Dynamic Treatment Regimes

MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

Editors: F. Bunea, R. Henderson, N. Keiding, L. Levina, R. Smith, W. Wong

Recently Published Titles

Asymptotic Analysis of Mixed Effects Models
Theory, Applications, and Open Problems
Jiming Jiang 155

Missing and Modified Data in Nonparametric Estimation With R Examples
Sam Efromovich 156

Probabilistic Foundations of Statistical Network Analysis
Harry Crane 157

Multistate Models for the Analysis of Life History Data
Richard J. Cook and Jerald F. Lawless 158

Nonparametric Models for Longitudinal Data
with Implementation in R
Colin O. Wu and Xin Tian 159

Multivariate Kernel Smoothing and Its Applications
José E. Chacón and Tarn Duong 160

Sufficient Dimension Reduction
Methods and Applications with R
Bing Li 161

Large Covariance and Autocovariance Matrices
Arup Bose and Monika Bhattacharjee 162

The Statistical Analysis of Multivariate Failure Time Data: A Marginal Modeling Approach
Ross L. Prentice and Shanshan Zhao 163

Dynamic Treatment Regimes
Statistical Methods for Precision Medicine
Anastasios A. Tsiatis, Marie Davidian, Shannon T. Holloway, and Eric B. Laber 164

For more information about this series please visit: *http://crcpress.com/go/monographs*

Dynamic Treatment Regimes

Statistical Methods for Precision Medicine

Anastasios A. Tsiatis
Marie Davidian
Shannon T. Holloway
Eric B. Laber

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-1-4987-6977-8 (Hardback)

Library of Congress Cataloging-in-Publication Data

Names: Tsiatis, Anastasios A. (Anastasios Athanasios), author.
Title: Dynamic treatment regimes : statistical methods for precision medicine / Anastasios A. Tsiatis, Marie Davidian, Shannon T. Holloway, and Eric B. Laber.
Description: Boca Raton : Chapman and Hall/CRC, 2020. | Series: Chapman & Hall/CRC monographs on statistics and applied probability | Includes bibliographical references and index. | Summary: "Precision medicine seeks to use data to construct principled, i.e., evidence-based, treatment strategies that dictate where, when, and to whom treatment should be applied. This book provides an accessible yet comprehensive introduction to statistical methodology for dynamic treatment regimes"-- Provided by publisher.
Identifiers: LCCN 2019038384 (print) | LCCN 2019038385 (ebook) | ISBN 9781498769778 (hardback) | ISBN 9780429192692 (ebook)
Subjects: LCSH: Medical statistics. | Medical records--Data processing.
Classification: LCC RA409 .T75 2020 (print) | LCC RA409 (ebook) | DDC 610.2/1--dc23
LC record available at https://lccn.loc.gov/2019038384
LC ebook record available at https://lccn.loc.gov/2019038385

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To my mother, Anna, and my son, Greg
– A.A.T.

To my mother, Dorothea, and my brother, John
– M.D.

To my mother, Carin, my husband, Jim, and my grandson, Lukas
– S.T.H.

To my wife, Sheela, and our stochastic processes, Madeline and Juliet
– E.B.L.

Contents

Preface

Treatment of a chronic disease or disorder almost always involves a series of decisions. For example, treatment of cancer involves a succession of decisions at key milestones in the disease progression; for example, selection of a first-line chemotherapy at the time of diagnosis, of a maintenance regimen for a patient who responds to first-line therapy or of a second-line/salvage chemotherapy for a patient who does not, and of additional intervention in the event of recurrence. Similarly, management of a behavioral or mental health disorder such as substance abuse or depression requires a series of decisions in which the clinician may start, stop, maintain, modify, or adjust interventions on the basis of a patient's response and other characteristics.

A dynamic treatment regime is a set of sequential decision rules, each corresponding to a key decision point in the disease or disorder process. Each rule maps information on a patient accrued to that point to a set of feasible treatment options, so basing selection of treatment at each decision point on a patient's baseline and evolving characteristics. Thus, a dynamic treatment regime formalizes the process by which a clinician synthesizes patient information to select treatment options in practice.

The emphasis on precision medicine, which involves tailoring treatment to a patient's characteristics in an evidence-based manner, has led to an increasing focus on statistical methodology for the discovery of dynamic treatment regimes from data. In particular, there is considerable interest in estimation, from suitable data, of an optimal dynamic treatment regime, one that, if used to select treatment for the patient population, would lead to the most beneficial outcome on average. The foundations of statistical methodology for this enterprise were pioneered by James Robins through several seminal publications starting in the mid-1980s, and fundamental advances on characterization and estimation of an optimal treatment regime by Susan Murphy and Robins in the mid-2000s formed the basis for a growing subsequent body of work by these researchers and others in the late 2000s and early 2010s.

Since that time, the literature on statistical methodology for dynamic treatment regimes has experienced a veritable explosion as the goal of precision medicine has become a central pillar of health sciences research. In addition, there are parallel developments on methodology for sequen-

tial decision making in different contexts in computer science and other disciplines. As a consequence, there is a vast and expanding literature relevant to discovery of dynamic treatment regimes from data, in which different terminology, notation, and perspectives abound, making this topic difficult to approach for the first time.

Our goal in writing this monograph is to address this challenge by presenting a unified and systematic introduction to methodology for dynamic treatment regimes. We do not intend for the book to be an exhaustive account of methodology in this area, however. Rather, we hope that the book will provide readers with foundational knowledge and a strong basis for studying the broader literature, including advances that post-date the book's publication. Accordingly, we present fundamental statistical frameworks along with selected methods that lay the groundwork for further study of this topic. Our ultimate objective is to enhance awareness of and appreciation for this body of work and its critical importance in the treatment of chronic diseases and disorders.

Our target audience is researchers and graduate students in statistics and related quantitative disciplines who are familiar with probability and statistical inference and popular statistical modeling approaches but have no prior exposure to dynamic treatment regimes or other relevant topics, such as causal inference. As discussed in the outline given in Section 1.5, the book includes both foundational and more advanced material. Thus, a reader can study selected portions of the book or the book in its entirety depending on his or her goals and background. Reading plans suited to those who seek a broad introduction to the foundations and key methods in this area and to those who desire in addition a deep, comprehensive understanding of the theoretical underpinnings are presented in Section 1.5.

As the focus of this book is on statistical methodology and its theoretical foundations for this audience, the book is likely not suitable for domain science investigators and other practitioners whose primary goal is to understand the methods at an applied level with an eye toward implementation in practice. We refer such readers to articles specifically designed for this purpose, such as Collins et al. (2004), Murphy et al. (2007a), Murphy et al. (2007b), Lei et al. (2011), Nahum-Shani et al. (2012b), Almirall et al. (2012a), Nahum-Shani et al. (2012a), Almirall et al. (2014), Kidwell (2014), and Kelleher et al. (2017).

The book is an outgrowth of notes developed by one of us (Tsiatis) for a PhD-level special topics course taught in the North Carolina State University (NC State) Department of Statistics over the past decade. We started with these notes as a foundation and have refined, expanded, and added to them to develop the book in its current form. As reviewed in Section 1.5, a subset of the material in the book is suitable as the basis for an introductory PhD-level course on this topic. We are indebted to

the Statistical and Applied Mathematical Sciences Institute (SAMSI) in Research Triangle Park, North Carolina, for hosting a PhD course taught by three of us (Davidian, Holloway, Laber) in spring 2019 as part of its Program on Statistical, Mathematical, and Computational Methods for Precision Medicine, which provided us with invaluable feedback from students at NC State, Duke University, and the University of North Carolina at Chapel Hill (UNC-CH).

One of us (Holloway) is the developer of a comprehensive R package, DynTxRegime, that implements a number of the methods for estimation of optimal dynamic treatment regimes from data reviewed in this book. The package is meant to be a "one-stop-shop" for methodology for dynamic treatment regimes and is available from the Comprehensive R Archive Network (CRAN). We gratefully acknowledge National Cancer Institute program project grant P01 CA142538, awarded to a consortium of NC State, Duke University, and UNC-CH, which has supported not only the development of this package but the efforts of the authors to conceive and complete this monograph.

We deliberately do not include in the book static accounts of application of methods covered in each chapter. Instead, detailed demonstrations of application of selected methods are presented on a dedicated, publicly accessible website, `http://dtr-book.com`. Most of these applications make use of the DynTxRegime package and are meant to assist a reader who has studied the methods in detail with their implementation and with gaining proficiency with the package. We intend for the website to be "dynamic," with periodic updates and modifications as the package and methods evolve. We also hope to post supplemental materials and other resources and encourage readers to check the website often.

We offer our profound thanks to John Kimmel, Executive Editor, Statistics, at Chapman & Hall/CRC Press of Taylor & Francis for encouraging us to take on this project and for his infinite patience with our frequent failure to meet our own self-imposed deadlines. Two of us (Tsiatis and Davidian) have had the privilege of working with John on previous books and are delighted to have had the opportunity to benefit from his extensive experience and guidance again. We also thank the several reviewers of the book-in-progress for their suggestions and feedback. We are grateful to Lisa Wong for creating the cover art. Finally, we thank our families, friends, and colleagues for their support.

