deep Learning architecture for causal time-to-event prediction. They propose a novel loss function, an augmented mean squared error (MSE) to account for censoring. They also use inverse propensity score weighting to account for selection bias of observational data, and try to correct for the second causal assumption.

$$l(x,y,delta) = MSE + I(delta = 0)\frac{1}{\lambda}MSE \tag{C.1}$$

A code implementation of the Deep Network architecture is provided, but must be adapted to our problem setting, we must assess adjust the model fit accordingly. The authors show experimental simulations of failure time prediction where underlying data is generated from a non-linear function, the square log-risk, and assess various proportions of censoring (10%, 40%, 70%). They perform 1000 simulations with train sample-size of $n \in (100, 500, 1000)$ and test sample-size of $n/2$, a typical order of magnitude magnitude for RCTs in oncology.

C.3 Neural Outcome Weighted Classifier

We propose a novel model that builds off from the classification-perspective literature. Zhao et al. (2015)'s ICO loss function is

Here we would like a shallow neural network, to be somewhat interpretable and to not find try to find rules that are too non-linear. How much do we really gain in terms of mean population survival? Would high value rule from our generalise well to other samples? i.e. not to overfit.

Dropout doesn't really work on a small network with only a few nodes per layer, does it ?

Algorithm 1 Neural Outcome Weighted Learning for Individualized Treatment

Input: $Data = (X, A, Y)\}$;

Output: $\mathscr{M}_{NN.ICO}(A|Data_{train}, w, Data_{new})$

Estimate Censoring Distribution $\hat{\Lambda}$ with Nelson-Aalen Estimator

Set NN Loss function to the ICO Loss function:

$l(X, A, Y, \delta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i Y_i}{\hat{S}_c(Y_i|A_i, X_i)} \frac{I A_i \neq D(X_i)}{\pi(A_i, X_i)}$

Learn ITR by optimizing over the loss function, iteratively reweighting

$\mathscr{M}_{\backslash\backslash\cdot\rangle\rfloor\wr(D,w,hp)} := F(X, A, Y) \rightarrow \hat{A}$

Initialize Weights w randomly

HyperParameter hp (intuition for specification to be discovered)

Batch backpropogation with dropout procedure (optimizer?), data $D = (X, A_{obs}, Y_{obs})$ test/train splits

Predict Counterfactual Outcomes: $Y^*(A = 0, X = x, M = m), Y^*(A = 1, X = x, M_m)$

Estimate Variance of the Estimator via Stochastic Dropout during Inference.

$for\ j \in [1, 100]\}$

$N_{dropout} = p_{dropout} * N_{total}$

$\Delta_{dropout} = \text{rbinom}(N_{dropout}, p_{dropout})$

$y_i j = \mathscr{M}(x, \text{output.layer} * \Delta_{dropout})\ se_{y_i} = \sqrt{\frac{(y_{ij} - \mu(Y_i))^2}{N_{dropout}}}$

Appendix D

Empirical Process Theory

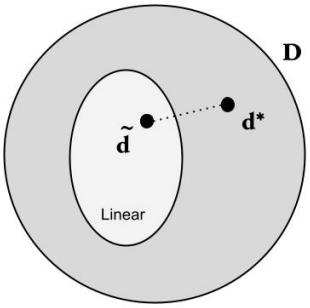


Figure D.1: Optimal linear rule $\tilde{d}$ is a projection of true optimal rule d^* onto linear subspace.

Empirical Processes. Empirical Process Theory analyses probabilistic convergence for empirical realisations of stochastic processes. The standard definition of empirical processes is as follows: Let $X_1, X_2, ..., X_n$ i.i.d. random variables from sample space X and probability measure P, and f a class of measurable function $f : X \to \mathbb{R}$. Let $f(x) = I\{x \leq t\}$ and write $\mathbb{F}_n(t), t \in \mathbb{R}$ as empirical process within class $\mathscr{F} = \{I\{x \leq t\} : t \in \mathbb{R}\}$.

$\{X(t) : t \in T\}$	Stochastic Process
$\{\mathbb{P}_n f : f \in \mathscr{F}\}$	Empirical Process
$\mathbb{P}_n f := \frac{1}{n}\sum_{i=1}^{n} f(D_i)$	Empirical Sample Average
$Pf := \int_D f dP = E\big(f(D)\big)$	Population Expectation

Empirical Cumulative Distribution Function:

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leq t\}, t \in \mathbb{R}$$

True Cum. Distriibution Function:

$$F(t) \equiv P(X \leq t) = P(I\{X \leq t\}) \equiv Pf, t \in \mathbb{R}$$

$\mathbb{F}_n(t)$ is pointwise strongly consistent for $F(t)$, i.e. $\mathbb{F}_n(t) \rightarrow^{as} F(t)$ for each $t \in \mathbb{R}$ by the Central Limit Theorem and the strong Law of Large Numbers. $\mathscr{F}$ is called P-Glivenko-Cantelli if:

$$||\mathbb{P}_n - P||_\infty \rightarrow^{as*} 0 \tag{D.1}$$

meaning that we can show uniform convergenge. In the case where F is also P-Donsker, the limiting process $\mathbb{G}$ of such convergence is a tight mean zero Gaussian process, $l^\infty(F)$ is space of bounded real functions on F,

$$\mathbb{G}_n f = \sqrt{(n)}(\mathbb{P}_n - P)f \tag{D.2}$$

$$\mathbb{G}_n = \sqrt{(n)}(\mathbb{P} - P) \rightsquigarrow 0 \text{ in } l^\infty(F) \tag{D.3}$$

Two conditions are required for weak convergence to tight random process:
C1. For all finite subsets $T_k = \{t_1, ..., t_k\} \subset T$

$$(X_n(t_1), ..., X_n(t_k)) \rightsquigarrow (X(t_1), ..., X(t_k)) \tag{D.4}$$

C2. For each $\varepsilon > 0$

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P^* \Big[\sup_{s,t \in T : \rho(s,t) < \delta} |X_n(s) - X_n(t)| > \varepsilon \Big] = 0 \tag{D.5}$$

If f is square-integrable in expecation $Ef^2(x) < \infty$ for all $f \in \mathscr{F}$ then Condition 1 is satisfied by multivariate CLT. And for the second condition, if we choose to set $\rho(f,g) = \{Var[f(X) - g(X)]\}^{1/2}$ then F becomes totally bounded and satisfies Condition 2.

Appendix E

Generating Survival Regression Data

We discuss the procedure to simulate survival times from cox regression formula and also AFT regression.

E.1 Generating Cox Regression Survival Data

Inverse Transform Sampling is a general method is applicable for sampling from distributions for which their CDF is invertible. This holds true in the case of the survival function. Bender et al. (2005) generate survival times from the cox regression model using this approach; We exposit their procedure for educational purposes in this section. Only a random number generator for the Uniform distribution is needed, which is elementary and universally available in statistical computing software. To generate our values from our distribution of interest, uniformly-randomly distributed values are plugged into the inverse CDF, whose domain is [0,1] by definition. The derivation below shows that we need only invert the Survival function.

$$
\begin{aligned}
S(t|x) &= 1 - F(t) && \text{Survival Function} \\
&= exp(-H(t)) && \text{by defn Eq. 2.7} \\
F(t) &= U \sim Unif(0,1) && \\
U &= 1 - S(t|x) && \\
&= S(t|x) && Unif(0,1) = 1 - Unif(0,1) \\
t &= S^{-1}(U) &&
\end{aligned}
$$

The cox regression is an estimator of the *hazard function*. To generate survival times from a specified Cox Regression, we must transform the expression into the *cumulative hazard func-*

tion and then to the *survival function*, as follows:

$$\begin{aligned}
\hat{h}(t|x) &= h_0(t) \times exp(\beta' x) && \text{Proportional Hazard Model} \\
\hat{H}(t|x) &= \int_0^{\infty} \Big(h_0(t) \times exp(\beta' x) \Big) && \text{Cumulative Hazard} \\
&= H_0(t) \times exp(\beta' x) \\
\hat{S}(t|x) &= exp(-H_0(t) \times exp(\beta' x)) && \text{Survival Function}
\end{aligned}$$

Inverting the Cox Survival estimator $\hat{S}$ yields a cox survival time generating function.

$$\begin{aligned}
U &= exp[-H_0(T) \times exp(\beta' x)] \\
T &= H_0^{-1}[-log(U) \times exp(-\beta' x)]
\end{aligned} \tag{E.1}$$

Baseline Hazard Function specified by Zhao et al. (2015) simulations is $h_0 = 2t$, therefore $H_0 = \int h_0 = \int_0^t 2t = t^2$, and the inverse is $H_0^{-1} = t^{1/2}$. Additionally, I would like to compare results given different baseline hazard functions, although this wouldn't be reflected in cox model parameter estimates as they are agnostic of baseline hazard function.

"The translation of the regression coefficients from hazard to survival time is easy if the baseline hazard function is constant, i.e. the survival times are exponentially distributed. This may be the reason why most simulation studies regarding the Cox model consider only the exponential distribution." according to Bender et al. (2005).

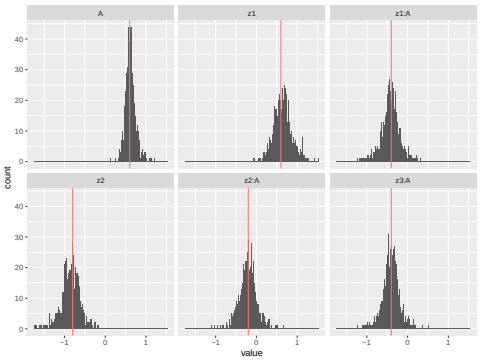
Verfication via simulation of patient survival data given the inverse-transformed Cox survival function (Eq. E.1) above is trivial. Generate survival times according to a specified Cox Regression formula, fit a cox model according to the same specification of terms, and verify that the parameter estimates reflect the specified parameters of our known data generating process. Because there is underlying randomness, repeat this procedure, say, $B = 100$ times, evaluate the mean and variance of parameter estimates. We can see how the model fits in small sample sizes ($N = 10, 50, 100$) for few number of covariates ($P = 1, 2, 3$) when regression formula specifies either full model including interaction terms, main effects only, and interaction-effects only.

We generate survival times according to the procedure above for the following three regression formulae, the "full" model with main and interaction effects, a model with main-terms only, and a model of interaction-terms only:

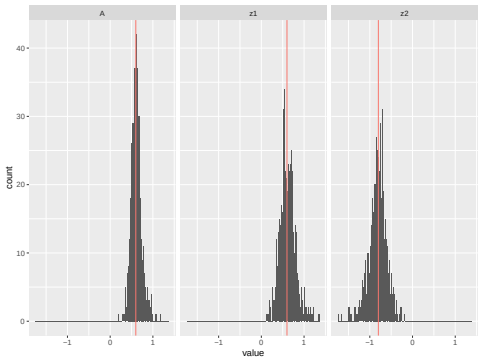
$$\begin{aligned}
full &= 0.6x_1 - 0.8x_2 + (0.6 - 0.4x_1 - 0.2x_2 - 0.4x_3)A \\
main &= 0.6x_1 - 0.8x_2 + 0.6A \\
intx &= (0.4x_1 + 0.2x_2)A
\end{aligned}$$

We then fit cox regressions with the same formulaic specification, to verify that the procedure indeed produces random data that follow the Cox Regression:

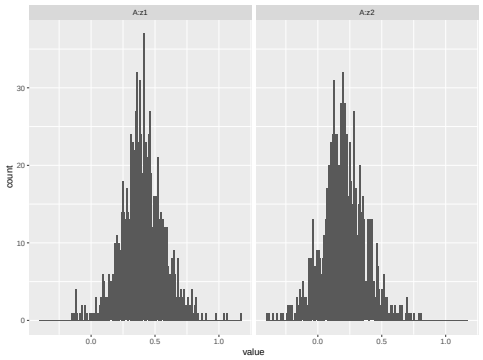
```
Full <- Surv(time=T, status) ~ x1 + x2 + x3 + A + A:x1 + A:x2 + A:x3")
Main <- Surv(time=T, status) ~ x1 + x2 + x3 + A ")
Intx <- Surv(time=T, status) ~ A:x1 + A:x2 + A:x3")
```



(a) Full



(b) Main



(c) Intx

Figure E.1: Parameter Recovery of Cox Regression Model

E.2 Accelerated Failure Time Regression

Leemis et al. (1990) show how to generate survival times from an AFT model.

Kosorok AFT regression formulae for survival times.

$$log(S_3) \sim -0.5 - 0.8x_1 + 0.7x_2 + 0.2x_3 + (0.6 - 0.4x_1 - 0.1x_2 - 0.4x_3)A + \varepsilon$$
$$log(S_4) \sim -0.2 - 0.5x_1 + 0.5x_2 + 0.3x_3 + (0.5 - 0.1x_1 - 0.6x_2 + 0.1x_3)A + \varepsilon$$

If we have generated times correctly from the AFT model, then we should be able to recover them by specifying a simple multiple linear regression model with log-transformed outcome variable.

```
fit = lm(log(time) ~ z1 + z2 + z3 + A + A:z1 + A:z2 + A:z3 , df)
```

In the special case of Weibull does there exist corresponding AFT and COX specifications that can produce the same distribution of survival data.

Scen	model	(Intercept)	z1	z2	z3	A	z1:A	z2:A	z3:A
1	DGF		0.6	-0.8	0	0.6	-0.4	-0.2	-0.4
1	AFT	-0.29 (0.09)	-0.3 (0.1)	0.4 (0.1)	0 (0.1)	-0.3 (0.09)	0.2 (0.1)	0.09 (0.1)	0.2 (0.1)
1	COX	NA (NA)	0.6 (0.16)	-0.81 (0.17)	-0.01 (0.16)	0.61 (0.15)	-0.4 (0.16)	-0.19 (0.16)	-0.41 (0.16)
2	DGF		-1.5	0.5		1	-0.7	-1.2	
2	AFT	-0.29 (0.09)	0.75 (0.1)	-0.26 (0.1)	0 (0.1)	-0.5 (0.09)	0.35 (0.1)	0.6 (0.1)	0 (0.1)
2	COX	NA (NA)	-1.52 (0.17)	0.52 (0.16)	-0.01 (0.16)	1.02 (0.16)	-0.71 (0.17)	-1.22 (0.17)	0 (0.16)
3	DGF	-0.5	-0.8	0.7	0.2	0.6	-0.4	-0.1	-0.4
3	AFT	-0.5 (0.14)	-0.8 (0.16)	0.69 (0.15)	0.21 (0.15)	0.6 (0.14)	-0.4 (0.16)	-0.1 (0.15)	-0.4 (0.16)
3	COX	NA (NA)	0.74 (0.17)	-0.65 (0.17)	-0.2 (0.17)	-0.56 (0.16)	0.35 (0.18)	0.11 (0.17)	0.37 (0.17)
4	DGF	-0.2	-0.5	0.5	0.3	0.5	-0.1	-0.6	0.1
4	AFT	-0.2 (0.14)	-0.5 (0.16)	0.5 (0.16)	0.3 (0.16)	0.5 (0.14)	-0.1 (0.15)	-0.6 (0.16)	0.1 (0.15)
4	COX	NA (NA)	0.47 (0.17)	-0.49 (0.16)	-0.28 (0.17)	-0.47 (0.15)	0.08 (0.17)	0.58 (0.17)	-0.08 (0.16)

