Spring 2022

Quantitative Methods for Improving Medical Decision-Making

Margret Erlendsdottir
Yale University Graduate School of Arts and Sciences, margret.erlendsdottir@yale.edu

Abstract

Quantitative Methods for Improving Medical Decision-Making

Margret Erlendsdottir

2022

Innovation in causal inference and implementation of electronic health record systems are rapidly transforming medical care. In this dissertation, we present three examples in which use of methods in causal inference and large electronic health record data address existign challenges in medical decision-making. First, we use principles of causal inference to examine the structure of randomized trials of biomarker targets, which have produced divergent results and controversial clinical guidelines for management of hypertension and other chronic diseases. We discuss four key threats to the validity of trials of this design. Second, we use methods in causal inference for adjustment of time-varying confounding to estimate the effect of time-varying treatment strategies for hypertension. We report the results of a study which used longitudinal electronic health record data from a prospective virtual cohort of veterans. Third, we use individual-level electronic health record data to predict the need for critical care resources during surges in COVID-19 cases, to aid hospital administrators with resource allocation in periods of crisis.

Quantitative Methods for Improving Medical Decision-Making

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Margret C. Erlendsdottir

Dissertation Director: Forrest W. Crawford

May 2022

Acknowledgments

First and foremost, I would like to thank my dissertation advisor, Forrest Crawford, for his support and inexhaustible patience. He has been a role model, leading by example as a scientist with the highest standards for rigor and critical thought. Forrest has gone above and beyond as a mentor, providing not only professional guidance based on his expertise in multiple fields, but also valuable advice on navigating the many challenges of academic life.

I would like to thank Amy Justice and Fan Li, who served on my committee. Amy has been a source of inspiration. Her passion for her research and clinical interests have energized my efforts to pursue a career as a physician-scientist. Fan has imparted an infectious enthusiasm for this work in our meetings. I am grateful for his expertise and energy in the completion of this work. I look forward to continuing my work with both Amy and Fan in the future.

I would also like to acknowledge my mentors and collaborators, Janet Tate and Aldo Peixoto for their support and contribution to the work presented in Chapter 3. I would also like to acknowledge the COVID-19 Statistics, Policy Modelling, and Epidemiology Collective for their contribution to the work presented in Chapter 4.

It has been a privilege to be a member of the Yale School of Public Health. I would like to extend my heartfelt gratitude to Chris Tschudi, who has followed my progress since my initial interview at the Yale MD/PhD program and has generously provided advice and support throughout my time in this program. I would like to thank my many mentors at the YSPH Modeling Unit, who have provided opportunities for collaboration and discussion. In particular, I would like to thank Gregg Gonsalves for serving as a reader for my dissertation, for sharing his expertise and passion for his work in our ongoing project, and for his kindness and encouragement.

I would like to thank my friends, Xiaoxuan Cai, Sammi Dean, Dan Eck, Soheil Esghi, Haidong

Lu, Natalia Kunst, Olya Morozova, Richard Li, Ottavia Prunas, Lilla Orr, Jinghao Sun, Tommy Thornhill and Anna York for all the wonderful memories in and out of the Crawford Lab and the YSPH Modeling Unit. In the MD/PhD program, I would like to thank Adriana Cherskov, Matt Dong, Laurel Kaye, and George Linderman for sharing the joys and trials of the long journey of physician-scientist training. I would also like to thank Rachel O'Connell, Neil Pathak, Joyce Cheng, Matt Griffin, and Mollie Tucker for their friendship and support.

These acknowledgments would not be complete without expressing my gratitude to my partner, Alex Luryi, and my family. I would like to thank Alex for his love and unending support and for providing fun and respite during the combined challenge of pandemic life and dissertation work. I would like to thank my parents and my brother for their encouragement and their unceasing confidence in me.

I would also like to acknowledge the Yale School of Public Health and NICHD grant 1DP2HD091799-01 for their support.

To my mentors, friends and family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Modern statistical and computational methods for analyzing big data are rapidly transforming medical care. First, a class of methodologies in the field of causal inference are at the forefront of statistical innovations. Causal inference provides a framework for rigorous evaluation of biomedical interventions. Principles of causal inference can be used to evaluate the properties of experimental and observational study designs. Methods in causal inference also permit us to create causal links between complex treatments and outcomes in large observational datasets derived from electronic health records. Second, a major advance has arisen with the implementation of electronic health record systems, which has increased the information collected and stored about every patient. The plethora of available data has created an opportunity to examine trends and evaluate practices in medicine at an unprecedented scale and scope. This data can be used in real time to inform and adjust decision-making to optimize the delivery of medical care.

In this dissertation, we utilize electronic health record data and tools in causal inference to address three unmet needs in medical decision-making. First, we use principles of causal inference to examine the structure of a class of randomized trials, trials of biomarker targets, which have produced divergent results and controversial clinical guidelines. We discuss our

findings in the setting of randomized controlled trials of blood pressure targets. Because randomized trials of blood pressure targets have not yielded coherent guidelines for management of hypertension, the second part of the dissertation presents the results of an evaluation of the effectiveness of common anti-hypertensives in a cohort of HIV-positive and matched HIV-negative veterans. In this analysis, we use methods in causal inference for adjustment of time-varying confounding to estimate the effect of these anti-hypertensives on the probability of adverse cardiovascular outcomes. In the third part of the dissertation, we describe a model of hospital capacity which we created to aid hospital administrators with critical care capacity planning during the COVID-19 pandemic. We use individual-level electronic health record data to simulate the expected need for critical care resources during surges in COVID-19 cases.

In Chapter 1, we describe four key threats to the validity of randomized controlled trials (RCT) of biomarker targets. RCTs are usually used to evaluate an intervention, such as a drug or procedure, which is used to treat a particular medical condition. However, RCTs of targets compare goal biomarker measurements that a physician attempts to achieve using a variety of treatment strategies. The choice of these treatment strategies is usually left to the physician. RCTs of biomarker targets are common in medicine, but clinical guidelines based on RCTs of biomarker targets have been controversial. One example is the controversy surrounding guidelines of systolic blood pressure targets; national organizations of physicians disagree regarding the optimal blood pressure target, despite the existence of many RCTs of systolic blood pressure targets. We identify four reasons, from the perspective of causal inference, that RCTs of biomarker targets might produce divergent results. We discuss these findings in a case study consisting of nine randomized trials of systolic blood pressure targets.

In Chapter 2, we evaluate the effect of two major classes of anti-hypertensives, angiotensin converting enzyme inhibitors (ACE-I) and hydrochlorothiazide (HCTZ), on cardiovascular outcomes in a population of HIV-positive and matched HIV-negative veterans with hypertension. The study population is selected from the Veterans Aging Cohort Study, a longitudinal and

prospective virtual cohort of HIV-positive veterans and their matched controls. We selected subjects from the VACS cohort who had hypertension as well as one or more additioanl cardiovascular risk factors. We use marginal structural Cox models with stabilized inverse probability of treatment weights to adjust for baseline and time-varying confounding. We find that after adjustment for possible confounders, treatment with ACE-I and HCTZ did not results in a decreased probability of occurrence of adverse cardiovascular events relative to other regimens of anti-hypertensives.

In Chapter 3, we describe the creation of a model designed to help hospital administrators meet the critical care needs of COVID-19 patients during surges in COVID-19 cases. The dynamic nature of the COVID-19 pandemic pose a substantial challenge for optimizing resource allocation to critical care units in hospitals, especially given the high demand for critical care resources among COVID-19 patients and mortality rates which would ensue if critical care capacity is exceeded. We use individual-level electronic health record data to build a model that simulates the movement of COVID-19 patients between the emergency department, the floor departments, and the critical care units of a hospital. The model can be tailored to other hospitals and hospital systems based on dynamics observed in those systems. We present the model structure and dynamics and procedure for parameter estimation. We also implemented the model in a publicly available web application and provide evidence that the model is well-calibrated using patient trajectories observed at Yale-New Haven Hospital.

Chapter 2

Randomized controlled trials of biomarker targets

Abstract

Randomized controlled trials (RCTs) are used to estimate the causal effect of a treatment on a health outcome of interest in a patient population. Often the specified treatment in an RCT is a medical intervention – such as a drug or procedure – experienced directly by the patient. Sometimes the "treatment" in an RCT is a target – such as a goal biomarker measurement – that the patient's physician attempts to reach using available medications or procedures. Large RCTs of targets are common in clinical research, and trials have been conducted to compare targets in the management of hypertension, diabetes, anemia, and acute respiratory distress syndrome. However, different RCTs intended to evaluate the same targets have produced conflicting recommendations and meta-analyses that aggregate results of trials of targets have been inconclusive. In this paper, we explain why RCTs of the same targets conducted in different patient and physician populations can arrive at starkly different results. We describe four key threats to the validity of trials of targets: 1) intention-to-treat analysis that

conflates the effects of assignment to a biomarker target with interventions delivered to the patient; 2) incomparability in results across trials of targets; 3) time-varying adaptive treatment strategies; and 4) Goodhart's law, "when a measure becomes a target, it ceases to be a good measure." We illustrate these findings using evidence from nine RCTs of blood pressure targets for management of hypertension.

Keywords: Intention-to-treat, randomized controlled trial, causal inference, meta-analysis, hypertension

2.1 Introduction

Randomized controlled trials (RCTs) reveal the effect of an intervention on health outcomes of patients who receive it, compared to those who do not. Random assignment of patients to treatments ensures that observed differences in patient outcomes following receipt of different treatments can be attributed to the effect of treatment [6]. Often, treatments compared in RCTs are well-defined clinical interventions, such as drugs or procedures, that are received directly by the patient. These RCTs report an estimate of the causal effect of the interventions tested in the trial. Other RCTs compare *targets* – such as biomarker measurements – that the patient's physician attempts to reach by administering drugs or procedures that may or may not be specified in the trial protocol. Randomized trials of targets measure the causal effect of assignment to a particular target, and not the effect of any particular intervention [7].

Trials of clinical targets are common in clinical medicine and have been used to establish management strategies in chronic diseases like hypertension, diabetes, and post-operative anemia [1, 8, 9]. For example, hypertension is treated by lowering blood pressure to a systolic blood pressure (SBP) target set in clinical guidelines published by national organizations [10, 11]. Large RCTs of SBP targets have been performed to determine the optimal blood pressure tar-

get for managing hypertension [12]. These RCTs assign patients to “standard” or “intensive” blood pressure targets. Physicians treating a patient assigned to a particular target choose treatments, which may include lifestyle modifications and pharmacologic therapy, to bring the patient’s SBP as close as possible to the assigned SBP target. Several studies have demonstrated that blood pressure control decreases risk of poor cardiovascular outcomes, such as stroke and myocardial infarction [13–18]. Similarly, RCTs of strategies for type 2 diabetes mellitus (T2DM) management have compared “standard” and “intensive” hemoglobin A1c (HbA1c) targets of above 7% or below 7% [19]. Glucose control has been shown in some studies to delay the development of complications due to diabetes mellitus [20–22]. For management of patients with severe anemia requiring blood transfusion, RCTs comparing liberal and restrictive blood transfusion targets corresponding to hemoglobin concentrations of 10 g/dL and 8 g/dL have also been conducted, showing no difference between the two strategies [9].

However, trials of targets have generated controversy because different randomized trials of the same targets have produced divergent results. Despite the existence of numerous RCTs of SBP, HbA1c, and hemoglobin targets, the optimal target and treatment strategies for these diseases remain unclear [8, 23]. For example, national organizations of physicians have published conflicting clinical recommendations for SBP targets. The American College of Cardiology (ACC) and the American Heart Association (AHA) published a new joint clinical guideline in 2017 recommending that physicians target a SBP of <130 mmHg in patients over the age of 60 with cardiovascular risk factors [10, 23]. The American College of Physicians (ACP) and American Association of Family Physicians (AAFP) declined to endorse these guidelines and continues to recommend targeting the previous target of SBP <140 mmHg in this patient population [11]. Physicians have expressed concerns about the side effects of aggressive treatment strategies for lowering blood pressure, including increased risk of falls in vulnerable elderly patients and adverse renal events due to high doses of anti-hypertensives [23, 24]. Recently epidemiologists and physicians have described the potential pitfalls of trials in which treatments are ill-defined [7, 25, 26].

In this paper, we use principles of causal reasoning to explain why trials of targets can produce inconclusive or contradictory results. First, trials of targets are analyzed according to the "intention-to-treat" principle to estimate the effect of a guideline – a biomarker target the physician attempts to reach with available treatments – rather than the treatment actually delivered by the physician. Biomarker target values are not well-defined medical interventions received by the patient, since physicians may choose a variety of treatment strategies in order to reach them. Second, effect estimates from trials of targets may not be comparable across studies and populations. Meta-analyses aggregating estimates from trials of targets into clinical guidelines using standard statistical methods may yield invalid estimates of the effect of the target, even when each trial is internally valid. Third, treatment regimens delivered to patients in trials of targets are typically adaptive and time-varying: future treatments depend both on past treatments and observed clinical response to past treatments. The adaptive treatment strategies used by physicians are not compared in trials of targets, and may not be amenable to comparison using randomization at all. Fourth, intervening to alter a clinical biomarker as a therapeutic target may compromise the prognostic value of any subsequent observation of that biomarker, including the achieved target value. Finally, we present a case study of evidence from nine randomized trials of blood pressure targets to illustrate the pitfalls of creating clinical guidelines using trials of targets.

2.2 Trials of targets are analyzed using the intention-to-treat principle

The intention-to-treat (ITT) principle posits that pragmatic trials should be analyzed using the randomized assignment to treatment as the exposure of interest [27, 28]. ITT analysis of randomized trials is justified by several related arguments. First, ITT analysis recovers the effect of the physician's intention to subject the patient to a given intervention. Real-world complica-

tions like non-compliance and protocol violations are built in to the estimate, possibly making the results more relevant to what might occur if the intervention were incorporated into routine clinical practice [29–31]. Second, because assignment is randomized, the ITT effect is unbiased in the sense that the comparison of outcomes under different assignments is unconfounded by common causes of assignment and outcome. Third, ITT analysis frees the investigator from the burden of collecting detailed information about actual treatments delivered to patients, and analysis consists of a simple comparison of outcomes under different assignments. Fourth, in traditional placebo-controlled double-blind RCTs, compliance with the assignment is not expected to differ across assignments, so ITT analysis has the additional virtue of recovering a null effect if the treatment truly has no effect on the outcome [7]. In placebo-controlled RCTs with differential noncompliance, ITT analysis is still expected to yield a conservative estimate of the effect of treatment (biased toward the null hypothesis of no treatment effect) because the group of patients assigned the active treatment is contaminated with those who actually receive no treatment [32].

Trials of targets are usually conducted and analyzed according to the ITT principle. For example, the Systolic Blood Intervention Trial (SPRINT), an influential trial of BP targets, randomly assigned patients to systolic BP targets of 120 mmHg or 140 mmHg and was analyzed using an ITT approach. Physicians targeted these BP goals by prescribing a combination of any major class of antihypertensive agents. Use of classes with strong evidence for cardiovascular benefits, like chlorthalidone and amlodipine, was encouraged but not required [17]. As prescribed by the ITT principle, SPRINT compared patients according their random assignment without regard for factors that could have affected the outcome, like compliance, loss to follow up, or medication prescribed to achieve the target. This analysis, which does not exclude any trial subjects, reveals the likely effect of the intervention if it were used outside of the trial context and avoids compromising the balance in possible confounding factors achieved with randomization [33].

To illustrate, consider a hypothetical randomized trial of the effect of SBP targets on cardiovascular (CV) events in patients with hypertension. Let Z denote the randomly assigned SBP target of 120 mmHg or 140 mmHg. Let Y denote the occurrence of an adverse CV event at a suitable follow-up time. Let X be the actual treatment chosen by the physician, e.g. lifestyle changes, anti-hypertensive medications, a procedure, or a combination of these. Let L denote physician and patient features that affect both the actual treatments delivered and the patient's eventual health outcomes. Patient features included in L could be demographics, co-morbidities, and baseline laboratory measurements; physician features could be medical specialty, patient load, or hospital affiliation.

Figure 2.1 shows a directed acyclic graph representing the causal structure of the variables Z, X, Y, and L [34, 35]. In a trial of targets, the assigned target Z is a guideline interpreted and implemented by the physician, who uses his or her discretion to choose the treatment regime X with which to reach the target. This treatment regime X is therefore a causal consequence of the randomized assignment Z to a SBP target and the patient and physician features L. The CV outcome Y is determined by these features L, the actual treatment X delivered to the patient, and possibly the assignment Z. The dashed arrow from Z to Y represents the possibility that the assigned target Z exerts a direct causal effect on Y. In a trial of targets, this $Z \rightarrow Y$ effect may exist because physicians are not blinded to the assignment Z, leaving open the possibility that there is an effect of Z not mediated by the particular treatment X received by the patient. ITT analysis of trials of targets amounts to comparison of Y under different assigned targets Z, ignoring the treatment regime X received by the patient. Typically some elements of L, including baseline patient demographics, are measured but not used in computing the primary ITT effect estimate.

Using principles of causal reasoning, critics of ITT analysis have pointed out that ITT analyses of RCTs can suffer from serious pitfalls [25, 26]. First, ITT analysis may not recover the causal effect of any particular intervention X [7, 27]. Second, subjects assigned to a

particular target $Z = z$, but who never reached that target, are nevertheless included in the "treated" group, despite the fact that they contribute no information about the effect of actually reaching the target [28]. In trials of targets, physicians cannot be blinded to assigned target Z because they must use this information to specify a treatment regime X. Moreover, trials of targets usually implicitly involve two or more active treatments. For example, physicians in SPRINT were encouraged to use chlorthalidone and amlodipine but were provided with all major anti-hypertensive classes and permitted to use any combination according to their individual judgment. As a result, physicians may employ starkly different treatment strategies for patients assigned to one target or the other, so that differential treatment strategies between assigned targets confound the ITT estimate. While the control arm in placebo-controlled RCTs is usually not contaminated by patients receiving the active treatment, the group of patients assigned to the control arm in a trial of targets may be treated using strategies also utilized in the treatment arm. For this reason, ITT analysis of trials of targets may not enjoy the desirable (conservative) properties of ITT analysis, as traditional placebo-controlled double-blind RCTs of interventions do.

When assignment Z exerts no direct effect on the CV outcome Y (the $Z \to Y$ arrow is absent in Figure 2.1, the ITT effect from a trial of targets is null when the treatment X has no effect on Y. However, when Z has an effect on Y and the actual treatment X delivered to the patient has no effect whatsoever on the outcome Y, the ITT effect from a trial of targets can nevertheless be non-null. Likewise, when biomarker target assignment directly affects the outcome but not the delivered treatment, the ITT effect can still be non-null. Formal proofs of these facts are given in Appendix 2.7.1.

Evidence from SPRINT suggests that differential direct effects of biomarker target on outcome can arise in trials conducted to compare BP targets [23]. In SPRINT, the group assigned to the "intensive" treatment arm had 30% more clinic visits during the study than the control arm. Increased frequency of clinic visits may affect health outcomes through mechanisms other than

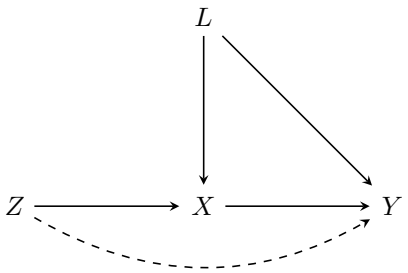

Figure 2.1: The causal structure of a hypothetical RCT of systolic blood pressure targets Z on cardiovascular outcomes Y. Assignment Z to a target SBP (e.g. 120 mmHg versus 140 mmHg) is randomized by the experimenter. The treatment actually received by the patient is X, and patient/physician features are represented by L. Trials of targets often collect limited information on actual treatments X, and instead analyze the trial using the intention-to-treat principle by comparing outcomes Y under different target assignments Z. The dashed arrow represents the possible "direct" effect of the assignment Z on the outcome Y. In a trial of targets, this $Z \rightarrow Y$ effect may exist because physicians are not blinded to the SBP target assignment.

management of blood pressure. For example, many patients increase compliance with prescribed treatments around the time of follow-up clinic visits and decrease compliance between visits [36]. Patients with hypertension who have immediate follow-up appointments after experiencing adverse effects of anti-hypertensives are significantly more likely to achieve blood pressure goals [37], and patients with more frequent clinic appointments may be screened more effectively. Assignment to the "intensive" arm of SPRINT could have affected patient outcomes directly by increasing clinic visits, through mechanisms other than anti-hypertensive medications.

2.3 Trials of targets in different populations may not be comparable

Consider two hypothetical trials of SBP targets Z (120 mmHg or 140 mmHg) on a CV outcome Y. Assume that in both trials: 1) the same target biomarker levels are being compared, 2)

randomization imposes an identical distribution on assignment Z to a target biomarker level, 3) the same outcome measure Y is used, and 4) the biological mechanism of the treatment effect on the outcome is the same within strata of L, so the outcome Y given X and L has the same distribution in both trials. In each trial, the investigators estimate the effect of the randomized assignment Z on the CV outcome Y using the same outcome measure. In this setting, each trial is internally valid, in the sense that it produces an unbiased estimate of the causal effect of assignment Z on the outcome Y, for the particular patient and physician group under study. But in what sense are the two estimates comparable? Can investigators aggregate evidence from trials 1 and 2 via meta-analysis to compute a "combined" estimate of the difference in average CV outcomes under the two SBP targets?

Unfortunately, the two trials of the same targets may not be comparable because of differences in the patient and physician populations. It is well known that ITT analyses of different RCTs may be incomparable, even when the biological effect of active treatment is the same [7, 38]. The patient and physician features are represented respectively by the distribution of L. Underlying health conditions may influence the physician's choice of X as well as outcome Y, so these features are included in L. For example, the ACC/AHA guidelines for systolic BP targets are based on trials which included patients with different co-morbidities. Some trials included only diabetics, some trials excluded diabetics, and others included only patients with chronic kidney disease [10, 14, 17, 39].

The distribution of X given Z and L represents the prescribing behavior of physicians in a given trial. Trials of targets may not establish a protocol specifying the medications which should be used. This flexibility allows physicians to use their individual judgment to tailor treatments to individual patients, but as a result prescribing behavior may vary across trials. However, assignment to a given target can obscure unmeasured variation in actual treatments delivered to patients. In trials of targets, this "consistency" assumption may be violated when there are multiple treatment mechanisms by which a given biomarker target z is reached, such

that the observed outcome is different for the same target z under two different treatment strategies x_1 and x_2 [40–43]. In the case of SBP targets, ACCORD and SPRINT compared the same blood pressure targets: 120 mmHg vs. 140 mmHg. At the end of the trial period of ACCORD, 41% of patients in the intensive treatment group and 16% in the standard treatment group were taking 4 or more anti-hypertensives [14]. At the end of SPRINT, 24.3% of patients in the intensive treatment group and 6.9% in the standard treatment group were taking that number [17]. Thus, patients in SPRINT were likely to receive a larger number of anti-hypertensive medications than patients in ACCORD even though the targets being compared were the same.

Because two trials of the same targets may be incomparable, aggregating or averaging results of trials of targets via meta-analysis may be misleading. Figure 2.2 shows an illustration of two hypothetical trials of the same SBP targets in which the baseline distribution of age L and physicians' treatment strategies X differ across the trials. Here, X is a simplified binary treatment strategy representing, for example, intensive versus standard anti-hypertensive medication. The patient populations in the two trials differ in their distributions of age L (top row), and physicians in the two trials employ different treatment strategies by age, given the assigned SBP target (second row). Physicians in Trial 1 are much more likely to treat patients assigned to the SBP target $Z =$140 mmHg using the intensive ($X = 1$) strategy compared to the standard ($X = 0$) strategy. The majority of patients in Trial 1 have ages L that lie in the range where these treatment differences are largest. In Trial 2, there is only a small difference in the probability of intensive treatment between SBP target assignments, in the age range of patients in that trial. The expected outcome functions represent the biological effect of the treatment within strata of L (third row). These outcome functions are the same in both trials for each value of treatment X. However, the trial results, represented as risk differences (fourth row) are starkly divergent: Trial 1 shows a clear beneficial effect, but Trial 2 shows a null result.

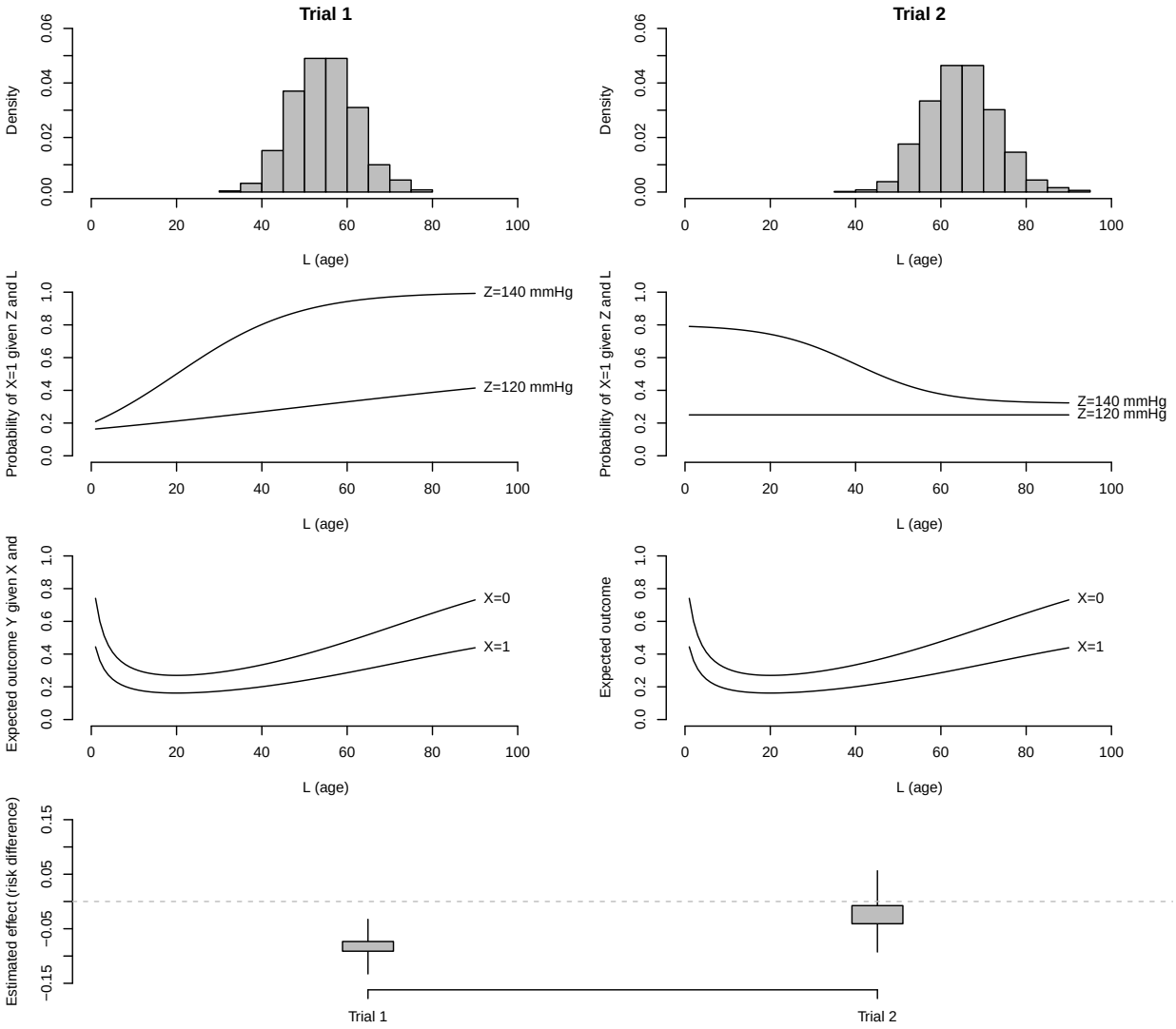

Figure 2.2: Illustration of two hypothetical trials of the same targets producing different effect estimates. Both trials seek to estimate the effect of the SBP target Z =140 mmHg versus Z =120 mmHg on an adverse cardiovascular outcome Y. In the top panel, the distribution of patient age L differs: participants in Trial 2 are older on average than those in Trial 1. In the second panel, the probabilities of intensive treatment under assignment to Z differ across age in the two trials. In the third row, the expected outcome functions, representing the biological effect of interventions and risk factors, are the same in both trials. The fourth row shows that the trials give starkly different results, shown as risk differences: Trial 1 shows a benefit of assignment to 140 mmHg versus 120 mmHg, while trial 2 is inconclusive.

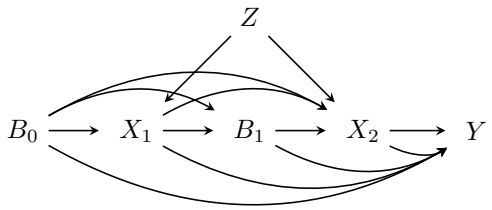

Figure 2.3: Illustration of the causal structure of time-varying treatments in a simplified trial of targets with two longitudinal time points. Biomarker measurements are taken at baseline (B_0) and following the first administration of treatment (B_1). Treatments are represented by X_1 and X_2, which are affected by prior biomarker measurements, prior treatments, and the target assignment Z. The outcome Y is measured following administration of X_2, and is a function of all prior biomarker measurements and treatments. Covariates L are omitted for simplicity. Because B_1 is simultaneously a causal consequence of X_0 and a cause of X_1, special adjustment approaches are required to estimate the effect of (X_1, X_2) on Y.

2.4 Trials of targets employ adaptive time-varying treatments

In trials of targets, physicians deliver a temporal sequence of treatments X whose types and doses are adjusted over time based on patient response and biomarker target goals. In the case of hypertension, physicians may initiate a regimen of anti-hypertensive medications using information obtained in their evaluation of the patient at time of diagnosis. At subsequent clinic visits, the dose, class, and number of medications prescribed to the patient may be adjusted by the physician following observed changes in the patient's health status and response to prior treatments. Trials of targets almost always involve time-varying adaptive treatments because the target biomarker measurement can only be reached through treatments that take time to administer, and whose effects on the patient's biomarker measurements take time to accrue. Unfortunately trials of targets typically do not collect detailed information on the time-varying adaptive treatment strategies that physicians use to reach an assigned target. SPRINT did not specify which anti-hypertensives physicians should prescribe to patients to achieve the target SBP, though the investigators did report aggregated data including the percentage of patients who received medications in each anti-hypertensive class at the last trial visit [17].

Figure 2.3 shows an illustration of the causal structure of a trial of targets in a simplified setting with two follow-up time points. Baseline patient and physician characteristics L are omitted for simplicity. Assignment to a biomarker target is represented by Z, which exerts an effect on the treatments X_1 and X_2 delivered to the patient by the physician. The baseline biomarker measurement B_0 (e.g. SBP at the time of hypertension diagnosis) is unaffected by assignment Z, but the follow-up biomarker measurement B_1 is a consequence of B_0 and the first treatment X_1. Because X_1 is affected by the assignment Z, B_1 is a causal consequence of Z as well. Together, all the biomarker measurements and treatments affect the outcome Y. In more complex scenarios with more time points, the target assignment may affect all treatment variables.

The effect of the biomarker target assignment Z on the outcome Y is unconfounded and estimable when Z is randomized, even if B_0, B_1, X_1, and X_2 are unobserved. But the effect of any particular sequence of treatments (X_1, X_2) on Y requires observing B_0, B_1, X_1, and X_2, and computing the causal effect of (X_1, X_2) on Y. This can be challenging when X_1 and X_2 do not arise in a sequentially randomized trial. Unfortunately, estimating the effect of a time-varying treatment X on Y by modeling Y as a function of the sequence of delivered treatments (X_1, X_2) and biomarkers (B_0, B_1) may be biased when X_1, B_0, and B_1 predict X_2 [44, 45]. As Figure 2.3 shows, B_1 is both a cause of treatment X_2 and a causal consequence of treatment X_1. The dual role of time-varying covariates like B_1 induces bias in naïve estimates of the causal effect of treatment regimens (X_1, X_2) on Y.

Fortunately, it is sometimes possible to estimate the effect of a time-varying treatment even when the sequence of treatments is not sequentially randomized. Under certain unconfoundedness assumptions, the G-formula and marginal structural models can be used to adjust for time-varying covariates to evaluate the effects of time-varying treatment strategies when detailed information about intermediate clinical measurements, outcomes, and treatments delivered to patients are available [46]. The g-formula decomposes the counterfactual causal

estimand of interest into a weighted average of stratum-specific expected treatment effects, conditional on both the treatment and covariate history of the patient. The weight for each stratum-specific treatment effect is the propensity score, the joint density of the observed covariate history conditional on treatment. Both the stratum-specific treatment effects and weights can be estimated from data using a variety of approaches. The g-formula identifies the probability that the event of interest occurs at a given time [47]. Marginal structural models use inverse probability of treatment weights (IPTW) to construct a "pseudo-population" in which the treatment of interest is assigned as if it were random. IPTW are often the inverse of the propensity score, such that subjects who are under-represented in either the intervention or non-intervention arms of the study receive higher weights in the analysis. Pooled logistic models or Cox regression approaches, weighted with IPTW, are also used to identify the hazard of the event [48]. Extensions of these methods have been developed for survival analysis, with the ability to account for censoring.

2.5 Goodhart's law may explain contradictory findings in trials of targets

The economist Charles Goodhart observed in the context of target measures for monetary policy that "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" [49]. The adage was later reformulated in its now-famous form, "when a measure becomes a target, it ceases to be a good measure" [50]. In the context of clinical medicine, we might rephrase Goodhart's law as follows:

> When the population distribution of a biomarker is forced to collapse onto a target value, the biomarker may cease to be a satisfactory predictor of the clinical endpoint of interest.

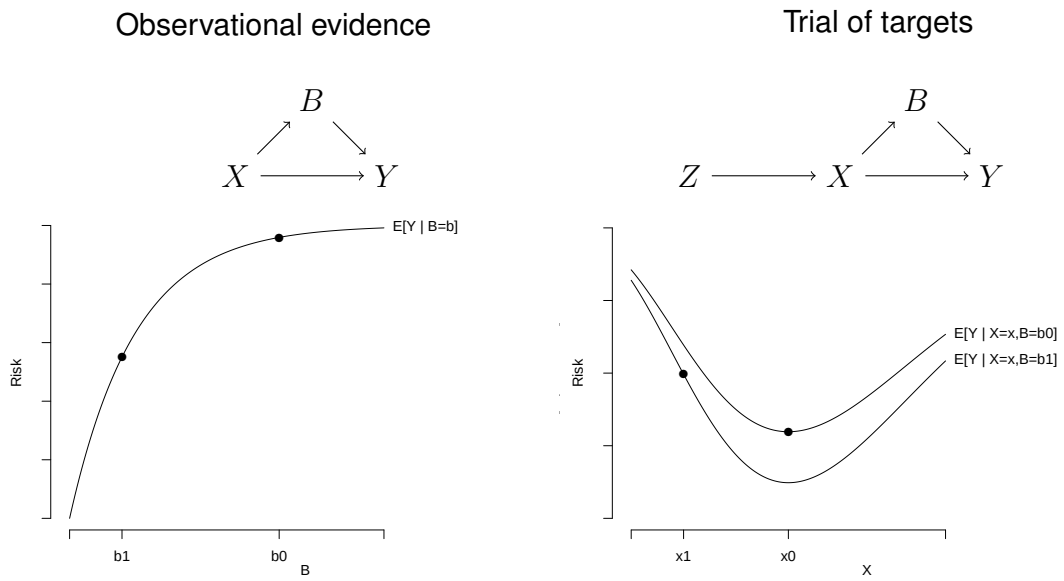

Figure 2.4: Illustration of Goodhart's law when the value of a biomarker B is used as a target for control. In the observational setting at left, larger values of B are associated with worse outcome. Target values $B = b_0$ and $B = b_1$ are chosen to be tested in an RCT. At right, successful targeting in the randomized trial means that physicians choose treatments $X = x_0$ and $X = x_1$ to achieve b_0 and b_1 respectively. The trial reports the risks evaluated at $X = x_0, B = b_0$ versus $X = x_1, B = b_1$, leading to the seemingly contradictory result that lower values of the biomarker cause worse outcome.

In observational (i.e. routine clinical) settings, lower (or higher) biomarker measures may be associated with better prognosis in a particular group of patients, and so the biomarker naturally becomes a target to be controlled. But the interventional means by which a target biomarker measurement is reached need not always improve prognosis, because the causal mechanism by which the intervention acts on the outcome need not pass solely through the biomarker [51].