Anastasios A. (Butch) Tsiatis, Marie Davidian,
Shannon T. Holloway, Eric B. Laber

Raleigh, North Carolina
September 2019

Chapter 1

Introduction

1.1 What Is a Dynamic Treatment Regime?

In the context of treatment of a chronic disease or disorder, a *dynamic treatment regime* is a set of sequential decision rules, each corresponding to a key point in the disease or disorder progression at which a decision on the next treatment action for a patient must be made. Each rule takes as input information on the patient to that point and returns the treatment he/she should receive from among the available, feasible options. A dynamic treatment regime thus formalizes the process by which a clinician treating a patient synthesizes information and selects treatments in practice. Dynamic treatment regimes are also referred to as *adaptive treatment strategies* or *adaptive interventions*, notably in the literature on treatment of mental health and behavioral disorders.

Precision medicine focuses on tailoring treatment decisions to a patient's characteristics and the incorporation of evidence in guiding these decisions. Dynamic treatment regimes thus provide a formal, principled framework for this enterprise. An optimal dynamic treatment regime can be defined as one that, if used to select treatment actions for the patient population, would lead to the most favorable outcome on average. Accordingly, formulation of dynamic treatment regimes and methodology for their development and evaluation based on data are of considerable interest to a growing community of clinical and intervention scientists wishing to develop optimal regimes for precision medicine and quantitative researchers seeking tools to support them in these efforts.

Accounts of methodological developments for dynamic treatment regimes are scattered across a vast literature in statistics, computer science, and medical decision making. The resulting differences in notation and terminology and the complex concepts involved can make this important topic difficult to approach. The purpose of this monograph is to provide a unified, systematic introduction to statistical methods for dynamic treatment regimes. Our focus is on their use in the health sciences; namely, for guiding treatment, prevention, and diagnosis of a disease or disorder. However, the ideas and concepts are relevant in other settings in which sequential decisions on interventions or policies must be made,

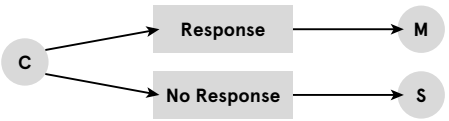

Figure 1.1 *Schematic depiction of two key decision points in treatment of a acute leukemia. At Decision 1, an induction chemotherapy* C *is selected. At Decision 2, for a patient for whom a response is induced, a maintenance therapy* M *is selected; for a patient for whom no response is induced, a salvage therapy* S *is selected.*

as in education, engineering systems, economics and finance, marketing, and resource management.

1.2 Motivating Examples

We begin by considering three applications that illustrate and motivate subsequent developments and to which we refer in later chapters.

1.2.1 Treatment of Acute Leukemias

Figure 1.1 presents schematically two key decision points in the treatment of acute leukemias, a setting to which we refer in subsequent chapters as a running example. When a patient diagnosed with a particular type and stage of the disease presents, the first decision a clinician faces is to select a chemotherapeutic regimen meant to induce a positive response, such a partial or complete remission. Assume that at Decision 1 there are two available induction therapy options, denoted as C_1 and C_2. Typically, following a cycle of induction therapy, a bone marrow biopsy is performed to assess whether or not the patient has achieved the desired response. If so, the patient is deemed a responder at that point; if not, the patient receives a second cycle of induction therapy, after which response status is again assessed. Thus, the desired response might be achieved sooner for some patients than others or not at all. The second key decision confronting the clinician is to determine the next step of treatment, which is dictated by the patient's response status. If the patient has achieved a satisfactory response, Decision 2 involves selecting a maintenance treatment, the goal of which is to sustain the response. If not, at Decision 2, a second line or salvage therapy must be chosen. Suppose there are two maintenance options, M_1 and M_2, and two salvage options, S_1 and S_2.

The goal of the clinician is thus to select an induction therapy from the set of available treatment options $\{C_1,C_2\}$ at Decision 1 and then a maintenance or salvage option from the set of all available treatment options $\{M_1,M_2,S_1,S_2\}$ at Decision 2, where, clearly, only maintenance options are feasible for patients who respond and only salvage options are feasible for those who do not. The clinician seeks to make these decisions so as to maximize the expected benefit to the patient with respect to a health outcome such as disease-free or overall survival time. In making these decisions in practice, the clinician uses his or her expert judgment to take account of accrued information on the patient available at each decision point, where this information may include demographics, prior medical history, genetic and genomic characteristics, initial and evolving physiologic and clinical variables, occurrence and timing of adverse reactions to induction therapy, and so on.

If attention is restricted to these two decision points, then a corresponding dynamic treatment regime comprises two *decision rules*. The first rule, associated with Decision 1, takes as input all information available on the patient at the time he or she presents (baseline) and returns an induction option selected from $\{C_1,C_2\}$. The second rule, corresponding to Decision 2, takes as input all of the information available at Decision 1 plus additional information evolving in the intervening period between Decisions 1 and 2, including response status, and returns a maintenance or salvage option as appropriate from $\{M_1,M_2,S_1,S_2\}$. Shortly, we introduce a mathematical representation of such rules and regimes and in subsequent chapters use it to characterize formally the clinician's goal of maximizing expected benefit to the patient as the problem of seeking an *optimal* set of such rules; that is, an optimal dynamic treatment regime.

Single and Multiple Decision Regimes

This situation exemplifies what is referred to as a *multiple decision* or *multistage* problem, where several, in this case two, sequential decision points can be identified; and selection of treatment over the entire sequence is of interest. In a multiple decision problem, dynamic treatment regimes thus comprise multiple decision rules, and the focus is on an outcome, in this case, survival time, that may be ascertained subsequent to the final decision.

In a *single decision* or *single stage* problem, a single decision point is identified. In this example, if the focus is on the role of induction chemotherapy in inducing remission (the desired response), then it is natural to restrict attention to Decision 1, where response is now the outcome of interest. Here, a dynamic treatment regime consists of a

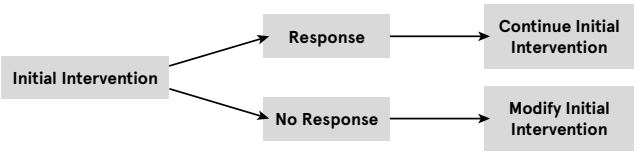

Figure 1.2 *Schematic depiction of two key decision points in treatment of attention deficit hyperactivity disorder. At Decision 1, an initial intervention is selected. At Decision 2, for a child who responds to his initial intervention, there is a single option, to continue the initial intervention. For a child who does not respond, the initial intervention either can be augmented by adding the other intervention or intensified by increasing the dose.*

single rule taking all information available at the time a patient presents as input and returning an option in $\{\mathsf{C}_1,\mathsf{C}_2\}$.

It is conventional in much of clinical chronic disease research, and from a regulatory perspective, to focus on a single decision point in isolation, even if it is recognized that subsequent decisions are required in the intervening period between that decision point and ascertainment of the outcome. For example, interest may be in the choice of induction therapy at Decision 1 as it relates to survival outcome. It is of course possible to conceive of single decision dynamic treatment regimes in this situation, as above. The outcome that results from treatment selection according to the single, Decision 1 rule will also be affected by the inevitable subsequent decisions that will be made (without reference to further, formal decision rules). We defer discussion of this point to later chapters; for now, we emphasize only that, depending on the setting, single or multiple decision dynamic treatment regimes may be of interest.

1.2.2 Interventions for Children with ADHD

Figure 1.2 depicts two decision points in the treatment of children with attention deficit hyperactivity disorder (ADHD). For a child diagnosed with ADHD, Decision 1 involves selection of an initial intervention, where here we consider two options, low dose medication M or low intensity behavioral modification therapy B, which will be administered ideally for twelve weeks. During the next twelve weeks, the child is evaluated periodically for response status, which is based on a clinical assessment of ADHD severity. A child who shows satisfactory response over the entire period can be considered a responder to the initial treatment, and there is no reason to alter that treatment at the end of the twelve

weeks. Thus, the second decision for the clinician, at twelve weeks, involves choosing the single option C, continue the child on the initial intervention. Alternatively, if at any assessment during the twelve weeks the child does not or ceases to show a response, it is not beneficial to continue the initial treatment, so a second decision must be made by the clinician at that point, which involves two options. If the child is on medication, the clinician can choose to increase the dose, thereby intensifying the treatment; or she can maintain the dose but also place the child in behavior modification therapy, so augmenting the initial intervention. Likewise, for a child initially receiving behavioral therapy, the clinician can increase the intensity of the behavioral therapy or continue the low intensity therapy but augment it by prescribing medication. Thus, for a child with negative response status, the two options at Decision 2 are to intensify, I, or augment, A, the initial intervention. As in the leukemia example, the timing of Decision 2 can be different for different children, with all children reaching Decision 2 by twelve weeks.

Here, then, the clinician's objective is to select an initial intervention from the set of available options {M, B} at Decision 1 and, following ascertainment of response status, to select at Decision 2 an option from {C, I, A}, where C is feasible only for children who are responding to their current interventions and {I, A} are appropriate for children who are not. As in the acute leukemia setting, the clinician would like to make these decisions to maximize the expected benefit to the child, where benefit is reflected in a longer-term outcome such as a parent or teacher reported assessment or an academic achievement measure. With attention restricted to these two decision points, a dynamic treatment regime involves a decision rule at Decision 1 (baseline) that takes as input all information on the child, including demographic, socioeconomic, achievement, and clinical assessment variables, and outputs an initial intervention selected from {M, B}. The Decision 2 rule takes this information plus additional, intervening information ascertained up until the time of the decision, including response status, and returns an option from {C, I, A}. As in the acute leukemia setting, of obvious interest is development of an optimal regime.

One can envision extension of this scenario to additional decision points, where, following Decision 1, at each subsequent decision point children who continue to respond to the initial intervention receive C, and the clinician can choose I or A for those who cease to respond to the initial intervention. For children who fail to respond to both their initial and either augmented or intensified subsequent interventions, further options could be established and corresponding rules devised.

This exemplifies a common situation, particularly in behavioral and mental health contexts, where, for individuals who respond to initial treatment, there is one option, to continue that treatment on the grounds

that it would be pointless to alter an apparently efficacious intervention. For individuals for whom initial treatment does not induce the desired response, there may be multiple follow-up options. In some applications, the roles of response and nonresponse are reversed.

1.2.3 Treatment of HIV Infection

As a final example, consider treatment for infection with the human immunodeficiency virus (HIV). After a patient is diagnosed with HIV infection, it is typical for the patient to be evaluated at regular, for example, monthly, clinic visits and for decisions on administration of antiretroviral therapy to be made at each visit. Initially and at each subsequent clinic visit, measures of the immunologic and virologic status of the patient are obtained to monitor the progression of the disease. A standard measure of immunologic status is CD4 T cell count (cells/mm^3), where lower counts reflect compromised immunologic status; and virologic status is indicated by viral load, the amount of HIV genetic material (RNA) present in a blood sample (viral RNA copies/ml), where smaller viral loads reflect better control of the virus.

On the basis of current and past measures of CD4 count and viral load, as well as other accrued information on the patient, the goal of the clinician at these clinic visits is to make decisions on therapy with the goal of maximizing the expected benefit to the patient after some period, say one year. Benefit may be reflected in the binary outcome of whether or not, after one year, viral load is below the level that can be detected by the assay used to ascertain viral load measurements. For simplicity, suppose that the clinician can choose either to administer antiretroviral therapy for the next month, coded as 1, or not, coded as 0. Then a dynamic treatment regime comprises a decision rule at Decision 1, the time a patient presents with HIV infection; and decision rules at Decisions $2, \ldots, 12$, say, corresponding to the monthly visits. The rule at Decision 1 takes as input baseline CD4 count, viral load, and other information and selects an option from the set of possible options $\{0, 1\}$; those at Decisions $2, \ldots, 12$ take as input previous and current CD4 counts, viral load measures, and any other baseline and evolving information on the patient and recommend an option from $\{0, 1\}$. A complication is that the virus can become resistant to antiretroviral therapies over time, in which case administration of such a therapy henceforth is of no benefit to the patient. In this simple example, for a patient whose accrued information at the time of a decision includes an indication that resistance has developed, the only feasible option from among those coded as 0 and 1 at this and all future decision points is option 0, do not administer therapy.

Here, in contrast to the previous two examples, the timing of the de-

cisions is predetermined and fixed according to the clinic visit schedule. As before, development of an optimal treatment regime is of interest.

1.3 The Meaning of "Dynamic"

One of the first uses of the term "dynamic treatment regime" is by Murphy et al. (2001), who define dynamic treatment regimes as comprising "rules for how the treatment level and type should vary with time," where "these rules are based on time-varying measurements of subject-specific need for treatment." These authors go on to highlight the distinction between a dynamic and *nondynamic* treatment regime, stating that the latter "is a special case of a dynamic treatment regime in which the treatment assignments do not vary by posttreatment observations." An interpretation of this is that a nondynamic regime involves rules that do not take into account patient information in assigning treatment. Such regimes have been called "static" rather than nondynamic in later literature.

For instance, consider again treatment of HIV-infected patients with antiretroviral therapy as in Section 1.2.3. Using this definition in this context, an example of a nondynamic or static treatment regime is one for which the rule at each monthly decision point dictates that therapy always be administered, so that a patient following this regime would always receive therapy regardless of his evolving virologic and immunologic status, side effects, or possible development of resistance of the virus to the antiretroviral agents. Another nondynamic regime might dictate that therapy should be administered for the first six months after the patient is diagnosed and then withdrawn for the next six months, again regardless of the progression of her disease. In contrast, a dynamic treatment regime involves rules that incorporate virologic, immunologic, and other information, so that the resulting treatment options administered vary according to values of the input information and are thus "dynamic," allowing treatment decisions to be responsive to the progression of the patient's disease. Clearly, the foregoing static/nondynamic regimes are of little relevance in practice, as it is unlikely that a clinician would be willing to, for example, withdraw therapy from a patient whose viral levels are not under control or to continue therapy for a patient whose virus has become resistant to it.

More recently, some authors have used "dynamic treatment regime" to refer to the fact that a regime involves multiple decision points, without regard to the nature of its rules, thus tacitly implying that a single decision regime is "nondynamic." In their definition above, Murphy et al. (2001) do allow a nondynamic regime with any number of decision points to involve rules that incorporate baseline information, so that "dynamic" refers to the additional dependence of the rules following the

initial decision point in a multistage situation on evolving information. This definition is thus consistent with the view of any single stage regime as "nondynamic."

Other authors have used "dynamic treatment regime" to refer to any regime, single or multistage, whose rules dictate treatment decisions that vary according to the values of the accrued patient information the rule takes as input, including at Decision 1, where the accrued information is that available at baseline.

A reader who is confused at this point has every right to be so. The convention we adopt in this book is that a dynamic treatment regime is as defined in the last paragraph; that is, a regime involving one or more decision points for which the decision rules potentially incorporate baseline and evolving patient information and thus lead to treatment selections that vary according to this information, including at Decision 1. A "static" regime is a special case of a dynamic treatment regime with one or more decision points whose rules do not incorporate such information. This definition of a dynamic treatment regime is aligned with the goal of precision medicine, even if attention is focused on a single decision point only.

As illustrated by the HIV example, static regimes are of limited interest in and are inconsistent with clinical practice and precision medicine. Consequently, we often refer to dynamic treatment regimes simply as "treatment regimes," without qualification.

1.4 Basic Framework

1.4.1 Definition of a Dynamic Treatment Regime

We now define more precisely the notion of a dynamic treatment regime and present the basic notational framework and conventions we adopt throughout this book.

In most of this book, we consider the situation where we can identify a finite number $K \geq 1$ decision points at which a treatment must be selected from among a set of available, feasible options. Indexing the decision points by $k = 1, \ldots, K$, we let $\mathcal{A}_k$ denote the set of available treatment options at Decision k, and let a_k represent an option in $\mathcal{A}_k$. We restrict attention mainly to the case where the number of options in $\mathcal{A}_k$ for each $k = 1, \ldots, K$ is finite and possibly different for different k. It is also possible for $\mathcal{A}_k$ to be an infinite set, as would be the case when the treatment options are doses of a drug in a continuous range of possible doses, which we mention at some points later in the book.

For definiteness, consider the setting of acute leukemia in Section 1.2.1, which involves $K = 2$ decision points. At Decision 1, $\mathcal{A}_1 = \{\mathsf{C}_1, \mathsf{C}_2\}$, so comprises the two induction chemotherapy options available

when an individual is diagnosed. At Decision 2, $\mathcal{A}_2 = \{\mathsf{M}_1,\mathsf{M}_2,\mathsf{S}_1,\mathsf{S}_2\}$, so contains the two maintenance and two salvage therapies. As noted previously, only the maintenance options are feasible for individuals who respond to the induction therapy initiated at Decision 1, and only the salvage options are appropriate for nonresponders to induction therapy. In general, $\mathcal{A}_k$ includes all available options, where of necessity some options may not be feasible for some individuals because of their past histories; we formalize this consideration later in the book.

Let x_1 denote the collection of information available on an individual at Decision 1, so at baseline. Let $a_1 \in \mathcal{A}_1$ be a treatment option at Decision 1, and let x_2 be additional information arising between Decisions 1 and 2. Let $a_2 \in \mathcal{A}_2$ be a treatment option at Decision 2, and let x_3 be additional information ascertained between Decisions 2 and 3. Continuing in this fashion, letting x_k be additional information collected between Decisions $k-1$ and k after receipt of option $a_{k-1} \in \mathcal{A}_{k-1}$ at Decision $k-1$, at Decision k, we denote by h_k the accrued information, or *history*, available on an individual. At Decision 1, the accrued information or history is simply the baseline information x_1; at subsequent decisions, the history consists of the additional information arising between previous decisions and the treatment options administered at those decisions. Thus, we define the history h_k formally as

$$\begin{aligned} h_1 &= x_1, \\ h_k &= (x_1, a_1, \ldots, x_{k-1}, a_{k-1}, x_k), \quad k = 2, \ldots, K, \end{aligned} \tag{1.1}$$

where $a_1, \ldots, a_{k-1}$ are the treatment options administered at Decisions 1 to $k-1$. Let $\mathcal{H}_k$ denote the support of h_k, $k = 1, \ldots, K$.

Thus, from (1.1), in the leukemia example, the history $h_1 = x_1$ available at Decision 1 might include baseline demographic, physiologic, and clinical variables; prior medical history; and genetic and genomic information, and $a_1 \in \mathcal{A}_1$ is the induction therapy option administered at Decision 1. The information x_2 collected intermediate to Decisions 1 and 2 might include updated measures of clinical variables, evolving marker values, indicators of occurrence of and timing of adverse events, and response status. Then the history available at Decision 2 is $h_2 = (x_1, a_1, x_2)$. The option $a_2 \in \mathcal{A}_2$ is the maintenance or salvage therapy administered at Decision 2, which clearly depends on the value of response status contained in x_2 and thus in h_2.

It proves convenient later in the book to define

$$\overline{x}_k = (x_1, \ldots, x_k), \quad \overline{a}_k = (a_1, \ldots, a_k), \quad \text{so that} \quad h_k = (\overline{x}_k, \overline{a}_{k-1}). \tag{1.2}$$

The "overbar" notation in (1.2) is standard and allows us to reference the components of the history, namely, all information ascertained on

an individual and the treatment options administered to an individual up to the current decision, separately when we discuss multiple decision problems. If we define $\overline{\mathcal{A}}_k = \mathcal{A}_1 \times \cdots \times \mathcal{A}_k$, $k = 1, \ldots, K$, to be the set of all possible combinations of treatment options that could be administered through Decision k, then clearly $\overline{a}_k \in \overline{\mathcal{A}}_k$.

Armed with this notation, we now define formally a dynamic treatment regime. At Decision k, a *decision rule* $d_k(h_k)$ is a function that maps an individual's history to a treatment option in $\mathcal{A}_k$, that is, $d_k : \mathcal{H}_k \rightarrow \mathcal{A}_k$, $k = 1, \ldots, K$. This definition makes precise that, at Decision k, a decision rule is a function that takes as input the accrued information or history for an individual and returns a treatment option from among the available options. Then a dynamic treatment regime d is defined to be a collection of such rules; that is, with K decision points,

$$d = \{d_1(h_1), \ldots, d_K(h_K)\}. \tag{1.3}$$

When $K = 1$, corresponding to a single stage setting, a dynamic treatment regime is a single rule, so that $d = \{d_1(h_1)\}$. For brevity, we often suppress the arguments of the rules and refer to a regime with K decision points as

$$d = (d_1, \ldots, d_K).$$

Corresponding to the convention set forth in Section 1.3, a dynamic treatment regime is one for which the rules $d_k(h_k)$ return different treatment options $a_k \in \mathcal{A}_k$ depending on the value of h_k for at least one of $k = 1, \ldots, K$. A static regime is then one for which each rule $d_k(h_k)$ returns the same treatment option in $\mathcal{A}_k$ regardless of the value of the history h_k. For example, if we restrict attention to the single stage problem involving Decision 1 in the leukemia context, so that all dynamic treatment regimes involve a single rule $d_1(h_1)$, and $h_1 = x_1$ comprises baseline information, an example of a static regime is one with rule

$$d_1(h_1) = \mathsf{C}_1 \quad \text{for all } h_1.$$

That is, the rule is of the form "give C_1 regardless of the baseline information on an individual." Note that, if we consider both decisions, $K = 2$, it is impossible to conceive of a plausible static regime for this problem, because any reasonable decision rule at Decision 2 must take into account response status, as, for example, a rule that assigns salvage therapy to all individuals regardless of whether or not they responded to induction therapy would contradict acceptable clinical practice and indeed be unethical.

These definitions of decision rules and treatment regimes are formulated in terms of information that would be realized by an individual over the course of the K decisions. Given a set of rules, treatment decisions for an individual can be made according to them as information

accrues. We emphasize that nowhere in these definitions do we refer to observed data, for example, from a clinical trial or observational study. A dynamic treatment regime can be defined independently of any data, as it is certainly possible to conceive of decision rules purely on the basis of knowledge of the information that would be available at each decision, subject matter expertise, and practical considerations.

In the foregoing formulation of a treatment regime, the decision rules can be viewed as deterministic or *nonrandom* in the sense that, at Decision k, given history h_k, the rule $d_k(h_k)$ assigns one and only one treatment option from among those in $\mathcal{A}_k$, $k = 1, \ldots, K$. It is also possible to define the notion of a *random* dynamic treatment regime, which is a regime for which the rule $d_k(h_k)$ assigns treatment options in $\mathcal{A}_k$ according to some prespecified probabilities depending on h_k. In the vast majority of applications, interest focuses on nonrandom regimes. Accordingly, in this book, we restrict attention to methodology for nonrandom dynamic treatment regimes. See Murphy et al. (2001) for discussion of random regimes.

1.4.2 Data for Development and Evaluation of Dynamic Treatment Regimes

In the classical study of the effect of a specific treatment option, a fundamental goal is to deduce the expected outcome if it were to be used in a population of interest and how that expected outcome compares to that for a competing option. This is based on estimation of the expected outcome for each option from data, where it is widely accepted that data from a clinical trial in which participants are assigned at random to receive the treatment options under consideration are most suitable for this purpose. As reviewed in detail in Chapter 2, data from randomized studies allow apparent differences in expected outcome to be attributed to the treatments under study. In contrast, data from observational studies in which the treatments received by participants are at their and their physicians' discretion can lead to biased inferences. This is due to the possibility that the effects of treatments are *confounded* with individual characteristics; for example, such that sicker individuals are more likely to receive one treatment option over another.

Given a particular dynamic treatment regime involving $K \geq 1$ decision points, it is of similar interest to estimate the expected outcome if the population were to receive treatment according to its rules and to compare the expected outcome to that associated with another regime. Data appropriate for this purpose are thus required. More generally, it is clear that an infinitude of possible regimes can be conceived, depending on different choices of rules. As we have noted, a fundamental objective is to identify an optimal set of rules; that is, an optimal regime, from

among all possible regimes that can be conceived. Estimation of an optimal regime thus must be based on suitable data. For these objectives, when $K = 1$, the issues are the same as in the classical setting.

When $K > 1$, estimation of the expected outcome if the population were to receive treatment according to the rules in a particular treatment regime and of an optimal regime involves additional considerations. As discussed in detail in Chapter 5, it is not possible to "piece together" analyses of data from K separate randomized or observational studies, each focused on a single decision point in isolation and each involving different sets of subjects, for this purpose. A major reason for the failure of this approach is that treatments administered at earlier decisions may have effects that do not manifest immediately and thus have implications for selection of treatments at later decision points. Accounting appropriately for such "delayed effects" requires that data be available on the same set of subjects through all K decisions. Data from a longitudinal, observational study in which the same patients are followed through all K decisions are one possible such resource, but could be subject to significant potential for confounding, now in a time-dependent fashion as detailed in Chapter 5, leading to biased inferences. Intuitively, data from a study in which individuals are randomly assigned to the treatment options at each decision point should be preferable to data from a such a longitudinal, observational study.

These considerations have led to great interest in the sequential multiple assignment randomized trial (SMART) design (e.g., Lavori and Dawson, 2004; Murphy, 2005). Given a set of $K > 1$ decision points, in a SMART, participants are randomized repeatedly to the available, feasible options at each decision point. Accordingly, discussion of methodology for dynamic treatment regimes of necessity must include a discussion of this study design, which is the topic of Chapter 9. In Chapters 5 and 6, a formal statistical framework for multiple decision problems is presented in which the conditions under which observed data are appropriate for estimation of dynamic treatment regimes can be clarified. This framework demonstrates that the SMART design yields a "gold standard" data source for this purpose.

1.5 Outline of this Book

The goal of this book is to present detailed, systematic overview of methodology for dynamic treatment regimes. With efforts to collect and curate patient-level data that can be used to inform estimation of treatment regimes rapidly evolving, we hope to make the foundations of the area and both fundamental techniques and newer advances accessible to statisticians and other researchers, who can use them to exploit these rich data resources.

The book is appropriate for graduate students and researchers in statistics and related quantitative fields and assumes no prior exposure to dynamic treatment regimes, causal inference, or associated topics. Chapters 1–7 and 9 present foundational material that could comprise an introductory PhD-level course on this topic. Familiarity with probability and statistical inference and popular statistical modeling approaches at a graduate (PhD) level should provide adequate background for much of this material, although parts of Chapters 6, 7, and 9 are at a somewhat higher mathematical level. On a first reading, technical arguments in these chapters can be omitted without loss of continuity. Chapters 8 and 10 cover more specialized material, which we recommend be approached after the reader feels comfortable with Chapters 1–7.

Central to this area is *causal inference*, particularly in the setting of time-dependent treatment assignment. In Chapter 2, we review the fundamentals of causal inference in a situation analogous to the setting of a single decision point. This chapter also reviews standard statistical modeling strategies and associated large sample theory, as these play an essential role in methodology for dynamic treatment regimes. Chapters 3 and 4 focus on treatment regimes in the single decision setting. Chapter 3 presents a causal inference framework in which an optimal single decision regime can be defined precisely, which provides the groundwork for methodology for estimation of optimal regimes from data. Several key approaches are presented in which a decision rule can be represented through a finite set of parameters that are estimated from data. Chapter 4 covers additional approaches, including those motivated by viewing estimation of an optimal regime as a classification problem, allowing techniques from machine learning to be exploited.

Chapters 5–7 cover the fundamentals of multiple decision dynamic treatment regimes. Because a formal account is rather involved, Chapter 5 presents a high level, less technical introduction to the multiple decision problem, which serves as a "roadmap" to the detailed, precise treatment in the next two chapters. Chapter 6 begins with a rigorous account of an appropriate statistical framework and of key assumptions that are required to deduce multiple decision regimes from observed data based on causal inference concepts in a time-dependent setting. Methodology for estimating the expected outcome of a given, specified, multistage treatment regime from data is reviewed. Chapter 7 presents a formal characterization of an optimal, multiple decision treatment regime and an overview of methodology for estimation of an optimal regime. This includes methods based on viewing the problem from a classification perspective. Chapter 8 extends the framework and methods to the setting where the outcome is a time to an event, which involves additional challenges owing to the fact that individuals may experience the

event before reaching all K decision points and that the outcome may be censored for some individuals in the observed data.

Chapter 9 provides an overview of SMARTs, including methods for sample size determination for these studies. Statistical inference on quantities associated with optimal dynamic treatment regimes is challenging. This is because this is an inherently nonsmooth statistical problem, so that standard asymptotic theory does not apply. Chapter 10 starts with an explicit demonstration of this problem and then presents an account of approaches to achieving valid large sample inference in single and multiple decision settings. Of necessity, this chapter is technical in nature.

Research on methodology for dynamic treatment regimes is still undergoing vigorous development. Chapter 11 provides a brief account of some additional topics. Supplemental material on these topics is available on the website dedicated to this book, which is noted in the Preface.

Many of the methods presented in this book involve quantities such as regression relationships between an observed outcome and patient characteristics and treatment options received and probabilities of receiving treatment options at a decision point given an individual's history. In most of the book, we focus on methods based on specification of parametric models for these relationships. This is consistent with how the methods are most often implemented in practice and in some cases leads to consideration of regimes whose rules are of a relatively straightforward form. Although this supports presentation of the methods in terms of approaches and theory that are likely to be familiar to most readers, it is not a requirement. As is noted in later chapters, in most cases it is possible to use instead more flexible modeling approaches, and, indeed, some methods are predicated on flexible representation of relationships.

Readers seeking a broad introduction to the foundations of and key methods in this area should read Chapters 1–5, the first two sections of Chapter 6, and Chapters 9 and 11. Those interested in a deeper, more comprehensive review should also read the remainder of Chapter 6, Chapter 7, and the first two sections of Chapter 10. Those familiar with survival analysis may also wish to read Chapter 8; the material on single decision problems is more accessible than that on multiple decisions, which is rather involved. Readers interested in the theoretical underpinnings of inference should cover Chapter 10, recognizing that this chapter is rather technical, as noted above.

Detailed accounts of application of some methods reviewed in Chapters 2–4, 5–7, and 8, are available on the dedicated website for this book given in the Preface. Many of these analyses are demonstrated using the R package `DynTxRegime`, developed by the authors, which is available on the website and at the Comprehensive R Archive Network (CRAN).

Remark. In this book, we use the term *fixed treatment regime* to refer

to a given, specified treatment regime. That is, we refer to a particular regime d with a given set of rules as a fixed regime. In the context of the HIV infection example in Section 1.2.3, where the history h_k at Decision k contains current CD4 count $\mathsf{CD4}_k$, say, an example of a fixed regime is the regime d with rules d_k such that $d_k(h_k) = 0$, do not administer antiretrovial therapy for the next month, if $\mathsf{CD4}_k > 200$ cells/mm^3; and $d_k(h_k) = 1$, do administer antiretroviral therapy for the next month, if $\mathsf{CD4}_k \leq 200$ cells/mm^3, $k = 1, \ldots, K$. Here, the rules characterizing d are given, specified functions of h_k. Another example of a fixed regime might involve different CD4 thresholds at different decision points; e.g., if there are $K = 12$ monthly visits, $d_k(h_k) = 0(1)$ if $\mathsf{CD4}_k > (\leq) 200$ for $k = 1, \ldots, 6$ and $d_k(h_k) = 0(1)$ if $\mathsf{CD4}_k > (\leq) 400$ for $k = 7, \ldots, 12$.

A fixed regime as defined here should not be confused with a static regime. As defined in Section 1.3, a static regime is such that all of its rules do not use any patient information to select treatment. As the examples in the previous paragraph demonstrate, a fixed regime need not involve static rules. Our use of the term "fixed" also should not be construed to imply that the timing of the decision points is according to a predetermined, or "fixed," schedule. In the acute leukemia and ADHD examples in Sections 1.2.1 and 1.2.2, the timing of the second decision point in any particular regime in these contexts is not predetermined but rather depends on a patient's response status.

Remark. Throughout the book, we present formal arguments justifying theoretical and methodological developments. To avoid having measure-theoretic considerations distract from appreciation of the conceptual foundations embodied in some of these arguments, particularly in early chapters, we often treat random variables that may be continuous or discrete as discrete without comment. The arguments of course can be generalized under appropriate conditions.

Bibliography

Almirall, D., Compton, S. N., Gunlicks-Stoessel, M., Duan, N., and Murphy, S. A. (2012a). Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Statistics in Medicine* **31,** 1887–1902.

Almirall, D., Lizotte, D., and Murphy, S. A. (2012b). SMART design issues and the consideration of opposing outcomes. Discussion of "Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer" by Wang et al. *Journal of the American Statistical Association* **107,** 509–512.

Almirall, D., Nahum-Shani, I., Sherwood, N. E., and Murphy, S. A. (2014). Introduction to SMART designs for the development of adaptive interventions: With application to weight loss research. *Translational Behavioral Medicine* **4,** 260–274.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Methods Based on Counting Processes.* Springer, New York.

Anstrom, K. J. and Tsiatis, A. A. (2001). Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data. *Biometrics* **57,** 1207–1218.

Arcones, M. A. and Giné, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annales de l'IHP Probabilités et Statistiques* **25,** 457–481.

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* in press.

Bai, X., Tsiatis, A. A., Lu, W., and Song, R. (2017). Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classfication perspective. *Lifetime Data Analysis* **23,** 584–604.

Bai, X., Tsiatis, A. A., and O'Brien, S. M. (2013). Doubly-robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. *Biometrics* **69,** 830–839.

Bartlett, P. L., Jordan, M. I., and McAuliffe, J. D. (2003). Large margin classifiers: Convex loss, low noise, and convergence rates. In *Advances in Neural Information Processing Systems 16 (NIPS 2003)*.

Bartlett, P. L., Jordan, M. I., and McAuliffe, J. D. (2006). Convexity, classification, and risk bounds. *Journal of the American Statistical Association* **101,** 138–156.

Bembom, O. and van der Laan, M. J. (2007). Statistical methods for analyzing sequentially randomized trials. *Journal of the National Cancer Institute* **99,** 1577–1582.

Beran, R. (1990). Calibrating prediction regions. *Journal of the American Statistical Association* **85,** 715–723.

Berry, D. A. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*. Springer, New York.

Bertsekas, D. P. (2015). *Convex Optimization Algorithms*. Athena Scientific, Belmont, Massachusetts.

Bickel, P. J. and Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica* **18,** 967–985.

Biernot, P. and Moodie, E. E. M. (2010). A comparison of variable selection approaches for dynamic treatment regimes. *The International Journal of Biostatistics* **6**.

Blatt, D., Murphy, S., and Zhu, J. (2004). A-learning for approximate planning. *Technical Report 04-63, The Methodology Center, Pennsylvania State University* .

Boos, D. D. and Stefanski, L. (2013). *Essential Statistical Inference*. Springer, New York.

Booth, J. G. and Hall, P. (1994). Monte Carlo approximation and the iterated bootstrap. *Biometrika* **81,** 331–340.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Breiman, L. (2001). Random forests. *Machine Learning* **45,** 5–32.

Breiman, L., Freidman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Breslow, N. E. (1972). Discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B* **34,** 216–217.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Annales de l'IHP Probabilités et Statistiques* **19,** 281–296.

Butler, E. L., Laber, E. B., Davis, S. M., and Kosorok, M. R. (2018). Incorporating patient preferences into estimation of optimal indi-

vidualized treatment rules. *Biometrics* **74,** 18–26.

Carlin, B. P., Berry, S. M., Lee, J. J., and Müller, P. (2010). *BAyesian Adaptive Methods for Clinical Trials*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Chaffee, P. and van der Laan, M. (2012). Targeted maximum likelihood estimation for dynamic treatment regimes in sequentially randomized controlled trials. *The International Journal of Biostatistics* **8**.

Chakraborty, B., Ghosh, P., Moodie, E. E. M., and Rush, A. J. (2016). Estimating optimal shared-parameter dynamic regimens with application to a multistage depression clinical trial. *Biometrics* **72,** 865–876.

Chakraborty, B., Laber, E. B., and Zhao, Y. (2014a). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics* **69,** 714–723.

Chakraborty, B., Laber, E. B., and Zhao, Y.-Q. (2014b). Inference about the expected performance of a data-driven dynamic treatment regime. *Clinical Trials* **11,** 408–417.

Chakraborty, B., Murphy, S. A., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* **19,** 317–343.

Chen, G., Zeng, D., and Kosorok, M. R. (2016). Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association* **111,** 1509–1521.

Chen, P. Y. and Tsiatis, A. A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57,** 1030–1038.

Cheung, Y. K., Chakraborty, B., and Davidson, K. W. (2015). Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program. *Biometrics* **71,** 450–459.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, Second Edition*. L. Erlbaum Associates, Hillsdale, New Jersey.

Collins, L., Murphy, S., and Bierman, K. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science* **5,** 185–196.

Collins, L. M., Nahum-Shani, I., and Almirall, D. (2014). Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clinical Trials* **11,** 426–434.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* **20,** 273–297.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **69,** 269–276.

Dawid, A. P. (1994). Selection paradoxes of Bayesian inference. In *Multivariate Analysis and Its Applications*, pages 211–220. Institute of Mathematical Statistics, Hayward, California.

Dean, A., Morris, M., Stufken, J., and Bingham, D. (2015). *Handbook of Design and Analysis of Experiments*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Diáz, I., Savenkov, O., and Ballman, K. (2018). Targeted learning ensembles for optimal individualized treatment rules with time-to-event outcomes. *Biometrika* **105,** 723–738.

Ertefaie, A. and Strawderman, R. L. (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika* **105,** 963–977.

Ertefaie, A., Wu, T., Lynch, K. G., and Nahum-Shani, I. (2016). Identifying a set that contains the best dynamic treatment regimes. *Biostatistics* **17,** 135–148.

Fan, A., Lu, W., and Song, R. (2016). Sequential advantage selection for optimal treatment regime. *Annals of Applied Statistics* **10,** 32–53.

Fan, C., Lu, W., Song, R., and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society, Series B* **79,** 1565–1582.

Fava, M., Rush, A. J., Trivedi, M. H., Nierenberg, A. A., Thase, M. E., Sackeim, H. A., Quitkin, F. M., Wisniewski, S., Lavori, P. W., Rosenbaum, J. F., and Kupfer, D. J. (2003). Background and rationale for the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. *Psychiatric Clinics of North America* **26,** 457–494.

Foster, J. C., Taylor, J. M., and Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine* **30,** 2867–2880.

Geng, Y., Zhang, H. H., and Lu, W. (2015). On optimal treatment regimes selection for mean survival time. *Statistics in Medicine* **34,** 1169–1184.

Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability* **18,** 851–869.

Glick, N. (1974). Consistency conditions for probability estimators and integrals of density estimators. *Utilitas Math* **6,** 61–74.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, Reading, Massachusetts.

Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *Annals of Statistics* **40,** 529–260.

Goldberg, Y., Song, R., Kosorok, M. R., et al. (2013). Adaptive Q-learning. In *From Probability to Statistics and Back: High-Dimensional Models and Processes–A Festschrift in Honor of Jon A. Wellner*, pages 150–162. Institute of Mathematical Statistics, Hayward, California.

Gunter, L., Chernick, M., and Sun, J. (2011a). A simple method for variable selection in regression with respect to treatment selection. *Pakistan Journal of Statistics and Operation Research* **7,** 363–380.

Gunter, L., Zhu, J., and Murphy, S. A. (2007). Variable selection for optimal decision making. In *Artificial Intelligence in Medicine. AIME 2007. Lecture Notes in Computer Science*, pages 149–154, Berlin Heidelberg. Springer.

Gunter, L., Zhu, J., and Murphy, S. A. (2011b). Variable selection for qualitative interactions. *Statistical Methodology* **8,** 42–55.

Gunter, L., Zhu, J., and Murphy, S. A. (2011c). Variable selection for qualitative interactions in personalized medicine while controlling the family-wise error rate. *Journal of Biopharmaceutical Statistics* **21,** 1063–1078.

Gurobi Optimization (2019). Gurobi Optimizer Reference Manual, Version 8.1. `www.gurobi.com/`.

Haberman, S. J. (1989). Concavity and estimation. *Annals of Statistics* **17,** 1631–1661.

Hager, R., Tsiatis, A. A., and Davidian, M. (2018). Optimal two-stage dynamic treatment regimes from a classification perspective with censored survival data. *Biometrics* **74,** 1180–1192.

Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics* **14,** 1431–1452.

Hall, P., Peng, L., and Tajvidi, N. (1999). On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika* **86,** 871–880.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Springer-Verlag, New York.

Henderson, H. V. and Searle, S. (1979). Vec and vech operators for

matrices, with some uses in Jacobians and multivariate statistics. *Canadian Journal of Statistics* **7,** 65–81.

Henderson, R., Ansell, P., and Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics* **66,** 1192–1201.

Hernán, M., Brumback, B., and Robins, J. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11,** 561–570.

Hernán, M. A., Lanoy, E., Costagliola, D., and Robins, J. M. (2006). Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology* **98,** 237–242.

Hirano, K. and Porter, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica* **80,** 1769–1790.

Hjort, N. L. and Pollard, D. (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806* .

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

Huang, X., Goldberg, Y., and Xu, J. (2019). Multicategory individualized treatment regime using outcome weighted learning. *Biometrics* in press.

Huang, X., Ning, J., and Wahed, A. S. (2014). Optimization of individualized dynamic treatment regimes for recurrent diseases. *Statistics in Medicine* **33,** 2363–2378.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association* **103,** 832–842.

IBM ILOG CPLEX (2019). IBM ILOG CPLEX Optimization Studio. `www.ibm.com/products/ilog-cplex-optimization-studio`.

Jiang, B., Song, R., Li, J., and Zeng, D. (2019). Entropy learning for dynamic treatment regimes. *Statistica Sinica* in press.

Jiang, R., Lu, W., Song, R., and Davidian, M. (2017a). On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society, Series B* **79,** 1165–1185.

Jiang, R., Lu, W., Song, R., Hudgens, M. G., and Napravavnik, S. (2017b). Doubly robust estimation of optimal treatment regimes for survival data with application to an HIV/AIDS study. *Annals of Applied Statistics* **11,** 1763–1786.

Jones, B. and Kenward, M. G. (2014). *Design and Analysis of Cross-*

Over Trials. Chapman & Hall/CRC Press, Boca Raton, Florida.

Kang, C., Janes, H., and Huang, Y. (2014). Combining biomarkers to optimize patient treatment recommendations. *Biometrics* **70,** 695–720.

Karatzoglou, A., Meyer, D., and Hornik, K. (2006). Support vector machines in R. *Journal of Statistical Software* **15,** 1–28.

Kelleher, S. A., Dorfman, C. S., Vilardaga, J. C. P., Majestic, C., Winger, J., Gandhi, V., Nunez, C., Van Denburg, A., Shelby, R. A., Reed, S. D., Murphy, S. A., Davidian, M., Laber, E. B., Kimmick, Gretchen, G., Westbrook, K. W., Abernathy, A. P., and Somers, T. J. (2017). Optimizing delivery of a behavioral pain intervention in cancer patients using a sequential multiple assignment randomized trial (SMART). *Contemporary Clinical Trials* **57,** 51–57.

Kidwell, K. M. (2014). SMART designs in cancer research: Learning from the past, current limitations and looking toward the future. *Clinical Trials* **11,** 445–456.

Kidwell, K. M. and Wahed, A. S. (2013). Weighted log-rank statistic to compare shared-path adaptive treatment strategies. *Biostatistics* **14,** 229–312.

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., and Murphy, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology* **34,** 1220–1228.

Kosorok, M. R. (2007). *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag, New York.

Krauth, W. (2006). *Statistical Mechanics: Algorithms and Computations*, volume 13. Oxford University Press, Oxford.

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

Laber, E. B., Linn, K. A., and Stefanski, L. A. (2014a). Interactive model building for Q-learning. *Biometrika* **101,** 831–847.

Laber, E. B., Lizotte, D. J., and Ferguson, B. (2014b). Set-valued dynamic treatment regimes for competing outcomes. *Biometrics* **70,** 53–61.

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E., and Murphy, S. A. (2014c). Dynamic treatment regimes: Technical challenges and applications. *Electronic Journal of Statistics* **8,** 1225–1271.

Laber, E. B. and Murphy, S. A. (2011). Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association* **106,** 904–913.

Laber, E. B. and Zhao, Y. Q. (2015). Tree-based methods for individualized treatment regimes. *Biometrika* **102,** 501–514.

Laber, E. B., Zhao, Y. Q., Regh, T., Davidian, M., Tsiatis, A., Stanford, J. B., Zeng, D., and Kosorok, M. R. (2016). Using pilot data to size a two-arm randomized trial to find a nearly optimal personalized treatment strategy. *Statistics in Medicine* **35,** 1245–1256.

Lange, K. (2010). *Numerical Analysis for Statisticians.* Springer-Verlag, New York.

Lavori, P. W. and Dawson, R. (2000). A design for testing clinical strategies: Biased adaptive within-subject randomization. *Journal of the Royal Statistical Society, Series A* **163,** 29–38.

Lavori, P. W. and Dawson, R. (2004). Dynamic treatment regimes: Practical design considerations. *Clinical Trials* **1,** 9–20.

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., and Murphy, S. A. (2011). A SMART design for building individualized treatment sequences. *Annual Review of Clinical Psychology* **8,** 21–48.

Letham, B., Rudin, C., McCormick, T. H., and Madigan, D. (2015). Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *Annals of Applied Statistics* **9,** 1350–1371.

Liao, P., Klasnja, P., Tewari, A., and Murphy, S. A. (2016). Sample size calculations for micro-randomized trials in mHealth. *Statistics in Medicine* **35,** 1944–1971.

Lin, R., Thall, P. F., and Yuan, Y. (2019). An adaptive trial design to optimize dose-schedule regimes with delayed outcomes. *Biometrics* in press.

Linn, K. A., Laber, E. B., and Stefanski, L. A. (2017). Interactive Q-learning for quantiles. *Journal of the American Statistical Association* **112,** 638–649.

Lipkovich, I., Dmitrienko, A., Denne, J., and Enas, G. (2011). Subgroup identification based on differential effect search: Recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics in Medicine* **30,** 2601–2621.

Liu, Y., Wang, Y., Kosorok, M. R., Zhao, Y., and Zeng, D. (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in Medicine* **37,** 3776–3788.

Liu, Y., Wang, Y., and Zeng, D. (2017). Sequential multiple assignment randomized trials with enrichment design. *Biometrics* **73,** 378–390.

Lizotte, D. J. and Laber, E. B. (2016). Multi-objective Markov decision processes for data-driven decision support. *Journal of Machine*

Learning Research **17,** 1–28.

Lizotte, D. J. and Tahmasebi, A. (2017). Prediction and tolerance intervals for dynamic treatment regimes. *Statistical Methods in Medical Research* **26,** 1611–1629.

Lou, Z., Shao, J., and Yu, M. (2018). Optimal treatment assignment to maximize expected outcome with multiple treatments. *Biometrics* **74,** 506–516.

Lu, W., Zhang, H. H., and Zeng, D. (2013). Variable selection for optimal treatment decision. *Statistical Methods in Medical Research* **22,** 493–504.

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E., and Kosorok, M. R. (2019a). Estimating dynamic treatment regimes in mobile health using V-learning. *Journal of the American Statistical Association* in press.

Luckett, D. J., Laber, E. B., and Kosorok, M. R. (2019b). Estimation and optimization of composite outcomes. *arXiv preprint arXiv:1711.10581v3* .

Luedtke, A. R. and van der Laan, M. (2016a). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics* **44,** 713–742.

Luedtke, A. R. and van der Laan, M. (2016b). Super-learning of an optimal dynamic treatment rule. *The International Journal of Biostatistics* **12**.

Mark, S. and Robins, J. (1993). Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model. *Statistics in Medicine* **12,** 1605–1628.

Matts, J. P. and Lachin, J. M. (1988). Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials* **9,** 327–344.

McKeague, I. W. and Qian, M. (2015). An adaptive resampling test for detecting the presence of significant predictors. *Journal of the American Statistical Association* **110,** 1422–1433.

Mebane, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: The rgenoud package for R. *Journal of Statistical Software* **42,** 1–26.

Mi, X., Zou, F., and Zhu, R. (2019). Bagging and deep learning in optimal individualized treatment rules. *Biometrics* in press.

Moodie, E. (2009). Risk factor adjustment in marginal structural model estimation of optimal treatment regimes. *Biometrical Journal* **51,**

774–788.

Moodie, E. and Richardson, T. (2010). Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics* **37,** 126–146.

Moodie, E. E. M., Dean, N., and Sun, Y. (2014). Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences* **6,** 223–243.

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

MOSEK (2019). MOSEK, Version 9.1. `www.mosek.com/`.

Muessig (2019). AllyQuest Adherence App Intervention for HIV-positive Men Who Have Sex With Men and Transgender Women: PilotTrial (AQ2). Clinical Trials.gov Identifier NCT03916484, `https://clinicaltrials.gov/ct2/show/NCT03916484`.

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

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine* **24,** 1455–1481.

Murphy, S. A., Lynch, K. G., Oslin, D., Mckay, J. R., and TenHave, T. (2007a). Developing adaptive treatment strategies in substance abuse research. *Drug and Alcohol Dependence* **88,** s24–s30.

Murphy, S. A., Oslin, D., Rush, A. J., and Zhu, J. (2007b). Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders. *Neuropsychopharmacology* **32,** 257–262.

Murphy, S. A., van der Laan, M. J., Robins, J. M., and CPPRG (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association* **96,** 1410–1423.

Mustanski (2018). A Pragmatic Trial of an Adaptive eHealth HIV Prevention Program for Diverse Adolescent MSM (SMART). ClinicalTrials.gov Identifier NCT03511131, `https://clinicaltrials.gov/ct2/show/NCT03511131`.

Nahum-Shani, I., Qian, M., Almiral, D., Pelham, W., Gnagy, B., Fabiano, G., Waxmonsky, J., Yu, J., and Murphy, S. A. (2012a). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods* **17,** 457–477.

Nahum-Shani, I., Qian, M., Almiral, D., Pelham, W., Gnagy, B., Fabiano, G., Waxmonsky, J., Yu, J., and Murphy, S. A. (2012b). Q-

learning: A data analysis method for constructing adaptive interventions. *Psychological Methods* **17,** 478–494.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., and Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* **52,** 446–462.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay in principles. Section 9 (translation published in 1990). *Statistical Science* **5,** 472–480.

Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization. *Annals of Statistics* **20,** 1514–1533.

Oetting, A. I., Levy, J. A., Weiss, R. D., and Murphy, S. A. (2011). Statistical methodology for a SMART design in the development of adaptive treatment strategies. In Shrout, P., Shrout, P., Keyes, K., and Ornstein, K., editors, *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*, pages 179–205, Arlington, VA. American Psychiatric Publishing, Inc.

Orellana, L., Rotnitzky, A., and Robins, J. M. (2010a). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: Main content. *The International Journal of Biostatistics* **6**.

Orellana, L., Rotnitzky, A., and Robins, J. M. (2010b). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part II: Proofs and additional results. *The International Journal of Biostatistics* **6**.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling.* Springer-Verlag, New York.

Qian, M. and Murphy, S. (2011). Performance guarantees for individualized treatment rules. *Annals of Statistics* **39,** 1180–1210.

Rich, B., Moodie, E. E., and Stephens, D. A. (2014). Simulating sequential multiple assignment randomized trials to generate optimal personalized warfarin dosing strategies. *Clinical Trials* **11,** 435–444.

Richardson, T. S. and Rotnitzky, A. (2014). Causal etiology of the research of James M. Robins. *Statistical Science* **29,** 459–484.

Rivest, R. L. (1987). Learning decision lists. *Machine Learning* **2,** 229–246.

Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure periods: Application to control of the healthy worker survivor effect. *Mathematical Modelling* **7,** 1393–

1512.

Robins, J. (1987). Addendum to: A new approach to causal inference in mortality studies with sustained exposure periods: Application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications* **14,** 923–945.

Robins, J. (1989). The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In Sechrest, L., Freeman, H., and Mulley, A., editors, *Health Service Research Methodology: A Focus on AIDS*, pages 113–159, New York. NCHSR, U.S. Public Health Service.

Robins, J. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association* pages 24–33.

Robins, J. (1997). Causal inference from complex longitudinal data. In Berkane, M., editor, *Latent Variable Modeling and Applications to Causality: Lecture Notes in Statistics*, pages 69–117, New York. Springer-Verlag.

Robins, J. (1999a). Marginal structural models versus structural nested models as tools for causal inference. In Halloran, M. E. and Berry, D., editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, volume 116 of *IMA*, pages 95–134, New York. Springer.

Robins, J. M. (1999b). Association, causation, and marginal structural models. *Synthese* **121,** 151–179.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In Lin, D. Y. and Heagerty, P., editors, *Proceedings of the Second Seattle Symposium on Biostatistics*, pages 189–326, New York. Springer.

Robins, J. M., Hernán, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11,** 550–60.

Robins, J. M., Orellana, L., and Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine* **27,** 4678–4721.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In Jewell, N., Dietz, K., and Farewell, V., editors, *AIDS Epidemiology–Metholdological Issues*, pages 297–331, Boston. Birkhäuser.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of re-

gression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89,** 846–866.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90,** 106–121.

Rockafellar, R. T. and Wets, R. J.-B. (2009). *Variational Analysis*. Springer Verlag, Berlin Heidelberg.

Romano, J. P., Shaikh, A. M., et al. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *Annals of Statistics* **40,** 2798–2822.

Rose, E., Laber, E., Davidian, M., Tsiatis, A., Zhao, Y., and Kosorok, M. (2019). Sample size calculations for SMARTs. *NC State University Department of Statistics Technical Report* **1,** 1–30.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* **79,** 516–524.

Rubin, D. and van der Laan, M. (2012). Statistical issues and limitations in personalized medicine research with clinical trials. *International Journal of Biostatistics* **8**.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66,** 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* **6,** 34–58.

Rubin, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics* **36,** 293–298.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association* **100,** 322–331.

Rush, A., Fava, M., Wisniewski, S., Lavori, P., Trivedi, M., Sackeim, H., Thase, M., Nierenberg, A., Quitkin, F., Kashner, T., Kupfer, D., Rosenbaum, J., Alpert, J., Stewart, J., McGrath, P., Biggs, M., Shores-Wilson, K., Lebowitz, B., Ritz, L., and Niederehe, G. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials* **25,** 119–142.

Saarela, O., Arjas, E., Stephens, D. A., and Moodie, E. E. M. (2015).

Predictive Bayesian inference and dynamic treatment regimes. *Biometrical Journal* **57,** 941–958.

Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Annals of Mathematical Statistics* **18,** 434–438.

Schrijver, A. (1998). *Theory of Linear and Integer Programming.* John Wiley & Sons, Hoboken, New Jersey.

Schuler, M. S. and Rose, S. (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *American Journal of Epidemiology* **185,** 65–73.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Statistical Science* **29,** 640–661.

Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society* **122,** 1251–1262.

Shen, J., Wang, L., and Taylor, J. M. G. (2017). Estimation of the optimal regime in treatment of prostate cancer recurrence from observational data using flexible weighting models. *Biometrics* **73,** 635–645.

Shi, C., Song, R., and Lu, W. (2016). Robust learning for optimal treatment decision with np-dimensionality. *Electronic Journal of Statistics* **10,** 2894–2921.

Shortreed, S. M., Laber, E., Stroup, T. S., Pineau, J., and Murphy, S. A. (2014). A multiple imputation strategy for sequential multiple assignment randomized trials. *Statistics in Medicine* **33,** 4202–4214.

Song, R., Luo, S., Zeng, D., Zhang, H. H., Lu, W., and Li, Z. (2017). Semiparametric single-index model for estimating optimal individualized treatment strategy. *Electronic Journal of Statistics* **11,** 364–384.

Song, R., Wang, W., Zeng, D., and Kosorok, M. R. (2015). Penalized Q-learning for dynamic treatment regimens. *Statistica Sinica* **25,** 901–920.

Stefanski, L. A. and Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician* **56,** 29–38.

Stephens, D. A. (2015). G-estimation for dynamic treatment regimes in the longitudinal setting. In Kosorok, M. R. and Moodie, E. E. M., editors, *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*, pages 89–117, Philadelphia, PA. SIAM.

Su, X., Tsai, C. L., Wang, H., Nickerson, D. M., and Li, B. (2009).

Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research* **10,** 141–158.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction, Second Edition*. MIT Press, Cambridge.

Swartz, M. S., Perkins, D. O., Stroup, T. S., McEvoy, J. P., Nieri, J. M., and Haal, D. D. (2003). Assessing clinical and functional outcomes in the clinical antipsychotic of intervention effectiveness (CATIE) schizophrenia trial. *Schizophrenia Bulletin* **29,** 33–43.

Tao, Y. and Wang, L. (2017). Adaptive contrast weighted learning for multi-stage multi-treatment decision-making. *Biometrics* **73,** 145–155.

Taylor, J. M. G., Cheng, W., and Foster, J. C. (2015). Reader Reaction to "A robust method for estimating optimal treatment regimes" by Zhang et al. (2012). *Biometrics* **71,** 267–273.

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research* **21,** 55–75.

Therneau, T., Atkinson, B., and Ripley, B. (2015). Package `rpart`: Recursive partitioning and regression trees. *Available at* `http://cran.us.r-project.org/` .

Tian, L., Alizadeh, A. A., Gentles, A. J., and Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association* **109,** 1517–1532.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer-Verlag, New York.

Tsiatis, A. A. (2014). Competing risks: Theory. In *Wiley StatsRef: Statistics Reference Online*. doi: 10.1002/9781118445112.stat05131.

van der Laan, M. J. (2010). Targeted maximum likelihood based causal inference: Part I. *The International Journal of Biostatistics* **6**.

van der Laan, M. J. and Luedtke, A. R. (2015). Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of Causal Inference* **3,** 61–95.

van der Laan, M. J. and Petersen, M. L. (2007a). Causal effect models for realistic individualized treatment and intention to treat rules. *The International Journal of Biostatistics* **3**.

van der Laan, M. J. and Petersen, M. L. (2007b). Statistical learning of origin-specific statically optimal individualized treatment rules. *The International Journal of Biostatistics* **3**.

van der Laan, M. J. and Rose, S. (2018). *Targeted Learning in Data Sci-*

ence: Causal Inference for Complex Longitudinal Studies. Springer, New York.

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics* **2**.

van der Vaart, A. (1991). On differentiable functionals. *Annals of Statistics* **19,** 178–204.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Vansteelandt, S. and Joffe, M. (2014). Structural nested models and g-estimation: The partially realized promise. *Statistical Science* **29,** 707–731.

Vapnik, V., Golowich, S., and Smola, A. (1997). Support vector regression for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems* **9,** 281–287.

Wallace, M., Moodie, E. E. M. M., and Stephens, D. A. (2019). Model selection for g-estimation of dynamic treatment regimes. *Biometrics* in press.

Wallace, M. P., Moodie, E. E. M., and Stephens, D. A. (2016). SMART thinking: A review of recent developments in sequential multiple assignment randomized trials. *Current Epidemiology Reports* **3,** 225–232.

Wallace, M. P. and Moodie, E. M. M. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* **71,** 636–644.

Wallace, M. P., Moodie, E. M. M., and Stephens, D. A. (2015). Model assessment in dynamic treatment regimen estimation via double robustness. *Biometrics* **71,** 636–644.

Wang, L., Rotnitzky, A., Lin, X., Millikan, R., and Thall, P. (2012). Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer. *Journal of the American Statistical Association* **107,** 493–508.

Wang, L., Zhou, Y., Song, R., and Sherwood, B. (2018). Quantile-optimal treatment regimes. *Journal of the American Statistical Association* **113,** 1243–1254.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning* **8,** 279–292.

Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical*

Trials, Second Edition. John Wiley & Sons, Chichester.

Wolsey, L. A. and Nemhauser, G. L. (2014). *Integer and Combinatorial Optimization*. John Wiley & Sons, Hoboken, New Jersey.

Wu, C. J. and Hamada, M. S. (2011). *Experiments: Planning, Analysis, and Optimization*, volume 552. John Wiley & Sons, Hoboken, New Jersey.

Wu, F. (2015). *Adaptive projection intervals*. PhD thesis, North Carolina State University.

Xu, Y., Müller, P., Wahed, A. S., and Thall, P. F. (2016). Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *Journal of the American Statistical Association* **111,** 921–950.

Xu, Y., Yu, M., Zhao, Y., Li, Q., Wang, S., and Shao, J. (2015). Regularized outcome weighted subgroup identification for differential treatment effects. *Biometrics* **71,** 645–653.

Zajonc, T. (2012). Bayesian inference for dynamic treatment regimes: Mobility, equity, and efficiency in student tracking. *Journal of the American Statistical Association* **107,** 80–92.

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

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012b). A robust method for estimating optimal treatment regimes. *Biometrics* **68,** 1010–1018.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100,** 681–694.

Zhang, B. and Zhang, M. (2018a). C-learning: A new classification framework to estimate optimal dynamic treatment regimes. *Biometrics* **74,** 891–899.

Zhang, B. and Zhang, M. (2018b). Variable selection for estimating the optimal treatment regimes in the presence of a large number of covariates. *Amnals of Applied Statistics* **12,** 2335–2358.

Zhang, Y. and Laber, E. B. (2015). Comment on "An adaptive resampling test for detecting the presence of significant predictors" by McKeague and Qian. *Journal of the American Statistical Association* **110,** 1451–1454.

Zhang, Y., Laber, E. B., Davidian, M., and Tsiatis, A. A. (2018). Interpretable dynamic treatment regimes. *Journal of the American Statistical Association* **113,** 1541–1549.

Zhang, Y., Laber, E. B., Tsiatis, A. A., and Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics* **71,** 895–904.

Zhao, Y., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015a). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association* **110,** 583–598.

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

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

Zhao, Y., Zeng, D., Socinski, M. A., and Kosorok, M. R. (2011). Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics* **67,** 1422–1433.

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

Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association* **112,** 169–187.

Zhu, R., Zhao, Y. Q., Chen, G., Ma, S., and Zhao, H. (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics* **73,** 391–400.

Zhu, W., Zeng, D., and Song, R. (2019). Proper inference for value function in high-dimensional Q-learning for dynamic treatment regimes. *Journal of the American Statistical Association* **114,** 1404–1417.