Table E.1: Survival Regression Parameter Estimates

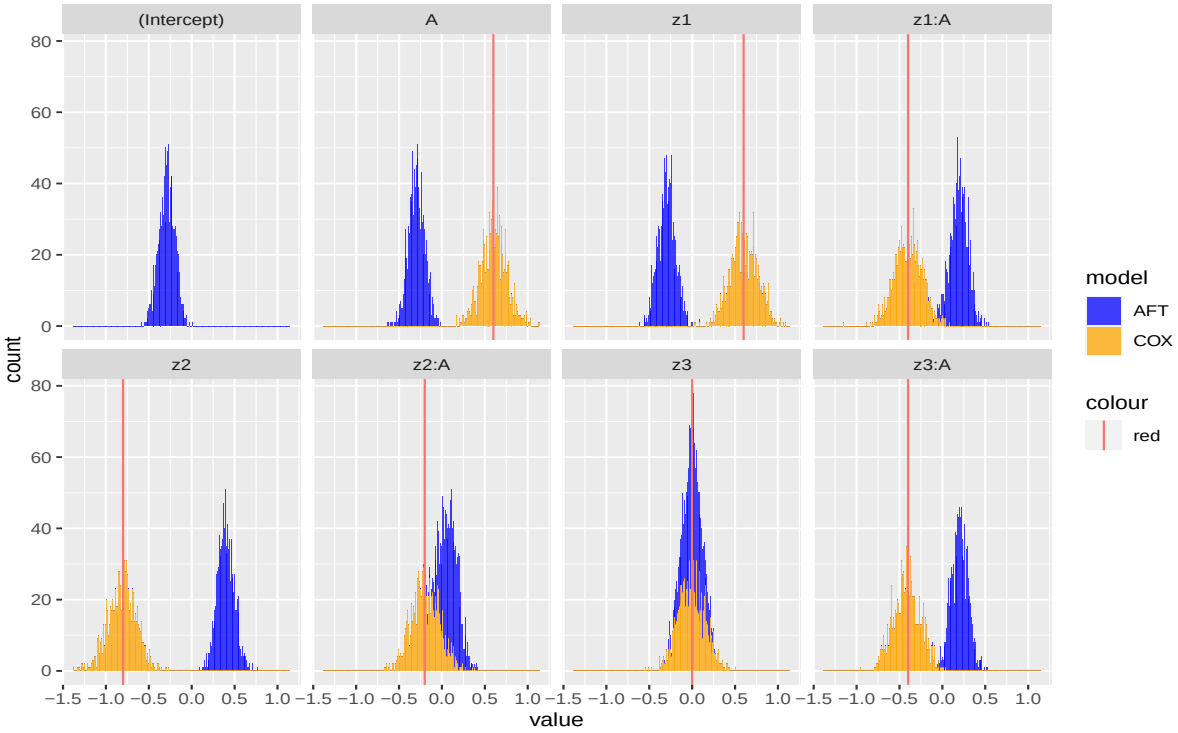


Figure E.2: (Zhao et al. (2015) Survival Scenario 1 (COX)

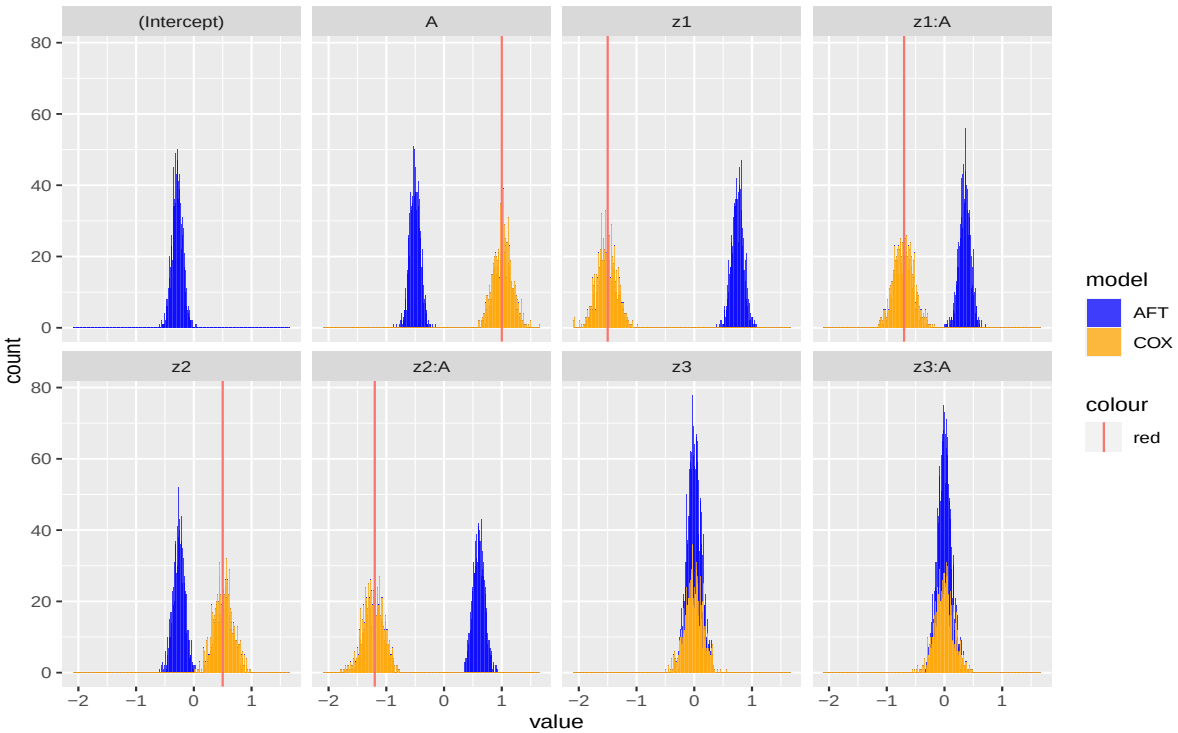


Figure E.3: (Zhao et al. (2015) Survival Scenario 2 (COX)

Parameter Estimates from Cox Fitting of Survival Times simulated from Cox Regression

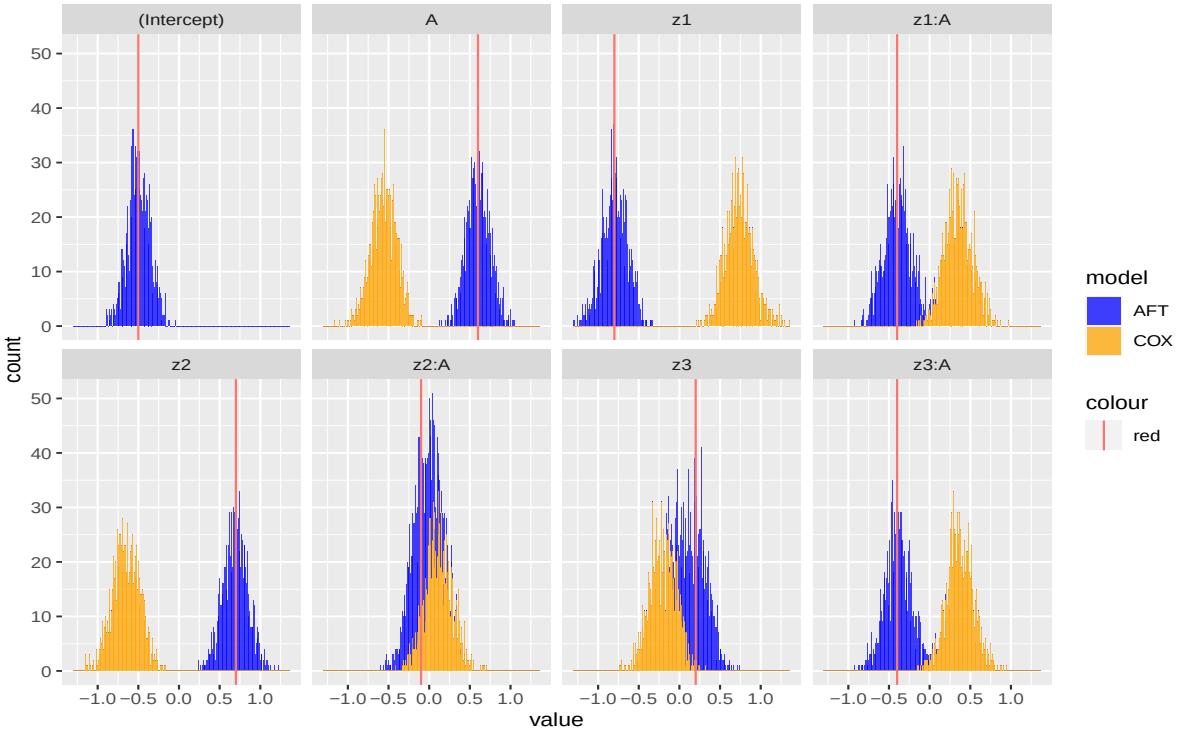


Figure E.4: (Zhao et al. (2015) Survival Scenario 3 (AFT)

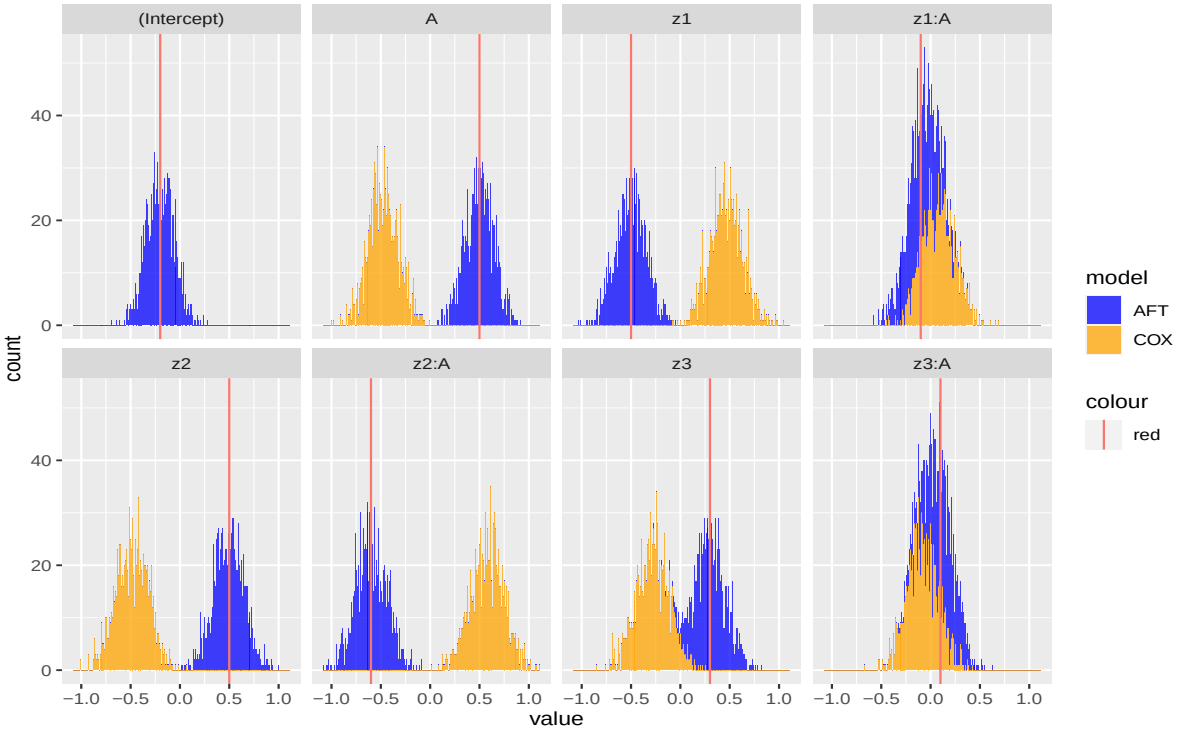


Figure E.5: (Zhao et al. (2015) Survival Scenario 4 (AFT)

	n	cr	A	z1	z2
1	100.00	-1.20	0.62 (0.14)	0.61 (0.21)	-0.86 (0.25)
2	100.00	-0.70	0.63 (0.15)	0.63 (0.24)	-0.82 (0.26)
3	100.00	-0.30	0.63 (0.17)	0.64 (0.26)	-0.85 (0.29)
4	200.00	-1.20	0.59 (0.08)	0.63 (0.15)	-0.8 (0.16)
5	200.00	-0.70	0.62 (0.11)	0.59 (0.17)	-0.79 (0.18)
6	200.00	-0.30	0.61 (0.12)	0.59 (0.18)	-0.8 (0.18)
7	400.00	-1.20	0.6 (0.06)	0.6 (0.1)	-0.81 (0.11)
8	400.00	-0.70	0.6 (0.07)	0.6 (0.12)	-0.79 (0.12)
9	400.00	-0.30	0.6 (0.08)	0.6 (0.13)	-0.79 (0.14)

Table E.2: Simulation from Cox Regression main

	n	cr	A:z1	A:z2
1	100.00	-1.20	0.4 (0.19)	0.19 (0.22)
2	100.00	-0.70	0.42 (0.21)	0.24 (0.23)
3	100.00	-0.30	0.39 (0.25)	0.2 (0.23)
4	200.00	-1.20	0.4 (0.14)	0.19 (0.17)
5	200.00	-0.70	0.4 (0.18)	0.21 (0.16)
6	200.00	-0.30	0.39 (0.17)	0.19 (0.18)
7	400.00	-1.20	0.4 (0.11)	0.2 (0.1)
8	400.00	-0.70	0.4 (0.11)	0.19 (0.11)
9	400.00	-0.30	0.4 (0.12)	0.19 (0.11)

Table E.3: Simulation from Cox Regression intx

Appendix F

Simulation Experiments

We are interested to compare ITR estimators in scenarios involving small sample sizes, high rates of censoring, and non-proportional hazard ratios between treatment groups. We follow a simulation experiment design of Zhao et al. (2015) in which RCT survival data are generated from both Cox and AFT models, and compare the various ITR estimators. Next, in Chapter 5, two data sets from real cancer trials are evaluated.

Estimator	**Paper**	**Description**
ZOM	N/A	Baseline, non-parametric, One-Size-Fits-All
CoxPH	Zhao et al. (2015)	Backwards Induction, Outcome Regression, Semi-Parametric
DR	Zhao et al. (2015)	Direct Value Search, Outcome Weighted Learning
ICO	Zhao et al. (2015)	
MSOWL	Bakoyannis (2023)	

Table F.1: ITR Estimators Under Comparison

F.1 Bakoyannis-Adapted ITR Simulation

Bakoyannis (2023) performs simulation for multi-state survival analysis, of which the states a patient can occupy are: (1) initial disease, (2) tumor recession, (3) disease progression/death, and (4) censored. Each state transition intensity is exponential distributed with rate parameter characterized by a linear model, with exponential or logarithmic terms in some cases. The relative order of transition times then determines the state-space trajectory of a simulated patient. ITR estimation is evaluated on small sample sizes $n \in \{100, 200, 300, 400\}$, with varying rates of censoring $C \sim Exp(exp(\theta)), \theta \in \{-1, -1.2, -1.6\}$.

We adapt the simulation for simple survival data, in which the transition intensity from alive to dead is exponentially distributed similar to the multistate model. Survival Times are simulated from exponential distribution with rate parameter λ, that is defined according to treatment effect function $f^* = f : Z \in \mathbb{R}^{\mathbb{P}} \rightarrow \mathbb{R}$ to $|C| = 4$ cases, two of which define linear

treatment effects, and the other two, non-linear.

$$
\begin{aligned}
&X \sim Uniform(n,-1,1) && \text{Patient Covariates}\\
&A \sim Binomial(n,0.5) && \text{Treatment Assignment}\\
&T \sim Exponential(n,\lambda) = \lambda e^{-\lambda t} && \text{Survival Distribution}\\
&\lambda = exp(-0.5*z1+0.5*z2+A*f^*(Z,C) && \text{Transition Intensity}\\
&f^*(Z) = \begin{cases} -3.5*Z_1+2.5*Z_2 & c=1\\ 2(Z_1-Z_2) & c=2\\ 1+Z_2-exp(-Z_1) & c=3\\ 2log(2-Z_1-Z_2)-1.4 & c=4 \end{cases} && \text{4 Treatment Effect Scenarios}
\end{aligned}
$$

A cox-regression with exponential baseline hazard is equivalent to exponentially distributed survival function. In this experiment, model misspecification from a fitted cox outcome regression model comes non-linear coefficients in the proportional hazards regression formula of the DGF. The cox regression is not considered in the multi-state case, although even in simple survival analysis if terms are non-linearly specified, then the model would not be expected to hold true.

In scenarios 1 & 2, health outcomes are linear transformations of the covariates and treatment. The Proportional Hazards Regression (under unconfoundedness of RCT data) should be correctly specified and thus find the optimal rule of equivalent value to that of MSOWL (SVM with linear kernel). We expect $Var(ITR_{cox}) < Var(ITR_{msowl})$ due to the parametric nature of cox vs non-parametric of SVM.

In scenarios 3 & 4, health outcomes are non-linear transformations of covariates and treatment. We expect MSOWL to uncover the optimal *linear* rule given enough data, whereas the proportional hazards regression should be misspecified and produce an ITR of lesser value. Linearity and Proportional Hazards are critical assumptions of the Cox Proportional Hazards model, the violation of which invalidates model inference .

F.2 Zhao Kosorok 2015 Simulation

Zhao et al. (2015) exemplify OWL superior to Cox Regression when the proportional hazard assumption does not hold, and competitive when it does. We replicate the simulation experiment with a few amendments. We replicate the four survival scenarios across three increasing rates of censoring, in order to evaluate the relative benefit between MSOWL and ICO of utilizing information from censored cases. The more heavily the data are censored, the greater we expect the disparity in value between MSOWL and ICO / Doubly-Robust methods.

Survival DGF In scenarios 1 and 2, survival times are generated from a Cox Proportional Hazards Regression model, with baseline hazard function specified as $h_0(t) = 2t$. In scenarios 3 and 4, survival times are generated from an Accelerated Failure Time (AFT) Model, which

does not obey the proportional-hazard assumption.

$$
\begin{aligned}
S_1 &\sim COX(h_0;\psi), \psi = 0.6x_1 - 0.8x_2 + (0.6 - 0.4x_1 - 0.2x_2 - 0.4x_3)A) \\
S_2 &\sim COX(h_0;\psi), \psi = -1.5x_1 + 0.5x_2 + (1 - 0.7x_1 - 1.2x_2)A \\
S_3 &\sim exp(-0.5 - 0.8x_1 + 0.7x_2 + 0.2x_3 + (0.6 - 0.4x_1 - 0.1x_2 - 0.4x_3)A + \varepsilon) \\
S_4 &\sim exp(-0.2 - 0.5x_1 + 0.5x_2 + 0.3x_3 + (0.5 - 0.1x_1 - 0.6x_2 + 0.1x_3)A + \varepsilon)
\end{aligned}
$$

Censoring Zhao et al. (2015) specify covariate-dependent censoring distributions, generated twice from a Cox regression and twice from an AFT. This showcases the behavior of the OWL models under misspecification, and the advantage in such cases of the Doubly-Robust model, which posits a cox regression model on the censoring distribution. The scenarios generated data that were right-censored at rates of approximately 35–55%.

$$f(C) = \theta e^{-\theta x}, \quad \theta \in exp(-1, -0.5, -0.2)$$

For our purposes, we simplify the censoring data generating function to be independent of covariates and exponentially distributed such that the hazard rate is constant (Eq. **??**). Our generated data scenarios were censored at approximately 25 – 55 % on average.

Sample Size Understanding small sample-size performance is $n_{\text{train}} \in 100, 200, 400$ were assessed by Zhao et al. (2015). ITR value estimation is done via Monte Carlo Integration on a very large, independent sample $N_{\text{test}} = 10,000$, without censoring. The variability of the estimators is assessed over $B = 1000$ repetitions for each combination of sample size, censoring rate, and survival DGF.

The tuning parameter λ was chosen using 5-fold cross validation over a pre-specified grid with the criterion being the empirical pseudo value function. Specifically, for each tuning parameter, we partitioned the training data into 5 parts, each of which serves as the validation set once while the other 4 parts of the data are utilized for estimation. We sum up the empirical pseudo values calculated across the validation sets from the corresponding trained decision rules, and choose the optimal tuning parameter as the one maximizing the summed value.

Computation The main simulation loop we wish to then repeat $B = 1000$ times, for every combination of each of the following hyper-parameters of the parameters $log(\theta_C) \in \{-1, -0.5, -0.2\}$, $n.train \in \{100, 200, 400\}$, and $scenario \in \{1, 2, 3, 4\}$. This is 3x3x1000 = 9000 loops which evaluate a monte Carlo test set of size $n.test = 10000$.

The value function has computational complexity on the order of $O(n^3)$, which starts to be prohibitive to run simulations with data sets larger than 2000. To address this inefficiency, we implement a divide & conquer approach to value computation; randomly splitting the data into samples of 200 and computing the value of each separately, and then averaging the results. 1000 loops take about 1 hour on MacBook M2 processor 9 cores.

Algorithm 2 Treatment Effect Survival Simulation

Loop Hyper-parameters: $\theta_C, n.train, id.scen$

▷ Generate causal survival data

$T \sim rbinom(n, 0.5)$

$X \sim matrix(rnorm(n \times 3, 0, 1), nrow = 3)$

$Y \sim \text{treatment\_effect\_function}(X, T, id.scen)$

$C \sim rexp(n, \exp(\theta_C))$

$Y_{\text{obs}} = \min(Y, C)$

$\Delta = (Y \leq C)$

▷ Create a train and test set from the above script

$n_{\text{train}} \in \{100, 200, 400\}$

$n_{\text{test}} = 10,000$ (no censoring)

▷ Compare estimators

$\text{itr.1.fit} = \text{itr.1.estimator}(X, T, Y)$

$\text{itr.1.val} = \text{value.estimator}(\text{ITR.Val.1}, \text{test.data})$

▷ Repeat for several estimators

return {ITR.val.1, ITR.val.2, ...}

Expected Behavior. From the mathematical definitions and statistical theory of the models explained in Section 4, we expect to observe the following behavior: Estimates from all estimators should exhibit improved efficiency as sample size increases. Cox should underperform the OWL methods when misspecified. MSOWL should outperform ICO when censoring occurs. MSOWL and ICO are mathematically equivalent when there is no censoring. MSOWL should always outperform both ZOM treatment rules.

In all scenarios, a modest outperformance is expected between MSOWL and ICO, which should increase as the prevalence of censoring increases. Increasing sample size is expected to improve to improve performance of all estimators, due to the Law of Large Numbers. In scenarios 1 & 2, survival times generated from Cox regression. The ICO estimator shows noticeable underperformance in the median estimate with respect to MSOWL as censoring rate increases. In Scenario 1, individualized estimators do not exceed the value the ZOM. In scenario 2, MSOWL modestly outperforms all other models, including ZOM, when n=200, but does so decidedly when n=400. With respect to COX, MSOWL matches performance at n=100, and exceeds at n=200. An increased variance in the COX estimator for the AFT-generated data is expected.

When improperly fitted, the variability of the OWL estimators can explode. The Jackknife and Cross-Validation procedures are used to evaluate optimal predictive specification of the learning rate parameter λ and the variability of finite-sample estimates.

F.2.1 Zhao et al., 2015 Plots

Plots from Zhao et al. (2015) ITR simulation experiment, attached here for ease of reference.

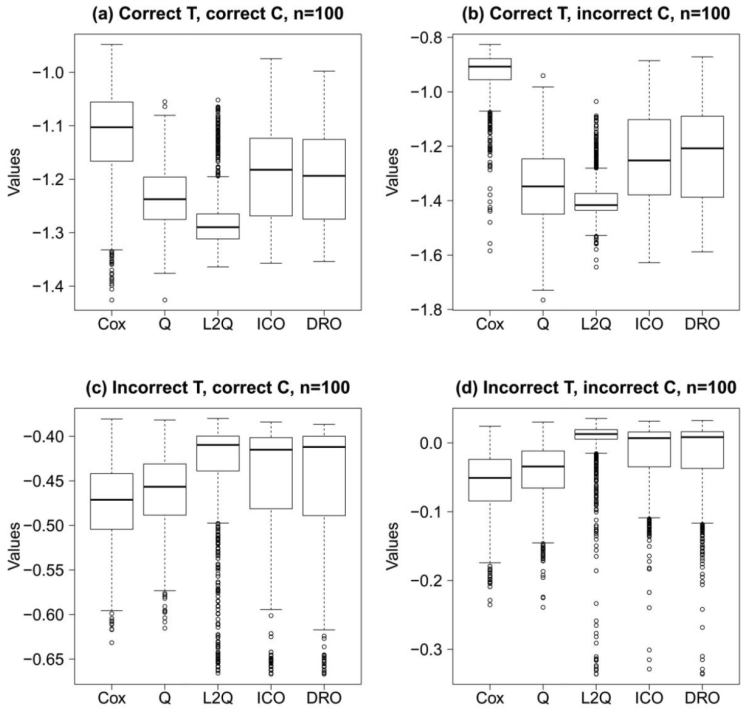


Figure F.1: Zhao et al. (2015) ITR Comparison Box Plot n=100

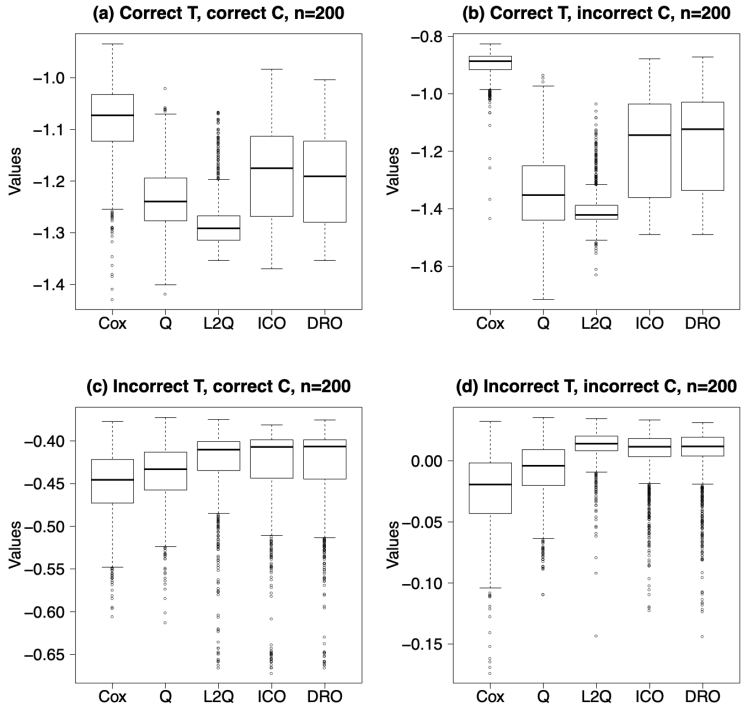


Figure F.2: Zhao et al. (2015) ITR Comparison Box Plot n=200

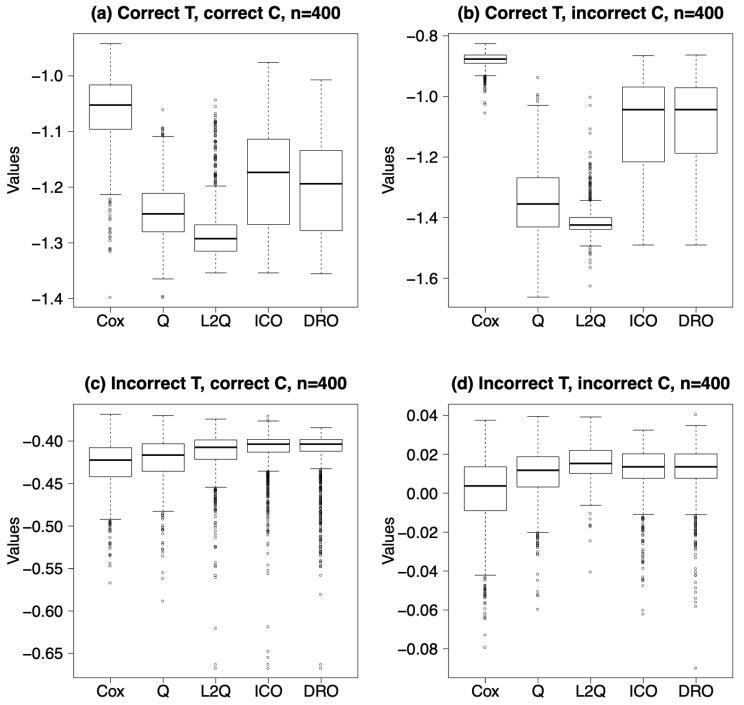


Figure F.3: Zhao et al. (2015) ITR Comparison Box Plot n=400

Appendix G

MSOWL Code

Algorithm 3 MSOWL ITR Estimation

$Wt = \text{reward}(X, Y, \Delta)$
$\hat{\beta}, \hat{\alpha} = \text{wvsm\_solve}(X, A, X_t, k, \sigma, \lambda)$
$\hat{A}_i = \beta_0 + \sum_{j=0}^{n} \left(\alpha_j \beta_j k(X_j, X_j)\right)$
$\hat{V} = \frac{1}{n} \sum Y I(\hat{A} A) \geq 0)$

```
msowl.itr <- function() {
    # Censoring Distr
    c.delta <- 1*(data$s1==data$s2 & data$t2<tau)
    fit <- coxph( as.formula(
    "Surv(time=t1,time2=t2,event=c.delta,type='counting')~1"
    ), data=data)
    # get interesting points in time (event or censoring)
    tt <- sort( unique(c(
        survfit(fit, se.fit = FALSE)$time, # censoring times
        data[data$s1!=data$s2, "t2"]  # observed events
    )))
    tt <- c(0,tt[tt<=tau])
    # zeta_t: for each interesting time point
    event <- matrix(
        sapply( tt, zeta_t, data=data, w=w, fit=fit, S=S, T=T),
        byrow=TRUE, nrow=length(tt)) # cl=num_cores
    # per-individual integration: hazard(t) = events(t) * dm(t)
    dm <- diff(c(tt, tau), lag=1) # dm
    xi <- colSums(event*dm)
    # this just transposes from row to column ...
    xi <- matrix(tapply(X=xi, INDEX = data$id, FUN=sum))
}
```

Algorithm 4 Reward Function

N patients,M events, where $O(M) = O(N)$.

For each event, integrate over the entire observed life-time of each individual we compute hazard across

```
reward <- function(t, data, w=c(0, 1, 0), fit, S, T){
    # estimated hazard function
    hazard <- basehaz(fit, centered=FALSE)
    # per state events before time t
    events <- list()
    for(j in 1:length(S)){
    if(!(S[j] %in% T)){
      # non-terminal state
      events[[j]] <- (data$s1==S[j] & data$t1<=t & data$t2>t)
    } else {
      # terminal state
      events[[j]] <- (data$s2==S[j] & data$t2<=t)
    }
    }
    # weight the events by patient-preference of state
    event <- do.call(cbind, events)%*%w
    # T^t
    Tt <- (1*!(data$s2 %in% T))*t  # non-terminal
    + (1*(data$s2 %in% T))*mapply(min, data$t2, t)  # terminal
    # Inverse Probability of Censoring Weighting
    # Get elem corresponding to maximum t <= "time"
    element <- function(t){
        max(1,max(1:length(hazard$time)*(hazard$time<=t)))
    }
    elements <- sapply(Tt, element)
    G_t <- exp(-hazard$hazard[elements])
    res <- event/G_t
    return(res)
}
```

Appendix H

Additional PRIME Analysis

This chapter contains more exhaustive computational results to supplement the analysis presented in the main text.

PRIME Survival Analysis

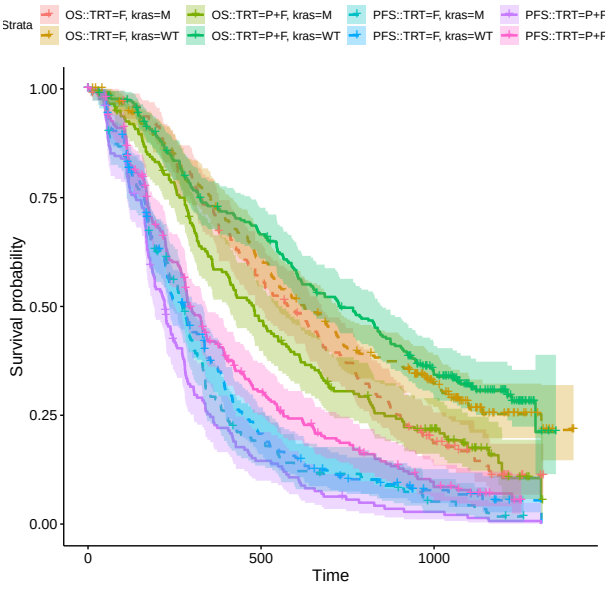
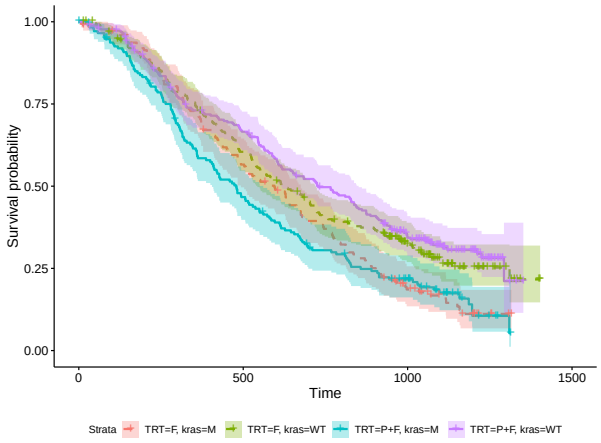
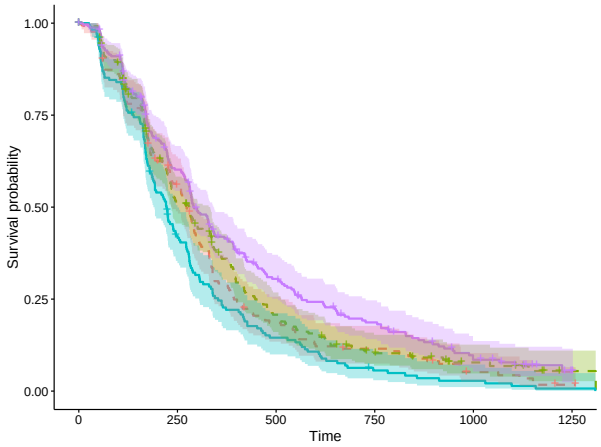


Figure H.1: PRIME: PFS & OS Survival.
Solid Line: Panitumumab + Folfax, Dotted Line: Folfax Only.



(a) Overall Survival (OS)



(b) Progression Free Survival (PFS)

Figure H.2: PRIME: Survival Curves OS & PFS Times by KRAS type

PRIME MSOWL & ICO Tuning

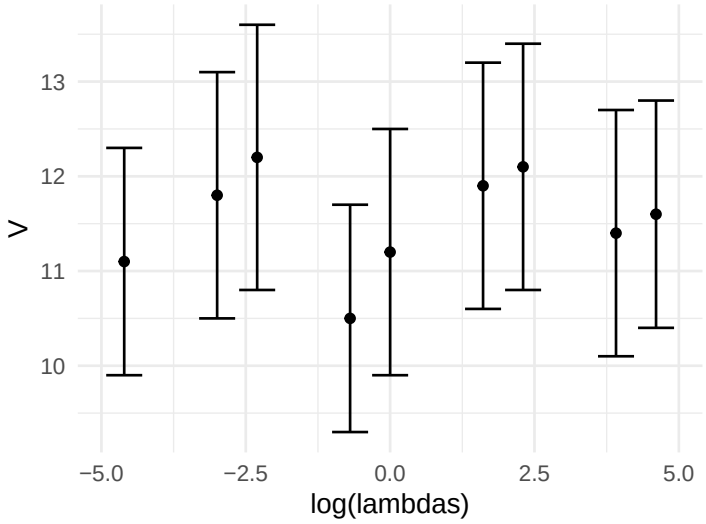


Figure H.3: PRIME RCT PFS: MSOWL 5-Fold Cross Validation

	lambdas	V	SD
1	0.01	11.10	1.20
2	0.05	11.80	1.30
3	0.10	12.20	1.40
4	0.50	10.50	1.20
5	1.00	11.20	1.30
6	5.00	11.90	1.30
7	10.00	12.10	1.30
8	50.00	11.40	1.30
9	100.00	11.60	1.20

Table H.1: RCT PRIME PFS: MSOWL 5-Fold Cross Validation

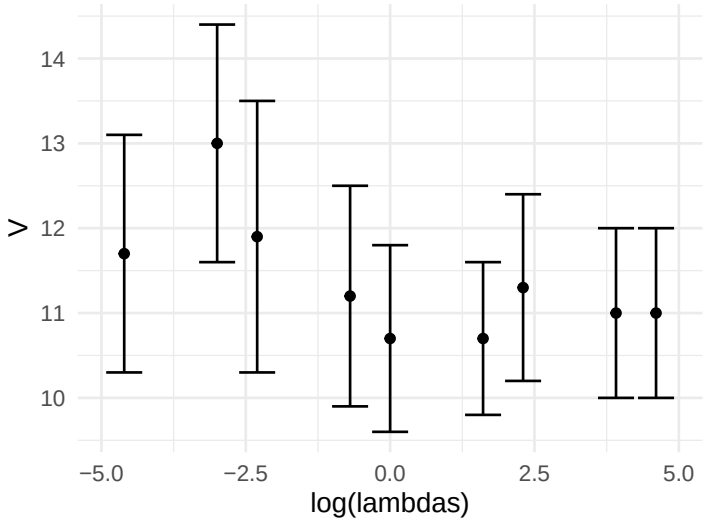


Figure H.4: PRIME RCT PFS: ICO 5-Fold Cross Validation

PRIME Cox Regression Fit

	lambdas	V	SD
1	0.01	11.70	1.40
2	0.05	13.00	1.40
3	0.10	11.90	1.60
4	0.50	11.20	1.30
5	1.00	10.70	1.10
6	5.00	10.70	0.90
7	10.00	11.30	1.10
8	50.00	11.00	1.00
9	100.00	11.00	1.00

Table H.2: RCT PRIME PFS: ICO 5-Fold Cross Validation

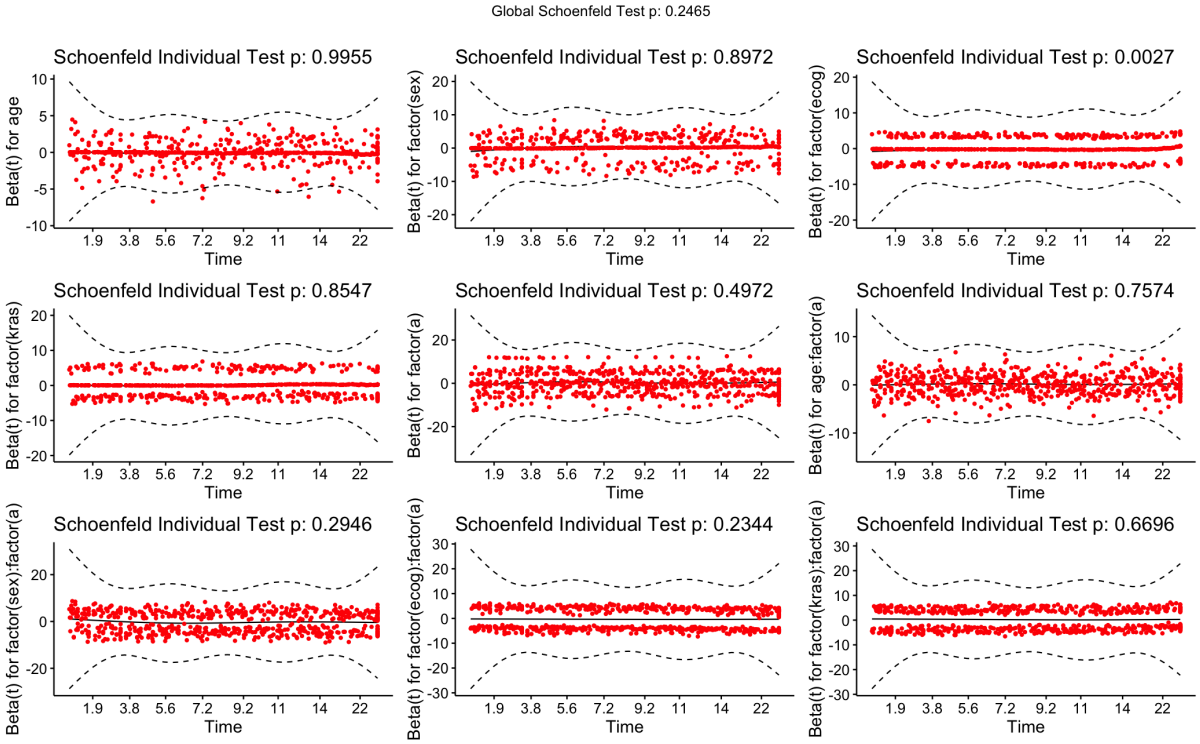


Figure H.5: PRIME: Cox Schoenfeld Residuals

Bibliography

Artman, W. J., Nahum-Shani, I., Wu, T., Mckay, J. R., and Ertefaie, A. (2018). Power analysis in a SMART design: sample size estimation for determining the best embedded dynamic treatment regime. Biostatistics, 21(3):432–448.

Ba, J. and Caruana, R. (2014). Do deep nets really need to be deep? Advances in neural information processing systems, 27.

Bakoyannis, G. (2023). Estimating optimal individualized treatment rules with multistate processes. Biometrics.

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4):962–973.

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

Bickel, P. J. and Doksum, K. A. (1977). Mathematical Statistics: Basic Ideas and Selected Topics. Prentice Hall.

Boghog (2008). 1nql. `https://commons.wikimedia.org/wiki/File:1NQL.png`. Public domain, via Wikimedia Commons.

Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., and Kasneci, G. (2022). Deep neural networks and tabular data: A survey. IEEE transactions on neural networks and learning systems.

Brigato, L. and Iocchi, L. (2020). A Close Look at Deep Learning with Small Data. ieee. arXiv:2003.12843 [cs, stat].

Brown, M. and Green, A. (2015). Quality of adverse event reporting in clinical trials. Clinical Trials, 12(3):234–242.

Butler, E. L., Laber, E. B., Davis, S. M., and Kosorok, M. R. (2018). Incorporating Patient Preferences into Estimation of Optimal Individualized Treatment Rules. Biometrics, 74(1):18–26.

Chakaborty, B. and Moodie, E. E. (2013). Statistical Methods for Dynamic Treatment Regimes. Springer.

Chen, X., Cheng, J. Q., and Xie, M.-g. (2021). Divide-and-conquer methods for big data analysis. arXiv preprint arXiv:2102.10771.

Cho, H., Holloway, S. T., Couper, D. J., and Kosorok, M. R. (2022). Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring.

Ciardiello, F. and Tortora, G. (2008). Egfr antagonists in cancer treatment. The New England Journal of Medicine, 358(11):1160–1174.

Cox, D. and Oakes, D. (1984). Analysis of Survival Data. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Datta, S. and Satten, G. A. (2001). Validity of the aalen–johansen estimators of stage occupation probabilities and nelson–aalen estimators of integrated transition hazards for non-markov models. Statistics & probability letters, 55(4):403–411.

Delgado, A. and Guddati, A. K. (2021). Clinical endpoints in oncology-a primer. American journal of cancer research, 11(4):1121.

Demiris, N. (2010). Introduction to survival analysis. Lecture slides.

Douillard, J.-Y., Siena, S., Cassidy, J., Tabernero, J., Burkes, R., Barugel, M., Humblet, Y., Bodoky, G., Cunningham, D., Jassem, J., et al. (2010). Randomized, phase iii trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (folfox4) versus folfox4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the prime study. Journal of clinical oncology, 28(31):4697–4705.

Efron, B. (1977a). The efficiency of cox's likelihood function for censored data. Journal of the American Statistical Association, 72(359):557–565.

Efron, B. (1977b). The efficiency of the kaplan-meier estimator: An empirical study. Journal of the American Statistical Association, 72(359):557–565.

Efron, B. (1977c). Self-consistency and the kaplan-meier estimator. Journal of the American Statistical Association, 72(359):557–565.

Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. Journal of the American statistical Association, 83(402):414–425.

Fong, E. and Holmes, C. (2019). On the marginal likelihood and cross-validation. arXiv preprint arXiv:1905.08737.

Gal, Y. and Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning.

Gill, R. D. (1984). Understanding cox's regression model: a martingale approach. Journal of the American Statistical Association, 79(386):441–447.

Honvoh, G. D., Cho, H., and Kosorok, M. R. (2022). Model selection for survival individualized treatment rules using the jackknife estimator. BMC medical research methodology, 22(1):328.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47(260):663–685.

Jeong, J.-H. and Jia, Y. (2022). Causaldeepcent: Deep learning for causal prediction of individual event times.

Kalbfleisch, J. D. and Prentice, R. L. (1978). The statistical analysis of failure time data. Journal of the Royal Statistical Society: Series B (Methodological), 40(3):281–297.

Kosorok, M. R. (2019). Introduction to Empirical Processes and Semiparametric Inference. Springer, New York.

Kosorok, M. R. and Moodie, E. E. M. (2015). Dynamic Treatment Regimes: Statistical Methods for Precision Medicine. Society for Industrial and Applied Mathematics (SIAM).

Laber, E. B. and Qian, M. (2018). Generalization error for decision problems. arXiv preprint arXiv:1812.08696.

Law, M. (2015). Introduction to cox regression. MRC Biostatistics Unit.

Leemis, L. M., Shih, L.-H., and Reynertson, K. (1990). Variate generation for accelerated life and proportional hazards models with time dependent covariates. Statistics & probability letters, 10(4):335–339.

Li, X. and Moodie, E. E. M. (2023). Adapting dynamic treatment regimes to real-time data: A bayesian approach. Journal of the American Statistical Association, 118(532):1769–1781.

Liestøl, K., Andersen, P. K., and Andersen, U. (1994). Survival analysis and neural nets. Statistics in Medicine, 13(12):1189–1200.

Lu, W., Zhang, Y., and Li, X. (2013). Estimating optimal treatment strategies with right-censored data. Journal of the American Statistical Association, 108(502):1038–1046.

Lunceford, J. K. and Davidian, M. (2004). Stratified estimation of treatment effect in randomized clinical trials with survival data. Biometrics, 60(4):877–882.

Misiukiewicz, K. and Posner, M. (2013). The spectrum of findings in treatment options for recurrent/metastatic head and neck cancer. Journal of Comparative Effectiveness Research, 2(6):533–535.

Moodie, E. E. M., Richardson, T. S., and Stephens, D. A. (2007). Demystifying optimal dynamic treatment regimes. Biometrics, 63(2):447–455.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(2):331–366.

Neyman, J. and Rubin, D. B. (1978). Statistical inference for causal effects. Journal of the American Statistical Association, 73(364):126–136.

Peeters, M., Siena, S., Van Cutsem, E., Sobrero, A., Hendlisz, A., Cascinu, S., Kalofonos, H., Devercelli, G., Wolf, M., and Amado, R. G. (2009). Association of progression-free survival, overall survival, and patient-reported outcomes by skin toxicity and kras status in patients receiving panitumumab monotherapy. Cancer, 115(7):1544–1554.

Ristl, R., Ballarini, N. M., Götte, H., Schüler, A., Posch, M., and König, F. (2021). Delayed treatment effects, treatment switching and heterogeneous patient populations: How to design and analyze rcts in oncology. Pharmaceutical statistics, 20(1):129–145.

Robins, J. M. (1999). Association, causation, and marginal structural models. Synthese, 121(1-2):151–179.

Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. Journal of the American Statistical Association, 90(429):122–129.

Royston, P. and Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Medical Research Methodology, 13(1):152.

Satten, G. A. and Datta, S. (2001). The kaplan–meier estimator as an inverse-probability-of-censoring weighted average. The American Statistician, 55(3):207–210.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. Statistica Sinica, 14(3):809–834.

Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. The Annals of Statistics, 32(1):135–166.

Van Ness, M., Bosschieter, T., Din, N., Ambrosy, A., Sandhu, A., and Udell, M. (2023). Interpretable survival analysis for heart failure risk prediction. In Machine Learning for Health (ML4H), pages 574–593. PMLR.

Vermorken, J. B., Stöhlmacher-Williams, J., Davidenko, I., Licitra, L., Winquist, E., Villanueva, C., Foa, P., Rottey, S., Skladowski, K., Tahara, M., Pai, V. R., Faivre, S., Blajman, C. R., Forastiere, A. A., Stein, B. N., Oliner, K. S., Pan, Z., and Bach, B. A. (2013). Cisplatin and fluorouracil with or without panitumumab in patients with recurrent or metastatic

squamous-cell carcinoma of the head and neck (SPECTRUM): An open-label phase 3 randomised trial. The Lancet Oncology, 14(8):697–710.

Watkins, C. J. C. H. and Dayan, P. (1989). Q-Learning. Machine Learning.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. (2012). Estimating optimal treatment regimes from a classification perspective: Treatment regimes and classification. Stat, 1(1):103–114.

Zhang, M. J. and Klein, J. P. (1995). Delay in effect of immunotherapy: The proportional hazards model. Statistics in Medicine, 14(24):2645–2657.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association, 107(499):1106–1118.

Zhao, Y.-Q., Laber, E. B., Ning, Y., Saha, S., and Sands, B. E. (2019). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. Journal of Machine Learning Research, 20(48):1–23.

Zhao, Y. Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2015). Doubly robust learning for estimating individualized treatment with censored data. Biometrika, 102(1):151–168.

Zhong, C. and Tibshirani, R. (2019). Survival analysis as a classification problem. arXiv preprint arXiv:1909.11171.