Figure 2.4 shows an illustrative example where the biomarker (such as systolic blood pressure) is represented by B. It is affected by the treatment X, and affects the outcome Y. The treatment X also directly affects the health outcome Y. The biomarker B is called a "mediator" because it mediates the effect of X on Y via the pathway $X \to B \to Y$ [52]. At left, observational evidence suggests that the biomarker may be a good target for control because the risk of the outcome Y increases with the value of the biomarker b. The natural distribution

of X in the observational setting is chosen by physicians, and is not assigned randomly. When X takes its natural "observational" distribution, B appears to be a good predictor of Y. In other words, B and Y exhibit an association, marginally across the distribution of X. In particular, some biomarker measurements b_1 and b_0 (e.g. low and standard SBP targets) yield the risk relationship $\mathbb{E}[Y|B = b_1] < \mathbb{E}[Y|B = b_0]$. To study the causal effect of low versus standard biomarker measurements investigators would need to randomize B directly, but they do not have the means to do so. Instead, they conduct an RCT of targets Z, represented by the trial of targets b_1 and b_0 at right. In the trial, physicians use the randomization of the assignment Z to assign patients to a target biomarker level b_1 or b_0.

In Figure 2.4, the causal structure of X, B, and Y remains the same in both settings, but the distribution of the actual treatment delivered X is determined by Z in a trial of targets. Ideally, the assignment Z results in physicians using different treatment strategies, say x_0 and x_1, to successfully reach the target. As a result, patients assigned to b_1 receive treatment x_1 to achieve $B = b_1$; likewise patients assigned to b_0 receive treatments x_0 and achieve $B = b_0$. In other words, the physician's treatment decision X in the trial of biomarker targets is tuned to deliver these targets. But since X exerts a direct effect on Y that is not mediated by B, this change in the distribution of X can have unintended consequences. On the lower right of Figure 2.4, the risk under under $B = b_0$ is indeed always greater than under $B = b_1$ for every particular value of treatment $X = x$. But because the trial alters the distribution of treatment X depending on the target Z, the trial measures these risk curves at different places in the dose-response function, resulting in the erroneous finding that "assignment to b_1 results in higher risk than assignment to b_0". This conclusion is false because the risk under $B = b_0$ dominates the risk under $B = b_1$ for every value of $X = x$. A formal characterization of Goodhart's law in trials of targets is given in Appendix 2.7.2.

This paradoxical relationship between target and outcome is suggested by the case of HbA1c targets and guidelines for management of diabetes. A lower HbA1c is associated with a lower

risk of adverse cardiovascular events in large observational studies, such as UKPDS and EPIC-Norfolk [53, 54]. RCTs of HbA1c targets have compared “intensive” and “standard” HbA1c control for management of T2DM, and clinical guidelines published by the ACP disagree with those published by the American Association of Clinical Endocrinologists (AACE) and the American Diabetes Association (ADA) [55, 56]. The ACP recommends standard control, targeting HbA1c between 7% and 8%, while the AACE and the ADA recommend intensive control targeting HbA1c below 7% or 6.5% [57]. However, recent evidence also demonstrated increased risk of cardiovascular risk in patients with hypoglycemia and low HbA1c, suggesting that there may be a J-shaped curve relating HbA1c control and CV events [58, 59]. Furthermore, some randomized trials have shown an increase in heart failure events following use of one of the newer medications used for glucose control, dipeptidyl peptidase-4 inhibitors [60, 61]. This increase in heart failure events occurred despite successful control of blood glucose levels to the recommended targets [8]. Successfully achieving an HbA1c target may not be sufficient to ensure improved outcomes, although some evidence for improved outcomes at the lower target does exist [53, 54]. It is possible that the causal effects of treatments for diabetes on cardiovascular events are mediated partly through lower HbA1c, but also that off-target effects of these treatments at higher levels of HbA1c control are responsible for higher risk of poor outcomes.

2.6 Case study: Trials of blood pressure targets

2.6.1 Existing controversy in hypertension guidelines

Clinical guidelines for management of hypertension published by the ACC/AHA in 2017 are based on meta-analysis of fifteen randomized trials of blood pressure targets, finding a significant effect of lower SBP target on major cardiovascular events (RR: 0.84, 95% CI: [0.73,

0.99]) and stroke (RR: 0.82, 95% CI: [0.70, 0.96]) [1, 10]. Table 2.2 shows the endpoints reported by the trials included in the meta-analysis and the definitions of those endpoints. The meta-analysis also reported a “marginally significant reduction” in all-cause mortality and risk of MI [1]. The effects of intensive treatment to a SBP target below 130 mmHg were mixed in these trials. Of the trials that were included, only one – SPRINT – reported reductions in major adverse cardiovascular events, all-cause mortality, cardiovascular death, and heart failure. None of the trials reported reductions in the risk of myocardial infarction, and only ACCORD reported a statistically significant reduction in the risk for stroke [1, 23]. Based on this meta-analysis, the resulting ACC/AHA guidelines recommend that for patients over 60 years old with cardiovascular risk factors, physicians target a SBP of 130 mmHg.

However, the ACP and AAFP declined to endorse these guidelines, and clinical experts have expressed concerns about the possibility of adverse events of aggressive treatment. Intensive treatment with anti-hypertensive medications to low SBP targets raises concern for increased risk of falls due to hypotension and adverse renal events due aggressive anti-hypertensive regimens [23, 62–64]. The meta-analysis did not assess an aggregate risk of falls and did not identify an increased risk of renal events, despite reports in SPRINT that “intensive” blood pressure control was associated with higher rates of acute kidney injury. In particular, some studies have observed higher rates of adverse cardiovascular and renal events at both high and low blood pressures, also called a J-curve, indicating that any benefits of lowering BP may be outweighed by adverse effects of treatment or hypotension [64]. These risks are of especial concern in elderly patients, who may be at risk of increased rates of myocardial infection with low diastolic pressures [65].

2.6.2 Data

We constructed a dataset based on results from the fifteen trials included in the meta-analysis of BP targets by Reboussin et al. [1] that informed the ACC/AHA guidelines. Table 2.1 shows the number of subjects in the trial, key inclusion and exclusion criteria, the targets compared, the primary outcome of interest, the percentage of women in the trial, the percentage of non-white subjects, and baseline age, SBP, and DBP. Reboussin et al. [1] performed an additional analysis to evaluate a strict SBP target less than 130 mmHg using a subset of nine of these trials: ACCORD [14], SPRINT [17], Cardio-Sis [66], SPS3 [67], HALT-PKD [15], HOMED-BP [16], REIN-2 [68], AASK [39], and MDRD [69]. These nine trials each included at least one arm in which the SBP target was less than 130 mmHg. We calculated pooled estimates of the mean and standard deviation of the age and baseline SBP and DBP for each trial population. We calculated relative risks and 95% confidence intervals for each outcome using the number of events reported by each study. We compared each of these relative risks to those found in the Supplementary materials of the ACC/AHA meta-analysis [1] to identify any differences in endpoint construction. We used principal components analysis and forest plots to examine heterogeneity in the results reported by the trials.

We identified several instances in which our extracted dataset differed from that used in the meta-analysis by Reboussin et al. [1]. AASK included both an experimental phase and a period of observational follow-up [39]. The meta-analysis by Reboussin et al. [1] included adverse events observed across both phases. We included only events which occurred in the trial phase. Reboussin et al. [1] included only subjects with CKD at baseline in their estimate of the risk of renal events in SPRINT. We included adverse renal events in both subjects with and without CKD at baseline [17]. Reboussin et al. [1] included only ischemic strokes in the estimate of risk of stroke in JATOS. We included both ischemic and hemorrhagic stroke events to be consistent with the combined stroke endpoints in the other trials [1, 70]. We excluded

the estimate of relative risk of MI in HOMED-BP because only the total number of MIs in the trial, rather than the events per arm, were reported in the publication [16]. We were not able to recover event definitions or to compute relative risks from REIN-2 for endpoints other than all-cause mortality. Instead, we utilized those reported by Reboussin et al. [1] [68]. For unclear reasons, our computed relative risk of adverse renal events in ACCORD, in which we included development of ESRD or need for dialysis, differed from the relative risk reported by Reboussin et al. [1] [14].

2.6.3 Heterogeneity in trial features, management strategies, and endpoints

We discovered substantial variation in the baseline health and co-morbidities of the subjects included in the fifteen RCTs. Three of the trials - ABCD-N, ACCORD, and UKPDS - included only subjects with diabetes (ABCDN, ACCORD, UKPDS). Two trials, Cardio-Sis and SPRINT, excluded all diabetics. Three trials - AASK, MDRD, and REIN-2 - included only subjects with nephropathies or chronic kidney disease, and HALT-PKD included only subjects with autosomal dominant polycystic kidney disease (ADPKD). The average age of subjects included in the trials ranged from 36.6 to 76.6 years. The average baseline SBP of subjects ranged from 126.7 to 169.5 mmHg. Only four of the fifteen studies were composed of predominantly women. Eight trials reported the percentage of non-white subjects included in the trial, which ranged from all white subjects (Cardio-Sis [66]) to all African-American subjects (AASK [39]). MDRD did not report the total percentage of non-white participants, but 7.9% of the subjects were African-American. VALISH [71] and the study by Wei, et al. [72] did not report the racial composition of their sample, but were performed in Japanese and Chinese populations respectively.

The fifteen trials reported only limited information about the treatment regimens utilized by

physicians to achieve the targets. We discovered that the distribution of drug classes used in each trial varied considerably. Many trials reported the percentage of patients on ACE-I, diuretics, BB, and CCB, either one year post-randomization or at the end of the study. The percentage of patients on an ACE-I, diuretic, BB, or CCB ranged from 2.1-100%, 8.9-54.9%, 5-92%, and 7-79% respectively. Furthermore, we found that the decision rules and algorithms used to select the treatments varied substantially, and that treatment choices were often left to the discretion of treating physicians. ACCORD, MDRD, SPS3, and SPRINT encouraged treatment according to previously established guidelines but otherwise did not specify either first-line drugs or randomize use of particular drugs [14, 17, 67, 69]. AASK, HOMED-BP, and Wei et al. randomized both blood pressure target, and use of a combination of ACE-I, ARB, BB, diuretics, or CCB as the first-line anti-hypertensive [16, 39, 72]. ACBD-N randomized patients in the intensive control group to either nisoldipine or enalapril as first-line treatment, and those in the standard control group to placebo [13]. UKPDS randomized those in the intensive treatment group to either captopril or atenolol at baseline, and those in the standard control group were treated throughout the study without ACE-I or beta-blockers, if possible [18]. All patients in HOT and JATOS received a dihydropyridine CCB at baseline, with addition of other medications to achieve the desired BP control [70, 73]. All patients in REIN-2 received ramipril at baseline, with addition of felodipine in the intensive treatment group [68]. HALT-PKD utilized a factorial design in which patients were assigned to either intensive or standard control as well as a regimen of either lisinopril and telmisartan or lisinopril and placebo [15]. Patients in VALISH received valsartan as the first-line anti-hypertensive [71].

We also noted variation in the definition of cardiovascular endpoints across trials. Table 2.1 lists the endpoints analyzed by Reboussin et al. [1]: all-cause mortality, CV death, MACE, MI, stroke, heart failure, and renal events. All the trials included reported ACM, and at least half of the studies reported each of the other endpoints. The construction of the composite endpoint, MACE, varied the most across trials. All of the studies included CV death and non-

fatal MI in their definitions. Seven studies included non-fatal stroke, and five studies included heart failure. CardioSis, JATOS, and SPRINTconsidered angina pectoris to be a MACE, and JATOS additionally included any obstructive arterial disease, abdominal aortic aneurysm, aortic aneurysm, and sudden death as MACE [17, 66, 70]. Definitions of renal events were similarly variable. While all the trials included diagnosis of end-stage renal disease (ESRD) in their definitions of renal events, three studies additionally included initiation of dialysis, and three also included either 1.5 or 2-fold increases in serum creatinine. Renal endpoints in SPRINT differed for subjects based on baseline renal health: in subjects with CKD, a decrease in estimated glomerular filtration rate (eGFR) of >50% was considered a significant adverse event, and in those without CKD, a decrease in eGFR of >30% was sufficient [17].

Reboussin et al. [1] reported a significant effect of lower SBP targets on the relative risk of several cardiovascular endpoints, but the individual trials reported heterogeneous, and mostly null, treatment effects. Reboussin et al. [1] reported that overall, lower BP targets protected against stroke, MACE, and MI, with a marginally significant effect for ACM [1]. However, none of the individual trials reported a significant improvement in rates of MI. In the subsample of nine trials examining intensive SBP control below 130 mmHg, lower BP targets were significantly associated with improvements in rates of stroke and MACE, but not MI. Of these nine trials, ACCORD was the only trial which reported a significant improvement in stroke risk. SPRINT, HALT-PKD and Cardio-Sis reported improvements in risk of MACE with intensive SBP control below 130 mmHg. However, SPRINT and Cardio-Sis excluded diabetics from their study, a population with increased risk of poor cardiovascular outcomes. In HALT-PKD, both trial arms were assigned to targets <130 mmHg, and the patients were young and all had ADPKD. These results may not be generalizable to a broader population.

Trial	N	Criteria	Targets	Primary outcome	Age	SBP	DBP	%F	%NW
AASK [39]	1094	HTN & CKD	MAP: <92 v. 102-107	Progression of CKD	60.1 (10.2)	148 (23.5)	94.5 (13.5)	27.1	100
ABCD-N [13]	480	T2DM & no HTN	DBP: <75 v. 80-89	Change in 24-hr Cr clearance	59.1 (0.6)	136.4 (0.9)	42.8 (60.1)	45.4	26.5
ACCORD [14]	4734	T2DM, risk factors for CVD	SBP: <120 v. <140	[C] MI, non-MI ACS, stroke, HF, CV death	62.2 (6.9)	139.2 (15.8)	76 (10.4)	47.7	39.5
CardioSis [66]	1110	Without T2DM	SBP: <130 v. <140	LVH after 2 years	67 (7)	163.3 (11.2)	89.6 (8.8)	58.8	0
HALT-PKD [15]	558	ADPKD	120/70-130/80 v. 95/60-110/75	% change in total kidney volume	36.6 (8.3)	126.7 (13.9)	80.1 (11.1)	49.3	7.3
HOMED-BP [16]	3518	Mild-moderate HTN	125/90 v. 125-134/80-84	[C] CV death, non-fatal MI, non-fatal stroke	59.6 (10.1)	151.6 (12.5)	90 (10.1)	50.1	NR
HOT [73]	1879(	HTN, elevated DBP	DBP: <80/85 v. <90	[C] MI, stroke, CV death	61.5 (7.5)	170 (98.8)	105 (98.8)	47.0	NR
JATOS [70]	4418	HTN, SBP > 160 mmHg	SBP: <140 v. 140-160	[C] Cerebral hemorrhage or infarction, TIA, SAH	73.6 (5.3)	171.6 (9.8)	89.1 (9.5)	61.1	NR
MDRD [69]	840	CKD	MAP: <92 v. <107	Rate of decline of GFR	51.7 (12.4)	130.5 (17)	80 (10)	39.5	NR
REIN2 [68]	335	Non-DM nephropathy, PU	DBP <90 v. <130/80	[C] ESRD, GFR decline, residual proteinuria	53.8 (15.3)	136.7 (16.9)	84.1 (9.7)	25.1	NR
SPRINT [17]	9361	Without DM, risk factors for CVD	SBP: <120 v. <140	[C] MI, non-MI ACS, stroke, HF, CV death	67.9 (9.5)	139.7 (15.6)	78.1 (12)	35.6	42.3
SPS3 [67]	3020	Lacunar stroke	SBP: <130 v. 130-149	All stroke	63 (10.8)	143 (19)	78.5 (10.5)	37.0	49
UKPDS [18]	1148	T2DM	<150/85 v. <180/105	Adverse event or death related to DM, ACM	56.4 (8.1)	159.3 (19.3)	94 (9.7)	44.5	13.3
VALISH [71]	3079	HTN, age 70-85 years	SBP: 140 v. 140-149	[C] Sudden or CV death, stroke, MI, hospitalization, renal dysfunction	76.1 (4.1)	169.5 (7.9)	81.5 (6.7)	62.5	NR
Wei [72]	724	HTN, age >70 years	SBP: 140 v. 150	Stroke, MI, CV death	76.6 (4.6)	159.5 (16.5)	84.2 (9.6)	33.7	NR

Table 2.1: Trials included in the ACC/AHA meta-analysis, including the number of subjects, BP targets, primary outcome, average age, baseline SBP, baseline DBP, the percentage of women, and the percentage of non-white (NW) participants in the trial. *Abbreviations*: CV - cardiovascular; CKD - chronic kidney disease; MAP - mean arterial pressure; T2DM - Type 2 diabetes mellitus; CVD - cardiovascular disease; ACS - acute coronary syndrome; MI - myocardial infarction; HF - heart failure; ADPKD - autosomal dominant polycystic kidney disease; TIA - transient ischemic attack; SAH - subarachnoid hemorrhage; GFR - glomerular filtration rate; ESRD - end-stage renal disease; ACM - all-cause mortality; LVH - left ventricular hypertrophy, PU - proteinuria; NR - not reported. Superscript [C]: the primary outcome was a composite of the endpoints listed.

Endpoint	ACC/AHA definition	References
All-cause mortality (ACM)	Death due to any cause	[13–18, 39, 66–73]
Cardiovascular (CV) death	Death due to cardiovascular causes	[13, 14, 16, 17, 39, 67, 68, 70–73]
Major adverse cardiovascular events (MACE)	Composite including a combination of: CV death, stroke, MI, heart failure	[14, 16, 17, 39, 66, 70–73]
Myocardial infarction	Fatal or nonfatal MI, not including angina pectoris	[13, 14, 16–18, 66–68, 70–73]
Stroke	Fatal and nonfatal stroke, not including transient ischemic attacks	[13, 14, 16–18, 39, 66–68, 70–73]
Heart failure	Composite: any of acute decompensated heart failure, CHF of New York Heart Association class II or higher, echocardiography determining left ventricular ejection fraction <40%	[13, 14, 17, 18, 39, 66, 70, 72]
Renal events	Composite: any of End-stage renal disease, doubling of serum creatinine, 50% reduction in estimated glomerular filtration rate, kidney transplantation, progression of CKD, renal failure	[14, 17, 18, 39, 68–71]

Table 2.2: Endpoints analyzed by the ACC/AHA meta-analysis [1] and trials reporting these endpoints.

2.6.4 Principal components analysis

The trials did not reveal any obvious relationships between their underlying population and design features and their reported results. We conducted principal components analysis to investigate the trial features - the mean age, baseline systolic blood pressure, percentage of women, and the SBP targets compared - and the relative risks of each endpoint. Figure 2.6 shows the projection of the first two principal components in relation to the trial results. The first two principal components explained between 81.5 and 96.2% of the variance between trials. Lower and upper SBP targets were highly correlated in all analyses, as expected. The age and the percentage of women in the trial appear to vary independently, except in the case of trials which reported stroke outcomes. The reported effect estimates from the trials are most heterogeneous with respect to these trial features in the case of all-cause mortality. Although most trials reported possible (but mostly not statistically significant) improvements in risk of stroke, MACE and MI with a lower SBP target (Fig. 2.6C, 2.6D), the absence of clustering in the data along any specific axis indicates that the observed features do not necessarily explain the variance in the results.

The limited data reported by the trials on the frequency of adverse events caused by intensive blood pressure lowering makes it difficult to investigate the possible side effects of aggressive blood pressure management. Clinical experts have identified possible adverse events attributable to aggressive regimens of anti-hypertensives which include hypotention, myocardial infarction, and other adverse cardiovascular and renal outcomes [62–65]. Eight trials reported adverse renal events, none of which reported a significant increase in rates of progression to end-stage renal disease, initiation of dialysis, or renal transplantation (Fig. 2.9) [1]. However, SPRINT did report an increased frequency of acute kidney injury, hypotension, syncope, electrolyte abnormalities and kidney failure in the treatment arm, targeted to SBP 120 mmHg

Trial	ACM	Stroke	MACE	MI
AASK*	0.89 (0.58, 1.35)	0.92 (0.549, 1.54)	0.93 (0.693, 1.26)	
ABCDN	0.92 (0.50, 1.70)	**0.32 (0.104, 0.95)**		1.30 (0.676, 2.50)
ACCORD*	1.05 (0.84, 1.30)	**0.58 (0.388, 0.88)**	0.94 (0.800, 1.11)	0.87 (0.687, 1.09)
CardioSis*	0.79 (0.21, 2.94)	0.44 (0.137, 1.42)	**0.53 (0.296, 0.94)**	0.66 (0.188, 2.33)
HALTPKD*			**0.26 (0.074, 0.91)**	0.52 (0.047, 5.68)
HOMEDBP*	0.87 (0.52, 1.45)	1.25 (0.650, 2.40)	1.04 (0.603, 1.79)	
HOT (80 vs. 85)	1.07 (0.88, 1.29)	0.80 (0.608, 1.06)	0.93 (0.774, 1.11)	0.95 (0.673, 1.35)
HOT (80 vs. 90)	1.10 (0.91, 1.34)	0.95 (0.710, 1.26)	0.94 (0.780, 1.12)	0.73 (0.523, 1.01)
HOT (85 vs. 90)	1.03 (0.85, 1.26)	1.18 (0.899, 1.55)	1.01 (0.844, 1.21)	0.76 (0.551, 1.05)
JATOS	1.28 (0.86, 1.91)	1.04 (0.687, 1.59)	0.93 (0.545, 1.57)	1.00 (0.322, 3.09)
MDRD*	1.62 (0.64, 4.07)			
REIN2*	0.67 (0.11, 3.96)	1.01 (0.063, 15.95)		1.01 (0.063, 15.95)
SPRINT*	**0.74 (0.60, 0.91)**	0.89 (0.631, 1.24)	**0.76 (0.649, 0.90)**	0.84 (0.641, 1.09)
SPS3*	1.06 (0.82, 1.38)	0.83 (0.664, 1.04)		0.91 (0.584, 1.42)
UKPDS	0.83 (0.65, 1.06)	0.58 **(0.368, 0.90)**		0.80 (0.605, 1.05)
VALISH	0.79 (0.47, 1.35)	0.69 (0.366, 1.30)		1.24 (0.334, 4.61)
WEI13	**0.58 (0.43, 0.80)**	**0.58 (0.346, 0.97)**	**0.59 (0.413, 0.85)**	0.99 (0.399, 2.48)
ACC/AHA (<130)	0.92 (0.79, 1.06)	**0.82 (0.700, 0.96)**	**0.84 (0.730, 0.99)**	0.85 (0.730, 1.00)
ACC/AHA (all)	0.89 (0.77, 1.02)	**0.77 (0.650, 0.91)**	**0.81 (0.700, 0.94)**	**0.86 (0.760, 0.99)**

Table 2.3: Relative risks of the four major outcomes included in the ACC/AHA meta-analysis with 95% confidence intervals. The relative risk is the ratio of the proportion of patients who experienced the endpoint in the treatment arm (numerator, lower BP target) versus the control arm (denominator, higher BP target). Starred (*) trials include at least one arm in which subjects were targeted to SBP < 130 mmHg. Effects significantly different from the null value of 1 are bolded. Effect estimates not reported by a particular trial are left blank.

[17]. ACCORD reported a decrease in estimated glomerular filtration rate and an increase in macroalbuminuria in the group targeted to SBP 120 mmHg, interpreted as “signals of possible harm...but the implications of these changes..are uncertain” [14]. The estimate of overall effect reported by each trial of targets incorporates both the potential benefit of reaching a lower blood pressure target and the potential harm arising from off-target effects of larger doses of anti-hypertensives. These differing effects have not been sufficiently disentangled by trials of targets, or a meta-analysis of these trials.

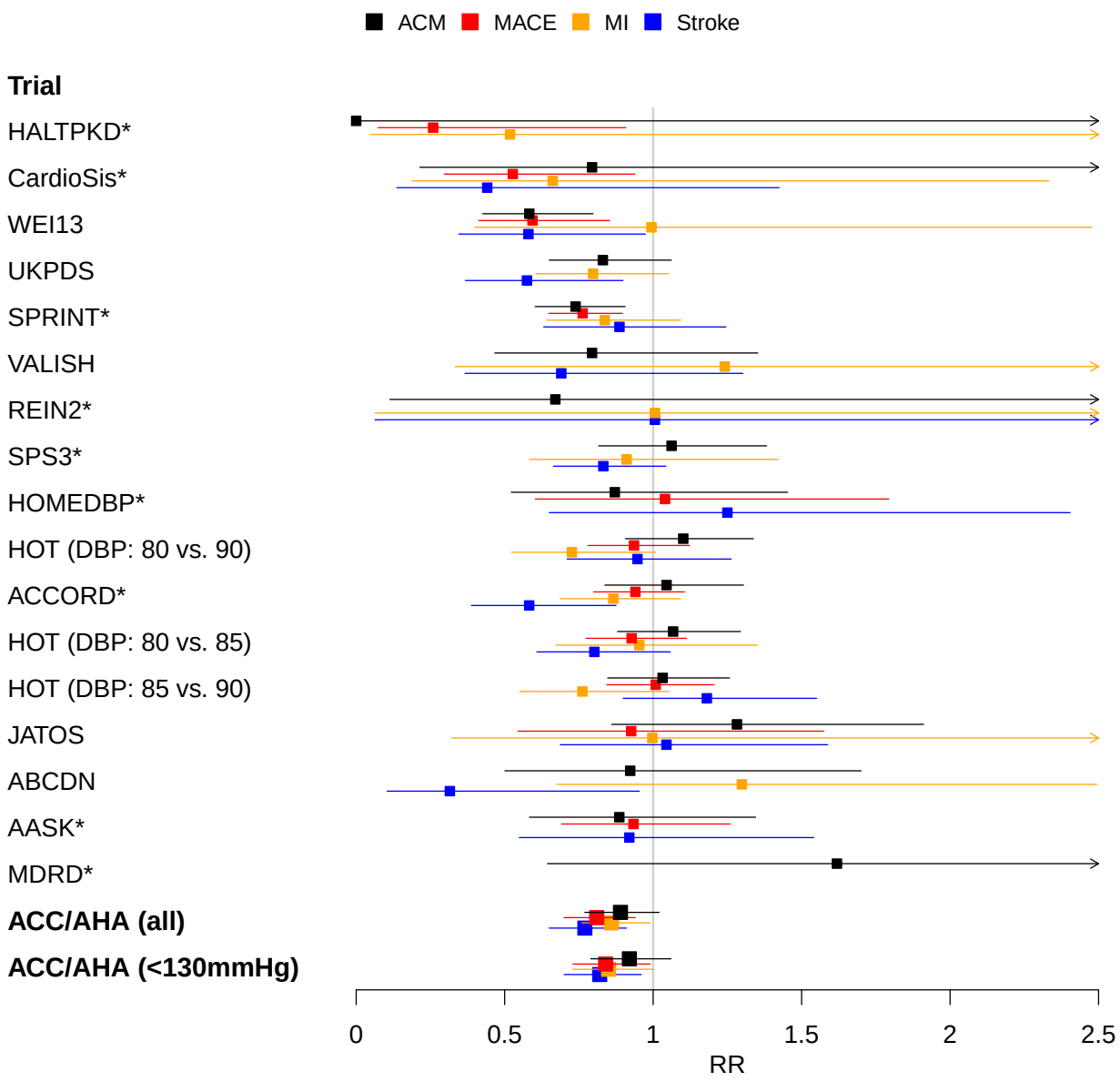

Figure 2.5: Relative risks of four major outcomes for trials included in the ACC/AHA meta-analysis [1]: all-cause mortality (ACM, black), stroke (red), major adverse cardiovascular events (orange), and myocardial infarction (blue). The aggregated results reported by the ACC/AHA meta-analyses are included in the last two entries. Trials are sorted by the magnitude of their average relative risk across the four outcomes (smallest to largest).

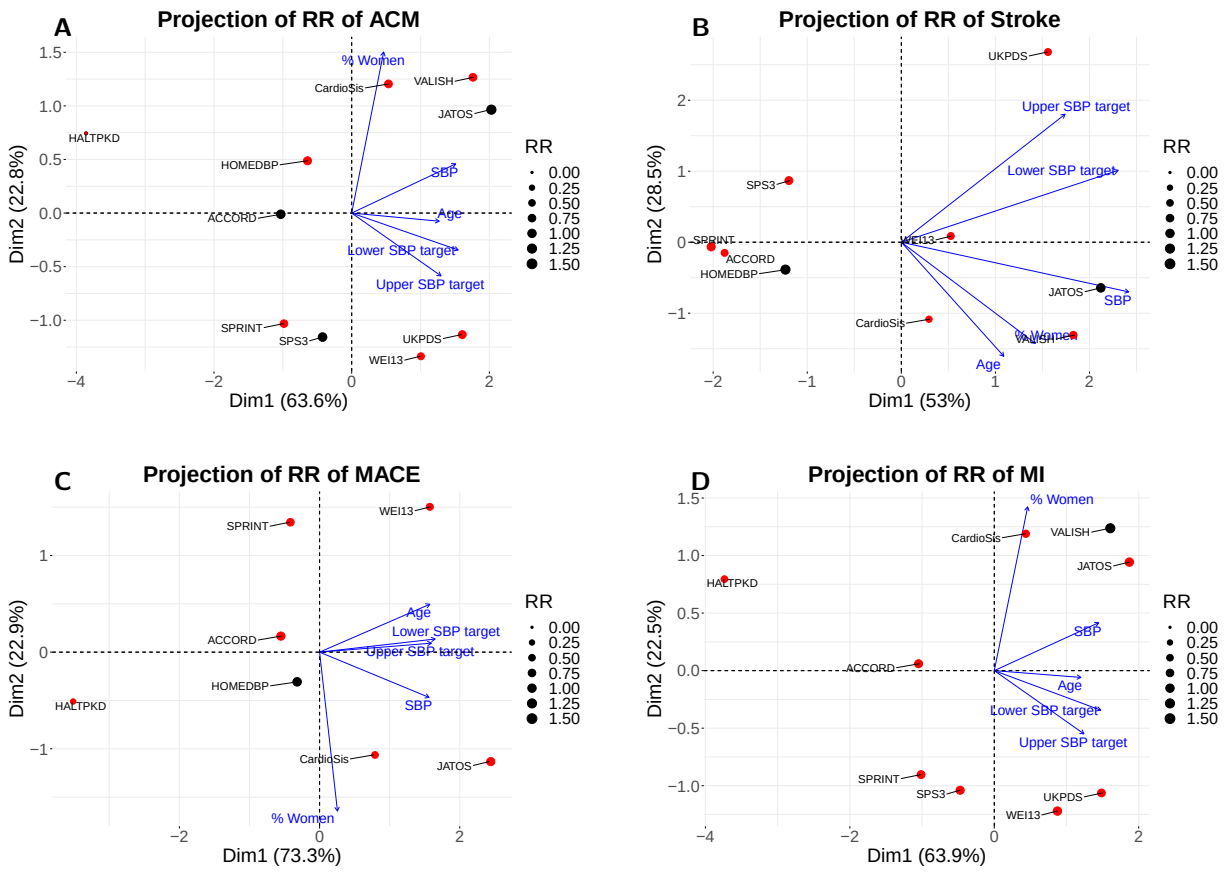

Figure 2.6: Principal components analysis of trial average age, baseline SBP, percentage of women, lower SBP target, and upper SBP target for trials comparing at least one SBP target and reporting relative risk of all-cause mortality (A), stroke (B), major adverse cardiovascular event (C), and myocardial infarction (D). Trials reporting relative risk less than one are shown in red.

2.7 Discussion

Trials of targets measure the effect of assignment to a biomarker target on a health outcome of interest. In this paper, we have shown that trials of targets can give misleading or ambiguous results and that estimated effects may not generalize to other patient populations. Randomized trials of targets, such as blood pressure targets in the context of hypertension, are in fact ITT trials which estimate the causal effect of a guideline on the outcomes of interest. For this reason, trials of BP targets report an estimate of the causal effect of *assignment to a BP target*, rather than the effect of treatments administered to achieve that target or the effect of having that blood pressure as one's baseline health state. We have demonstrated that an ITT trial of targets can produce a non-null result even when the treatments delivered in the trial are assumed to have no effect on the outcome. This problematic result occurs when the distribution of treatments is different in each trial arm. Further, trials investigating the effects of the same targets on the same health outcome in similar patient populations may nevertheless yield different results when physicians' treatment decisions or protocols differ. We have also shown that trials of targets may not be comparable due to variation in the patient and physician populations. Therefore the effect estimates reported by trials of targets are valid estimates of the causal effect of assignment to a target, but are not generalizable to another patient population or population of physicians with different treatment behavior. We have described the role that time-varying treatments play in management of chronic conditions and in trials of targets, and we have shown that the effects of time-varying treatment strategies are not revealed by trials of targets. This phenomenon can also be explained as an illustration of Goodhart's law - "when a measure becomes a target, it ceases to be a good measure" - in clinical practice. The continuing controversy surrounding management of hypertension, encapsulated by the competing guidelines set forth by national organizations, serve as a case study in which we can observe the ambiguities that arise when we base clinical decision-making on evidence obtained from trials of targets.

The weaknesses of trials of targets discussed here may be shared with any randomized trial that shares the same structure, in which the unit of randomization is a guideline and practitioners are free to administer the intervention that they believe will be most useful in achieving the goal. The criticisms set forth illustrate the importance of specifying a well-defined treatment, even in the context of randomized experiments, which are usually considered to be free of confounding biases due to randomization. However, trials without a well-specified treatment protocol, like trials of targets, may remain subject to these biases, which persist due to the multitude of ways that treatment may be assigned depending on characteristics of the population of subjects and practitioners included in the trial. The four challenges to the validity of trials targets listed here also explain the possible pitfalls of conducting randomized trials using surrogate outcomes. For example, as illustrated in the discussion of Goodhart's law, achieved blood pressure could be considered to be a surrogate outcome for adverse cardiovascular events due associations between blood pressure and cardiovascular health in population-level observational studies. A randomized trial using achieved blood pressure targets as the outcome of interest could illustrate that anti-hypertensives successfully lower blood pressure to these targets. However, Figure 2.4B shows that this surrogate outcome is on the causal path between treatment for hypertension and adverse cardiovascular outcomes. Achievement of the target does not guarantee that we have identified an optimal treatment strategy for reducing adverse cardiovascular outcomes.

Addressing the weakness of trials of targets is possible through better trial data reporting, better trial design, and use of modern methods for causal inference. First, future trials of targets should report stratum-specific effect estimates for clinically relevant patient strata. These strata should be standardized as much as possible across trials. The trials should also report detailed information regarding the treatments utilized by physicians in the trial for each stratum of patients. The data reported by the trial should be sufficient to characterize the joint distribution of treatments and patient features in each arm of the trial that meta-analytic methods could control for relevant differences between trials.

Second, treatment strategies, rather than targets, should be compared experimentally when possible. While the reporting requirements discussed above may make aggregation of trials of targets possible, they may pose an insurmountable logistical challenge in the design of future randomized controlled trials of targets. Therefore, we consider careful specification of treatment strategies as the randomized intervention to be a more desirable solution. Beyond data from trials of targets, trials of first-line anti-hypertensive medications have been conducted to compare first-line treatments. Current guidelines recommend initiation of thiazide diuretics, angiotensin-converting enzyme inhibitors, and calcium channel blockers [10]. A large observational study of 4.9 million patients, LEGEND-HTN, compared monotherapies for incident hypertension, concluding that thiazide and thiazide-like diuretics were superior to other therapies [74]. While important, these findings do not guide physician prescribing choices after initiation of treatment. While sequentially randomized trials are difficult and expensive to conduct, several such trials have been implemented, including one which has been used to develop a treatment algorithm for prostate cancer [75]. Four treatments were tested in this trial. Patients were evaluated at 8-week intervals. Patients who responded to treatment were continued on their treatments, and non-responders were randomly assigned to another treatment. Based on this evidence, the researchers determined that a specific regimen of paclitaxel, estramustine, and carboplatin was of interest for further Phase 3 studies. A sequentially randomized trial for hypertension might similarly characterize responders and non-responders at pre-determined clinic visits in a similar way; random assignment to different treatment regimes could occur at these visits for non-responders. Such a study might efficiently identify characteristics of responders, and treatments which are efficient for non-responders.

Third, methods in causal inference have been developed to observationally address the time-varying confounding which results from the use of time-varying treatments. The accumulation of large electronic health record databases increases the potential of these methods to provide meaningful causal conclusions. These databases allow extensive adjustment for sets of possible confounders. Methods which have been used to adjust for time-varying confounding

include inverse probability of treatment weighting [48] and the parametric g-formula [76–78]. These methods have been used in large observational datasets to compare sequences of treatments. For example, Zhang et al. [79] used electronic health records (EHR) to compare erythropoietin dosing strategies targeting low, middle, and high hematocrit targets in patients with end-stage renal disease. The study replicated the findings of RCTs testing the low and high targets. It also estimated the effects of the mid-range target on renal outcomes . IPTW can also be used in the context of dynamic treatment regimes, and development of methods for evaluation of dynamic treatment regimes is ongoing [80]. Application of these methods in the context of hypertension could be used to identify algorithms for deciding on a course of treatment while taking into account a patient's previous response to prescribed medications.

Acknowledgements: This work was supported by NIH grant NICHD DP2 HD091799-01. We are grateful to Amy Justice, Fan Li, Peter M. Aronow, Winston Lin, Janet Tate, and Aldo Peixoto for their input on this work.

Appendix

2.7.1 Intention-to-treat analysis of trials of targets

In this Appendix we study ITT analysis of results from trials of targets under different hypotheses about the causal relationships depicted in Figure 2.1.

No direct effect of Z on Y

When the biomarker assignment in a trial of targets exerts no direct effect on the outcome, ITT analysis recovers the null hypothesis of no treatment effect when this hypothesis is true. Suppose that there exists no direct effect of Z on Y, so the arrow $Z \rightarrow Y$ in Figure 2.1 is

absent. Formally, this means that the distribution of Y given X and L is invariant to the value of $Z = z$, or

$$\mathbb{E}[Y|X = x, Z = z, L = l] = \mathbb{E}[Y|X = x, L = l]$$

for all values of z, x, and l. Suppose further that the treatment X received by the patient has no effect whatsoever on the outcome Y, or

$$\mathbb{E}[Y|X = x, L = l] = \mathbb{E}[Y|X = x', L = l]$$

Then the ITT analysis of the effect of Z on Y is also null. The ITT effect is

$$\begin{aligned}
\mu &= \mathbb{E}[Y|Z = z_1] - \mathbb{E}[Y|Z = z_0] \\
&= \sum_l (\mathbb{E}[Y|Z = z_1, L = l] \Pr(L = l|Z = z_1) - \mathbb{E}[Y|Z = z_0, L = l] \Pr(L = l|Z = z_0)) \\
&= \sum_l (\mathbb{E}[Y|Z = z_1, L = l] - \mathbb{E}[Y|Z = z_0, L = l]) \Pr(L = l)) \\
&= \sum_l \sum_x (\mathbb{E}[Y|X = x, Z = z_1, L = l] \Pr(X = x|Z = z_1, L = l) \\
&\quad - \mathbb{E}[Y|X = x, Z = z_0, L = l] \Pr(X = x|Z = z_0, L = l)) \Pr(L = l) \\
&= \sum_l \sum_x (\mathbb{E}[Y|X = x, L = l] \Pr(X = x|Z = z_1, L = l) \\
&\quad - \mathbb{E}[Y|X = x, L = l] \Pr(X = x|Z = z_0, L = l)) \Pr(L = l) \\
&= \sum_l \sum_x \mathbb{E}[Y|X = x, L = l](\Pr(X = x|Z = z_1, L = l) - \Pr(X = x|Z = z_0, L = l)) \Pr(L = l)
\end{aligned}$$

But since $\mathbb{E}[Y|X = x, L = l] = \mathbb{E}[Y|X = x', L = l]$ for all l, then

$$\begin{aligned}
\mu &= \sum_l \mathbb{E}[Y|X = x', L = l] \Pr(L = l) \sum_x (\Pr(X = x|Z = z_1, L = l) - \Pr(X = x|Z = z_0, L = l)) \\
&= \sum_l \mathbb{E}[Y|X = x', L = l] \Pr(L = l)(1 - 1) \\
&= 0
\end{aligned}$$

Therefore ITT analysis of a trial of targets recovers a null effect when neither the biomarker target assignment Z nor the treatment delivered to the patient X directly affects the CV outcome Y.

Direct effect of Z on Y exists

When the treatment received by the patient has no effect whatsoever on the outcome, but the assignment has an effect on the outcome, ITT analysis of a trial of targets can still produce a non-null result. To understand why, note that if the treatment has no effect on the outcome, this means that the distribution of the outcome Y given the assignment $Z = z$, treatment $X = x$, and coviarate $L = l$ is invariant to x within strata of $L = l$ and $Z = z$. In particular, the expected outcomes under two different treatments must be equal,

$$\mathbb{E}[Y|Z = z, X = x, L = l] = \mathbb{E}[Y|Z = z, X = x', L = l]$$

where $x \neq x'$, for every value of z and l. The ITT analysis of the effect of Z on Y measures

$$\begin{aligned}
\mu &= \mathbb{E}[Y|Z = z_1] - \mathbb{E}[Y|Z = z_0] \\
&= \sum_l (\mathbb{E}[Y|Z = z_1, L = l] \Pr(L = l|Z = z_1) - \mathbb{E}[Y|Z = z_0, L = l] \Pr(L = l|Z = z_0)) \\
&= \sum_l (\mathbb{E}[Y|Z = z_1, L = l] - \mathbb{E}[Y|Z = z_0, L = l]) \Pr(L = l)) \\
&= \sum_l \sum_x (\mathbb{E}[Y|X = x, Z = z_1, L = l] \Pr(X = x|Z = z_1, L = l) \\
&\quad - \mathbb{E}[Y|X = x, Z = z_0, L = l] \Pr(X = x|Z = z_0, L = l)) \Pr(L = l) \\
&= \sum_l \sum_x (\mathbb{E}[Y|Z = z_1, X = x, L = l] \Pr(X = x|Z = z_1, L = l) \\
&\quad - \mathbb{E}[Y|Z = z_0, X = x, L = l] \Pr(X = x|Z = z_0, L = l)) \Pr(L = l) \\
&= \sum_l \sum_x \mathbb{E}[Y|X = x, L = l](\Pr(X = x|Z = z_1, L = l) - \Pr(X = x|Z = z_0, L = l)) \Pr(L = l)
\end{aligned}$$

But since $\mathbb{E}[Y|Z=z, X=x, L=l] = \mathbb{E}[Y|Z=z, X=x', L=l]$ for all z and l, then

$$\mu = \sum_l (\mathbb{E}[Y|Z=z_1, X=x', L=l] - \mathbb{E}[Y|Z=z_0, X=x', L=l]) \Pr(L=l)$$

where x' is an arbitrary treatment. Now if $\mathbb{E}[Y|Z=z_1, X=x', L=l] \neq \mathbb{E}[Y|Z=z_0, X=x', L=l]$ for any x' and l, then μ may not be equal to zero. In particular, if $\mathbb{E}[Y|Z=z_1, X=x', L=l] > \mathbb{E}[Y|Z=z_0, X=x', L=l]$ for every x' and l for example, then $\mu > 0$ even though the actual treatment has no effect on the outcome.

Likewise, if the biomarker target assignment exerts no effect on the physician's treatment decisions, ITT analysis of a trial of targets can nevertheless produce a non-null result. Suppose the probability of treatment $X=x$ is invariant to the assignment $Z=z$, within strata of L, or

$$\Pr(X=x|Z=z_1, L=l) = \Pr(X=x|Z=z_0, L=l).$$

This means that physicians do not take the randomly assigned biomarker target into account when deciding on a treatment strategy. Then the ITT result is

$$\mu = \sum_l \sum_x \Pr(X=x|Z=z, L=l)(\mathbb{E}[Y|Z=z_1, X=x', L=l] - \mathbb{E}[Y|Z=z_0, X=x', L=l]) \Pr(L=l)$$

where z is an arbitrary target assignment. Again, if $\mathbb{E}[Y|Z=z_1, X=x', L=l] \neq \mathbb{E}[Y|Z=z_0, X=x', L=l]$ for any x' and l, then μ may not be equal to zero.

2.7.2 Goodhart's law

To see why modifying treatments to successfully reach a biomarker target can have adverse consequences, let b_0 be the standard target, and let b_1 be the aggressive target. Let the assignment $Z=0$ correspond to b_0 and let $Z=1$ correspond to b_1. Suppose the assignment

Z always results in the patient reaching the target, so sthat $X = 0$ implies $B = b_1$ and $X = 1$ implies $B = b_0$ deterministically. Then the expected outcomes become

$$\mathbb{E}[Y(z = 1)] = \mathbb{E}[Y|X = x_1, B = b_1]$$

and

$$\mathbb{E}[Y(z = 0)] = \mathbb{E}[Y|X = x_0, B = b_0].$$

The RCT therefore reveals the risk difference

$$\delta = \mathbb{E}[Y(z = 1)] - \mathbb{E}[Y(z = 0)] = \mathbb{E}[Y|X = x_1, B = b_1] - \mathbb{E}[Y|X = x_0, B = b_0] \qquad (2.1)$$

in which the outcome is compared under x_1 and b_1 versus x_0 and b_0. Even if the aggressive target is always beneficial

$$\mathbb{E}[Y|X = x, B = b_1] < \mathbb{E}[Y|X = x, B = b_0]$$

for every fixed value of x, the trial result need not show $\delta < 0$.

2.7.3 Case study: Additional figures and tables

We present here several additional figures associated with the case study discussed in Section 2.6. As described in Section 2.6, these figures demonstrate the heterogeneity in results reported by trials included in the meta-analysis by Reboussin et al. [1]. Table 2.2 lists seven endpoints analyzed in the meta-analysis by Reboussin et al. [1]. Four endpoints - ACM, MACE, MI, and stroke - were included in the text. Here, we have included the results reported by the trials for the other three outcomes: CV death, heart failure, and renal events. Figures 2.7, 2.8 and 2.9 show the relative risks for CV death, heart failure, and adverse renal events re-

spectively reported by the trials and the meta-analysis by Reboussin et al. [1]. The relative risks represent the ratio of risk in the lower target group (numerator) versus the higher target group (denominator). Although the meta-analysis by Reboussin et al. [1] reported that targeting an SBP below 130 mmHg was beneficial in reducing the risk of heart failure, only AASK, ACCORD, CardioSis, and SPRINT were included in this subgroup analysis. Of these trials, only SPRINT found a significant benefit in reducing the risk of heart failure 2.8. While the meta-analysis by Reboussin et al. [1] failed to find an increase in adverse renal events, a limited number of trials reported the risk of renal events at all. Of these trials, many included as adverse renal outcomes only severe events such as development of ESRD, a need for kidney transplantation, and initiation of dialysis [14, 17, 18, 39, 68–71]. As discussed in Section 2.6, SPRINT reported an increased frequency of less severe renal events in the intensive treatment group, such as acute kidney injury, hypotension, syncope, electrolyte abnormalities and kidney failure, and ACCORD reported a decrease in estimated glomerular filtration rate and an increase in macroalbuminuria in the intensive treatment group, interpreted as “signals of possible harm” [14].

Figures 2.10, 2.11, and 2.12 present the results of principal components analysis investigating the relationships between trial features and the results reported by the trials, for the outcomes CV death, heart failure, and adverse renal events respectively. This analysis revealed an absence of obvious relationships between trial features and the magnitude and direction of the outcome. The trial features included in the analysis are the average age of the trial population, baseline SBP, percentage of women, lower SBP target, and upper SBP target. The figure displays the projection of these features into the plane of the first two principal components, as well as the result reported by the trial and its direction.

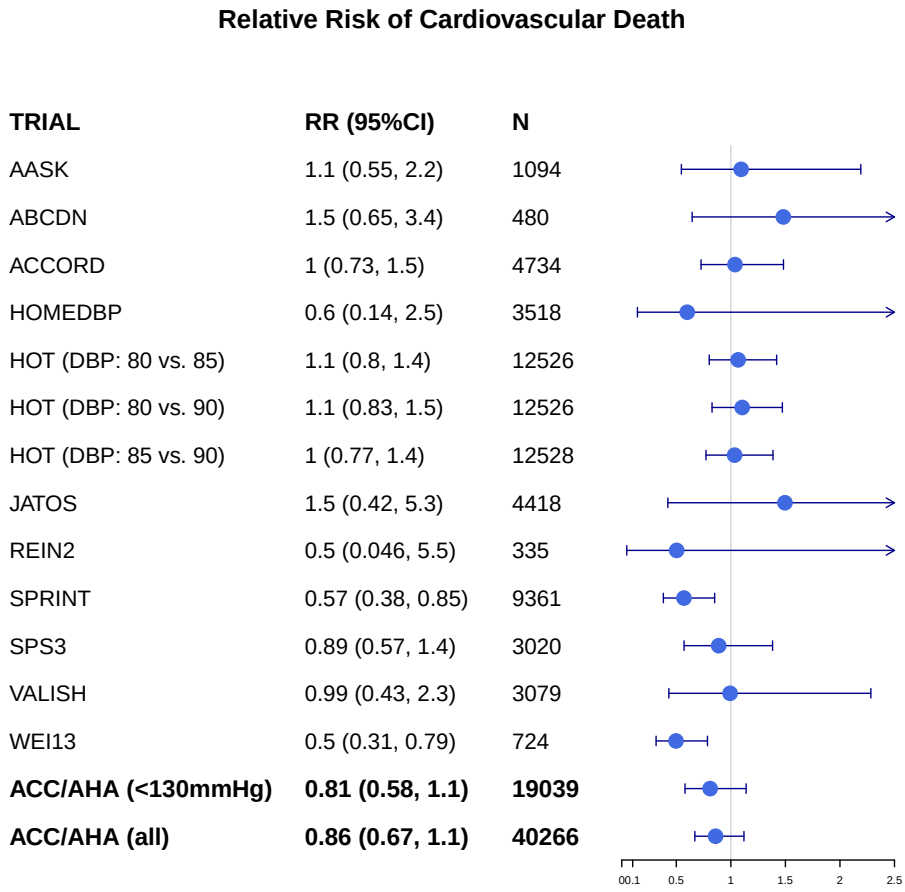

Figure 2.7: Relative risk of cardiovascular death reported by the trials (lower target (numerator) versus higher target (denominator)).

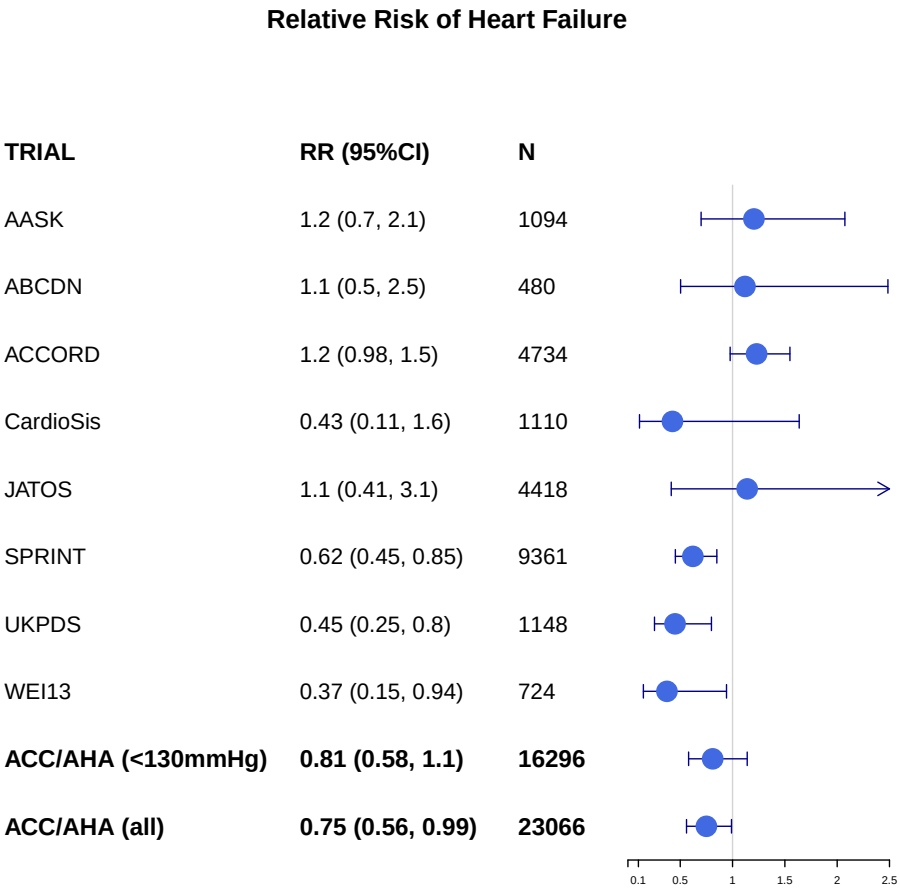

Figure 2.8: Relative risk of heart failure reported by the trials (lower target (numerator) versus higher target (denominator)).

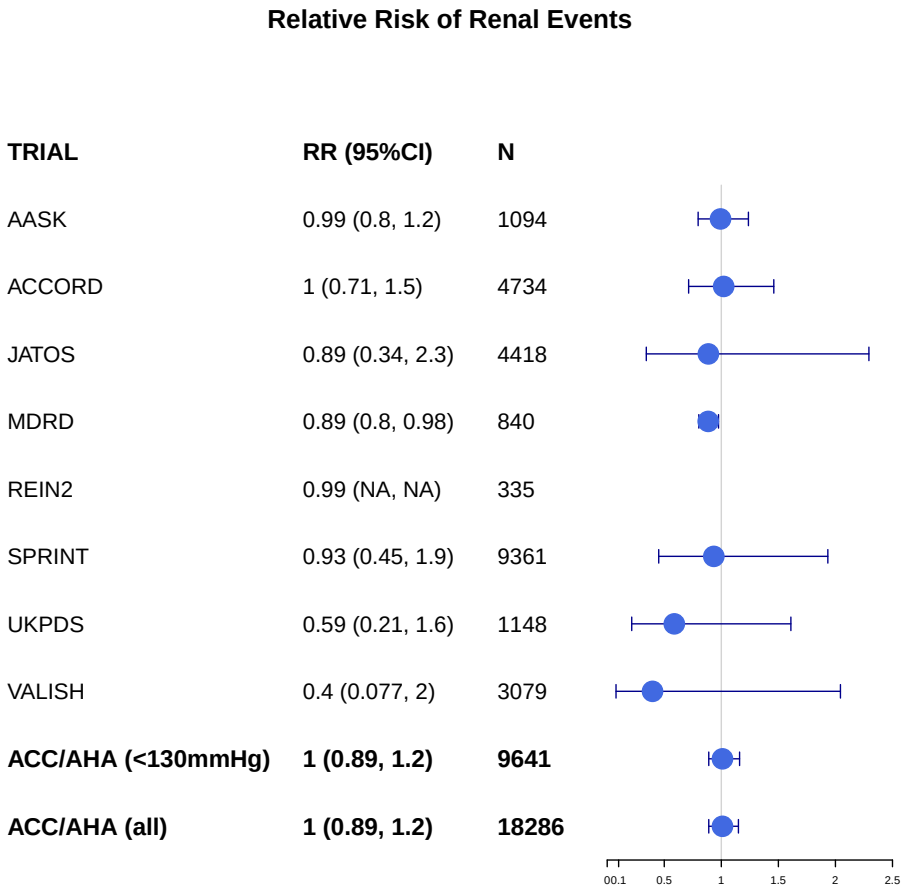

Figure 2.9: Relative risk of adverse renal events reported by the trials (lower target (numerator) versus higher target (denominator)).

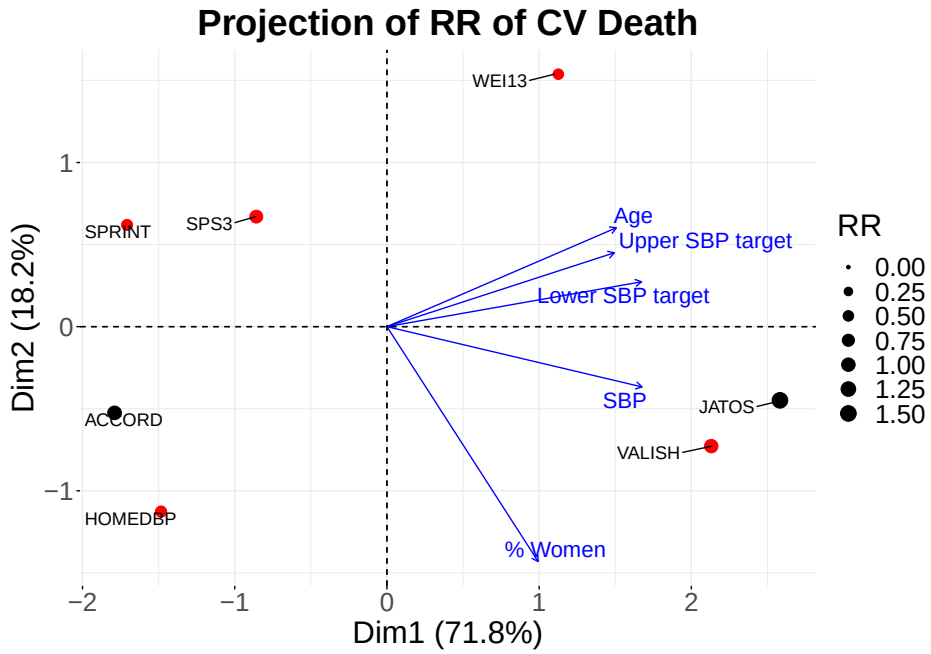

Figure 2.10: Principal components analysis of trial average age, baseline SBP, percentage of women, lower SBP target, and upper SBP target for trials comparing at least one SBP target and reporting relative risk of CV death. Trials reporting relative risk less than one are shown in red.

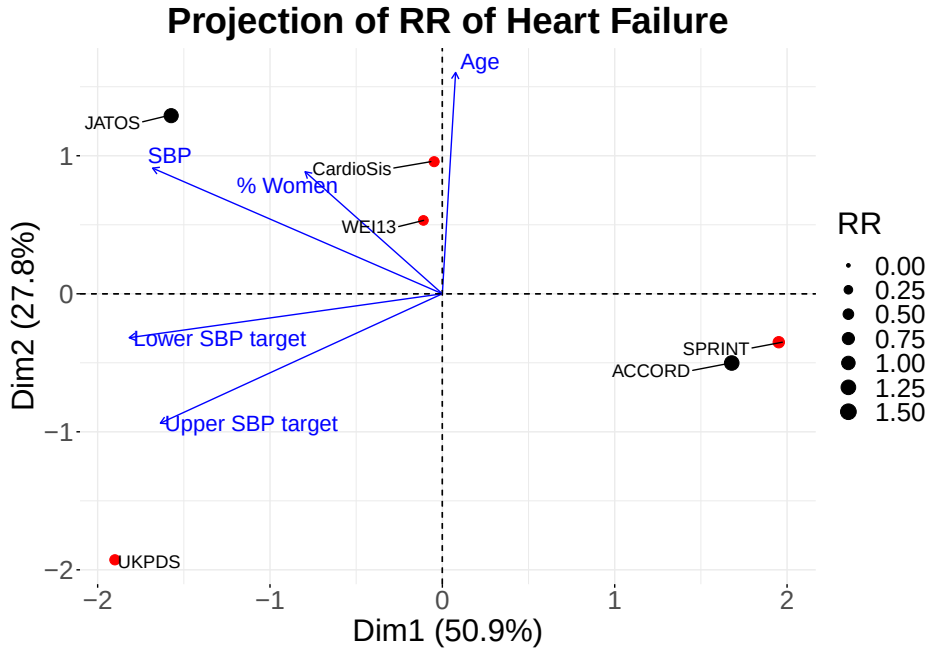

Figure 2.11: Principal components analysis of average age, baseline SBP, percentage of women, lower SBP target, and upper SBP target for trials comparing at least one SBP target and reporting relative risk of heart failure. Trials reporting relative risk less than one are shown in red.

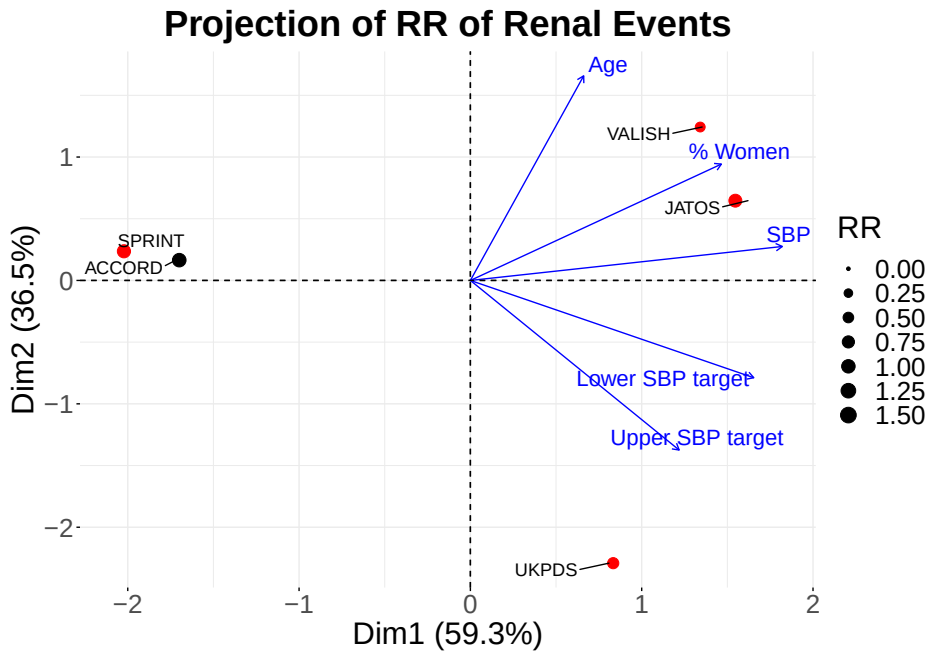

Figure 2.12: Principal components analysis of trial average age, baseline SBP, percentage of women, lower SBP target, and upper SBP target for trials comparing at least one SBP target and reporting relative risk of adverse renal events. Trials reporting relative risk less than one are shown in red.

Chapter 3

Estimating the effect of angiotensin converting enzyme inhibitors and hydrochlorothiazide on cardiovascular outcomes in HIV-positive and matched HIV-negative veterans with hypertension

Abstract

Hypertension is a well-established risk factor for adverse cardiovascular outcomes and is of especial importance in the HIV-infected and veteran populations, which are at increased risk for adverse cardiovascular outcomes relative to the general population. The current guidelines for hypertension provide blood pressure goals and recommendations for anti-hypertensives with which to initiate treatment. However, the optimal choice of anti-hypertensives following initiation remains unclear. In this paper, we present results of an analysis in which we evaluate the effect of angiotensin converting enzyme inhibitors (ACE-I) and hydrochlorothiazide (HCTZ) on adverse cardiovascular outcomes. The study population is selected from the Veterans Aging Cohort Study, a longitudinal and prospective virtual cohort of HIV-positive veterans and

their matched controls. We selected subjects from the VACS cohort who had hypertension as well as one or more cardiovascular risk factors. We use marginal structural Cox models with stabilized inverse probability of treatment weights to adjust for baseline and time-varying confounding. We find that after large-scale adjustment for possible confounders, treatment with ACE-I or HCTZ does not result in decreased probability of adverse cardiovascular events relative to other regimens of anti-hypertensives.

Keywords: Hypertension, marginal structural Cox models, angiotensin converting enzyme inhibitors, hydrochlorothiazide, time-varying confounding.

3.1 Introduction

Hypertension is a well-established risk factor for adverse cardiovascular outcomes [81, 82]. Hypertension is treated by using anti-hypertensive measures to lower blood pressure to a target specified by clinical guidelines. Randomized controlled trials of blood pressure (BP) targets have reported that lower systolic blood pressure (SBP) targets decrease risk of poor cardiovascular (CV) outcomes, such as stroke and myocardial infarction [1, 10, 11]. Guidelines based on these targets have been controversial due to their conflicting recommendations [23]. The American College of Cardiology (ACC) and the American Heart Association (AHA) recommend that physicians target a SBP of <130 mmHg in patients over the age of 60 with cardiovascular risk factors [10, 23]. The American College of Physicians (ACP) and American Association of Family Physicians (AAFP) recommend targeting an SBP <140 mmHg in this patient population [11]. This controversy is due to divergent results among trial of blood pressure targets, such as those reported in the Systolic Pressure Intervention Trial (SPRINT) [17], which reported a benefit of targeting a lower systolic blood pressure, and ACCORD [14], which reported a null result.

Furthermore, the optimal regimen of anti-hypertensives for reaching a blood pressure target remains unclear [74]. Some studies have compared therapies for initiation of treatment, and current guidelines recommend initial therapy with thiazide diuretics, angiotensin-converting enzyme inhibitors, or calcium channel blockers [10]. A large observational study of 4.9 million patients, LEGEND-HTN, concluded that thiazides and thiazide-like diuretics were optimal for initiation of anti-hypertensives [74]. However, management of hypertension requires administration of changing regimens of anti-hypertensives, where the regimen is adjusted according at clinic visits according to additional clinical information. Thus, treatments for hypertension are time-varying, and physicians require guidance regarding the optimal choice of treatment following initiation of treatment.

Management of hypertension in HIV-infected individuals is of particular importance because HIV-infected individuals are at higher risk than the general population for adverse cardiovascular outcomes, such as acute myocardial infarction (MI) [83–85]. This effect may be due to treatment with anti-retroviral medications, such as protease inhibitors, or due to increased immune activation due to infection [86]. As the HIV-infected population has aged with the success of anti-retroviral treatments, cardiovascular disease has become an important contributor to mortality in this population [87, 88]. However, HIV-infected individuals are often treated using the same strategies as non-HIV-infected individuals [85].

Methods in causal inference have made it possible to evaluate the effects of time-varying treatments using observational data. Because management of hypertension involves adaptive, time-varying treatment strategies, traditional adjustment by simply conditioning on longitudinally observed confounders can introduce bias into the estimate of treatment effect [46]. This bias occurs because measurements of blood pressure and other clinical data, as well as past treatment, are used by physicians to choose future treatment. These intermediate measurements of clinical data are called “time-varying confounders," and adjustment for time-varying confounding requires utilizing g-methods, which are designed to identify the causal effect of

a time-varying treatment regime. In particular, marginal structural models can be used to evaluate the effects of time-varying treatment strategies when detailed information about intermediate clinical measurements, outcomes, and treatments delivered to patients are available [46]. Marginal structural models use inverse probability of treatment weights (IPTW) to construct a "pseudo-population" in which the treatment of interest is assigned as if it were random [48, 89, 90]. Marginal structural Cox models have been used in several instances to examine the clinical utility of treatments, such as in the setting of multiple sclerosis and opioid use disorder [91, 92].

In this study, we use marginal structural Cox models, adjusted using stabilized IPTW, to evalute the effect of frequently used anti-hypertensive medications on the risk of adverse cardiovascular events in a cohort of non-diabetic veterans with cardiovascular risk factors. The medications of interest were angiotensin converting enzyme inhibitors (ACE-I) and hydrochlorothiazide (HCTZ). The study population was selected from an existing virtual cohort, the Veterans Aging Cohort Study, which has prospectively enrolled HIV-positive and matched HIV-negative patients since 1998. We adjusted for demographics, baseline measurements of blood pressure and other clinical data, and longitudinal observations of clinical markers and treatment with other anti-hypertensives to identify the difference in counterfactual survival probability under treatment or no treatment with ACE-I and HCTZ.

3.2 Methods

3.2.1 Data sources

The Veterans Health Administration (VA) is the largest integrated healthcare system in the United States and utilizes a nationally standardized EHR system [93]. Veterans Aging Cohort Study (VACS) is a longitudinal virtual cohort that has been prospectively enrolling patients

since 1998 [94]. The study includes HIV-infected veterans as well as non-infected veterans matched by key demographics, including age and clinical site. The institutional review board at the West Haven VA approved this study.

3.2.2 Subject selection

We selected subjects from the VACS virtual cohort for our study by applying a subset of the inclusion and exclusion criteria used in SPRINT, a large and influential trial of blood pressure targets [17]. These criteria were intended to select an older patient population with cardiovascular risk factors, excluding diabetics. We included subjects who were at least age 50 between October 1, 2007 and September 30, 2017. We required two SBP measurements greater than 130 mmHg within a 3 months-1 year period prior to enrollment. We followed the cohort until April 26, 2018.

We identified subjects with cardiovascular risk factors by requiring one of: age greater than 75 years, a history of cardiovascular disease (CVD), high-risk Framingham Risk Score, or chronic kidney disease (CKD). We used available demographic information to determine age, *International Classification of Diseases, Ninth Revision (ICD-9)* codes to determine history of cardiovascular disease, estimated glomerular filtration rate (eGFR) outside the range of 20-60 mL/min as computed by the CKD-epi equation to identify chronic kidney disease, and available laboratory panels to calculate Framingham Risk Scores. We defined a history of cardiovascular disease as observation of any ICD-9 code that corresponded to acute infarcts of any level of severety. As per SPRINT's inclusion criteria, all subjects with diabetes mellitus, a history of ischemic or hemorrhagic stroke, and a heart failure event within 6 months of study inclusion were excluded.

Covariate			Missing (%)
	Mean $\pm$ SD		
Age	61.7 $\pm$ 7.36		0
# of anti-hypertensive medications	1.13 $\pm$ 1.2		0
SBP	142 $\pm$ 10.9		0
DBP	83.8 $\pm$ 10.4		0
FHRS	6.88 $\pm$ 3.14		2.36
BMI	29.1 $\pm$ 5.57		7.8
Creatinine	1.14 $\pm$ 0.325		0
eGFR	77.6 $\pm$ 20.3		0
VACS Index	37.8 $\pm$ 11.6		10.8
	Category	**Frequency (%)**	
Sex	Female	1.76	0
	Male	98.2	
Race	White	45.4	0
	Black	44.2	
d	Hispanic	7.42	
	Other	3.01	
HIV status	HIV-	68.5	0
	HIV+	31.5	
HCV status	HCV-	67.8	0
	Chronic HCV	13.8	
	HCV genotype	0.0787	
	HCV Ab+, RNA -	2.93	
	HCV Ab+, RNA unknown	0.837	
	HCV Ab unknown, RNA-	0.333	
	HCV ICD9 code	0.148	
	Never tested	14.1	
Smoking	Never smoked	24.8	1.12
	Current smoker	53.3	
	Past smoker	20.8	

Table 3.1: **Summary of baseline features of the patients selected into this study from the VACS virtual cohort.** The table presents mean and standard deviation for continuous variables, frequency of each category for categorical variables, and percentage of missing observations of each variable. Age, number of anti-hypertensive medications, SBP, DBP, creatinine, and eGFR were required to evaluate inclusion and exclusion criteria for the study and thus have no missing values. *SD*: standard deviation; *SBP*: systolic blood pressure; *DBP*: diastolic blood pressure; *FHRS*: Framingham Risk Score; *BMI*: body mass index; *eGFR*: estimated glomerular filtration rate; *HCV* hepatitis C virus; *Ab*: antibody; *RNA*: ribonucleic acid.

3.2.3 Definition of treatment

We identified anti-hypertensive medications dispensed using prescription fill and refill data. The VACS virtual cohort data includes both medications dispensed through the VA Pharmacy Benefits Management program as well as through United States Medicare claims data [95]. Days for which medication was prescribed were determined from the prescription information. The medications which were considered to have anti-hypertensive effects are listed in Table 3.2. Each anti-hypertensive was counted as a separate medication, and also classified by mechanism of action. Anti-hypertensive classes of interest were angiotensin converting enzyme inhibitors (ACE-I), angiotensin receptor blockers (ARB), beta blockers (BB), calcium channel blockers (CCB), thiazide diuretics (of which hydrochlorothiazide (HCTZ) was the only medication), and potassium-sparing diuretics. We show the percentage of patients prescribed these classes of anti-hypertensives in Table 3.2. For the purposes of the analysis conducted in this study, we defined treatment as binary: either receipt or no receipt of a particular class at any given time.

3.2.4 Definition of outcomes

We based the outcome of interest in this study on the composite endpoint of SPRINT [17]. This trial utilized the composite outcome including "first occurrence of MI, non-MI acute coronary syndrome (non-MI ACS), stroke, heart failure (HF), or death attributable to CVD" [17]. The last observation of this cohort occurred on April 26, 2018. We used ICD-9 and ICD-10 codes recorded through this date to determine the first occurrence of MI, ACS, stroke, and heart failure. We obtained death dates from the VA Vital Status File, which has been validated through comparison with the National Death Registry [95, 96]. We defined censoring as the absence of observation of any blood pressures for 18 months, suggesting that the patient was no longer being regularly followed by a physician at the VA.

Year	0	1	2	3	4	5	6	7	8	9	10
# of patients	21612	10697	7232	5107	4047	3286	2645	2111	1611	1031	184
Percentage (%) of patients on:											
Alpha blocker	13	12	14	15	17	17	19	20	21	20	16
Doxazosin	0.94	0.78	0.83	0.84	0.79	0.85	1.1	1.2	0.93	1.1	0.54
Prazosin	1.4	1.6	1.8	2	2.3	2.4	2.2	2.1	2.7	2.1	1.6
Tamsulosin	4.6	4.3	4.6	5.5	7.2	7.3	8.8	9.8	11	9.5	6.5
Terazosin	6.5	5.8	6.8	7	7.2	7	7.3	7.6	7.4	7.8	7.1
ACE-I	27	22	25	26	28	27	26	27	28	26	22
Benazepril	0.74	0.63	0.73	0.78	0.86	0.76	0.45	0.24	0.37	0.58	0
Fosinopril	0.6	0.51	0.5	0.78	0.64	0.67	0.57	0.76	0.87	0.68	1.1
Lisinopril	24	21	23	24	25	25	24	26	26	24	19
Enalepril	0.97	0.64	0.62	0.76	0.79	0.79	0.91	0.57	0.68	0.58	1.6
ARB	4.6	4.6	4.8	5.3	6.2	6.5	7.3	7.5	8.8	10	9.2
Irbesartan	0.0046	0.0093	0.014	0	0.025	0	0	0	0	0	0
Losartan	3.6	3.7	3.9	4.4	5.4	5.7	6.4	6.5	7.5	8.4	7.1
Valsartan	1	0.93	0.91	0.96	0.84	0.79	0.98	0.99	1.2	1.6	2.2
Beta blocker	24	18	19	20	20	20	21	22	23	23	16
Atenolol	7.4	6.3	7.1	7.9	7.2	7.6	7.6	8.1	8	6.3	2.7
Carvedilol	2.6	1.6	1.6	1.4	1.6	1.8	1.8	1.7	2.6	2.9	2.7
Labetalol	0.24	0.22	0.29	0.22	0.35	0.21	0.38	0.33	0.37	0.39	0.54
Metoprolol	14	9.6	9.7	10	11	11	12	12	12	13	10
CCB	21	21	24	26	27	28	28	29	29	30	31
Amlodipine	16	17	19	20	21	22	22	23	23	24	24
Diltiazem	2	1.9	1.9	2	2.4	2.2	2.3	2.9	2.7	2.4	3.3
Felodipine	16	17	19	20	21	22	22	23	23	24	24
Nifedipine	2.1	1.7	1.8	2.1	1.7	2.1	2.1	1.9	1.6	2.3	1.6
Verapamil	1	1	1.1	1.3	1.3	1.4	1.2	1.6	2	1.7	2.2
HCTZ	20	19	22	22	23	22	21	20	22	19	9.8
K+ sparing	2.8	2.4	2.7	3	3	2.6	3	2.9	3	3.5	2.7
Spironolactone	1.1	0.71	0.76	0.9	1	0.76	1	1	1.1	1.4	1.6
Triamterene	1.8	1.7	1.9	2.1	2.1	1.8	2	1.9	1.9	2.2	1.1
Furosemide	4	2.9	3.1	3.3	3.4	3.3	3.7	3.4	4.5	4.2	3.8
Other											
Clonidine	1.2	1.1	1.3	1.7	1.6	1.7	1.5	1.4	1.6	1.9	0
Hydralazine	0.73	0.58	0.71	0.8	1.1	1.4	1.3	1.4	1.8	2.2	1.1

Table 3.2: **Table of anti-hypertensive medications.** *ACE-I:* angiotensin converting enzyme inhibitor; *ARB*: angiotensin receptor blocker; *CCB*: calcium channel blocker; *HCTZ*: hydrochlorothiazide; *K+sparing*: potassiume-sparing diuretic.

3.2.5 Covariates

Observed covariates in this study included demographic variables, baseline and longitudinal laboratory results, and the previously validated VACS Index [97, 98]. The VACS Index was developed by the VACS group to predict risk of all-cause mortality in the veteran population. The VACS Index aggregates several metrics of health, including HIV biomarkers and measures of liver and renal function. The index has been validated both in HIV-infected and non-HIV-infected individuals [99]. These variables were selected because of their potential role as confounders of the effect of treatment with anti-hypertensive medications on the composite cardiovascular outcome.

Table 3.1 shows mean values and observed frequencies of demographic and other variables measured at baseline, the time of inclusion into the study (t=0). Demographic variables were age, sex, and race (white, black, hispanic, or other). We also included smoking status (never smoked, current smoker, or past smoker), HIV status and hepatitis virus C (HCV) status (negative, HCV diagnosed through antibody or RNA levels, or HCV diagnosed via ICD9 code). Time-varyign covariates were SBP, DBP, VACS Index, hemoglobin A1c (hbA1c), cholesterol, triglyceride, high-density lipoprotein (HDL), low-density lipoprotein (LDL), creatinine, heart failure status, and body mass index (BMI) were obtained from laboratory test results and clinic visit records. Heart failure status was defined by receipt of furosemide and a beta-blocker simultaneously, since this combination is used to treat heart failure. Baseline covariates were measured at baseline, summarized in Table 3.1, and time-varying covariates were taken both at baseline and longitudinally, summarized in Figure 3.6 and the Appendix. Measurements of baseline features were considered to be taken at time 0, time of entry into the study. All observations of time-varying covariates were discretized into 30-day intervals. If multiple measurements were observed within a 30-day interval, the average of these measurements was taken to be the value for that interval. If no observation was obtained in an interval, the value

at the previous time point was carried forward. Subjects were censored if no observation of blood pressure occurred within 18 months, at which point the subject was considered lost to follow-up.

3.2.6 Statistical Analysis

Our goal was to estimate the effect of treatment with each class of anti-hypertensives on the counterfactual survival probability, with time to first adverse cardiovascular event as the outcome of interest. We used marginal structural Cox models (MSCM) with inverse probability of treatment weighting (IPTW) to estimate these counterfactual survival probabilities through adjustment for both baseline and time-varying confounding. MSCMs, as described by Hernán et al. [48], achieve this goal by reweighting the contribution of each patient to the risk set at a given time using IPTW, creating a pseudo-population in which treatment at each time point was assigned as if randomly.

Notation

Let T be the time to event. Let i denote an individual patient, with total sample size N. We denote maximum follow up time with τ, and t denotes a particular time point. Let $\overline{X}_t$ be the time-varying sequence of treatments $\{X_0, X_1, ..., X_t\}$ where t denotes the time at which the treatment was delivered and $t = 0, ..., \tau$. Let $\overline{L}_t$ be the time-varying sequence of covariates $\{L_0, L_1, ..., L_t\}$ where t denotes the time at which covariate was measured and $T = 0, ..., \tau$. Covariates measured at baseline are denoted using V. Let $\overline{C}_t$ be sequence of censoring indicators $\{C_1, C_2, ..., C_t\}$ where t denotes the time at which censoring status is observed. Values of X_t and L_t for $t < 0$ take the values of X_0 and L_0 respectively, and $C_t = 0$ for $t < 0$.

For two sequences of treatments, $\bar{x}_t$ and $\bar{x}_t^*$, we wish to estimate the counterfactual risk differ-

ence:

$$\Pr[T^{\bar{x}_t} > t] - \Pr[T^{\bar{x}_t^*} > t] \tag{3.1}$$

$$= \sum_v \Big(\Pr[T^{\bar{x}_t} > t | V = v] - \Pr[T^{\bar{x}_t^*} > t | V = v] \Big) \Pr(V = v) \tag{3.2}$$

$$= \sum_v \Big(\Pr[T > t | \overline{X}_t = \bar{x}_t, V = v] - \Pr[T > t | \overline{X}_t^* = \bar{x}_t^*, V = v] \Big) \Pr(V = v) \tag{3.3}$$

$$= \frac{1}{N} \sum_i \Big(\Pr[T > t | \overline{X}_t = \bar{x}_t, V = v_i] - \Pr[T > t | \overline{X}_t^* = \bar{x}_t^*, V = v_i] \Big) \tag{3.4}$$

where Eq. (3.2) is identified by Eq. (3.3) if conditions for causal inference are satisfied. The counterfactual $T^{\bar{x}_t}$ represents the time to event of interest, supposing that the treatment delivered had been $\bar{x}_t$. The difference in counterfactual survival probability is marginalized with respect to baseline features, V. Using Cox proportional hazards regression, we obtain an estimate of the conditional survival probability, $\Pr[T > t | \overline{X}_t = \bar{x}_t, V = v_i]$. The estimate of conditional survival probability is also an estimate of the counterfactual conditional survival probability $\Pr[T > t | \overline{X}_t = \bar{x}_t, V = v]$ after adjustment for time-varying confounding. This adjustment is performed using the weighting strategy described in Section 3.2.6.

Marginal structural Cox model

The marginal structural Cox model, as described by Hernán et al. [48] provides a strategy for estimating the hazard of an adverse cardiovascular event given treatment history and time-varying confounders [48]. We denote the hazard of an adverse cardiovascular event occurring for patient i at time t given treatment x_{it} at time t and baseline covariates V_i as $\lambda_i(t|x_{it}, V_i)$. We

model the hazard $\lambda_i(t|x_{it}, V_i)$ using a time-dependent Cox proportional hazard model [100]:

$$\lambda_i(t|x_{it}, V_i; \gamma_1, \gamma_2) = \lambda_{0t}\text{exp}(\gamma_1 x_{it} + \gamma_2 V_i) \tag{3.5}$$

where λ_{0t} is the baseline hazard, x_{it} is the binary treatment delivered to subject i at time t, γ_1 its associated coefficient and γ_2 is the vector of coefficients associated with the vector of baseline coefficients V_i. We perform adjustment for time-varying confounders L_t by weighting the contribution of each subject to the risk-set included in the partial likelihood of the Cox model at each time point. These weights are subject specific and are typically stabilized inverse probability of treatment weights. For subject i, the stabilized weight $sw_i(t)$ is:

$$sw_i(t) = \prod_{k=0}^{t} \frac{\Pr(X_{ik} = x_{ik}|\overline{X}_{i,k-1} = \bar{x}_{i,k-1}, V_i = v_i)}{\Pr(X_{ik} = x_{ik}|\overline{X}_{i,k-1} = \bar{x}_{i,k-1}, \overline{L}_{ik} = \bar{l}_{ik}, V_i = v_i)} \tag{3.6}$$

Some of the subjects in the dataset were right-censored before their outcomes were observed. To account for the effect of censoring, we used inverse probability of censoring weights (IPCW), which we denote as $sw_i^C(t)$.

$$sw_i^C(t) = \prod_{k=0}^{t} \frac{\Pr(C_{ik} = 0|\overline{C}_{i,k-1} = 0, \overline{X}_{ik} = \bar{x}_{ik}, V_i = v_i)}{\Pr(C_{ik} = 0|\overline{C}_{i,k-1} = 0, \overline{X}_{ik} = \bar{x}_{ik}, \overline{L}_{ik} = \bar{l}_{ik}, V_i = v_i)} \tag{3.7}$$

Fitting of the Cox proportional hazards regression model is then performed with a combined weight, $sw_i(t) \times sw_i^C(t)$.

We used logistic regression models to estimate the treatment and censoring weights for each analysis comparing receipt of treatment to no treatment. All observations of time-varying co-

variates and of treatment were discretized into 30-day intervals. The value of the covariate was carried forward if no clinic visit occurred, and the mean of the observations was taken if multiple observations occurred. In the logistic regression models, baseline covariates V were: age, baseline number of anti-hypertensive medications medications, systolic and diastolic blood pressure at time of inclusion into the study, sex, smoking status, age, Framingham risk score, race, estimated GFR, creatinine, HIV status, hepatitis C virus status, body mass index (BMI), and VACS index. The treatment variable x was binary, indicating receipt or no receipt of the class of interest in each 30-day interval in the analysis. Time varying variables L_t were systolic blood pressure, diastolic blood pressure, the VACS index, hbA1c, cholesterol, triglycerides, HDL, LDL, creatinine, body mass index, and binary variables indicating whether or not medications other than the class of interest were dispensed at each time point. For example, if the treatment of interest in a particular analysis was ACE-Is, indicator variables for receipt at each time point of alpha blockers, ARBs, beta-blockers, CCBs, HCTZ, potassium sparing diuretics, furosemide, clonidine, and hydralazine were included as time-varying covariates.

Thus, the risk difference estimated in each analysis reflects the causal effect of the assigned treatment, adjusted for the effect of other anti-hypertensive medications. The logistic regression model used to estimate the denominator of the weight $sw_i(t)$ in Eq. 3.6 had the form:

$$\log\left(\frac{\Pr(X_{ik}=1|\overline{X}_{i,k-1}=\bar{x}_{i,k-1},\overline{L}_{ik}=\bar{l}_{ik},V_i=v_i)}{\Pr(X_{ik}=0|\overline{X}_{i,k-1}=\bar{x}_{i,k-1},\overline{L}_{ik}=\bar{l}_{ik},V_i=v_i)}\right)=\beta_{id}\boldsymbol{Z}_{id}$$

where the predictor, with individual index i dropped for simplicity, are represented by $\boldsymbol{Z}_d$ and the vector of coefficients is β_d. The vector of predictors is $\boldsymbol{Z}_d = \{1, \bar{x}_{k-1}, \bar{l}_k, \bar{l^*}_k, v, v^*\}$, with v^* and $\bar{l^*}_k$ representing all pairwise interaction within the baseline covariates v and the time-varying covariates $\bar{l}_k$ respectively. Similarly, the model for the numerator of the IPTW, $sw_i(t)$, in Eq. 3.6 was fitted with predictors $\boldsymbol{Z}_n = \{1, \bar{x}_{k-1}, v, v^*\}$, with fitted coefficients β_n.

The logistic regression model used to estimate the denominator of the IPCW, $sw_i^C(t)$, in Eq. 3.7 had the form:

$$\log\left(\frac{\Pr(C_i=1|\overline{C}_{i,k-1}=0,\overline{X}_{ik}=\bar{x}_{ik},\overline{L}_{ik}=\bar{l}_{ik},V_i=v_i)}{\Pr(C_i=0|\overline{C}_{i,k-1}=0,\overline{X}_{ik}=\bar{x}_{ik},\overline{L}_{ik}=\bar{l}_{ik},V_i=v_i)}\right)=\beta_{id}^c\boldsymbol{Z}_{id}^c \tag{3.8}$$

where the predictors, again with individual index i dropped for simplicity, are $\boldsymbol{Z}_d^c$ and the vector of coefficients is β_d^c. In this model, the predictors are $\boldsymbol{Z}_d^c=\{1,\bar{c}_{k-1},\bar{x}_k,\bar{l}_k,\bar{l^*}_k,v,v^*\}$, with $\bar{c}_{k-1}$ representing the history of censoring. The analogous model for the numerator of the IPTW was fitted with predictors $\boldsymbol{Z}_n^c=\{1,\bar{c}_{k-1},\bar{x}_k,v,v^*\}$ and coefficients β_n^c.

We performed two separate analyses, investigating the effect of ACE-I and HCTZ. Missing covariates in the dataset were imputed five times using multiple imputation with chained equations [101, 102]. We included all covariates and the treatment variable in the imputation model. For each analysis, we truncated the most extreme weights at the following percentile thresholds: (.5, 99.5), (1, 99), (10, 90), (50, 50), where truncation at the 50% percentile corresponds to an unweighted analysis with no adjustment for time-varying confounding [89, 90]. We used the *survival* package in R to fit the time-dependent Cox proportional hazards model presented in Eq. 3.5, using these truncated weights and the Breslow estimator to compute the survival curve conditional on baseline covariates [103, 104]. We computed the survival curve marginalized with respect to the empirical distribution of baseline covariates to obtain an estimate of the causal estimand of interest introduced in Eq. 3.4.

We used 50 bootstrap samples in each imputed dataset to generate 95% confidence intervals at each time point and of the coefficients $\{\gamma_1,\gamma_2\}$ used to model the hazard $\lambda_i(t|x_{it},v_i)$ in Eq. 3.5. Bootstrapped variance estimates have been shown to be less likely to be biased than robust standard errors [105].

3.3 Results

3.3.1 Descriptive statistics

We included 21,612 patients in the cohort at the initiation of the study who were eligible according to the selection criteria. Table 3.1 shows summary statistics for the patient features measured at baseline. Figure 3.1 shows the distribution of the continuous features. The mean age of the cohort was 61.7, the mean baseline blood pressure at time of inclusion was 142/84 mmHg. Due to the inclusion criteria, the distribution of baseline systolic blood pressure and age were truncated. Because this was a cohort of veterans, the patient were mostly male, with 1.76% of the patients being female. Of the patients, 68.5% were HIV-infected, and 32.3% were HCV infected by some metric. Figure 3.2 shows the point-wise mean and 95% confidence interval of the systolic and diastolic blood pressures and the VACS index. The mean systolic blood pressure remained stable throughout the course of the analysis, and diastolic blood pressure decreased slightly. Mean VACS index increased throughout the study, indicating an increase in mortality risk over time, as expected. Additional summary tables and figures for other time-varying covariates, such as lipid panels and hbA1c, are included in the Appendix.

At time zero, the mean number of anti-hypertensives being prescribed to each patient was 1.13, increasing to a maximum of 1.7 ± 1.5 at 8.5 years. Table 3.2 shows the number of patients remaining in the study at each year of follow-up, the percentage of patients on each class of anti-hypertensives at each time point, and the medications included in each class. Patients depart from the study either because they were lost to follow up or because they experienced the composite outcome. The most common medications to be prescribed among this cohort over all time points were ACE-inhibitors (25.0%), calcium channel blockers (24.2%), and hydrochlorothiazide (20.4%). Lisinopril was the most commonly prescribed medication,

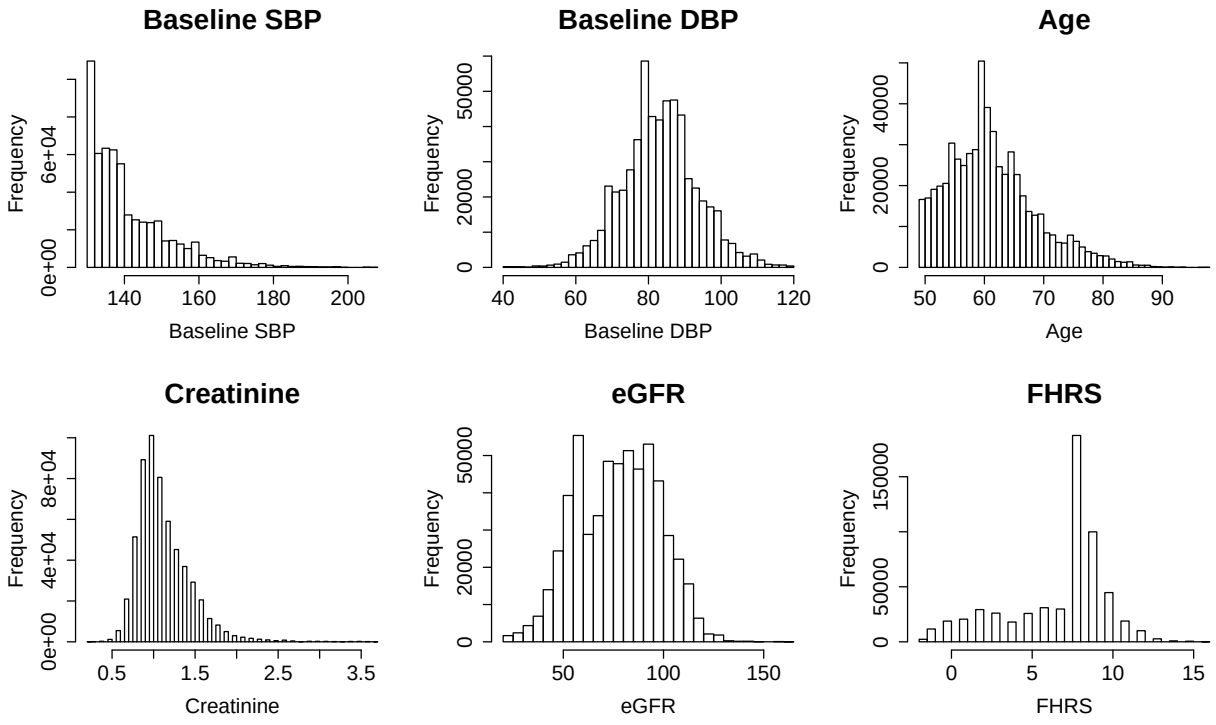

Figure 3.1: **Distribution of baseline features.** We show histograms of baseline SBP, DBP, age, creatinine, eGFR, and FHRS of patients selected into the study. As required by the selection criteria, the distribution of baseline SBP and age are truncated at 130 mmHg and 50 years respectively.

with 24% of patients receiving lisinopril at initiation of the study.

During the course of the 10-year follow-up, 7723 (35.7%) of patients incurred the composite outcome. Figure 3.3 shows descriptive Kaplan-Meier survival curves for the composite outcome and its individual components: acute coronary syndrome, all-cause mortality, myocardial infarction, heart failure, transient ischemic attack, and both ischemic and hemorrhagic stroke. Of these adverse cardiovascular events, heart failure and acute coronary syndrome were most common: 3,094 patients experienced a heart failure event as their first adverse event, and 2,476 experienced acute coronary syndrome as their first adverse event. Among patients who incurred a cerebrovascular event, 447 patients experienced temporary ischemic attack, 276 patients experienced ischemic stroke, and 10 patients experienced a hemorrhagic stroke. All-cause mortality was not a substantial contributor, with 182 patients; this was due to the fact that most patients experienced either a heart failure or acute coronary event before mortality.

3.3.2 Stabilized weights

We used several different logistic regression models with varying levels of complexity to determine the model with which to generate stabilized weights. Appendix table 3.4 presents the Akaike information criterion and deviance of each of the four models necessary to construct the stabilized weight, the numerator and denominator of the stabilized IPTW described in Eq. 3.6, and the numerator and denominator of the stabilized IPCW described in Eq. 3.7. Specifically, we evaluated four different model fits for the unstabilized IPTW and IPCW respectively: 1) a baseline model including only baseline features without any time-varying covariates, 2) a model including baseline features and time-varying clinical covariates such as systolic and diastolic blood pressure, lipid panels, hbA1C, and LDL, 3) a model with baseline features, time-varying clinical covariates, and receipt of medications other than the treatment of interest

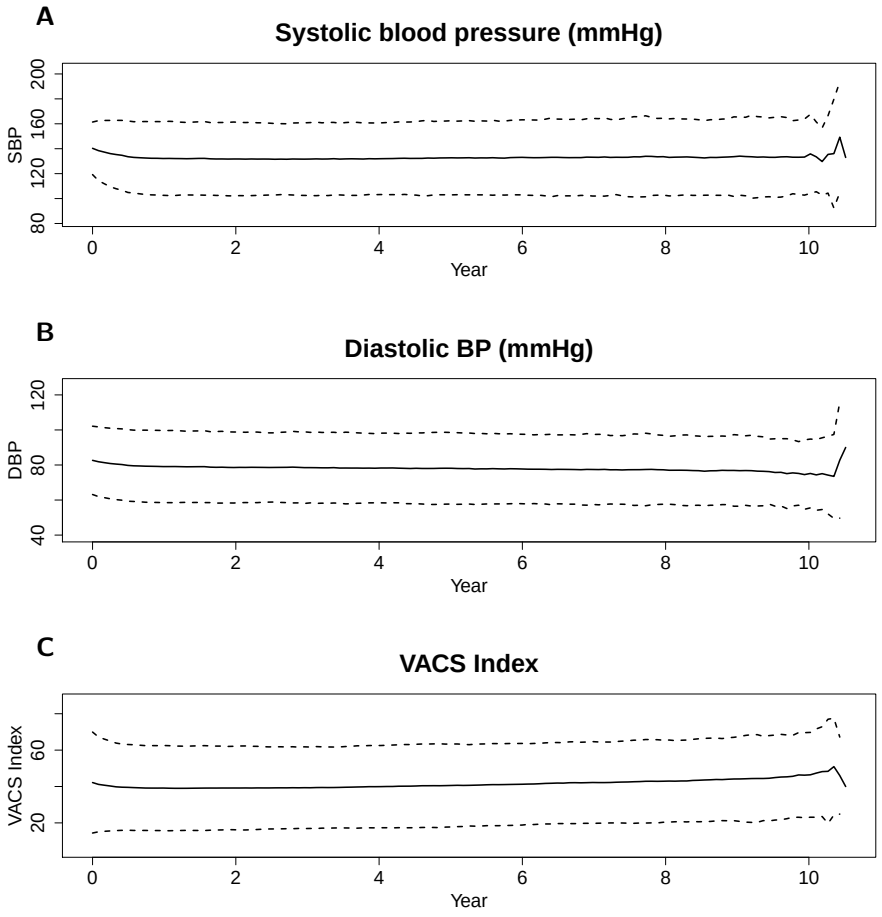

Figure 3.2: **Longitudinal observation of time-varying covariates: SBP, DBP, and the VACS index**. These plots show the point-wise mean and standard deviation of three of the time-varying covariates included in the analysis in each 30-day time interval. SBP remained largely stable over time. DBP decreased slightly with follow-up, reaching 75.0 $\pm$ 10.4 mmHg at 10 year follow-up. VACS Index increased with increasing follow-up time, reflecting increasing risk of mortality. Summary statistics at each year of follow-up and additional figures are included in the Appendix.

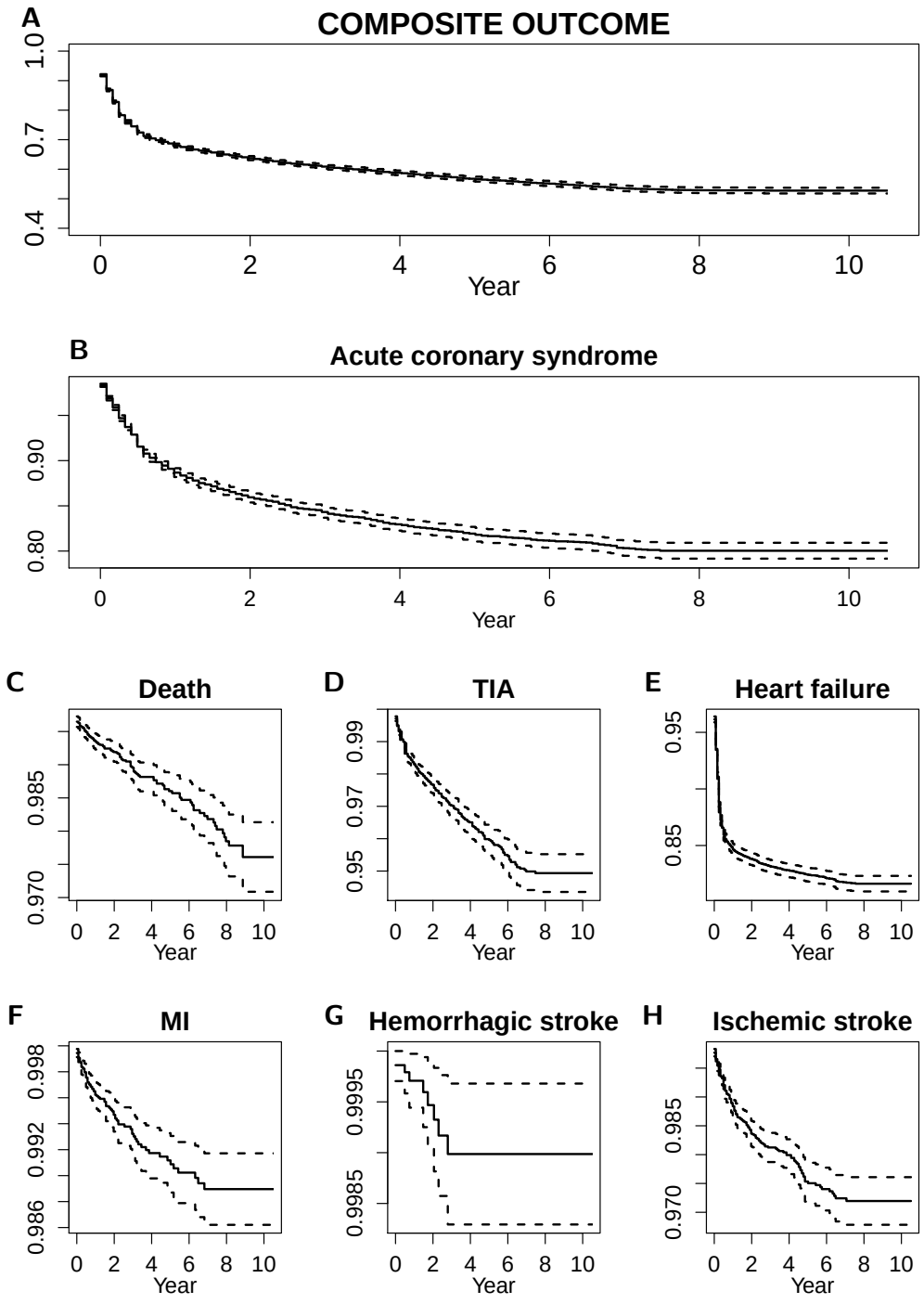

Figure 3.3: **Descriptive Kaplan-Meier curves for the primary endpoint, the composite outcome, and its individual components**: acute coronary syndrome (ACS, 2,476 events), all-cause mortality (182 events), transient ischemic attack (TIA, 447 events), heart failure (3,094 events), myocardial infarction (MI, 96 events), and hemorrhaghic (10 events) and ischemic (276 events) stroke.

included as time-varying covariates, and 4) all covariates included in model (3), as well as all pairwise interactions between baseline covariates age, FHRS, eGFR, baseline # of medications, and creatinine, all pairwise interactions between the time-varying clinical covariates. We observed decreasing model deviance with increasing model complexity, as expected. Importantly, we observed decreasing AIC with increased model complexity, suggesting that overfitting of the data was not a concern. We did not attempt to fit more fully saturated models in this analysis due to computational limitations arising from the necessity of utilizing both multiple imputation to address missingness and bootstrapping for variance estimation.

Figure 3.8 and 3.9 show the untruncated distribution of the logarithm of the stabilized weights computed using model (4), including all time points. As expected, the distribution of the stabilized weights was mean 1. We observed substantial overlap in the distribution of the stabilized weights for patients who received and did not receive an treatment in both analyses, suggesting that the positivity assumption necessary for fitting marginal structural models in this analysis was not obviously violated.

3.3.3 Marginal structure Cox model

The primary comparison of interest that we report is the pointwise difference in counterfactual survival probability between two static treatment regimes: “always treatment" with the anti-hypertensive of interest ($\bar{x}_t = \{1, 1, ..., 1\}$), or “never treatment" with the anti-hypertensive of interest ($\bar{x}_t = \{0, 0, ..., 0\}$). We conducted two analysis in which we examined the effect of ACE-I and HCTZ as the anti-hypertensive of interest. The counterfactual survival probability at time t under “always treatment" should be interpreted as the expected value of the probability that an individual has not experienced the composite outcome as of time t, under an anti-hypertensive regimen that includes the treatment of interest as well as other anti-hypertensives that patients in the cohort were prescribed alongside the treatment of interest. The counterfac-

tual survival probability at time t under “never treatment" should be interpreted as the expected value of the probability that an individual has not experienced the composite outcome as of time t, under an anti-hypertensive regimen that does not includes the treatment of interest at any point but includes other anti-hypertensives that subjects in the cohort received, in the absence of the treatment of interest. For example, for ACE-I, the treatment comparison is between anti-hypertensive regimens that must include an ACE-I and regimens that cannot include an ACE-I. The regimens that do not include an ACE-I still consist of compensatory regimens composed of other combinations of anti-hypertensive medications. A difference of 0 represents a null result, with regimens including and not including ACE-I exerting the same effect on the outcome. A difference greater than zero represents a benefit of regimens including ACE-I, and a difference less than zero represents greater benefits from regimens not including ACE-I.

In this section, we report the estimated pointwise counterfactual difference in survival probability for the overall population, the counterfactual difference in survival probability for the subgroup of HIV-positive patients, and the estimated hazard ratios associated with the covariates included in the weighted Cox proportional hazards regression. We also include the unadjusted risk difference, computed using weights all set to 1, to reflect the confounded difference in survival probability between the groups receiving and not receiving the treatment of interest. All confidence intervals were based on bootstrapped samples for each imputed dataset, combined using Rubin's rule [106].

ACE inhibitors

The marginal counterfactual risk difference of the composite outcome between treatment with ACE-I and no treatment with ACE-I was -0.035 (-0.47, 0.40) at 2 years and -0.065 (-0.62, 0.49) at 5 years. In comparison, the unadjusted risk difference was -0.06 (-0.14, 0.01) at 2 years and -0.09 (-0.12, -0.07) at 5 years. Figure 3.4A shows the adjusted risk difference

for the entire population across all 10 years of follow up, with weights truncated at the 0.5% and 99.5% percentiles, and Figure 3.4B shows the unadjusted risk difference. The results obtained after adjustment suggest that regimens with ACE-I do not result in decreased risk of the composite outcome relative to regimens without ACE-I. The unadjusted risk difference, which indicates higher risk of the composite outcome under treatment, is likely evidence of confounding, where patients with hypertension who are prescribed ACE-I have higher risk of adverse cardiovascular outcomes than patients who are not prescribed medication. The adjusted hazard ratio associated with treatment with ACE-I was 1.19 (0.06, 22.25), and the unadjusted hazard ratio was 1.40 (1.28, 1.53).

We also conducted a subgroup analysis by computing the marginal counterfactual risk difference for the HIV-positive individuals in the cohort. Figure 3.4C and D show the adjusted risk difference and unadjusted risk difference respectively. In the HIV-positive subgroup, the adjusted risk difference was -0.09 (-0.71, 0.53) at 2 years and -0.03 (-0.51, 0.46) at 5 years. The unadjusted risk difference was -0.12 (-0.14, -0.10) at 2 years and -0.06 (-0.19, 0.06) at 5 years. Thus, regimens with ACE-I did not show benefits relative to regimens without ACE-I after adjustment, while risk of the outcome was increased in the treated group at 2 years before adjustment.

Hydrochlorothiazide

The marginal counterfactual risk difference of the composite outcome between treatment with HCTZ and no treatment with HCTZ was 0.01 (-0.14 0.16) at 2 years and 0.02 (-0.15, 0.18) at 5 years. The unadjusted risk difference was -0.02 (-0.03, -0.004) at 2 years and -0.02 (-0.05, 0.001) at 5 years. Figure 3.5A shows the adjusted risk difference for the entire population across all 10 years of follow up, with weights truncated at the 0.5% and 99.5% percentiles, and Figure 3.5B shows the unadjusted risk difference. The results obtained after adjustment suggest that regimens with HCTZ are not more beneficial than regimens without HCTZ. The

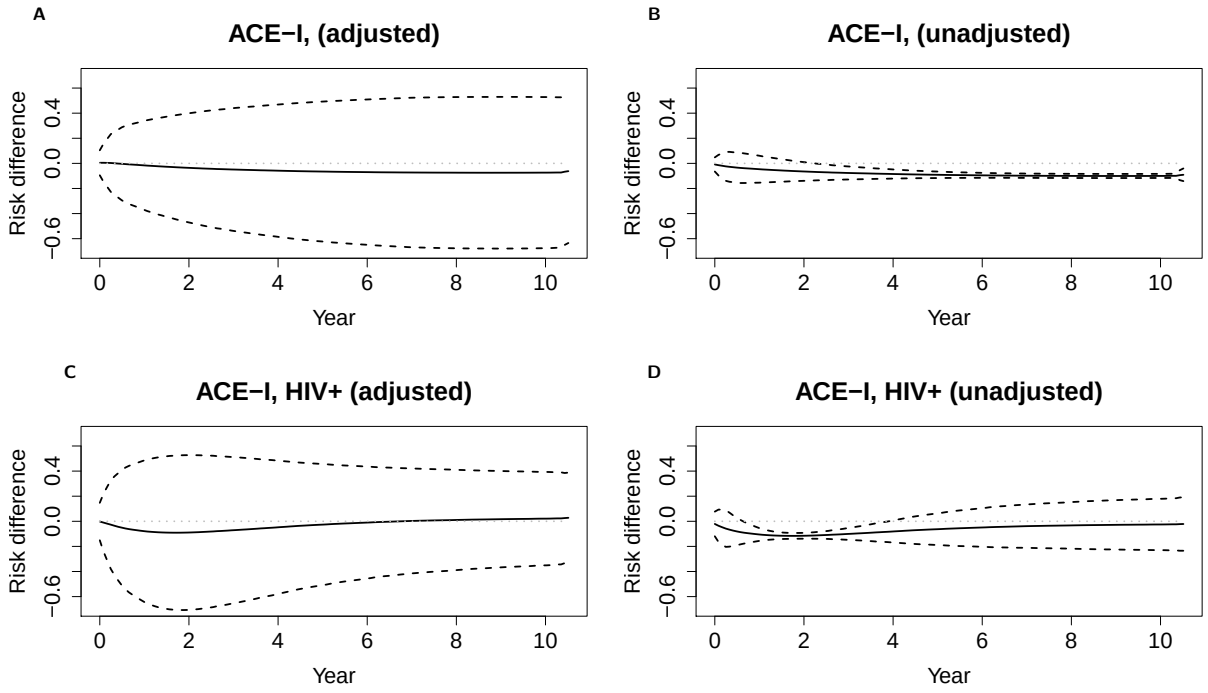

Figure 3.4: **Difference between the risk of counterfactual composite outcome under treatment with ACE-I and no treatment with ACE-I. A:** Adjusted risk difference in the whole cohort, with weights truncated at the 0.05 and 99.5 percentiles. **B:** Unadjusted risk difference in the whole cohort. **C:** Adjusted risk difference in the HIV-positive patients, with weights truncated at the 0.05 and 99.5 percentiles. **D:** Unadjusted risk difference in HIV-infected patients. The dotted grey line indicates the null, with risk of the outcome under treatment and no treatment being the same.Positive risk difference indicates probability of survival is higher under treatment. Negative risk difference indicates that probability of survival is higher under no treatment.

adjusted hazard ratio associated with treatment with HCTZ was 0.28 (0.02, 3.55), and the unadjusted hazard ratio was 0.46 (0.21, 1.03).

As in the case of ACE-I, we also conducted a subgroup analysis by computing the marginal counterfactual risk difference for the HIV-positive individuals in the cohort. Figure 3.5C and D show the adjusted risk difference and unadjusted risk difference respectively. In the HIV-positive subgroup, the adjusted risk difference was 0.02 (-0.15, 0.18) at 2 years and 0.02 (-0.16, 0.19) at 5 years. The unadjusted risk difference was -0.03 (-0.07, 0.003) at 2 years and -0.02 (-0.03, -.003) at 5 years. Thus, regimens with HCTZ did not appear to decrease risk of the composite outcome, but the analysis demonstrates adjustment for confounding. The unadjusted results demonstrate increased risk of the outcomeat 5 years under treatment with regimens containing HCTZ.

Additional estimated hazard ratios are included in the Appendix. Sensitivity analysis using different thresholds for truncating the stabilized weights was also conducted at the (1, 99) and (5,95) percentiles. Additional risk difference curves based on these different truncation thresholds are also included in the Appendix.

3.4 Discussion

In this work, we have evaluated the effect of ACE-I and HCTZ in a cohort of non-diabetic veterans with hypertension and cardiovascular risk factors. We have used stabilized IPTW and IPCW to adjust for the effect of various time-varying confounders, including intermediate blood pressure measurements which might have been used to adjust the treatment strategy and treatment with other anti-hypertensive medications. The cohort consists of HIV-positive veterans as well as matched HIV-negative controls, and we conducted a subgroup analysis to evaluate the effect of these anti-hypertensives in the HIV-positive patients. The outcome of

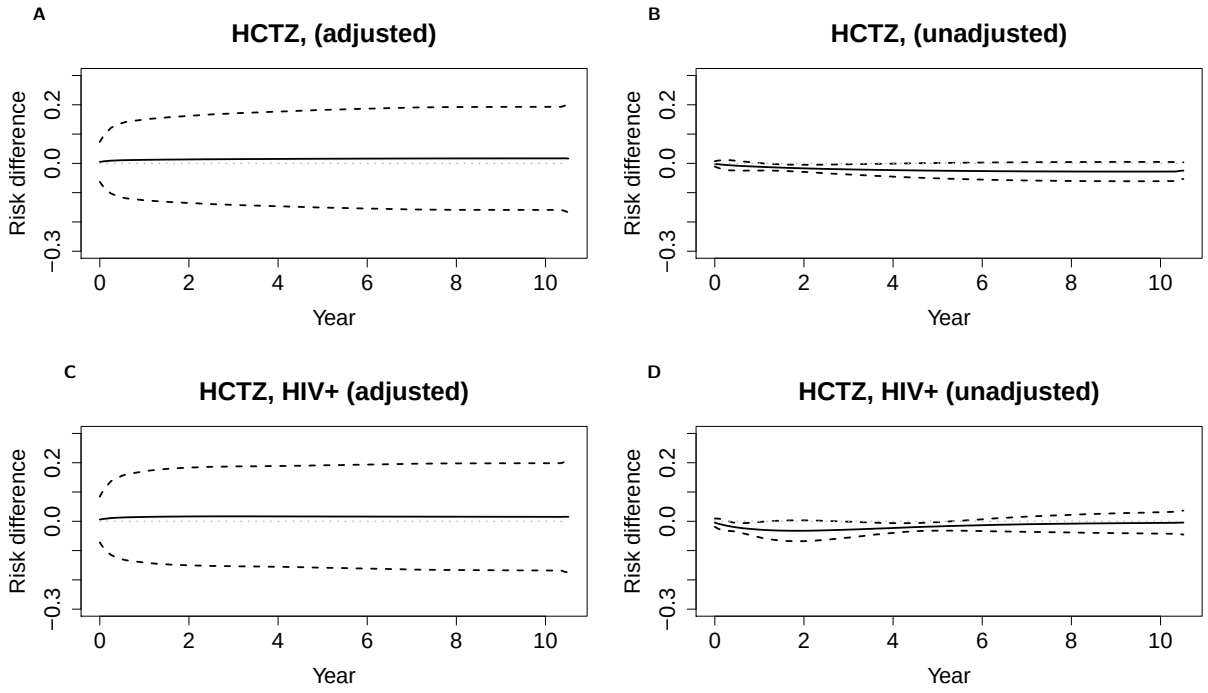

Figure 3.5: **Difference between the risk of counterfactual composite outcome under treatment with HCTZ and no treatment with HCTZ. A:** Adjusted risk difference in the whole cohort, with weights truncated at the 0.05 and 99.5 percentiles. **B:** Unadjusted risk difference in the whole cohort. **C:** Adjusted risk difference in the HIV-positive patients, with weights truncated at the 0.05 and 99.5 percentiles. **D:** Unadjusted risk difference in HIV-infected patients. The dotted grey line indicates the null, with risk of the outcome under treatment and no treatment being the same.Positive risk difference indicates probability of survival is higher under treatment. Negative risk difference indicates that probability of survival is higher under no treatment.

interest was time to first occurrence of a composite endpoint of adverse cardiovascular events.

After adjustment, our analysis revealed that treatment regimens containing ACE-I and HCTZ did not confer decreased risk of the composite outcome relative to regimens without ACE-I or HCTZ respectively, either in the overall cohort or in the HIV-positive subgroup. These results do not suggest that treatment with ACE-I and HCTZ are not useful, but rather that compensatory regimens without ACE-I and HCTZ exist which exert a similar effect on the probability of the composite outcome. The primary finding in this study is the difference in counterfactual survival probability under two static treatment regimens: always treating with a particular medication, and never treating with a particular medication. The survival probability at a specific time represents the probability that an individual does not experience the outcome before that time. The difference in counterfactual survival probabilities allow us to assess survival probabilities under these two treatment regimens.

The interpretation of these results in the context of treatment with other anti-hypertensives also warrants clarification. The presence of anti-hypertensives other than the treatment of interest is considered to be a source of time-varying confounding, which might influence both the choice of future treatment and the outcome. The counterfactual survival probability under a particular time-varying treatment regimen assumes that the time-varying confounders take the values that they would have under that treatment regimen. For example, in the comparison of the counterfactual outcome under always treatment or never treatment with ACE-I, we adjust for the effect of other anti-hypertensives, including HCTZ, which may vary while treatment with ACE-I is held constant. If we estimate the effect of always treatment with ACE-I, the estimated effect assumes that HCTZ was prescribed as it would have been had ACE-I been given at every point.

We also report the estimated hazard ratios from the weighted Cox proportional hazards regression. However, we caution against interpreting these hazard ratios as representative of the casual effect of the treatments of interest because of selection bias inherent to any esti-

mate of hazard ratio and the fact that the hazard ratio is averaged over the period of follow-up [107].

As in all observational studies, the results reported in this study may be subject to some limitations. The population of veterans is overwhelmingly male, and further study is required to determine if the effects of anti-hypertensive treatment would be the same in women. Unmeasured confounding for which we would be unable to adjust may be present in this study. For example, some aspects of a physician's decision-making process may not be recorded in the electronic health record, such as an assessment of a patient's motivation to adhere to medication or to lifestyle modifications to control blood pressure. While we observed both medications prescribed by the VA and those covered by Medicare for patients over the age of 65, we may not have observed all medications taken by patients under the age of 65. Furthermore, observational work alone cannot be used to change or influence guidelines for treatment of hypertension.

We encountered substantial methodological challenges during this work that should be addressed, providing motivation for innovation in the field of causal inference for time-varying treatments. We encountered extreme instability in the estimated IPTW and IPCW. In addition to logistic regression models, we attempted to utilized several other weighting models, including computation of optimal probability weights [108], estimation with random forests [109], and a weighted combination of estimated weights from logistic regression and random forests [110]. The optimization procedure necessary to compute the optimal weights on a dataset of this size was not a practical choice for analysis due to the computing time and resources required. Random forests yielded an estimate of probability zero for misclassified observations, leading to infinite weights when occurring in the denominator of the IPTW and IPCW. Combining the random forest and logistic regression estimates appeared to substantially stabilize the weights while correcting estimates of probability zero generated by the random forests, but computational limitations due to the combination of imputation and bootstrapping rendered

this method impossible to use in this analysis. In addition, the estimated weights in general become more unstable as we calculated the cumulative product of probabilities over many time points, eventually preventing the successful fitting of the weighted Cox proportional hazard models using the untruncated weights.

In parallel with addressing these methodological challenges, we recommend further evaluation of these treatments and others for management of hypertension in the veteran population. We excluded diabetic patients in this study and anticipate performing a similar analysis to determine if the results differ from those in non-diabetic patients. Many patients are initiated on multiple anti-hypertensives, and evaluation of different combinations of anti-hypertensives would inform physicians regarding which combination may be optimal. We also recommend repeating the analysis for other subgroups, including women and groups stratified by differences in the severity of baseline hypertension. Solving computational challenges described above will allow us to use these methods in larger cohorts and to build more complex models for IPTW, IPCW, and the weighted Cox proportional hazards outcome model. We anticipate that results obtained using these methods will become more precise and clinically useful as these methods improve and electronic health records become more comprehensive, with longer-term follow up.

Acknowledgments: I would like to thank Fan Li, Amy Justice, and Haidong Lu for their support in completing this project. I would also like to acknowledge NICHD 1DP2HD091799-01.

Data sharing: These data are protected health information and are not available for sharing. Code is available for sharing upon request.

Appendix

3.4.1 Data

We include several additional figures summarizing the features of the cohort included in the study. Figure 3.6 shows the pointwise mean values, with 95% confidence intervals, of hemoglobin A1c, cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein, creatinine and body mass index, which were used as time-varying predictors in the estimation of the stabilized weights. Table 3.3 shows the mean values with 95% confidence intervals of these time-varying covariates, as well as systolic blood pressure, diastolic blood pressure, and VACS index, at each year of follow-up in the study. Figure 3.7 shows the longitudinal change in medications prescribed: the pointwise mean number of medications, as well as the percentage of patients who were prescribed each major class of anti-hypertensives during the study period, including ACE-inhibitors, ARBs, beta blockers, CCBs, HCTZ, and potassium sparing diuretics.

3.4.2 Model Diagnostics

Figure 3.8 shows the distribution of estimated stabilized weights, including all weights from all time points. The estimated weights for those receiving ACE-I are shown in blue, and those for patients not receiving ACE-I are shown in red. The purple region indicates the area of overlap. The distribution of weights of those receiving and not receiving ACE-I are similar, suggesting that positivity violations may not pose a significant challenge in the dataset and that some randomness in physician behavior is present. Figure 3.9 shows a similar overlap in the distribution of estimated stabilized weights for those receiving HCTZ and not receiving HCTZ. Thus, we can draw a similar conclusion regarding positivity violations in our assessment of the effect of HCTZ.

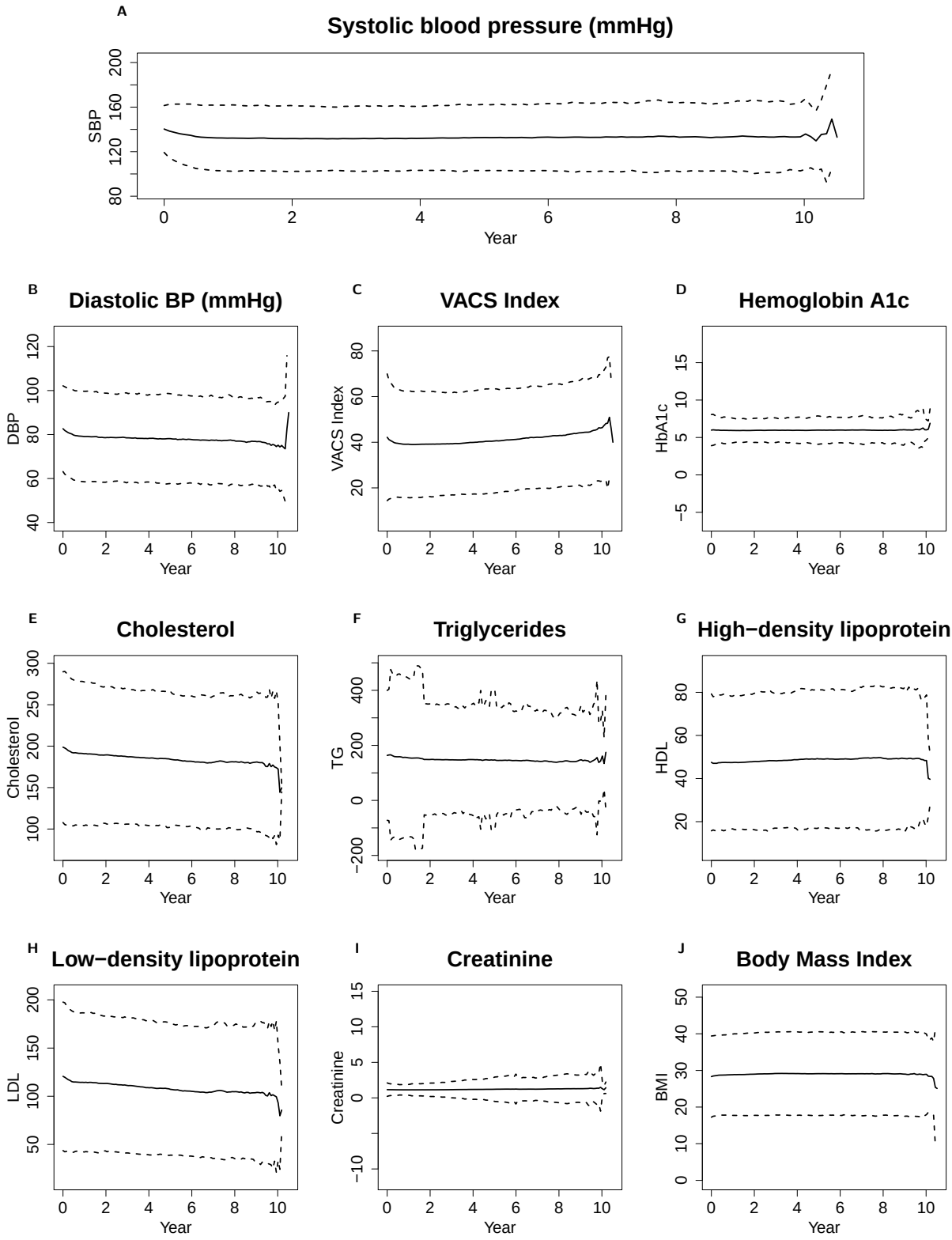

Figure 3.6: **Pointwise mean and 95% confidence intervals for time-varying covariates:** systolic blood pressure, diastolic blood pressure, VACS index, hemoglobin A1c, cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein, creatinine and body mass index.

Year	1	2	3	4	5	6	7	8	9	10	11
Systolic blood pressure (mmHg)	133.4 ± 14.8	132.0 ± 15.4	131.6 ± 15.3	131.9 ± 14.8	132.5 ± 15.2	132.9 ± 15.6	133.2 ± 16.0	133.4 ± 15.9	133.4 ± 16.3	133.1 ± 16.0	135.8 ± 17.4
Diastolic blood pressure (mmHg)	79.5 ± 10.4	78.8 ± 10.6	78.4 ± 10.6	78.1 ± 10.1	78.1 ± 10.5	77.8 ± 10.1	77.4 ± 10.4	77.3 ± 10.0	76.8 ± 10.2	76.1 ± 10.1	75.0 ± 10.4
Cholesterol	189.5 ± 45.2	188.6 ± 43.6	187.4 ± 42.3	186.0 ± 41.9	184.4 ± 41.9	181.8 ± 40.6	179.9 ± 40.7	180.5 ± 41.1	178.4 ± 42.0	178.8 ± 44.0	170.5 ± 42.8
High-density lipoprotein	47.1 ± 16.0	48.0 ± 16.2	48.3 ± 16.0	48.7 ± 16.2	49.1 ± 16.4	49.1 ± 16.5	49.4 ± 16.8	49.6 ± 16.9	49.8 ± 17.2	49.6 ± 16.9	50.0 ± 13.8
Triglycerides	158.4 ± 135.8	149.5 ± 134.0	147.6 ± 103.0	147.7 ± 103.3	146.4 ± 124.9	144.9 ± 95.8	142.0 ± 92.6	138.6 ± 82.9	140.4 ± 88.8	142.7 ± 100.5	158.2 ± 92.9
Low-density lipoprotein	112.9 ± 37.5	112.5 ± 36.5	111.4 ± 36.5	109.5 ± 36.0	107.6 ± 35.3	105.4 ± 34.8	104.1 ± 34.8	104.6 ± 35.4	102.0 ± 36.1	103.0 ± 38.4	91.2 ± 33.4
Hemoglobin A1c	5.9 ± 0.9	5.9 ± 0.8	5.9 ± 0.8	6.0 ± 0.9	6.0 ± 0.9	6.0 ± 0.9	6.0 ± 0.9	6.0 ± 0.9	6.0 ± 0.9	6.0 ± 1.0	6.1 ± 0.7
Creatinine	1.1 ± 0.4	1.1 ± 0.5	1.1 ± 0.6	1.2 ± 0.7	1.2 ± 0.8	1.2 ± 0.9	1.2 ± 0.9	1.3 ± 0.9	1.3 ± 1.0	1.3 ± 1.1	1.3 ± 0.5
Body Mass Index	28.3 ± 5.7	28.8 ± 5.7	28.9 ± 5.8	29.1 ± 5.8	29.0 ± 5.9	29.1 ± 5.9	29.0 ± 5.9	29.0 ± 5.9	29.0 ± 5.8	28.9 ± 5.8	29.0 ± 5.6
VACS Index	42.7 ± 14.3	39.9 ± 12.3	40.1 ± 12.1	40.2 ± 11.7	40.9 ± 12.0	41.6 ± 11.9	42.3 ± 11.7	43.1 ± 12.0	44.0 ± 11.8	44.9 ± 12.2	47.0 ± 12.3

Table 3.3: **Mean and 95% confidence intervals for time-varying covariates during each year of follow-up:** systolic blood pressure, diastolic blood pressure, VACS index, hemoglobin A1c, cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein, creatinine and body mass index.

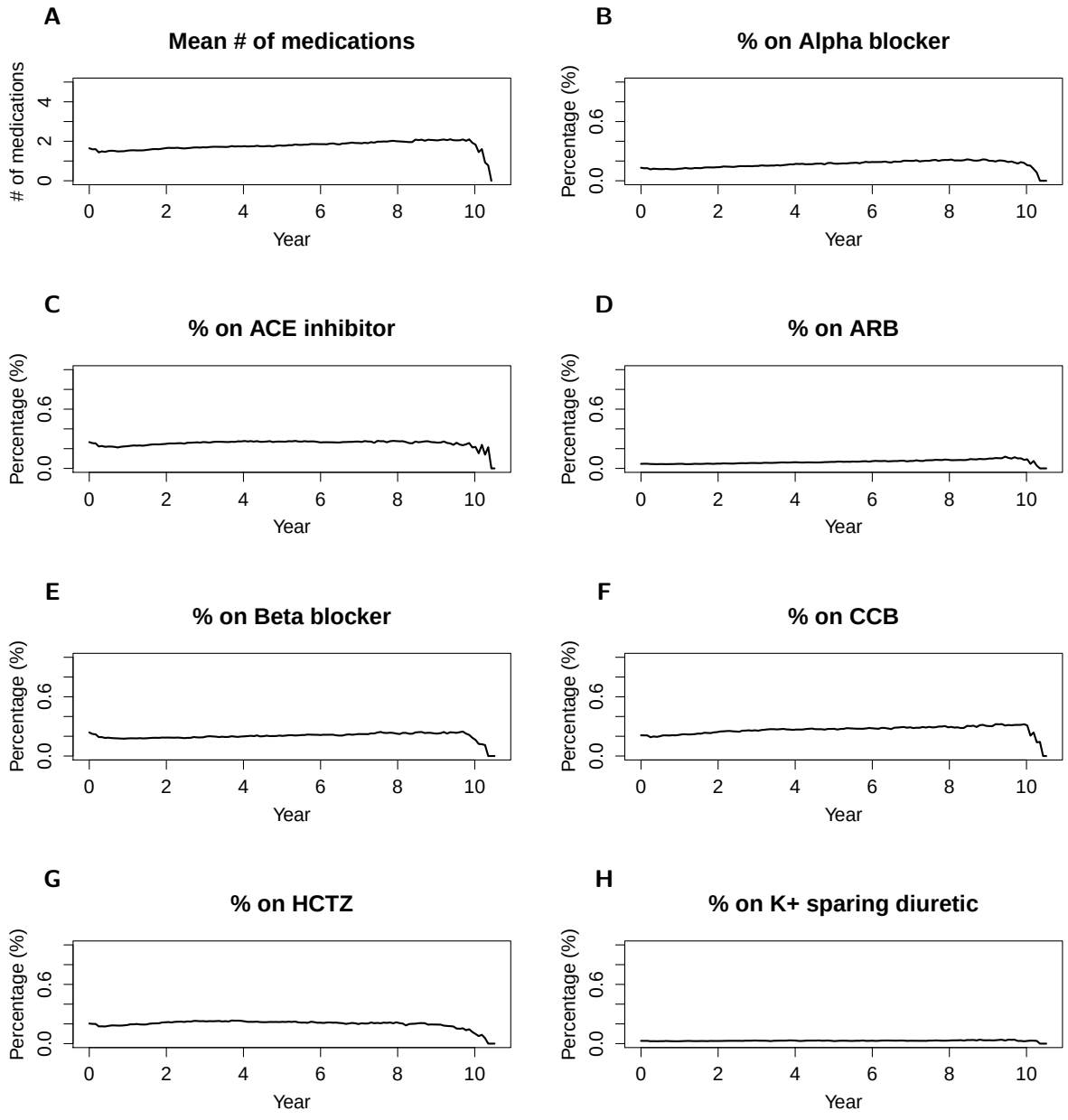

Figure 3.7: **Pointwise mean number of medications, and the percentage during each time interval of patients who were prescribed each major class or type of anti-hypertensives during the study period:** angiotensin converging enzyme-inhibitors (ACE-I), angiotensin receptor blockers (ARBs), beta blockers, calcium channel blockers (CCB), hydrochlorothiazide (HCTZ), and potassium sparing diuretics (K+ sparing).

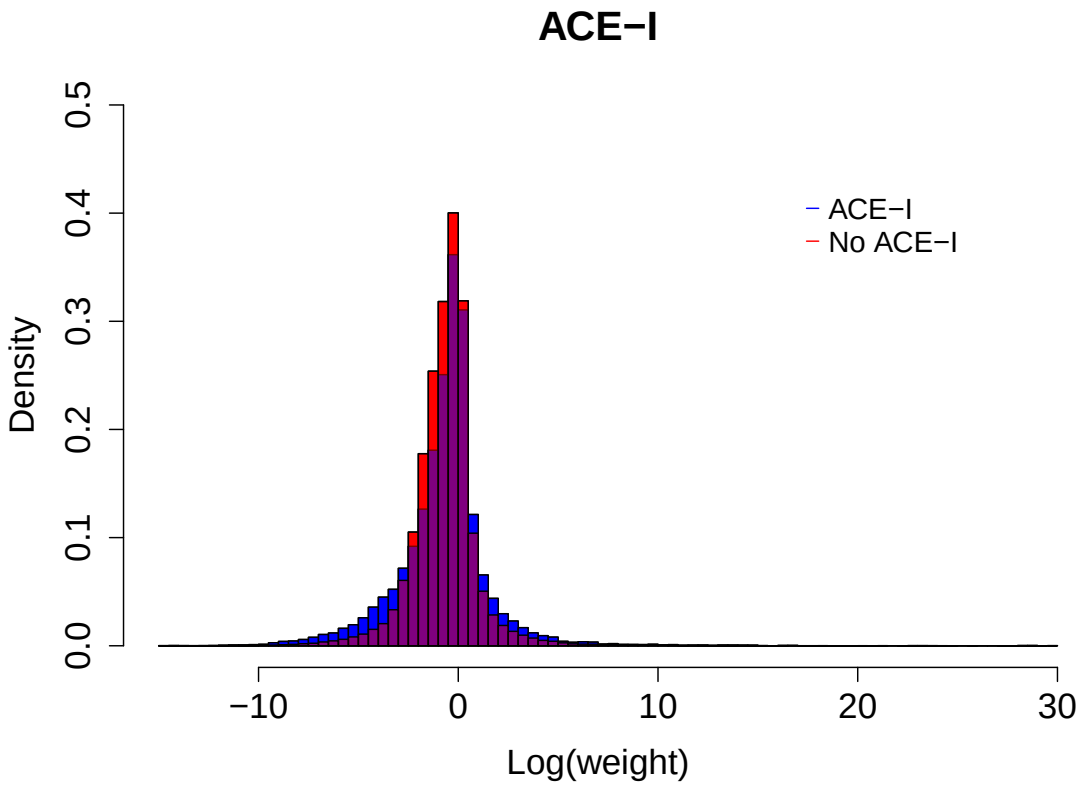

Figure 3.8: **Histogram of the logarithm of the estimated stabilized weights in the groups receiving and not receiving ACE-I.** This histogram shows the distribution of weights across all time points. The estimated weights for those receiving ACE-I are shown in blue, and those for patients not receiving ACE-I are shown in red. The purple region indicates the area of overlap.

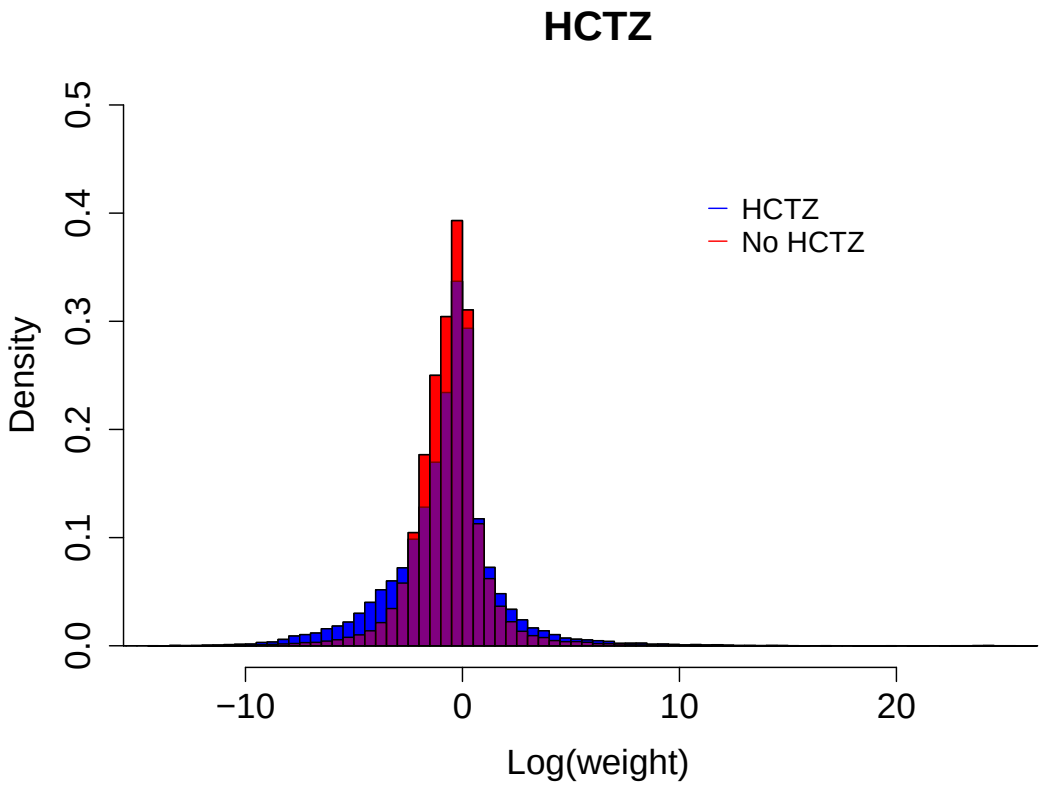

Figure 3.9: **Histogram of the logarithm of the estimated stabilized weights in the groups receiving and not receiving HCTZ.** This histogram shows the distribution of weights across all time points. The estimated weights for those receiving HCTZ are shown in blue, and those for patients not receiving HCTZ are shown in red. The purple region indicates the area of overlap.

Table 3.4 shows model diagnostics for weighting models of varying complexity; diagnostics displayed are deviance and Akaike Information Criteron. Several models were fit with different combinations of baseline covariates, baseline covariates with some interaction terms between baseline covariates, baseline covariates with time-varying clinical covariates, baseline covariates with time-varying clinical covariates and time-varying covariates for other medications, and all of the above in addition to interaction terms between baseline covariates and time-varying covariates.

Baseline covariates are: age, baseline number of anti-hypertensive medications medications, systolic and diastolic blood pressure at time of inclusion into the study, sex, smoking status, age, Framingham risk score, race, estimated GFR, creatinine, HIV status, HCV status, body mass index (BMI), and VACS index. Time varying covariates considered to be “Clinical(t)," representing clinical data which may have been used to adjust treatment were: systolic blood pressure, diastolic blood pressure, the VACS index, hbA1c, cholesterol, triglycerides, HDL, LDL, creatinine, and body mass index. The terms “Meds(t)" represent binary variables indicating whether or not medications other than the class of interest were dispensed at each time point. For example, if the treatment of interest in a particular analysis was ACE-Is, indicator variables for receipt at each time point of alpha blockers, ARBs, beta-blockers, CCBs, HCTZ, potassium sparing diuretics, furosemide, clonidine, and hydralazine were included as time-varying covariates. “Baseline variables, with interactions" indicates that the model was fit with baseline covariates and all pairwise interactions between age, Framingham risk score, estimated glomerular filtration rate, creatinine, and baseline number of anti-hypertensives. Interaction terms included in the most complex models were all pairwise interactions between the time-varying clinical covariates as well as the pairwise interactions between baseline covariates enumerated previously.

Four models were assessed for estimation of the denominator of the IPTW and IPCW, and two models were assessed for estimation of the numerator of IPTW and IPCW. AIC continued to

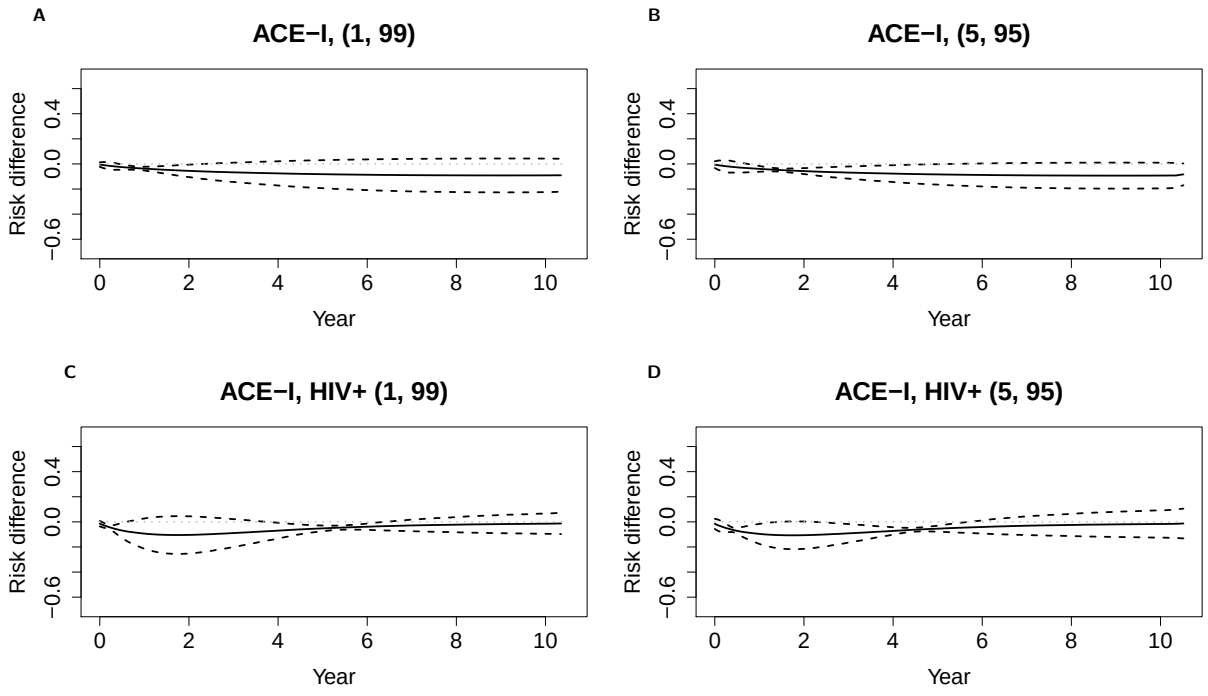

Figure 3.10: **The adjusted difference in survival probability between groups treated with ACE-I and not treated with ACE-I, under weight truncation thresholds at (1,99) and (5,95) percentile thresholds**. Results for the entire study population are shown in **A** and **B**. Results for the HIV positive subgroup are shown in **C** and **D**.

decrease with increasing model complexity across all four models, indicating an absence of overfitting even with the most complex model.

3.4.3 Supplementary Results

We conducted several sensitivity analyses using different percentile truncation thresholds for the stabilized weights: (1,99) and (5, 95). The adjusted difference in survival probability between groups treated with ACE-I and not treated with ACE-I, as well as subgroup analyses in HIV-positive individuals, under these different thresholds are presented in Figure 3.10. The results from analogous analyses of the effects of HCTZ are presented in figure 3.11. These analyses demonstrate that with increasing truncation, the results approach those from the unadjusted analyses presented in the main text.

Medication		Model	DoF	AIC	Deviance
ACE-I	IPTW (denominator)	Baseline	29	528371	528313
		Baseline + Clinical(t)	95	525354	525164
		Baseline + Clinical(t) + Meds(t)	158	476063	475747
		All, with interactions	213	473515	473089
	IPTW (numerator)	Baseline	29	528371	528313
		Baseline, with interactions	39	527044	526966
	IPCW (denominator)	Baseline	29	130721	130663
		Baseline + Clinical(t)	95	104914	104724
		Baseline + Clinical(t) + Meds(t)	158	104296	103980
		All, with interactions	213	104301	103875
	IPCW (numerator)	Baseline	29	130721	130663
		Baseline, with interactions	39	130719	130641
HCTZ	IPTW (denominator)	Baseline	29	469826	469768
		Baseline + Clinical(t)	95	466549	466359
		Baseline + Clinical(t) + Meds(t)	158	417813	417497
		All, with interactions	213	415499	415073
	IPTW (numerator)	Baseline	29	469826	469768
		Baseline, with interactions	39	467875	467797
	IPCW (denominator)	Baseline	29	130695	130637
		Baseline + Clinical(t)	95	104906	104716
		Baseline + Clinical(t) + Meds(t)	158	104259	103943
		All, with interactions	213	104263	103837
	IPCW (numerator)	Baseline	29	130695	130637
		Baseline, with interactions	39	130694	130616

Table 3.4: **Deviance and Akaike information criterion for several model fits of increasing complexity for estimation of IPTW and IPCW in the analysis of ACE-I and HCTZ.** Baseline covariates are: age, baseline number of anti-hypertensive medications medications, systolic and diastolic blood pressure at time of inclusion into the study, sex, smoking status, age, Framingham risk score, race, estimated GFR, creatinine, HIV status, HCV status, body mass index (BMI), and VACS index. Time varying covariates (Clinical(t)): systolic blood pressure, diastolic blood pressure, the VACS index, hbA1c, cholesterol, triglycerides, HDL, LDL, creatinine, and body mass index. Time-varying medication covariates (Meds(t)): binary variables for receipt of anti-hypertensives other than ACE-I and HCTZ respectively. Interactions between baseline covariates: all pairwise interactions between age, Framingham risk score, estimated glomerular filtration rate, creatinine, and baseline number of anti-hypertensives. All, with interactions: baseline covariates + Clinical(t) + Meds(t) + all pairwise interactions between the time-varying clinical covariates and all pairwise interactions between age, Framingham risk score, estimated glomerular filtration rate, creatinine, and baseline number of anti-hypertensives.

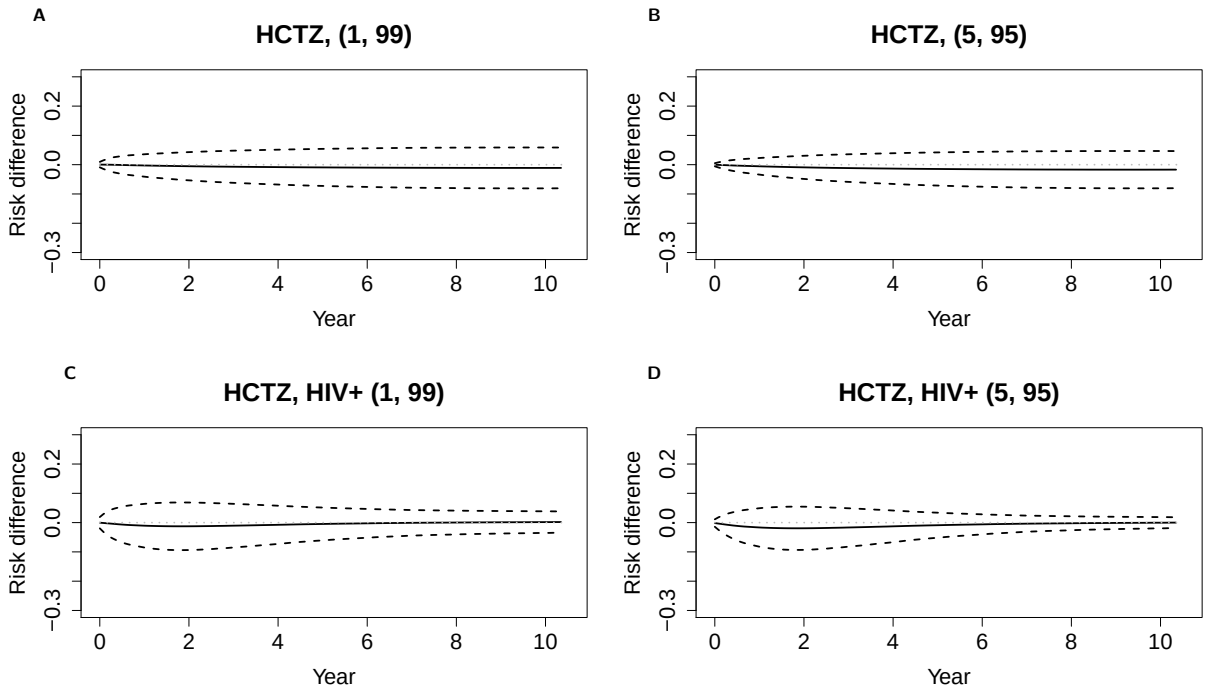

Figure 3.11: **The adjusted difference in survival probability between groups treated with HCTZ and not treated with HCTZ, under weight truncation thresholds at (1,99) and (5,95) percentile thresholds**. Results for the entire study population are shown in **A** and **B**. Results for the HIV positive subgroup are shown in **C** and **D**.

Tables 3.5 and 3.7 present the estimated coefficients from the marginal structural Cox model, weighted using the stabilized weights truncated at the (0.5, 99.5) percentiles, corresponding to the results discussed in the main text. The standard error of these estimates and the estimated hazard ratios, with bootstrapped 95% confidence intervals pooled across all 5 imputed datasets, are also shown. Tables 3.6 and 3.8 present the estimated coefficients from the unadjusted Cox proportional hazards regression. Similarly, the standard error of these estimates and the estimated hazard ratios with bootstrapped confidence intervals are shown.

	log(HR)	sd(log(HR))	HR (95% CI)
ACE-I	0.17	1.50	1.19 (0.06, 22.25)
Sex (male)	0.29	1.26	1.33 (0.11, 15.61)
Smoking - Current smoker	0.14	0.78	1.15 (0.25, 5.36)
Smoking - Past smoker	0.03	1.06	1.03 (0.13, 8.17)
Age	0.02	0.16	1.02 (0.74, 1.40)
FHRS	-0.09	0.78	0.91 (0.20, 4.23)
Race - Black	-0.26	1.73	0.77 (0.03, 22.67)
Race - Hispanic	-0.14	0.40	0.87 (0.40, 1.91)
Race - Other	0.01	1.05	1.01 (0.13, 8.00)
eGFR	0.01	0.10	1.01 (0.83, 1.23)
# HTN meds (t=0)	0.12	2.18	1.12 (0.02, 79.93)
Creatinine	-0.43	6.10	0.65 (0.00, 100807.34)
HIV+	1.44	2.05	4.24 (0.08, 234.61)
Chronic HCV	-0.08	0.34	0.92 (0.47, 1.81)
HCV genotype	-0.06	1.53	0.94 (0.05, 18.96)
HCV Ab+, RNA-	-0.15	1.54	0.86 (0.04, 17.68)
HCV Ab+, RNA unknown	0.46	2.38	1.58 (0.01, 167.36)
HCV Ab unknown, RNA-	-0.79	1.69	0.45 (0.02, 12.39)
HCV ICD9	-0.42	0.48	0.66 (0.26, 1.68)
HCV - never tested	-0.02	0.28	0.98 (0.57, 1.69)
VACS Index	0.02	0.04	1.02 (0.94, 1.10)
SBP (t=0)	0.00	0.01	1.00 (0.98, 1.03)
DBP (t=0)	0.00	0.01	1.00 (0.99, 1.02)
Age*FHRS	0.00	0.00	1.00 (0.99, 1.01)
Age*eGFR	0.00	0.00	1.00 (1.00, 1.00)
Age*# HTN meds (t=0)	-0.01	0.02	0.99 (0.95, 1.04)
Age*Creatinine	0.00	0.07	1.00 (0.86, 1.15)
FHRS*eGFR	0.00	0.01	1.00 (0.99, 1.01)
FHRS*# HTN meds (t=0)	0.01	0.03	1.01 (0.96, 1.06)
FHRS*Creatinine	0.11	0.18	1.12 (0.79, 1.60)
eGFR*# HTN meds (t=0)	0.00	0.01	1.00 (0.98, 1.02)
eGFR*Creatinine	0.00	0.08	1.00 (0.85, 1.16)
# HTN meds (t=0)*Creatinine	0.17	0.76	1.19 (0.27, 5.23)

Table 3.5: **Coefficients and hazards ratios estimated using Cox proportional hazards regression to evaluate the effect of ACE-I, adjusted using stabilized weights truncated at the 0.5% and 99.5% percentiles.**

	log(HR)	sd(log(HR))	HR (95% CI)
ACE-I	0.34	0.05	1.40 (1.28, 1.53)
Sex (male)	0.01	0.17	1.01 (0.72, 1.41)
Smoking - Current smoker	0.16	0.18	1.17 (0.82, 1.68)
Smoking - Past smoker	0.12	0.04	1.13 (1.04, 1.22)
Age	0.02	0.04	1.02 (0.94, 1.11)
FHRS	-0.15	0.18	0.86 (0.60, 1.22)
Race - Black	-0.04	0.13	0.96 (0.74, 1.25)
Race - Hispanic	-0.16	0.08	0.85 (0.73, 1.00)
Race - Other	-0.19	0.12	0.83 (0.66, 1.05)
eGFR	0.03	0.01	1.03 (1.00, 1.06)
# HTN meds (t=0)	0.36	0.46	1.43 (0.58, 3.52)
Creatinine	0.32	0.74	1.38 (0.32, 5.91)
HIV+	1.72	0.79	5.57 (1.19, 25.99)
Chronic HCV	-0.19	0.19	0.83 (0.57, 1.21)
HCV genotype	0.07	0.61	1.07 (0.33, 3.54)
HCV Ab+, RNA-	-0.06	0.08	0.94 (0.81, 1.10)
HCV Ab+, RNA unknown	0.12	0.26	1.12 (0.67, 1.87)
HCV Ab unknown, RNA-	-0.19	0.22	0.82 (0.54, 1.27)
HCV ICD9	0.12	0.52	1.13 (0.41, 3.12)
HCV - never tested	0.04	0.04	1.04 (0.96, 1.13)
VACS Index	0.01	0.02	1.01 (0.97, 1.04)
SBP (t=0)	0.00	0.00	1.00 (0.99, 1.01)
DBP (t=0)	0.00	0.01	1.00 (0.99, 1.01)
Age*FHRS	0.00	0.00	1.00 (1.00, 1.00)
Age*eGFR	0.00	0.00	1.00 (1.00, 1.00)
Age*# HTN meds (t=0)	0.00	0.00	1.00 (0.99, 1.00)
Age*Creatinine	0.00	0.01	1.00 (0.98, 1.02)
FHRS*eGFR	0.00	0.00	1.00 (1.00, 1.00)
FHRS*# HTN meds (t=0)	0.00	0.00	1.00 (0.99, 1.00)
FHRS*Creatinine	0.06	0.05	1.06 (0.97, 1.17)
eGFR*# HTN meds (t=0)	0.00	0.00	1.00 (0.99, 1.00)
eGFR*Creatinine	-0.01	0.01	0.99 (0.97, 1.02)
# HTN meds (t=0)*Creatinine	-0.08	0.15	0.93 (0.69, 1.25)

Table 3.6: **Coefficients and hazards ratios estimated using Cox proportional hazards regression to evaluate the effect of ACE-I, unadjusted**

	$\log(HR)$	sd($\log(HR)$)	HR (95% CI)
HCTZ	-0.12	0.66	0.89 (0.24, 3.28)
Sex (male)	1.14	3.37	3.13 (0.00, 2329.15)
Smoking - Current smoker	0.22	0.47	1.24 (0.49, 3.15)
Smoking - Past smoker	0.49	1.57	1.63 (0.08, 35.25)
Age	-0.02	0.31	0.98 (0.53, 1.81)
FHRS	0.01	1.60	1.01 (0.04, 23.10)
Race - Black	0.36	1.20	1.43 (0.14, 15.11)
Race - Hispanic	-0.12	0.55	0.89 (0.30, 2.60)
Race - Other	-0.01	0.35	0.99 (0.50, 1.97)
eGFR	-0.01	0.21	0.99 (0.66, 1.50)
# HTN meds (t=0)	-0.91	5.89	0.40 (0.00, 41446.93)
Creatinine	-0.80	10.23	0.45 (0.00, 229931903.21)
HIV+	1.56	2.22	4.78 (0.06, 372.08)
Chronic HCV	-0.22	0.30	0.80 (0.45, 1.43)
HCV genotype	-0.94	1.43	0.39 (0.02, 6.52)
HCV Ab+, RNA-	-0.19	0.51	0.83 (0.30, 2.27)
HCV Ab+, RNA unknown	0.15	0.79	1.16 (0.25, 5.47)
HCV Ab unknown, RNA-	0.11	0.58	1.11 (0.36, 3.49)
HCV ICD9	-0.09	0.65	0.92 (0.25, 3.31)
HCV - never tested	-0.16	0.48	0.85 (0.33, 2.19)
VACS Index	0.01	0.03	1.01 (0.95, 1.08)
SBP (t=0)	0.01	0.03	1.01 (0.96, 1.06)
DBP (t=0)	-0.01	0.06	0.99 (0.88, 1.11)
Age*FHRS	0.00	0.01	1.00 (0.99, 1.01)
Age*eGFR	0.00	0.00	1.00 (1.00, 1.00)
Age*# HTN meds (t=0)	0.00	0.03	1.00 (0.95, 1.06)
Age*Creatinine	-0.01	0.10	0.99 (0.81, 1.20)
FHRS*eGFR	0.00	0.01	1.00 (0.98, 1.01)
FHRS*# HTN meds (t=0)	-0.02	0.05	0.98 (0.90, 1.08)
FHRS*Creatinine	-0.03	0.61	0.97 (0.30, 3.20)
eGFR*# HTN meds (t=0)	0.01	0.03	1.01 (0.96, 1.06)
eGFR*Creatinine	-0.02	0.03	0.98 (0.93, 1.03)
# HTN meds (t=0)*Creatinine	0.36	2.02	1.43 (0.03, 75.58)

Table 3.7: **Coefficients and hazards ratios estimated using Cox proportional hazards regression to evaluate the effect of HCTZ, adjusted using stabilized weights truncated at the 0.5% and 99.5% percentiles.**

	$\log(HR)$	sd($\log(HR)$)	HR (95% CI)
HCTZ	0.09	0.05	1.10 (0.99, 1.21)
Sex (male)	0.04	0.18	1.04 (0.72, 1.49)
Smoking - Current smoker	0.15	0.17	1.17 (0.83, 1.64)
Smoking - Past smoker	0.11	0.05	1.12 (1.02, 1.23)
Age	0.02	0.04	1.02 (0.95, 1.10)
FHRS	-0.16	0.20	0.85 (0.57, 1.27)
Race - Black	-0.04	0.11	0.96 (0.78, 1.19)
Race - Hispanic	-0.16	0.09	0.85 (0.72, 1.01)
Race - Other	-0.18	0.10	0.84 (0.69, 1.02)
eGFR	0.03	0.01	1.03 (1.00, 1.05)
# HTN meds (t=0)	0.39	0.39	1.47 (0.68, 3.18)
Creatinine	0.30	0.60	1.35 (0.41, 4.40)
HIV+	1.73	0.78	5.62 (1.21, 26.05)
Chronic HCV	-0.20	0.18	0.82 (0.58, 1.17)
HCV genotype	0.00	0.67	1.00 (0.27, 3.73)
HCV Ab+, RNA-	-0.06	0.08	0.94 (0.81, 1.09)
HCV Ab+, RNA unknown	0.11	0.30	1.11 (0.62, 1.98)
HCV Ab unknown, RNA-	-0.16	0.23	0.85 (0.54, 1.33)
HCV ICD9	0.08	0.59	1.09 (0.34, 3.46)
HCV - never tested	0.05	0.05	1.05 (0.96, 1.15)
VACS Index	0.01	0.02	1.01 (0.97, 1.04)
SBP (t=0)	0.00	0.01	1.00 (0.99, 1.01)
DBP (t=0)	0.00	0.01	1.00 (0.99, 1.01)
Age*FHRS	0.00	0.00	1.00 (1.00, 1.00)
Age*eGFR	0.00	0.00	1.00 (1.00, 1.00)
Age*# HTN meds (t=0)	0.00	0.00	1.00 (0.99, 1.00)
Age*Creatinine	0.00	0.01	1.00 (0.98, 1.02)
FHRS*eGFR	0.00	0.00	1.00 (1.00, 1.00)
FHRS*# HTN meds (t=0)	0.00	0.00	1.00 (0.99, 1.00)
FHRS*Creatinine	0.06	0.05	1.07 (0.96, 1.18)
eGFR*# HTN meds (t=0)	0.00	0.00	1.00 (1.00, 1.00)
eGFR*Creatinine	-0.01	0.01	0.99 (0.97, 1.02)
# HTN meds (t=0)*Creatinine	-0.09	0.13	0.92 (0.72, 1.17)

Table 3.8: **Coefficients and hazards ratios estimated using Cox proportional hazards regression to evaluate the effect of HCTZ, unadjusted**

Chapter 4

Modeling COVID-19 care capacity in a major health system

Abstract

Hospital resources, especially critical care beds and ventilators, have been strained by additional demand throughout the COVID-19 pandemic. Rationing of scarce critical care resources may occur when available resource limits are exceeded. However, the dynamic nature of the COVID-19 pandemic and variability in projections of the future burden of COVID-19 infection pose challenges for optimizing resource allocation to critical care units in hospitals. Connecticut experienced a spike in the number of COVID-19 cases between March and June 2020. Uncertainty about future incidence made it difficult to predict the magnitude and duration of the increased COVID-19 burden on the healthcare system. In this paper, we describe a model of COVID-19 hospital capacity and occupancy that generates estimates of the resources necessary to accommodate COVID-19 patients under infection scenarios of varying severity. We present the model structure and dynamics, procedure for parameter estimation, and publicly available web application where we implemented the tool. We then describe calibration using

data from over 3,000 COVID-19 patients seen at the Yale-New Haven Health System between March and July 2020. We conclude with recommendations for modeling tools to inform decision-making using incomplete information during future crises.

Keywords: intensive care unit, emergency department, SARS-CoV-2

4.1 Introduction

The novel severe acute respiratory syndrome coronavirus (SARS COV-2), which causes coronavirus disease 2019 (COVID-19), emerged in 2019 in Hubei province in China. Early reports from China suggested that people infected with COVID-19 are at high risk for severe respiratory disease and serious complications [111]. Management of respiratory failure and acute respiratory distress syndrome (ARDS) often requires mechanical ventilation managed in an intensive care unit (ICU). A retrospective study of inpatients at the Jinyintan Hospital and Wuhan Pulmonary Hospital reported that 54% of patients experienced respiratory failure, 31% of inpatients developed ARDS, and 59% developed sepsis. Of the adults in this study, 26% required critical care [112]. A large outbreak of COVID-19 also occurred in Lombardy, Italy, which reported that 99% of critically ill patients required respiratory support, with mechanical ventilation in 88% of patients [113].

In 2020 and the spring of 2021, the surge of COVID-19 patients in the United States revealed that critical care resources available in some health systems were insufficient to address the COVID-19 outbreak. Early models of hospital capacity in New York City, which experienced a severe wave of COVID-19 infections beginning in March 2020, projected that the city would require 40,000 ICU beds to handle the peak number of COVID-19 cases, with 3,000 ICU beds available at baseline [114]. Subsequently, localized bed shortages occurred in parts of New York City [115, 116], spurring hospital systems [117–119] and governors [120–123] to order

immediate bed capacity expansion. This expansion was accomplished through the cancellation of elective procedures[124], bed re-purposing for critical care [125], and the addition of floor beds in overflow centers [123, 126, 127]. Failure to meet the critical care needs of COVID-19 patients has dire consequences; in Italy, reports of rationing of ventilators based on age cutoffs emerged from overwhelmed hospitals and ICUs [128–131]. During the fall and winter of 2020 and spring of 2021, waves of new COVID-19 cases strained critical care resources at hospitals across the United States and internationally. Rates of COVID-19 in the United States skyrocketed to record levels, with over 160,000 new cases of COVID-19 reported in a single day [132, 133]. Health systems across the country have reported that their critical care units were either near or at full capacity due to a surge of COVID-19 patients [134–137]. Furthermore, the spread of COVID-19 has affected the availability of staff, with 900 healthcare workers at Mayo Clinic testing positive and healthcare workers with asymptomatic COVID-19 continuing to provide care in North Dakota due to staff shortages and a high burden of COVID-19 patients [135, 138].

The development and administration of vaccines such as Pfizer BNT162b2 and Modern mRNA-1273 in high-income nations like the United States, Israel, and the United Kingdom has slowed rates of infection with SARS-Cov-2, but critical care capacity continues to be overwhelmed by the unvaccinated population and the development of aggressive variants. In the southern United States, one in four ICUs are above 95% capacity, with COVID19 patients accounting for approximately half of all ICU patients [139]. These high hospitalization rates are driven by evolving variants of COVID-19, such as the Delta variant, which the CDC has deemed more contagious than previous variants, more likely to cause severe illness, and capable of causing breakthrough infections in vaccinated people [140].

Tools for managing hospital capacity during surges in COVID-19 cases will be required as long as vaccination rates remain low in some countries and new variants continue to arise. Rapid practice guidelines have recommended several steps to increase critical care capacity to meet

the demands of COVID-19 patients [141]. Hospitals have increased their supply of ventilators and developed protocols to reserve ventilators by utilizing non-invasive ventilation and high-flow nasal oxygen [142–144]. Hospitals may also suspend elective medical and surgical procedures when necessary and re-allocate staff from other departments to serve in critical care or COVID-19 specific units [145, 146]. Physicians and nurses trained in critical care medicine have been given larger teams of non-critical care trained staff in an attempt to increase their ability to manage as many patients simultaneously as possible [146]. Procedures for cleaning and re-using personal protective equipment necessary to protect staff from infection, such as N95 masks and air-purifying respirators, have also been developed, although the efficacy of these procedures has not yet been rigorously studied [141].

Rapid practice guidelines recommend the use of mathematical modeling to guide surge capacity planning in hospital systems which expect to encounter potential shortages in critical care resources [141]. Guidelines state that the models should “be pragmatic and focus on the only relevant question for surge capacity: how many patients will need hospital and ICU resources on a given day?" More specifically, the models should provide early predictions, insight regarding both best and worst case scenarios, and the local rate of spread of infection and rate of hospitalization. Many modeling tools were created at the beginning of the COVID-19 pandemic to assist with predictions of incident COVID-19 cases and hospitalizations [147–162]. A comparison of four prominent models by Chin et al. [163] found that “for accuracy of prediction, all models fared very poorly." These tools used population-level epidemic projections as inputs to their model of hospital occupancy, which may have contributed to compounding errors in forecasting hospital bed occupancy due to uncertainties in the early epidemiological models of COVID-19. The authors concluded that “trustworthy models require trustworthy input data" and that the models “need to be subjected to pre-specified real time performance tests."

In this paper, we present model of COVID-19 hospital occupancy that uses data from health systems to generate its predictions. The model is independent of the uncertainty in population-

level epidemiological predictions of infection and can flexibly accommodate local variations important to decision-makers, such as significant differences in patient demographic distributions and hospital protocols. The model provides predictions of floor and ICU occupancy and mortality for infection scenarios specified by the hospital administrator or decision-maker. These infection scenarios can be informed both by observed presentations of COVID-19 patients to the health care system and epidemiological predictions of infection. The model can predict the effects of planned modifications to hospital capacity, and projections can be tailored to the dynamics of a specific hospital system using several parameters, including age-specific average lengths of stay in each hospital department, the probabilities of transitioning between those departments, and probability of discharge and death. We introduce the model structure, and describe both the model dynamics and the calibration procedure for the model parameters. The model was calibrated using observed patient trajectories from the 3000-bed Yale New-Haven Hospital system, collected and processed during the surge in COVID-19 cases in Connecticut between March and July 2020. We validate the model dynamics using the observed hospital census during this time period. We conclude with recommendations to guide scientists developing model-based recommendations for managing hospital capacity during the COVID-19 pandemic.

4.2 Methods

4.2.1 Model structure

In this model, we describe the flow of COVID+ patients through a hospital system using a system of ordinary differential equations. By COVID+ patients presenting to ED, we refer to patients who have either tested positive for COVID-19 prior to admission or those that test positive for COVID-19 within 14 days of admission. Thus, we exclude two ways patients within

the hospital could be COVID-19 positive, namely through nosocomial infection and through transfers from other hospitals post-triage. We model transitions between eight different compartments: 1) presentations to the health system or emergency department, where triage occurs P, 2) floor beds F, 3) ICU beds C, 4) MS, corresponding to a state post-discharge from the emergency department (ED) for those patients with mild symptoms, 5) R, corresponding to recovery post-discharge for patients admitted to the hospital, 6) WF, the queue for floor beds which would develop if floor beds are not available, 7) WC, the queue for ICU beds which would develop if ICU beds are not available, and 8) death. We model the following events in the health system.

- **Presentation to the health system:** COVID+ patients can present to their health system for asymptomatic COVID-19 screening, following a positive COVID-19 test, or following initial onset of symptoms. They may present either to outpatient clinics or to the ED of their local hospital. At these locations, *triage* occurs, such that patients are either discharged with mild symptoms, admitted to the floor, or admitted to the ICU.

- **Following admission to the floor:** Patients admitted to the floor can be discharged directly from the floor, require *stepping up* to an ICU due to a deterioration in their condition, or die on the floor. Patients may arrive on the floor from the ED or after leaving the ICU following an improvement in their condition.

- **Following admission to the ICU:** Patients in the ICU require frequent monitoring and intensive interventions. COVID+ patients are especially at risk of requiring critical care due to high rates of pneumonia and acute respiratory distress syndrome (ARDS) reported in COVID+ patients [2–5, 112, 130]. Following recovery, ICU patients *step down* to the floor for discharge. Patients may also die in the ICU.

- **Following discharge from the ED:** If a patient's condition is not severe, COVID+ patients are instructed to return home and to self-isolate. Most patients with mild disease

recover during their isolation. Some patients return to the ED in worsened condition, and a small number may die after discharge.

- **If the floor has reached capacity:** If floor beds are not available, patients may have to wait in the ED or elsewhere until a bed becomes available. During this time, patients are often monitored or receive care. They may be discharged if they recover, or die.

- **If the ICU has reached capacity:** If ICU beds are not available, patients may have to wait on the floor, in the ED, or elsewhere until a bed becomes available. Because these patients are critically ill, the probability of death if no care if received is high.

The rate parameters associated with each step of the model are shown in Table 4.1.

We stratify incoming patients into age tiers, based on known differences in patient outcomes by age [2]. Dynamics for each age group are governed by the following system of ordinary differential equations. The dynamics for each age group are coupled by the constraint on total hospital floor beds L and total ICU beds M. For age group i,

$$
\begin{aligned}
\frac{dP_i}{dt} &= \xi_{MS_i} MS_i - (\sigma_{MS_i} + \sigma_{C_i} + \sigma_{F_i} + \mu_{P_i}) P_i \\
\frac{dMS_i}{dt} &= \sigma_{MS_i} P_i - (\phi + \mu_{MS_i} + \xi_{MS_i}) MS_i \\
\frac{dWC_i}{dt} &= (\sigma_{C_i} P_i + \theta_{F_i} F_i + \theta_{WF_i} WF_i) \left(1 - \frac{1}{1 + e^{s(C-M)}}\right) - \mu_{WC_i} WC_i - \eta WC_i \left(\frac{1}{1 + e^{s(C-M)}}\right) \\
\frac{dC_i}{dt} &= (\sigma_{C_i} P_i + \theta_{F_i} F_i + \theta_{WF_i} WF_i + \eta WC_i) \left(\frac{1}{1 + e^{s(C-M)}}\right) - (\mu_{C_i} + \chi_{C_i}) C_i \\
\frac{dWF_i}{dt} &= (\sigma_{F_i} P_i + \chi_{C_i} C_i) \left(1 - \frac{1}{1 + e^{s(F-L)}}\right) - \zeta WF_i \left(\frac{1}{1 + e^{s(F-L)}}\right) - (\mu_{WF_i} + \theta_{WF_i}) WF_i \\
\frac{dF_i}{dt} &= (\sigma_{F_i} P_i + \zeta WF_i + \chi_{C_i} C_i) \left(\frac{1}{1 + e^{s(F-L)}}\right) - (\chi_{F_i} + \mu_{F_i} + \theta_{F_i}) F_i \\
\frac{dR_i}{dt} &= \phi MS_i + \chi_{F_i} F_i \\
\frac{dD_i}{dt} &= \mu_{MS_i} MS_i + \mu_{WC_i} WC_i + \mu_{C_i} C_i + \mu_{WF_i} WF_i + \mu_{F_i} F_i,
\end{aligned}
$$

Table 4.1: Rate parameters included in the model of hospital capacity

Parameter	Description
Rates of departure from ED	
σ_{MS}	Rate at which ED patients are triaged as having mild symptoms
σ_C	Rate at which ED patients are triaged to the ICU (boarding)
σ_F	Rate at which ED patients are triaged to the floor
Rates of departure from the floor or floor queue	
θ_F	Rate of stepping up from the floor to the ICU
χ_F	Rate of stepping down from the floor (discharge)
ζ	Rate at which patients are moved to the floor from the floor queue
χ_{WF}	Rate of stepping down from the floor queue (discharge)
θ_{WF}	Rate of stepping up from the floor queue to the ICU
Rates of departure from the ICU or ICU queue	
χ_C	Rate of stepping down from the ICU to the floor or the floor queue
η	Rate at which patients are moved to the ICU from the ICU queue
Rates for those with mild symptoms	
ϕ	Rate of recovery of patients triaged as having mild symptoms
ξ_{MS}	Rate at which patients with mild symptoms return to the ED
Death rates	
μ_{MS}	Death rate of patients triaged as having mild symptoms
μ_C	Death rate of patients in the ICU
μ_F	Death rate of patients on the floor
μ_{WC}	Death rate of patients waiting for an ICU bed
μ_{WF}	Death rate of patients waiting for a floor bed

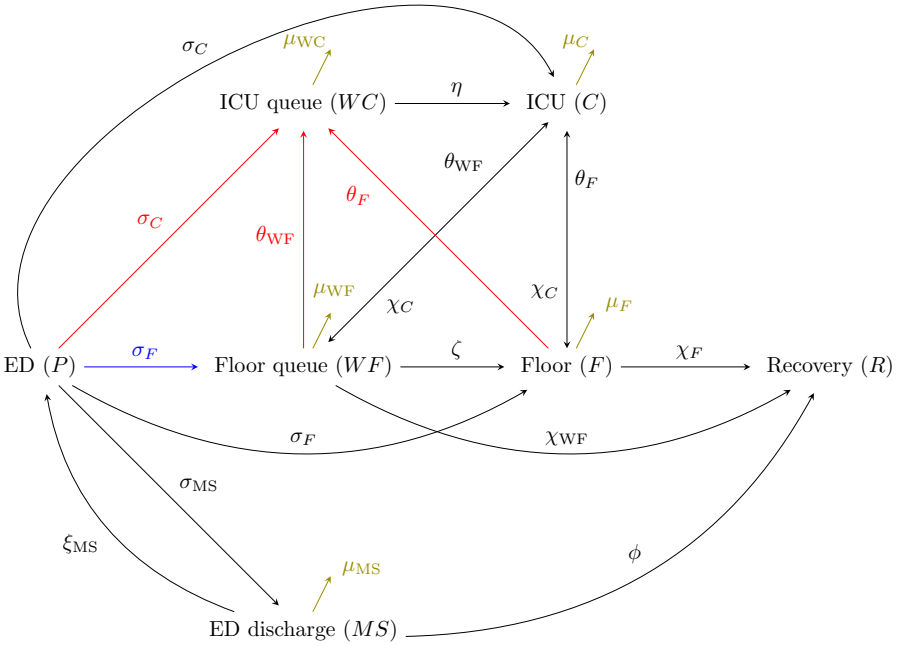

Figure 4.1: Model structure and parameters. Simplified dynamics are presented here for the case of one age-group for clarity. Patients present to the hospital system via the ED, where they are triaged and either discharged or admitted to the hospital. Patients who are admitted may go to the floor or directly to the ICU. The model captures patient flow from the floor to the ICU and back, as well as discharge dynamics from both the ED and the floor to recovery. Rate parameters which capture the speed at which patients transition between compartments are included here and described in Table 4.1. Arrows in red and blue represent patient-flow over-flow dynamics in the ICU and floor, respectively. Olive-colored arrows represent death rates from each compartment.

where $I_i(t)$ represents initial COVID+ presentations to the ED of age group i, $F = \sum_i F_i$ is the occupied floor capacity, $C = \sum_i C_i$ is the occupied ICU capacity, L and M are their respective available capacities, and D is the state of death. The exponential sigmoidal terms in the above system of equations represent approximations to the ideal on-off switch, i.e., $f(x) = \mathbf{1}_{x \geq i} \approx (1 - \frac{1}{1+e^{s(x-i)}})$, with the parameter s controlling the fidelity of the approximation, the higher, the better. The use of sigmoids speeds up the computation of the solution of the system of equations significantly, allowing the user to rapidly observe the effect of changes in inputs on the outputs of the model with minimal error. A graphical depiction of these dynamics is shown in Figure 4.1.

4.2.2 Specifying dynamics and capacity scenarios

The following features of the model are modifiable by the user in the web application in which the model is implemented.

The user begins by specifying an infection scenario: the number of COVID+ patients that present to a health system per day during a specified time horizon. The user chooses a number of days between 2 and 60 during which to generate capacity projections. They also select the initial number of COVID+ presentations at day zero of the projection and the expected behavior of the change in the number of COVID+ presentations during the time of the projection. For example, the user could choose an initial number of presentations of 50 patients, “exponential" change in patients, a doubling time of 14 days, and a time horizon of 14 days. This scenario translates to an exponential increase in the number of COVID+ presentations per day, such that 100 COVID+ patients present to the health system on day 14. Choices for the type of increase were: exponential, linear, saturated, and flat (no increase).

The goal of the model is to allow hospitals to look at projections of their occupancy and expected clinical outcomes. We allow the user to specify both the baseline number of available

beds in the ICU and on the floor and a possible policy response, an increase their number of floor and ICU beds dedicated to COVID+ patients. For this purpose, we allow the user to specify a one-time linear ramping of capacity for either/both type(s) of beds in an interval within the time-frame. In terms of the system of differential equations laid out above, in that case L and M would be replaced with the piece-wise linear $L(t)$ and $M(t)$ respectively.

The user may also modify several key parameters which reflect the patient population in a specific catchment area, allowing the user to tailor the model to their particular needs. These parameters include the age distribution of admitted COVID+ patients, the average length of stay of COVID+ patients in the ICU and on the floor, and the probability of death of COVID+ patients in the ICU and on the floor. The user may only specify length of stay and probability of death for two age groups: adults between 18-64 years old and 65 years and above. We assume that the proportion of COVID+ patients under the age of 18 is small. Thus, the output of the model should not be sensitive to variation in these parameters for this age group. Default values in the web application are based on the YNHHS patient population.

Key outputs of the model which inform decision-marking regarding resource allocation are: the number of days to overflow, extra beds needed for COVID+ patients, number of deaths in each compartment of the model, and predicted case-fatality rate. Time to overflow and number of extra beds needed for COVID+ patients are important metrics for hospital decision-makers. Increasing the number of beds dedicated to COVID-19 patients is one of the most significant policy levers available to hospital management. The ability to predict overflow could lead to dedication of non-COVID resources to COVID+ patients or the acquisition of additional space, for example. It is also clinically important, especially as the outcomes of patients that need ICU care are significantly worse if there are no available beds (a surge scenario). The predicted number of deaths among COVID+ patients helps decision-makers with understanding the possible consequences of allocating beds and resources in different ways. The number of deaths expected under the prescribed infection scenario and capacity limits are broken down

by location: deaths in the ICU, on the floor, and in the floor and ICU queues. The case fatality rate is computed using the number of predicted deaths and the total number of COVID+ patients entering the system during the time of the simulation.

4.2.3 Model calibration

Data sources

We used data from the Yale-New Haven Hospital System (YNHHS) collected between March 2020 and July 2020 to calibrate the model. YNHHS consists of five hospitals: Yale-New Haven Hospital (1,608 beds), Bridgeport Hospital (719 beds), Greenwich Hospital (304 beds), Lawrence and Memorial Hospital (260 beds), and Westerly Hospital (81 beds). For parameters which could not be estimated using available YNHHS data, we used population-level estimates from the Center for Disease Control's Morbidity Mortality Weekly Report (CDC MWWR) [2–5]. This study received approval from the Institutional Review Board of Yale University's Human Research Protection Program (IRB ID: 2000028666).

We used three YNHHS data sources to calibrate the model: individual-level records for patients who had tested positive for COVID-19 in the YNHHS Emergency Departments (ED), individual-level records for patients who had been admitted to the hospital, and hospital-level summaries of capacity. For patients presenting to a YNHHS ED, we had access to ED arrival time, the primary chief complaint, age, ED departure time, admission status, admission department, the time of the positive COVID-19 laboratory test, and the date and time of previous presentation to the ED. For patients who were admitted, we had access to daily records including inpatient departments, location, age, and the dates at which they were moved between departments, admitted to the ICU, initiated on ventilation, discharged, or died. Using these data, we reconstructed patient trajectories through the YNHHS hospitals. Figure 4.2A

shows the total census of hospitalized COVID+ patients in YNHHS, as well as the census on the floor and ICU specifically. Figures 4.2B-E show the survival probabilities of death and departure from the floor and ICU, without accounting for competing hazards.

Procedure for calibration of parameters

To fit the model, we used survival analysis with competing risks to estimate the following parameters governing rates of transition between hospital departments using the patient records available in the YNHHS dataset. We estimated three parameters describing rates of departure from the ED: from the ED to discharge (σ_{MS}), the ED to admission to the floor (σ_F), and the ED to admission to the ICU (σ_C). We used inpatient data to estimate the rate of transition from the floor to the ICU (θ_F), rate of discharge from the floor (χ_F), rate of transition from the ICU to the floor (χ_C), death while on the floor (μ_F) and death while in the ICU (μ_C). We performed a primary analysis in which time to all competing events were assumed to follow a gamma distribution. We estimated parameters in three age groups (0-17 years, 18-64 years, 65+ years). We used bootstrapping with 1,500 samples to generate estimates of the variance of these parameters. We performed two secondary analyses with simpler parameterizations to assess the performance of the estimated model parameters under different distributional assumptions: 1) time to each competing event is exponentially distributed, or 2) time to discharge follows a gamma distribution, and time to all other events is exponentially distributed. Additional statistical details of the procedure are described in the Supplementary Appendix.

Several parameters could not be estimated from available YNHHS data. In most cases, these parameters were estimated using data published in the CDC MMWR [2–5]. The YNHHS data did not provide information regarding probabilities or times to full recovery among individuals after they were discharged. Therefore, rates of recovery (ϕ), and death rates among individuals with mild symptoms (μ_{MS}) were calculated in each age group from population-level proportions provided by CDC MMWR [2]. Because queues for floor beds and ICU beds did

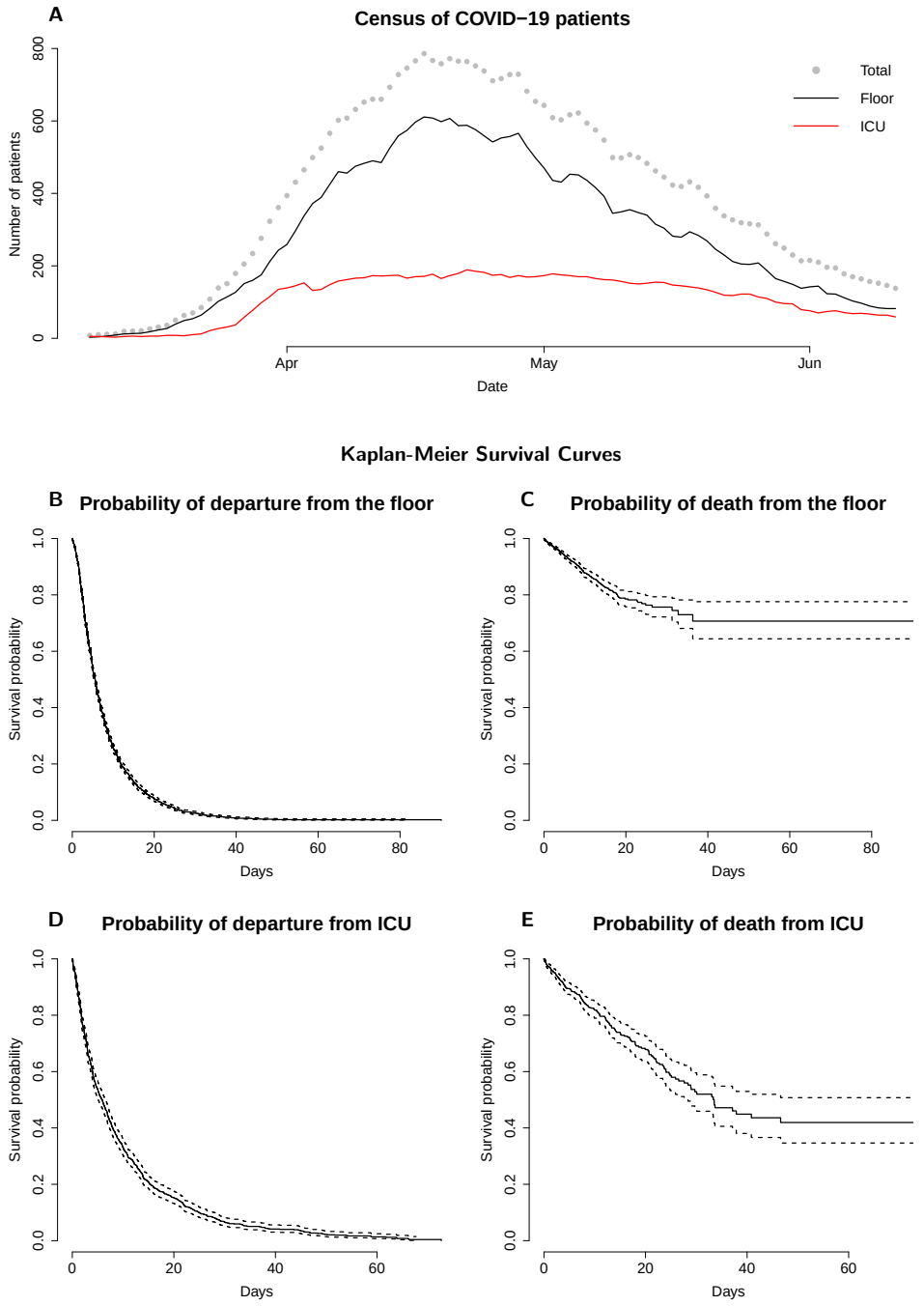

Figure 4.2: A: The census of hospitalized patients throughout YNHHS. The total census is shown with gray points, the census on the floor is shown in black, and the census in the ICU is shown in red. B and D: Kaplan-Meier curves describing the probability of remaining on the floor or ICU given time since arrival to the floor or ICU respectively. Patients are considered to be right-censored if they are still on the floor or in the ICU at the end of the observation period. Departure includes discharge, transfer to another department, and death. C and E: Kaplan-Meier curves describing the probability of survival, and not death, on the floor and in the ICU given time since arrival to the floor or ICU respectively. Patients are considered to be right-censored if they are transferred to another department, are discharged, or remain in the floor or in the ICU at the end of the observation period.

not occur at YNHHS hospitals at the time of model implementation, we were required to make several assumptions to generate the remaining parameters. We estimated that the rate of death, discharge, and step up to the ICU from the floor queue (μ_{WF}, χ_{WF}, and θ_{WF}, respectively) would be the same as death and discharge rates from the floor, due to the fact that many hospitals are able to provide care to patients who are waiting for a floor bed before one becomes available. We set the average time of death in the ICU queue without access to critical care resources to be 6 hours. We set the probability of movement from the floor queue to an open floor bed and the ICU queue to an open ICU bed to be 0.9, reflecting a 90% chance that a patient would move from the queue to an open bed before discharge, death, or transition to another department. Rates of movement out of the queues were calculated accordingly.

4.2.4 Parameter estimates

The observed time-series of patients entering each department included both COVID-19 patients admitted directly to each department and direct transfers from other hospitals. Figure 4.2A shows the total census of COVID-19 patients admitted to all YNHHS hospitals during the observation period. Parameter estimation used all observed YNHHS patient trajectories, and model fit was evaluated using the the largest hospital, YNHH, where decision-making regarding resource allocation was most critical. In total, 2,275 COVID-19 patients who met criteria were admitted or transferred to YNHH during the observation period. The estimated capacity of YNHH for COVID-19 patients was 180 beds in the ICU and 578 beds on the floor. YNHH neither reached capacity nor ran out of ventilators during the surge in COVID-19 cases. (Figure 4.2) We used individual-level patient trajectories and time-stamped transitions between departments to calculate the time spent by each patient in the ED, on the floor, and in the ICU. We determined the destination of each patient after each of their stays in these departments; discharge, death, or another hospital department. We also identified patients whose trajectories and outcomes were right-censored by the end of the observation period.

The results of one analysis, in which model parameters were estimated assuming that time to each event follows a gamma distribution, are listed in Table 4.2. Additional results are included in the Appendix. Using these estimated rates, we computed transition probabilities between the ED, floor, and ICU, and lengths of stay in each department; we include these transformed values in Table 4.3. As expected, a large percentage (73%, 95% CI: 56%-90%) of those under the age of 18 years were discharged from the ED with mild symptoms, while a majority (68%, 95% CI: 65%-71%) of those over the age of 65 were admitted to the floor. About 21% (95% CI: 19-24%) of patients over the age of 65 were admitted directly to the ICU from the ED. We estimate similarly that 20% of patients between 19-64 years were admitted directed to the ICU, but the variability in the estimate is higher than that of the older age group (95% CI: 9.5-31%). Of patients who completed their trajectories, 13% of adults between 18-64 years requiring critical care died in the ICU, and 27% of adults over 65 years requiring critical care died in the ICU. Average length of stay (LOS) on the floor was 10 days (95% CI: 9.0, 11) for those over the age of 64, as opposed to 3.3 days (95% CI: 2, 4.6) and 7.6 days (95% CI: 7.1, 8.1) for those between 0-18 years and between 18-64 years respectively. LOS in the ICU was longest on average, 14 days (95% CI: 12, 116)) for adults between 19-64 years, 11 days (95% CI: 9.7, 13) for adults over 65 years, and 8.5 days (95% CI: 1.5-15) for children under the age of 18.

4.2.5 Model fit

Accurate predictions of occupancy were the most important output of the model, as a tool to help hospital administrations with surge planning. We evaluated the ability of the model to generate predictions of hospital occupancy at YNHH that matched the observed occupancy during a surge in COVID-19 patients at Yale-New Haven Hospital (YNHH), the largest of the five hospitals in YNHHS. These predictions were generated using the observed time series of patients admitted to each YNHH department between March 8 and June 12, 2020, and were compared to the observed occupancy in each department during this time. Parameters were

	0-18 years	18-64 years	65+ years
Estimated parameters			
σ_{MS}	1.1 (0, 4.1)	0.081 (0, 1.9)	0.55 (0.43, 0.68)
σ_C	0.017 (0, 0.32)	0.075 (0, 0.67)	1.1 (0.87, 1.2)
σ_F	0.37 (0, 2.3)	0.035 (0, 2.4)	3.4 (3.2, 3.6)
χ_C	0.12 (0.011, 0.27)	0.061 (0.05, 0.072)	0.064 (0.052, 0.076)
χ_F	0.28 (0.17, 0.4)	0.11 (0.11, 0.12)	0.071 (0.068, 0.075)
χ_{WF}	0.28 (0.17, 0.4)	0.11 (0.11, 0.12)	0.071 (0.068, 0.075)
θ_F	0.0095 (0, 0.041)	0.014 (0.011, 0.017)	0.013 (0.011, 0.016)
θ_{WF}	0.0095 (0, 0.041)	0.014 (0.011, 0.017)	0.013 (0.011, 0.016)
ζ	2.7 (1.5, 4)	1.2 (1.1, 1.3)	0.87 (0.79, 0.95)
ξ_{MS}	0.012 (0.012, 0.012)	0.0054 (0.0054, 0.0054)	0.0039 (0.0039, 0.0039)
μ_{MS}	0.00066 (0.00066, 0.00066)	0.00066 (0.00066, 0.00066)	0.00066 (0.00066, 0.00066)
μ_C	0.0035 (0.0023, 0.0048)	0.0083 (0.0038, 0.013)	0.024 (0.018, 0.03)
μ_F	0.002 (0.0018, 0.0022)	0.00033 (0, 0.001)	0.012 (0.0092, 0.015)
μ_{WF}	0.002 (0.0018, 0.0022)	0.00033 (0, 0.001)	0.012 (0.0092, 0.015)
Fixed parameters			
ϕ	0.088*	0.094*	0.095*
μ_{WC}	4	4	4
η	36	36	36

Table 4.2: **Model parameters:** Included in this table are both estimated rates of transition with 95% confidence intervals, rates taken from external sources [2–5] and rates which we set after being unable to determine then either from data or the literature. The estimated rates are based on the assumption that the time to each competing risk follow a two-parameter gamma distribution, where the product of the two parameters yield the estimated mean of the distribution. These parameters were estimated separately for each age group. Rates labeled with (*) were taken from CDC MMWR [2].

	0-18 yrs	19-64 yrs	65+ yrs
Age distribution in ED, %	0.02 (0.02, 0.02)	0.58 (0.58, 0.58)	0.4 (0.4, 0.4)
% discharged from ED	0.73 (0.56, 0.9)	0.54 (0.33, 0.75)	0.11 (0.09, 0.13)
% admitted from ED to floor	0.25 (0.1, 0.4)	0.26 (-0.053, 0.57)	0.68 (0.65, 0.71)
% admitted from ED to ICU	0.037 (-0.02, 0.052)	0.2 (0.095, 0.31)	0.21 (0.19, 0.24)
% death on the floor	0.024 (0.015, 0.033)	0.0052 (-0.00013, 0.011)	0.13 (0.1, 0.15)
% death in the ICU	0.01 (0.0069, 0.013)	0.13 (0.069, 0.18)	0.27 (0.22, 0.32)
% step up from floor to ICU	0.52 (0.22, 0.81)	0.79 (0.73, 0.85)	0.65 (0.59, 0.7)
% step down to the floor	0.042 (-0.034, 0.12)	0.11 (0.089, 0.13)	0.14 (0.12, 0.16)
Triage time in ED (days)	0.81 (-0.17, 1.8)	4 (-1.2, 9.3)	0.2 (0.18, 0.21)
Average LOS on floor (days)	3.3 (2, 4.6)	7.6 (7.1, 8.1)	10 (9.6, 11)
Average LOS in ICU (days)	8.5 (1.5, 15)	14 (12, 16)	11 (9.7, 13)

Table 4.3: **Probabilities of transition between the ED, Floor, and ICU, and lengths of stay:** Included in this table are both estimated probabilities of transition with 95% confidence intervals, and lengths of stay in each hospital department. As with the estimated rates in Table 4.2, the estimated probabilities and lengths of stay are based on the assumption that the time to each competing risk follow a two-parameter gamma distribution, where the product of the two parameters yield the estimated mean of the distribution. These parameters were estimated separately for each age group.

estimated assuming gamma-distributed time to departure from a department. Figure 4.3 show the observed occupancy and occupancy predicted by the model according to this analysis. Appendix figures 4.4 and 4.5) show similar results for the secondary analyses using simpler distributional assumptions.

The model accurately predicts occupancy in YNHH. Figure 4.3B shows that the ICU occupancy predicted by the model closely matches the increase in the observed YNHH ICU occupancy between March and April, the period of peak occupancy between April and May, and the decrease in occupancy between May and June. Similarly, Figure 4.3A demonstrates that the model also accurately predicts the increase in floor occupancy between March and April and the date of peak occupancy, April 18, 2020. The model predicts that floor occupancy on April 18 would be 343 patients (95% CI: 323, 365). The observed YNHH occupancy on April 18 was 405 patients. The model's prediction for deaths among COVID-19 patients on the floor and in the ICU were also reasonable but are not shown for privacy reasons.

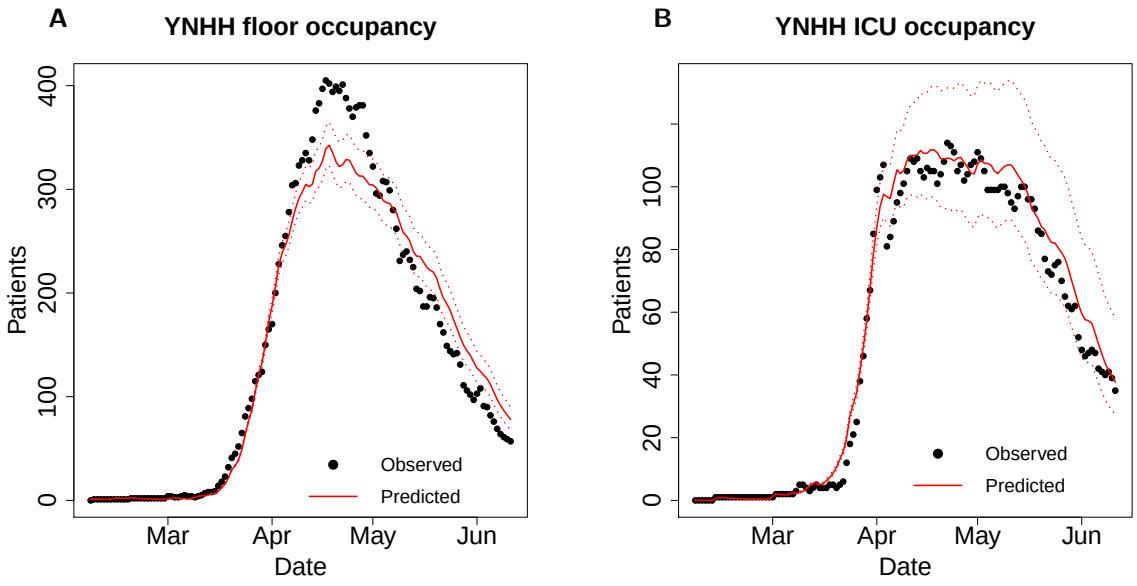

Figure 4.3: Inpatient predicted and observed COVID19 floor occupancy (A) and ICU occupancy (B). Parameters describing rates of transition between hospital departments were estimated assuming gamma-distributed time to event. The dotted line represents occupancy at YNHH. The solid red line represents model output based on parameters calculated using our fitting procedure and capacity estimates from YNHH. The dotted red lines represent estimated occupancy according to the bounds of 95% confidence intervals for each parameter.

4.3 Discussion

In this paper, we have presented a model of hospital capacity which was utilized by YNHHS to inform decision-making regarding resources which would be necessary to handle the surge in COVID-19 patients at YNHHS. As requested by rapid response guidelines [141], the model predicts the number of patients requiring hospital resources under a variety of scenarios. We have described the process that we used to fit the model using data from YNHHS collected as the surge in COVID-19 patients was occurring. We have provided evidence that our estimation procedure provides a reasonable estimate of the true dynamics in a system by comparing the performance the model to the observed dynamics in YNHH during the surge in COVID-19 cases.

The model successfully predicted the most important quantity: occupancy in the ICU, which is the most scarce resource and most important for helping severely ill COVID-19 patients. However, this method may have several weaknesses. The model slightly underpredicted floor occupancy. This could be due to inaccurate assumptions used to estimate the model parameters. The model parameters were estimated for 3 age groups. The model fit could be improved in future versions by creating additional subgroups according to age and other risk factors for hospitalization and severe disease in COVID-19 patients. The model also assumes that rates of transition between departments remains constant over time. However, several factors, including changing hospital protocols for triage and treatment, may have resulted in fluctuations in the rates of transition over time. Such non-stationary behavior is challenging to replicate and would not be captured by the model or parameter estimation procedure. In addition, the model is deterministic, and the estimates of variance in occupancy are based on uncertainty in parameter estimation rather than inherent stochasticity in the model. Improved estimates of variance might be achieved by making the model fully stochastic.

We expect hospital dynamics for COVID-19 patients to continue to evolve in the future with

wide-spread vaccination and the rise of new variants. Rates of transition between departments, lengths of stay on the floor and in the ICU, and rates of death are likely to be substantially different in the current era of widespread vaccination relative to the early days of the pandemic due to decreased risk of severe illness in vaccinated individuals and well-established treatment protocols developed since the beginning of the pandemic. Thus, we recommend continual re-estimation of the model parameters using patient trajectories observed in the setting of wide-spread vaccination. Hospital dynamics with the development of a new variant would likely revert suddenly to dynamics similar to those observed in the beginning of the pandemic. Model parameters estimated using the dataset described here could be used to predict model capacity until enough new data are observed to re-estimate model parameters specific to the new variant.

The urgency of the crisis caused by surging COVID-19 patients contributed substantially to the challenge of developing a useful model. We would like to conclude to providing a few recommendations for creating a model in a crisis. First, early collaboration with end-users of the product was essential. After creation of the model structure and early implementation of model as an interactive web application, we met several times with administrators at YNHHS who were in charge of capacity planning. They provided feedback on the model and the web application, in addition to crucial perspective on the most urgent unmet needs which could be addressed by the model. Second, we recommend reducing the dependence of these models on unverifiable assumptions. Instead of constructing a population-level model which would predict hospitalizations without explicitly modeling dynamics within a hospital, we chose to use observed ED visits at YNHH to project best and worst case scenarios and individual-level patient trajectories to capture observed patient dynamics in the model. Therefore, almost all of the parameters used to fit this model are based on observable data from electronic health records. Despite using observable data to construct the model, we still had to carefully consider the implications of the incomplete nature of the dataset. The model was sensitive to differences in parameters that were estimated using different procedures. The fidelity in the

predictions of ICU occupancy was only achieved after we took into account right-censoring of patient trajectories, with sicker patients remaining in the hospital at the time of the analysis. Third, interactive implementations of any model results should be streamlined to involve the smallest possible number of parameters for ease of use, and the rest should be reasonable defaults. Despite the large number of parameters necessary to use the model, we included only a limited number for users of the web application to manipulate.

Acknowledgments: This work was supported by NIH grant NICHD DP2 HD091799-01. We would like to thank Amento Annette, Thomas Balcezak, Keisha Boykin, Matthew Comerford, Philip Corso, Rick D'Aquila, Gary Desir, Edward Kaplan, Brian Keane, Thomas Prem, Rema Seeram, and Paul Taheri at the Yale New Haven Hospital system for providing access to data. We would also like to acknowledge the COVID-19 Statistics, Policy Modelling and Epidemiology Collective (C-SPEC), which included the following individuals: David Paltiel, Gregg Gonsalves, Jeffrey Eaton, Josh Salomon, Meagan Fitzpatrick, Nick Menzies, Reza Yaesoubi, Adam Beckman, Alyssa Bilinski, Anna York, Anne Williamson, Bianca Mulaney, Christian Testa, Emma Clarke-Deelder, EvanMacKay, Hanna Ehrlich, Jinyi Zhu, John Giardina, Katie Rich, Kayoko Shioda, Lin Zhu, Elizabeth White, Luke Massa, Maile Phillips, Melanie Chitwood, Monica Farid, Nicole Swartwood, Nikhil Deshmukh, Raphael Sherak, Ruthie Birger, Samantha Burn, Stephanie Perniciaro, Suzan Iloglu, Thomas Thornhill, Tyler Copple, Yu-Han Kao, and Yuli-Lily Hsieh. The web application was written by Soheil Eshghi, Margret Erlendsdottir, Maile Thayer Phillips, Suzan Iloglu, Christian Testa and Forrest W. Crawford using the R shiny framework. We are especially grateful to Gregg Gonsalves, David Paltiel, Hanna Ehrlich, Raphael Sherak, Melanie Chitwood, Thomas Thornhill, Nicole Swartwood, and Stephanie Perniciaro for advice and comments.

Conflicts of interest: We have no conflicts of interest to report.

Data statement: All data are protected health information and are not available for sharing. Code is available on request.

Appendix

4.3.1 Parameter estimation

To estimate model rate parameters, we estimated the exit rate of patients from each compartment of the model, assuming exponentially-distributed or gamma-distributed length of stay (LOS) in each compartment. We estimated exit rates for each competing hazard.

We estimated the rates of each of these competing hazards by constructing a likelihood function in the following way. For each competing hazard k, $k = 1, .., K$, the time until event k is T_k, and is exponentially distributed with rate λ_k. We assume independence between all T_k. We observe $T = \min_k T_k$, an indicator δ which takes value k if the subject experienced event k, and a censoring indicator C which takes value 1 if a subject remains in the hospital at the end of the study period and is 0 otherwise. We have N total subjects in our dataset, and for each individual i, $i = 1, ..., N$, the observed data $O_i = (t_i, \delta_i, C_i)$.

We constructed a likelihood function to estimate the rate parameters λ_k for each of $T_k \sim \exp(\lambda_k)$ or shape and scale parameters α_k and β_k if $T_k \sim$ Gamma(α_k, β_k). Each set of distributional assumptions utilizes a different set of parameters, so we denote generally the vector of parameters used in a particular likelihood with Θ and the parameters describing the distribution of T_k with θ_k. We performed three analyses using different distributional assumptions for the density $f_k(t; \theta_k)$. The first analysis assumed that all T_k were exponentially distributed. The second analysis assumed that all T_k were exponentially distributed, except for those corresponding to discharge, which were gamma distributed. The third analysis assumed two-parameter gamma distributions for all T_k. As is typical in survival analysis, the density of time to event T_k is $f_k(t; \theta_k)$, the survival function for event type k is $S_k(t; \theta_k) = (1 - F_k(t; \theta_k))$, the hazard of event k is $h_k(t; \theta_k) = f_k(t; \theta_k)/S_k(t; \theta_k)$, the overall survival function given inde-

pendence between events is $S(t;\Theta) = \prod_{k=1}^{K}(1 - F_k(t;\theta_k))$. The contribution of an uncensored individual i with an observed outcome is:

$$f_k(t_i;\theta_k) = h_k(t_i;\theta_k)\prod_{k=1}^{K} S_k(t_i;\theta_k) = f_k(t_i;\theta_k)\prod_{j\neq k}(1 - F_j(t_i;\theta_k))$$

The contribution of a censored individual is simply $\prod_{k=1}^{K}(1 - F_j(t_i;\theta_k))$.

Thus, the likelihood function for the observed data is:

$$L(t_1, ..., t_N, \Theta) = \prod_{i=1}^{N}\Big[\prod_{k=1}^{K} f_k(t_i;\theta_k)\prod_{j\neq k}(1 - F_j(t_i;\theta_k))\Big]^{\mathbb{I}(\delta_i=k)(1-C_i)}\prod_{k=1}^{K}(1 - F_j(t_i;\theta_k))^{C_i} \quad (4.1)$$

For all instances in which a gamma distribution was assumed, we computed the estimated mean of each T_k using the maximum likelihood parameter estimates. These estimates, in addition to estimates obtained assuming exponential distributions, were together used to estimate the probability of transition out of each department and length of stay in the ICU and on the floor in the following away. We assumed all T_k to be exponentially distributed random variables with rate λ_k, including those parameters estimated assuming gamma distribution in the log likelihood. If X is the random variable which denotes the event which occurs, the probability of transition to a particular event k is $\mathbb{P}(X = k) = \frac{\lambda_k}{\sum_k \lambda_k}$. The length of stay within a department is also exponentially distributed, with $\mathbb{E}[T]$ = $\mathbb{E}[\min_k T_k] = \frac{1}{\sum_k \lambda_k}$. For the purposes of parameter estimation, discharge from the ICU and step down from the ICU were considered to be separate competing hazards. The estimated probability of step down from the ICU used to parameterize the model is the sum of the probabilities of step down and discharge. Due to the relatively small number of people discharged directly from the ICU, we considered this to be a reasonable modification.

Figures and tables for alternative distributional assumptions in parameter estimation procedure.

1. Exponential model:

	0-18 years	18-64 years	65+ years
ϕ	0.088 (0.088, 0.088)	0.094 (0.094, 0.094)	0.095 (0.095, 0.095)
σ_{MS}	1.4 (0, 5)	2.4 (2.2, 2.6)	0.43 (0.36, 0.49)
σ_C	0.072 (0, 0.33)	0.4 (0.34, 0.47)	0.48 (0.42, 0.55)
σ_F	0.48 (0, 1.8)	2.1 (2, 2.3)	3.1 (2.9, 3.2)
χ_C	1.6 (1.6, 1.6)	0.071 (0.06, 0.081)	0.072 (0.061, 0.082)
χ_F	1.6 (1.6, 1.6)	0.11 (0.1, 0.12)	0.065 (0.062, 0.068)
χ_{WF}	1.6 (1.6, 1.6)	0.11 (0.1, 0.12)	0.065 (0.062, 0.068)
θ_F	7 (7, 7)	0.017 (0.014, 0.02)	0.015 (0.014, 0.017)
θ_{WF}	7 (7, 7)	0.017 (0.014, 0.02)	0.015 (0.014, 0.017)
η	36	36	36
ζ	78 (78, 78)	1.2 (1.1, 1.3)	0.86 (0.8, 0.92)
ξ_{MS}	0.012 (0.012, 0.012)	0.0054 (0.0054, 0.0054)	0.0039 (0.0039, 0.0039)
μ_{MS}	0.00066 (0.00066, 0.00066)	0.00066 (0.00066, 0.00066)	0.00066 (0.00066, 0.00066)
μ_C	2.7e-07 (0, 1.1e-06)	0.0083 (0.0061, 0.011)	0.03 (0.025, 0.035)
μ_F	5.1e-07 (4.7e-07, 5.6e-07)	0.0024 (0.0015, 0.0033)	0.016 (0.014, 0.018)
μ_{WC}	4	4	4
μ_{WF}	5.1e-07 (4.7e-07, 5.6e-07)	0.0024 (0.0015, 0.0033)	0.016 (0.014, 0.018)

Table 4.4: Parameters assuming exponential distributions for each competing risk.

	0-18 yrs	19-64 yrs	65+ yrs
Age distribution in ED	0.02 (0.02, 0.02)	0.58 (0.58, 0.58)	0.4 (0.4, 0.4)
% discharged from ED	0.71 (0.62, 0.81)	0.49 (0.47, 0.51)	0.11 (0.092, 0.12)
% admitted from ED to floor	0.25 (0.16, 0.34)	0.43 (0.41, 0.45)	0.77 (0.75, 0.79)
% admitted from ED to ICU	0.037 (-0.0032, 0.078)	0.081 (0.07, 0.092)	0.12 (0.11, 0.14)
% death on the floor	9.6e-07 (8.6e-07, 1.1e-06)	0.019 (0.012, 0.025)	0.16 (0.15, 0.18)
% death in the ICU	1.5e-07 (4.6e-08, 2.5e-07)	0.11 (0.079, 0.13)	0.29 (0.25, 0.33)
% step up from floor to ICU	0.46 (0.46, 0.46)	0.72 (0.68, 0.76)	0.59 (0.54, 0.63)
% step down to the floor	0.014 (0.014, 0.014)	0.13 (0.11, 0.15)	0.16 (0.14, 0.18)
Triage time in ED (days)	0.51 (-0.19, 1.2)	0.2 (0.19, 0.22)	0.25 (0.24, 0.26)
Average LOS on floor (days)	2.2 (2.2, 2.2)	7.5 (7.1, 8)	10 (9.9, 11)
Average LOS in ICU (days)	0.13 (0.13, 0.13)	12 (11, 14)	9.7 (8.7, 11)

Table 4.5: Probabilities of transition and lengths of stay assuming exponential distributions on time to each competing risk.

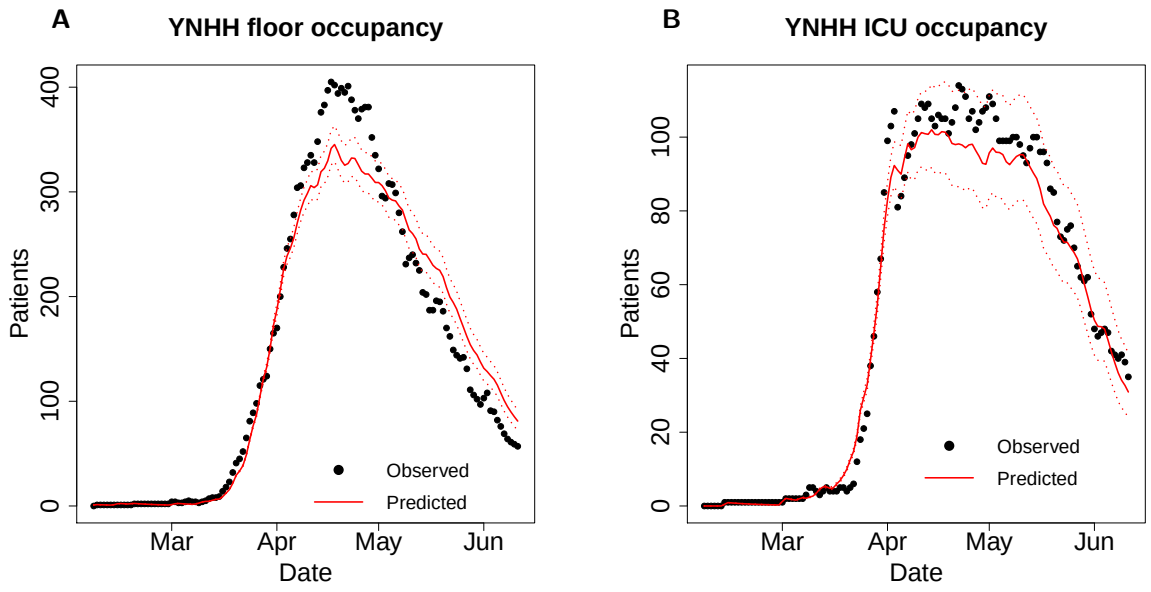

Figure 4.4: Inpatient predicted and observed COVID19 floor occupancy (A) and ICU occupancy (B). Parameters describing rates of transition between hospital departments were estimated assuming exponentially distributed time to event. The dotted line represents occupancy at YNHH. The solid red line represents occupancy predicted by the model based on parameters calculated using our fitting procedure and capacity estimates from YNHH. The dotted red lines represent estimated occupancy according to the bounds of 95% confidence intervals for each parameter.

2. Gamma distribution for time to discharge from ED, floor, or ICU; exponential otherwise:

	0-18 years	18-64 years	65+ years
ϕ	0.088 (0.088, 0.088)	0.094 (0.094, 0.094)	0.095 (0.095, 0.095)
σ_{MS}	1.1 (0, 5)	0.37 (0, 2.3)	0.55 (0.42, 0.68)
σ_C	0.072 (0, 0.32)	4.1 (1.7, 6.6)	0.48 (0.42, 0.55)
σ_F	0.48 (0, 1.7)	4 (1.4, 6.6)	3.1 (2.9, 3.2)
χ_C	0.13 (0.13, 0.13)	0.063 (0.055, 0.071)	0.066 (0.057, 0.075)
χ_F	0.027 (0, 0.098)	0.11 (0.11, 0.12)	0.071 (0.068, 0.075)
χ_{WF}	0.027 (0, 0.098)	0.11 (0.11, 0.12)	0.071 (0.068, 0.075)
θ_F	0.52 (0.52, 0.53)	0.017 (0.014, 0.02)	0.015 (0.014, 0.017)
θ_{WF}	0.52 (0.52, 0.53)	0.017 (0.014, 0.02)	0.015 (0.014, 0.017)
η	36	4 36	36
ζ	5 (4.7, 5.6)	1.2 (1.1, 1.3)	0.92 (0.86, 0.99)
ξ_{MS}	0.012 (0.012, 0.012)	0.0054 (0.0054, 0.0054)	0.0039 (0.0039, 0.0039)
μ_{MS}	0.00066 (0.00066, 0.00066)	0.00066 (0.00066, 0.00066)	0.00066 (0.00066, 0.00066)
μ_C	6.3e-07 (6.1e-07, 6.6e-07)	0.0083 (0.0061, 0.011)	0.03 (0.025, 0.035)
μ_F	7.6e-08 (0, 3.4e-06)	0.0024 (0.0015, 0.0033)	0.016 (0.014, 0.018)
μ_{WC}	4	4	4
μ_{WF}	7.6e-08 (0, 3.4e-06)	0.0024 (0.0015, 0.0033)	0.016 (0.014, 0.018)

Table 4.6: Parameters assuming exponential distributions for each competing risk, except for gamma distributed time to discharge.

4.3.2 Design and construction of the R shiny web application

A version of the model described in this report was implemented in the R shiny web application, deployed at `https://forrestcrawford.shinyapps.io/covid19_icu/`. Figure 4.6 shows the default page of the web application. This interactive web application was intended to be a tool for scenario analysis, used by hospital administrators and departments involved in capacity planning during the COVID-19 pandemic. This tool takes as an input an infection scenario which can be specified by the user, the capacity and current occupancy of the hospital system, and parameters describing basic features of the patient population served by the hospital. The tool allows the user to specify a strategy for capacity expansion, observing the effect of adding beds to the system. The outputs of the tool include projections of expected

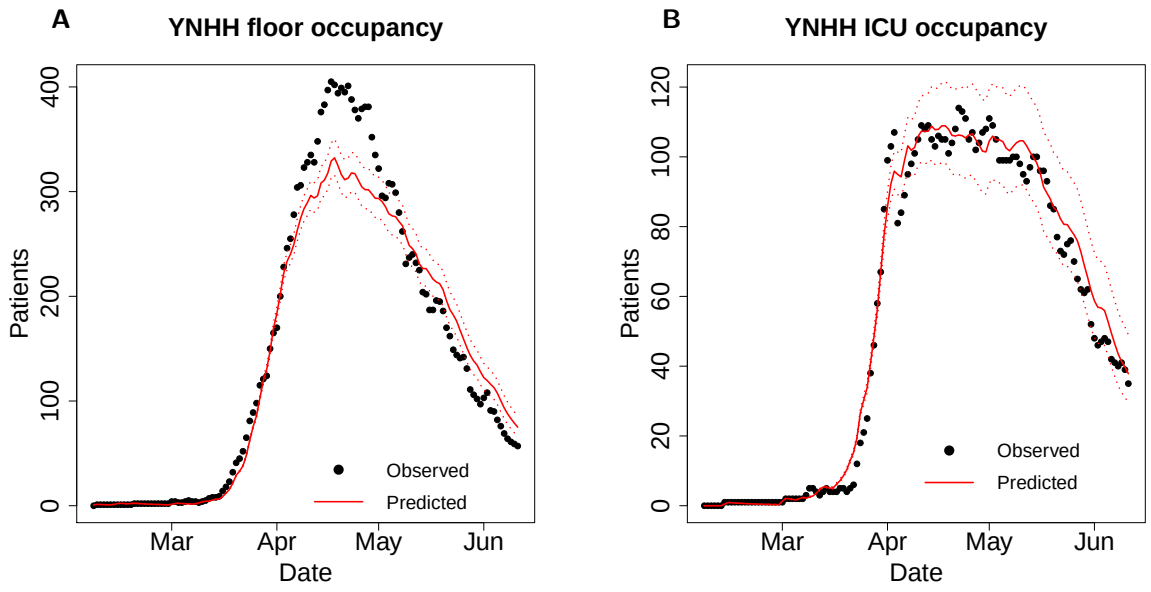

Figure 4.5: Inpatient predicted and observed COVID19 floor occupancy (A) and ICU occupancy (B). Parameters describing rates of transition between hospital departments were estimated assuming exponentially-distributed time to event, except for time to discharge in the ICU and on the floor, which were assumed to be gamma-distributed. The dotted line represents either occupancy or cumulative observed at YNHH. The solid red line shows occupancy predicted by the model based on parameters calculated using our fitting procedure and capacity estimates from YNHH. The dotted red lines represent estimated occupancy according to the bounds of 95% confidence intervals for each parameter.

	0-18 yrs	19-64 yrs	65+ yrs
Age distribution in ED	0.02 (0.02, 0.02)	0.58 (0.58, 0.58)	0.4 (0.4, 0.4)
% discharged from ED	0.69 (0.56, 0.81)	0.17 (-0.26, 0.59)	0.13 (0.11, 0.16)
% admitted from ED to floor	0.27 (0.16, 0.39)	0.56 (0.41, 0.71)	0.75 (0.72, 0.78)
% admitted from ED to ICU	0.037 (-0.0043, 0.086)	0.27 (-0.00055, 0.54)	0.12 (0.1, 0.13)
% death on the floor	9.4e-07 (-2.5e-07, 2.1e-06)	0.018 (0.012, 0.024)	0.15 (0.14, 0.17)
% death in the ICU	5.4e-07 (-1.2e-07, 1.2e-06)	0.12 (0.089, 0.15)	0.31 (0.27, 0.35)
% step up from floor to ICU	0.28 (0.28, 0.28)	0.8 (0.77, 0.83)	0.62 (0.58, 0.66)
% step down to the floor	0.31 (0.3, 0.31)	0.13 (0.11, 0.14)	0.15 (0.13, 0.17)
Triage time in ED (days)	0.59 (-0.22, 1.4)	0.16 (0.074, 0.24)	0.24 (0.23, 0.26)
Average LOS on floor (days)	0.35 (0.34, 0.36)	7.3 (6.9, 7.8)	9.6 (9.2, 10)
Average LOS in ICU (days)	23 (23, 23)	14 (12, 15)	10 (9.2, 11)

Table 4.7: Probabilities of transition and lengths of stay assuming exponential distributions on time to each competing risk, except for gamma-distributed time to discharge.

occupancy, deaths, an estimate of the time at which the system would reach capacity, and an estimate of the number of extra beds which would be necessary to relieve the overflow.

Because exact time series of COVID-19 presentations were not always easily accessible, we created an opening dashboard on the “Scenario" tab which allows the user to specify an infection scenario (Fig. 4.7). On this dashboard, the user can specify a time horizon for projections, the shape of the infection curve, and parameters which control the shape of this curve. The infection curve represents the number of new COVID-19 presentations to the health system each day, rather than the number of new COVID-19 infections in the population. Thus, the inputs to this model could be based on model fits to observed COVID-19 ED presentations at a particular hospital. This feature allows users to test various scenarios without having access to exact time series of ED presentations. Furthermore, these scenarios can be studied without knowledge of highly uncertain dynamics of new infections in the population.

ED presentation dynamics

The functions for ED presentations and capacity are intentially left unspecified and can be set by the user. The user can choose among the following options:

Figure 4.6: **Home page**: The default page of the web application, available at `https://forrestcrawford.shinyapps.io/covid19_icu/`

- **Exponential increase:** The number of ED presentations rises exponentially over the time-period in question from its initial value, with an exponent of $\frac{\ln(2)}{T_{\text{doubling}}}$, with T_{doubling} being the user-specified doubling time. This is especially relevant early-on in the epidemic.

- **Linear increase:** The number of ED presentations rises linearly over the time-period in question from its initial to its final value. This is relevant when the epidemic has not peaked but is being kept somewhat in check by various interventions.

- **Saturating:** In this case, the number of ED presentations plateaus to its final value at the end of the time-horizon, rising as a logistic function centered in the middle of the time period under study. This is relevant when the epidemic is close to peaking.

- **Flat:** The number of ED presentations remains equal to day zero across the time-frame.

The user can also model the effect of exponential, linear, and saturating decrease in ED pre-

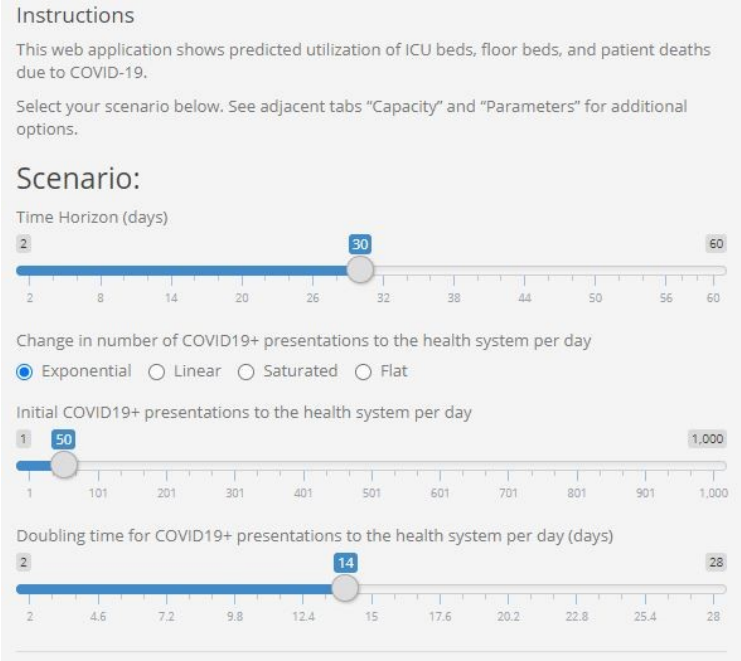

Figure 4.7: **Scenario tab:** allows the user to select a timeseries of daily COVID-19 presentations to a health system.

sentations using the tool, for example due to the implementation of non-pharmaceutical interventions.

On the "Capacity" and "Strategy" tabs, the user can specify the capacity of their healthcare system and a strategy for capacity expansion. Defaults in the web application are based on YNHH capacity. The "Capacity" tab includes inputs for the number of floor and ICU beds available at baseline, as well as the percentage occupied at time zero for the intended simulation (Fig. 4.8). The "Strategy" tab can be toggled on and off depending on whether a strategy for further surge is planned (Fig. 4.9). When on, the user can specify a target number of beds for the floor and ICU, as well as the time frame within which the capacity expansion is expected to occur.

The "Parameters" tab allows the user to adjust several parameters which we determined to be easily estimated and important for tailoring model dynamics to specific patient populations. For example, overall admission and death rates are significantly different between age groups,

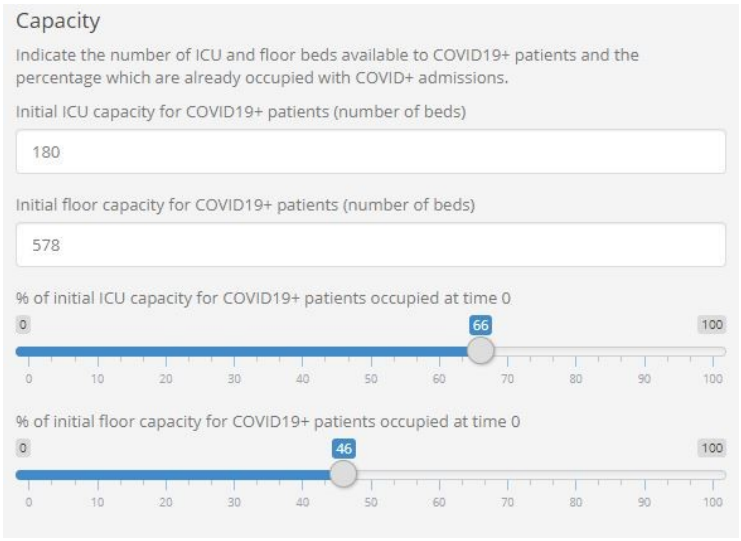

Figure 4.8: **Capacity tab:** allows the user to specify the number of beds available to COVID-19 patients, and the number of beds occupied at time zero.

Figure 4.9: **Strategy tab:** allows the user to specify whether or not the health system plans to surge, when the surge would occur, and the number of beds planned.

and we have included a two-part slider which allows the user to control the age distribution of the population in the simulation. Furthermore, average LOS on the floor and in the ICU is tracked by hospitals and influences both occupancy and the time at which the hospital would reach capacity.

The model outputs are presented on the “Plots" and “Summary” tabs. We included “Key point” which include the most important outputs to hospital administrators, chosen based on feedback from YNHHS administrators. Of especial interest were the time to full capacity and the number of additional beds, if any, which would be necessary to accommodate all COVID-19 patients. On the “Plots” tab (Fig. 4.10), Plots A and B show daily and cumulative COVID-19 presentations to the health system. Plots B and C show cumulative deaths predicted by the model, as well as a breakdown of these deaths by department. Plot D shows the occupancy in the ICU and the floor predicted by the model, as well as flags which denote the time at which capacity is reached, if it is exceeded during the simulation. Exact values for anticipated deaths, days to ICU and floor overflow, extra beds needed, and predicted case-fatality rates are summarized on the “Summary” tab.

To aid new users in the use of the model, we have included three tabs - “Input”, “Outputs”, and “About” - which qualitatively describe the model inputs, outputs, and structure.

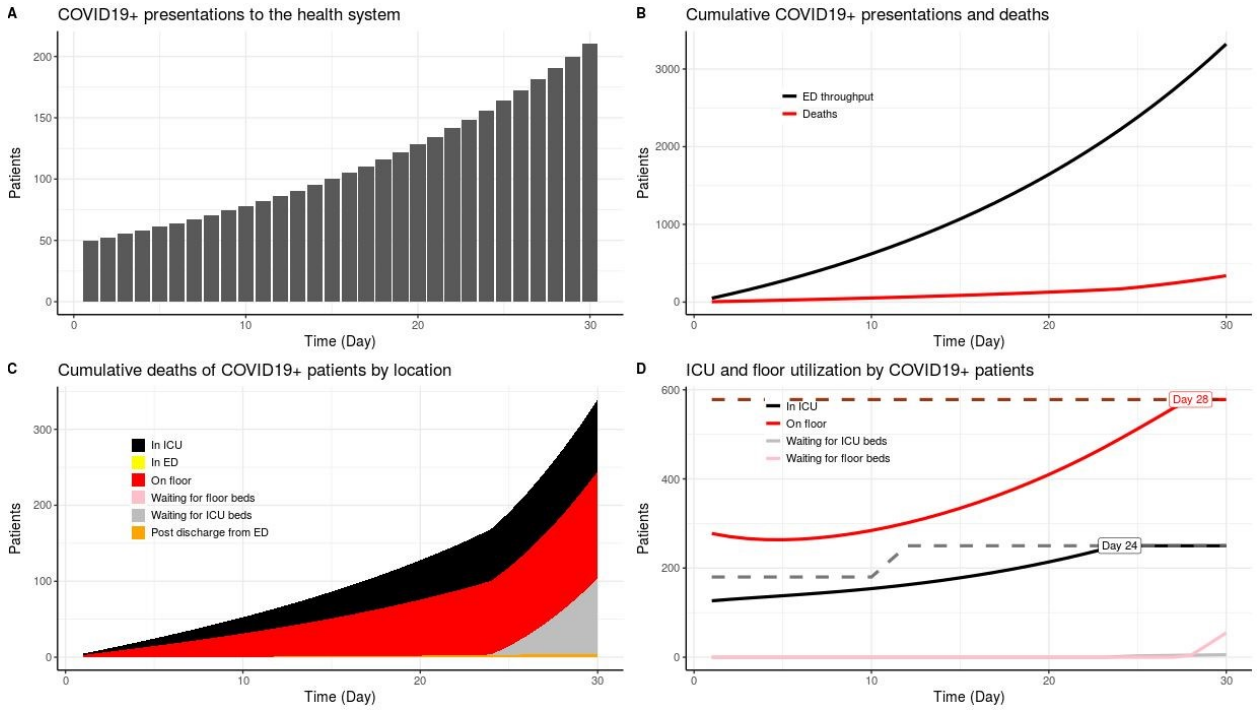

Figure 4.10: **Plots tab:** These plots are the main output of the model. Plot A shows the time series of presentations to the ED, Plot B shows cumulative ED presentations and projected cumulative deaths, and Plot C shows cumulative deaths by location. Plot D demonstrates projected occupancy on the floor and in the ICU, as well as a flag which appears with the time at which capacity of the healthcare system would be overwhelmed given the scenario and capacity settings entered by the user.

Chapter 5

Conclusion

In the dissertation, we have presented three cases in which the causal inference and electronic health record data have allowed us to reexamine the way that we evaluate the effects of medical interventions and the system by which we deliver medical care. We used principles of causal inference to demonstrate that trials of biomarker targets may not be the optimal study design by which to establish clinical guidelines for hypertension. We attempted to remedy this problem by employing electronic health records and methods from causal inference simultaneously to assess time-varying strategies for managing hypertension. We also used electronic health record data to address challenges in resource allocation during the COVID-19 pandemic, demonstrating that such a data-driven strategy could be used to successfully meet the needs of hospital administrators in crisis.

Two main themes emerged during the completion of this work. First, we demonstrate the importance of writing down a clear and unambiguous causal estimand when the goal is to draw causal conclusions about an intervention. Second, we demonstrate the importance of addressing computational limitations in order to successfully conduct analyses and build models using individual-level electronic health record data.

In Chapters 2 and 3, we demonstrate that rigorous assessment of treatments for hypertension requires a clear definition of a target causal estimand. In medical practice, treating to a target may be useful because the target is associated with improvements in some other outcome. There are some cases in which this is obviously true. For example, management of HIV involves administrating anti-retroviral therapy until viral load falls below a threshold, such that viral replication and chances of transmission become vanishingly small. Therefore, regardless of the other effects of the anti-retroviral therapies themselves, reaching this target is beneficial. However, treating to a target in management of chronic diseases like hypertension, in which the target has a complex “J-shaped" relationship with outcome and the medications themselves may have some adverse effects, may not be the best strategy. Use of a directed acyclic graph allowed us to clearly delineate the difference between the randomized target and the multitude of treatments used to achieve the target, as well as to describe the limitations of randomization in this setting. Furthermore, our work on assessing the effects of time-varying treatments required that we define a counterfactual difference in survival probability and an adjustment strategy which would allow us to estimate that estimand. We could not simply report a hazard ratio, even after appropriate adjustment, due to ambiguity in the interpretation of that ratio. In both these projects, rigor in causal inference was necessary to ensure that studies and analyses generate estimates of the true causal quantities of interest.

In Chapters 3 and 4, we encountered substantial challenges which demonstrate the importance of both obtaining sufficient computational resources to carry out analyses using electronic health records and innovating to reduce the computational burden of these analyses. While statistical innovations will continue to be essential, the obvious limitation on our analysis in Chapter 3 was computational power, rather than the absence of available statistical methodology or data. Existing methods in regression analysis, such as multinomial regression modeling, and machine learning, such as boosted classification and regression trees, could have been used to build more complex models for estimation of stabilized weights and enable additional treatment comparisons. However, we were unable to implement these methods due

to the computational demands of imputation and bootstrapping for variance estimation. Over time, as patient follow-up in electronic health records increases and methods for extracting additional information from written physician records improves, the available data will only grow in size. In Chapter 4, we encountered similar challenges. Estimation of the model parameters and their variances also required that we use bootstrapping. Thus, continuous updating of the parameters based on new electronic health records data, essential to ensuring that the model reflects current hospital dynamics, was limited by the speed with which we could perform this estimation. Furthermore, the current implementation of the model described in this work includes variance estimates based on uncertainty in the parameter estimates. An assessment of the variance of our capacity predictions taking into account stochasticity in the transitions between model compartments is computationally intensive. This analysis would have required computational resources which were not available to us in the early days of the COVID-19 pandemic, when rapid generation of predictions of hospital capacity was required.

Bibliography

[1] David M. Reboussin, Norrina B. Allen, Michael E. Griswold, Eliseo Guallar, Yuling Hong, Daniel T. Lackland, Edgar (Pete) R. Miller, Tamar Polonsky, Angela M. Thompson-Paul, and Suma Vupputuri. Systematic review for the 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension*, 71(6):e116–e135, 2018. x, xviii, 5, 21, 22, 23, 24, 25, 27, 28, 30, 39, 40, 47

[2] Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12 - March 28, 2020. *Centers for Disease Control and Prevention, MMWR Morbidity Mortality Weekly Report*, April 2020. xxi, 94, 95, 100, 101, 105

[3] Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12 - March 16, 2020. *Centers for Disease Control and Prevention, MMWR Morbidity Mortality Weekly Report*, March 2020.

[4] Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *Centers for Disease Control and Prevention, MMWR Morbidity Mortality Weekly Report*, April 2020.

[5] Characteristics and Clinical Outcomes of Adult Patients Hospitalized with COVID-19 - Georgia, March 2020. *Centers for Disease Control and Prevention, MMWR Morbidity Mortality Weekly Report*, May 2020. xxi, 94, 100, 101, 105

[6] Kari Lock Morgan and Donald B. Rubin. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2):1263–1282, 04 2012. 5

[7] Miguel A Hernán and Sonia Hernández-Díaz. Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials*, 9(1):48–55, 2012. 5, 6, 8, 9, 12

[8] Masayuki Ikeda and Rumiko Shimazawa. Challenges to hemoglobin A1C as a therapeutic target for type 2 diabetes mellitus. *Journal of General and Family Medicine*, 20 (4):129–138, 2019. 5, 6, 20

[9] Jeffrey L Carson, Frederick Sieber, Donald Richard Cook, Donalrd Hoover, Helaine Noveck, Bernard Chaitman, Lee Fleisher, Lauren Beaupre, william Macaulay, George G Rhoads, Barbara Paris, Aleksandra Zagorin, David W Sanders, Khwaja J Zakriya, and

Jay Magaziner. Liberal versus restrictive blood transfusion strategy: 3-year survival and cause of death results from the FOCUS randomised controlled trial. *The Lancet*, 385 (9974):1183 – 1189, 2015. 5, 6

[10] Paul K. Whelton, Robert M. Carey, Wilbert S. Aronow, Donald E. Casey, Karen J. Collins, Cheryl Dennison Himmelfarb, Sondra M. DePalma, Samuel Gidding, Kenneth A. Jamerson, Daniel W. Jones, Eric J. MacLaughlin, Paul Muntner, Bruce Ovbiagele, Sidney C. Smith, Crystal C. Spencer anWd Randall S. Stafford, Sandra J. Taler, Randal J. Thomas, Kim A. Williams, Jeff D. Williamson, and Jackson T. Wright. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension*, 71(6):e13–e115, 2018. 5, 6, 12, 21, 34, 47, 48

[11] A Qaseem, TJ Wilt, R Rich, LL Humphrey, J Frost, MA Forciea, and the Clinical Guidelines Committee of the American College of Physicians and the Commission on Health of the Public and Science of the American Academy of Family Physicians. Pharmacologic treatment of hypertension in adults aged 60 years or older to higher versus lower blood pressure targets: A Clinical Practice Guideline From the American College of Physicians and the American Academy of Family Physicians. *Annals of Internal Medicine*, 166:430–437, 2018. 5, 6, 47

[12] Dena Ettehad, Connor A Emdin, Amit Kiran, Simon G Anderson, Thomas Callender, Jonathan Emberson, John Chalmers, Anthony Rodgers, and Kazem Rahimi. Blood pressure lowering for prevention of cardiovascular disease and death: A systematic review and meta-analysis. *The Lancet*, 387(10022):957 – 967, 2016. 6

[13] Robert W. Schrier, Raymond O. Estacio, Anne Esler, and Philip Mehler. Effects of aggressive blood pressure control in normotensive type 2 diabetic patients on albuminuria, retinopathy and strokes. *Kidney International*, 61(3):1086 – 1097, 2002. 6, 24, 26, 27

[14] ACCORD. Effects of intensive blood-pressure control in type 2 diabetes mellitus. *New England Journal of Medicine*, 362(17):1575–1585, 2010. 12, 13, 22, 23, 24, 26, 27, 29, 40, 47

[15] Robert W. Schrier, Kaleab Z. Abebe, Ronald D. Perrone, Vicente E. Torres, William E. Braun, Theodore I. Steinman, Franz T. Winklhofer, Godela Brosnahan, Peter G. Czarnecki, Marie C. Hogan, Dana C. Miskulin, Frederic F. Rahbari-Oskoui, Jared J. Grantham, Peter C. Harris, Michael F. Flessner, Kyongtae T. Bae, Charity G. Moore, and Arlene B. Chapman. Blood pressure in early autosomal dominant polycystic kidney disease. *New England Journal of Medicine*, 371(24):2255–2266, 2014. 22, 24, 26

[16] Kei Asayama, Takayoshi Ohkubo, Hirohito Metoki, Taku Obara, Ryusuke Inoue, Masahiro Kikuya, Lutgarde Thijs, Jan A Staessen, and Yutaka Imai. Cardiovascular outcomes in the first trial of anti-hypertensive therapy guided by self-measured home blood pressure. *Hypertension Research*, 35:1102–1110, 2012. 22, 23, 24, 26, 27

[17] SPRINT. A randomized trial of intensive versus standard blood-pressure control. *New England Journal of Medicine*, 373(22):2103–2116, 2015. 8, 12, 13, 15, 22, 24, 25, 26, 27, 29, 40, 47, 50, 52

[18] UKPDS. Tight blood pressure control and risk of macrovascular and microvascular com-

plications in type 2 diabetes: UKPDS 38. *BMJ*, 317(7160):703–713, 1998. 6, 24, 26, 27, 40

[19] S. Seidu, F. A. Achana, L. J. Gray, M. J. Davies, and K. Khunti. Effects of glucose-lowering and multifactorial interventions on cardiovascular and mortality outcomes: A meta-analysis of randomized control trials. *Diabetic Medicine*, 33(3):280–289, 2016. 6

[20] P. REICHARD, B. BERGLUND, A. BRITZ, I. CARS, B. Y. NILSSON, and U. ROSENQVIST. Intensified conventional insulin treatment retards the microvascular complications of insulin-dependent diabetes mellitus (IDDM): the Stockholm Diabetes Intervention Study (SDIS) after 5 years. *Journal of Internal Medicine*, 230(2):101–108, 1991. 6

[21] The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England Journal of Medicine*, 329(14):977–986, 1993.

[22] Yasuo Ohkubo, Hikedi Kishikawa, Eiichi Araki, Takao Miyata, Satoshi Isami, Sadatoshi Motoyoshi, Yujiro Kojima, Naohiko furuyoshi, and Motoaki Shichiri. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. *Diabetes Research and Clinical Practice*, pages 103–117, 04 1995. 6

[23] Sanjay Kaul. Are guidelines for treatment of hypertension trustworthy? *Journal of the American College of Cardiology*, 73(23):3027–3030, 2019. 6, 10, 21, 47

[24] Fabio Kaul, Gianpaolo Reboldi, and Paolo Verdecchia. Hypertension and the J-curve phenomenon: Implications for tight blood pressure control. *Hypertension Research*, 36: 109–111, 2013. 6

[25] Els Goetghebeur and Tom Loeys. Beyond intention to treat. *Epidemiologic Reviews*, 24 (1):85–90, 2002. 6, 9

[26] Ian Shrier, Russell J Steele, Evert Verhagen, Rob Herbert, Corinne A Riddell, and Jay S Kaufman. Beyond intention to treat: What is the right question? *Clinical Trials*, 11(1): 28–37, 2014. 6, 9

[27] Lewis B Sheiner and Donald B Rubin. Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics*, 57(1):6–15, 1995. 7, 9

[28] Sandeep K Gupta. Intention-to-treat concept: A review. *Perspectives in Clinical Research*, 2(3):109, 2011. 7, 10

[29] Robert T Wertz. Intention to treat: Once randomized, always analyzed. *Clinical Aphasiology*, 23:57–64, 1995. 8

[30] Stephane R Heritier, Val J Gebski, and Anthony C Keech. Inclusion of patients in clinical trial analysis: The intention-to-treat principle. *Medical Journal of Australia*, 179(8):438–440, 2003.

[31] Victor M Montori and Gordon H Guyatt. Intention-to-treat principle. *Cmaj*, 165(10): 1339–1341, 2001. 8

[32] William F Rosenberger and John M Lachin. *Randomization in Clinical Trials: Theory and Practice*. John Wiley & Sons, 2015. 8

[33] Jack L. Lee, Jonas H. Ellenberg, Deborah G. Hirtz, and Karin B. Nelson. Analysis of clinical trials by treatment actually received: Is it really an option? *Statistics in Medicine*, 10:1595–1605, 10 1991. 8

[34] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009. 9

[35] Tyler J. VanderWeele and James M. Robins. Signed directed acyclic graphs for causal inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):111–127, 2010. 9

[36] Joyce A. Cramer, Richard D. Scheyer, and Richard H. Mattson. Compliance Declines Between Clinic Visits. *Archives of Internal Medicine*, 150(7):1509–1510, 07 1990. 11

[37] Diana I. Brixner, Carrie McAdam-Marx, Xiangyang Ye, Helen Lau, and Mark A Munger. Assessment of time to follow-up visits in newly-treated hypertensive patients using an electronic medical record database. *Current Medical Research and Opinion*, 26(8): 1881–1891, 2010. 11

[38] Miguel A Hernán and James M Robins. Per-protocol analyses of pragmatic trials. *New England Journal of Medicine*, 377(14):1391–1398, 2017. 12

[39] Jr JT Wright, G Bakris, and T and Green. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: Results from the AASK trial. *Journal of the American Medical Association*, 288(19):2421–2431, 2002. 12, 22, 23, 24, 26, 27, 40

[40] Stephen R. Cole and Constantine E. Frangakis. The Consistency Statement in Causal Inference: A Definition or an Assumption? *Epidemiology*, 20:3–5, 01 2009. 13

[41] Miguel A Hernán and Tyler J VanderWeele. Compound treatments and transportability of causal inference. *Epidemiology (Cambridge, Mass.)*, 22(3):368, 2011.

[42] Tyler J. VanderWeele and Miguel A. Hernán. Causal Inerence Under Multiple Versions of Treatment. *Journal of Causal Inference*, 1:1–20, 05 2013.

[43] David H. Rehkpof, Maria M. Glymour, and Theresa L. Osypuk. The consistency assumption for causal inference in social epidemiology: When a rose is not a rose. *Current Epidemiology Reports*, 3:63–71, 03 2016. 13

[44] Paul R Rosenbaum. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series A*, 147(5):656–666, 1984. 16

[45] James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9-12):1393–1512, 1986. 16

[46] MA Hernán and JM Robins. Causal inference: What if. *Chapman & Hall/CRC*, 2020. 16, 48, 49

[47] James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393 – 1512, 1986. 17

[48] MA Hernán, B Brumback, and JM Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5): 561–570, 2000. 17, 35, 49, 55, 56

[49] Charles AE Goodhart. Problems of monetary management: The UK experience. In *Monetary Theory and Practice*, pages 91–121. Springer, 1984. 17

[50] Marilyn Strathern. 'Improving ratings': Audit in the British University system. *European review*, 5(3):305–321, 1997. 17

[51] David Manheim and Scott Garrabrant. Categorizing variants of Goodhart's Law. *arXiv preprint arXiv:1803.04585*, 2018. 18

[52] Tyler J. Vanderweele and Stijn Vansteelandt. Conceptual issues concerning mediation, interventions, and composition. *Statistics and Its Interface*, 2:457–468, 01 2009. 18

[53] Irene M Stratton, Amanda I Adler, H Andrew W Neil, David R Matthews, Susan E Manley, Carole A Cull, David Hadden, Robert C Turner, and Rury R Holman. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): A prospective observational study. *BMJ*, 321(7258):405–412, 2000. 20

[54] Kay-Tee Khaw, Nicholas Wareham, Sheila Bingham, Robert Luben, Ailsa Welch, and Nicholas Day. Association of hemoglobin A1c with cardiovascular disease and mortality in adults: The European prospective investigation into cancer in Norfolk. *Annals of Internal Medicine*, 141:413–320, 2004. 20

[55] Yehuda Handelsman, Zachary T. Bloomgarden, George Grunberger, Guillermo Umpierrez, Robert S. Zimmerman, Timothy S. Bailey, Lawrence Blonde, George A. Bray, A. Jay Cohen, Samuel Dagogo-Jack, Jaime A. Davidson, Daniel Einhorn, Om P. Ganda, Alan J. Garber, W. Timothy Garvey, Robert R. Henry, Irl B. Hirsch, Edward S. Horton, Daniel L. Hurley, Paul S. Jellinger, Lois Jovanovič, Harold E. Lebovitz, Derek LeRoith, Philip Levy, Janet B. McGill, Jeffrey I. Mechanick, Jorge H. Mestman, Etie S. Moghissi, Eric A. Orzeck, Rachel Pessah-Pollack, Paul D. Rosenblit, Aaron I. Vinik, Kathleen Wyne, and Farhad Zangeneh. American Association of Clinical Endocrinologists and American College of Endocrinology - Clinical Practice Guidelines for Developing a Diabetes Mellitus Comprehensive Care Plan - 2015. *Endocrine Practice*, 21(Supplement 1):1–87, 2015. 20

[56] American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care*, 43:S66–S76, 2020. 20

[57] Amir Qaseem, Timothy J. Wilt, Devan Kansagara, Carrie Horwitch, Michael J. Barry, Mary Ann Forciea, and for the Clinical Guidelines Committee of the American College of Physicians. Hemoglobin A1c Targets for Glycemic Control With Pharmacologic Therapy for Nonpregnant Adults With Type 2 Diabetes Mellitus: A Guidance Statement Update From the American College of Physicians. *Annals of Internal Medicine*, 168(8):569–576, 04 2018. 20

[58] Elizabeth Selvin, Michael W. Steffes, Hong Zhu, Kunihiro Matsushita, Lynne Wagenknecht, James Pankow, Josef Coresh, and Frederick L. Brancati. Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *New England Journal of Medicine*, 362(9):800–811, 2010. 20

[59] April P. Carson, Caroline S. Fox, Darren K. McGuire, Emily B. Levitan, Martin Laclaustra, Devin M. Mann, and Paul Muntner. Low hemoglobin A1c and risk of all-cause mortality among US adults without diabetes. *Circulation: Cardiovascular Quality and Outcomes*, 3(6):661–667, 2010. 20

[60] Benjamin M. Scirica, Deepak L. Bhatt, Eugene Braunwald, P. Gabriel Steg, Jaime Davidson, Boaz Hirshberg, Peter Ohman, Robert Frederich, Stephen D. Wiviott, Elaine B. Hoffman, Matthew A. Cavender, Jacob A. Udell, Nihar R. Desai, Ofri Mosenzon, Darren K. McGuire, Kausik K. Ray, Lawrence A. Leiter, and Itamar Raz. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *New England Journal of Medicine*, 369(14):1317–1326, 2013. 20

[61] Heart failure and mortality outcomes in patients with type 2 diabetes taking alogliptin versus placebo in EXAMINE: A multicentre, randomised, double-blind trial. *The Lancet*, 385(9982):2067 – 2076, 2015. 20

[62] Quang T. Nguyen, Scott R. Anderson, Lindsay Sanders, and Loida Nguyen. Managing Hypertension in the Elderly: A Common Chronic Disease with Increasing Age. *American Health and Drug Benefits*, 5:146–153, 05 2012. 21, 28

[63] Giuseppe Mancia and Guido Grassi. Aggressive Blood Pressure Lowering Is Dangerous: The J-Curve. *Hypertension*, 64:29–36, 01 2014.

[64] Franz H. Messerli, Giuseppe Mancia, Richard Conti, Ann Hewkin, Stuart Kupfer, Annette Champion, Rainer Kolloch, Athanase Benetos, and Carl J. Pepine. Dogma Disputed: Can Aggressively Lowering Blood Pressure in Hypertensive Patients with Coronary Artery Disease Be Dangerous. *Annals of Internal Medicine*, 144:884–893, 2006. 21

[65] John M. Cruickshank, Jeffrey M. Thorp, and James F. Zacharias. Benefits and Potential Harm of Lowering High Blood Pressure. *Lancet*, 329:581–584, 03 1987. 21, 28

[66] Paolo Verdecchia, Jan A Staessen, Fabio Angeli, Giovanni de Simone, Augusto Achilli, Antonello Ganau, Gianfrancesco Mureddu, Sergio Pede, Aldo P Maggioni, Donato Lucci, Gianpaolo Reboldi, and Cardio-Sis investigators. Usual versus tight control of systolic blood pressure in non-diabetic patients with hypertension (Cardio-Sis): An open-label randomised trial. *The Lancet*, 374(9689):525 – 533, 2009. 22, 23, 25, 26, 27

[67] SPS3. Blood-pressure targets in patients with recent lacunar stroke: The SPS3 randomised trial. *The Lancet*, 382(9891):507–515, 2010. 22, 24, 26, 27

[68] Piero Ruggeneti, Annalisa Perna, Giacomina Loriga, Maria Ganeva, Bogdan Ene-Iordache, Marta Turturro, Elena Perticucci, Ivan Nediyalkov Chakarski, Daniela Leonardis, giovanni Garini, Adalberto sessa, Carlo Basile, Mirella Alpa, Renzo Scanziani, Sorba Gianbattista, Carmine Zoccali, Remuzzi Giuseppe, and REIN-2 Study Group. Blood-pressure control for renoprotection in patients with non-diabetic chronic renal disease (REIN-2): Multicentre, randomised controlled trial. *The Lancet*, 365(9463): 939 – 946, 2005. 22, 23, 24, 26, 27, 40

[69] Saulo Klahr, Andrew S. Levey, Gerald J. Beck, Arlene W. Caggiula, Lawrence Hunsicker, John W. Kusek, and Gary Striker. The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. *New England Journal of Medicine*, 330(13):877–884, 1994. 22, 24, 26

[70] JATOS. Principal Results of the Japanese Trial to Assess Optimal Systolic Blood Pressure in Elderly Hypertensive Patients (JATOS). *Hypertension Research*, 31:2115–2127, 2008. 22, 24, 25, 26, 27

[71] Toshio Ogihara, Takao Saruta, Hiroaki Matsuoka, Kazuaki Shimamoto, Toshiro Fujita, Kazuyuki Shimada, Yutaka Imai, and Masaru Nishigaki. Valsartan in elderly isolated systolic hypertension (VALISH) study: Rationale and design. *Hypertension Research*, 27(9):657–661, 2004. 23, 24, 26, 27, 40

[72] Yong Wei, Zhimin Jin, Guoying Shen, Xiaowei Zhao, Wanhua Yang, Ye Zhong, and Jiguang Wang. Effects of intensive antihypertensive treatment on Chinese hypertensive patients older than 70 years. *The Journal of Clinical Hypertension*, 15(6):420–427, 2013. 23, 24, 26, 27

[73] