Single-stage Constrained Optimal Treatment Regimes chap-one Introduction Precision medicine tailors medical treatments to each patient's own characteristics. It categorize individuals into subpopulations based on, for example, their response to a specific treatment, or their susceptibility to a certain disease, etc. Hence, it targets therapeutic or preventive interventions to those who may benefit, and save those who may not benefit from unnecessary side effects and costs. Given a patient state, such as genetic information, demographics, results of diagnostic test, and so on, dynamic treatment regimes determine what treatment should be assigned next. These are data-driven decision rules that map patient characteristics to recommended treatments.

There is a rich body of research on estimating optimal treatment regimes using data from randomized clinical trials or observational studies. In most cases, a dynamic treatment regime is defined to be optimal if it maximizes the expected value of a certain cumulative clinical outcome when applied to a population of interest. Methods to estimate an optimal treatment regime include Q-learning Nahum2012, penalized Q-learning Song2011, interactive Q-learning Linn2014, A-learning Schulte2014, regret-regression henderson2010, g-estimation gestimation, and policy search methods Zhao2012,Zhao2015,Zhang2012,Zhang2012b,Orellana2010a,Zha
However, these estimators seek to maximize the expectation of a single scalar outcome, and therefore, neglect the clinical need to balance several competing outcomes. For example, a clinician may have to balance treatment effectiveness, side-effect burden, and cost while developing a treatment strategy for a patient with a chronic disease; or maximize the expected time to an adverse event while controlling the variance of the time to the adverse event.

Despite its practical importance, very little work has been done on handling multiple competing outcomes. Lizotte et al. considered linear combinations of two competing outcomes indexed by a trade-off parameter and compute the optimal treatment regime for all combination Lizotte2010. However, it may not be realistic to assume that a linear trade-off is sufficient to describe all possible patient preferences LaberTwo2014. Wang et al. used a compound score or "expert score" by numerically combining information on treatment efficacy, toxicity, and the risk of disease progression Wang2012. Unfortunately, the elicitation of a good composite outcome can be difficult and the misspecification of a composite outcome may severely affect the quality of the estimated treatment regime Laber2014. There are also some methods to avoid formation of composite outcomes. Laber et al. proposed set-valued dynamic treatment regimes LaberTwo2014. This method inputs current patient information and outputs a set of recommended treatments. This set contains multiple treatments unless there exists a treatment that is best across all outcomes. This method may not be able to recommend a single treatment and needs expertise for tie breaking when a set of several treatments are recommended. Also, it needs to specify "clinically significant differences" for competing outcomes. Linn at el. proposed constrained interactive Q-learning algorithm Linn2014a, which provides an algorithm to find the optimal regime under constraints in the two-stage setting.

We propose a new statistical framework to tackle the problem of balancing multiple competing outcomes using constrained estimation. By constraining the values of secondary outcomes, we search for the optimal feasible regimes for the primary outcome, there by finding constrained optimal regimes. This type of framework is useful in scenarios such as where the clinicians desire to find a treatment strategy that maximize the effectiveness of a treatment regime while controls the side-effect burden and cost. In this chapter, we consider the single-stage scenario. The constrained optimal regime estimator is developed and demonstrated through simulations. Its consistency and asymptotic normality are proven. For demonstration, data from single-stage randomized trials are assumed. Observational data also fit in our framework provided additional assumptions about the treatment assignment mechanism are reasonable, specifically the no unmeasured confounder assumptions. However, data from observational studies should be used with caution, as the no unmeasured confounder assumption is often unverified Chakraborty2013.

Methodology Define single-stage constrained optimal regimes Dataset There is only one decision point in the single stage setting. The data from a randomized trial are denoted as

$$\{(X^i, A^i, Y^i)\}_{i=1}^n,$$

consisting of n identically, independently distributed trajectories of (X, A, Y), whose distribution are often unknown. Capital letters, X, A, Y, are used to denote the random variables; lower case letters x, a, y to denote realized values of these random variables. $X \in X$ represents the patient information collected up to the decision point, where $X \subseteq R^p$ is the support of X. $A \in A$ represents the treatment assignment, where

$A = \{1, 2, \cdots, m\}$ is the set of all possible treatments. The vector variable $Y \in R^J$ denotes the outcomes of interest. Let Y_1, the first component of Y, be the primary outcome of interest. It is coded so that higher values are desirable. Meanwhile, $Y_2, \cdots, Y_J$ are the secondary outcomes of interest, coded so that the lower values are better.

Potential outcome framework To identify the causal effect of a certain regime, we take on the potential outcome or counter-factual framework established by Neyman, Rubin and Robins for assessing treatment effects from either randomized or observational studies Neyman,Rubin2005, Rubin1980, Robins1997, Hernan2006. The set of potential outcomes is $W^* = \{Y^*(a), foralla \in A\}$, where $Y^*(a)$ is the vector-valued outcome that would have been observed if the subject was assigned treatment a. The assumptions made in this framework are as follows. itemize

A1. Consistency:

$$Y = Y^*(A).$$

This means that actual observed outcome vector Y for an individual who received treatment A is the same as the potential outcome for that individual assigned with the same treatment, regardless of the experimental conditions used to assign treatment. It also implies that there is no interference among individuals Rubin1980.

A2. No unmeasured confounders:

$$W^* A \mid X.$$

This means that the set of potential outcomes, $\{Y^*(a), foralla \in A\}$, are conditionally independent of treatment assignment A given patient information X. In randomized study, this condition is satisfied by construction in randomized studies. However, it can not be verified in observational studies Robins1997.

A3. Positivity assumption: There exists $\epsilon > 0$, so that

$$Pr(A = a \mid X) > \epsilon, foralla \in A$$

with probability one Hernan2006. This ensures that there is a positive probability of receiving every possible treatment assignment for every value of patient covariates in the population. This assumption is satisfied in well-designed randomized studies. It can also be empirically verified in observational studies. Yet, if it is violated, estimating of regimes for certain subsets of patients can be impossible.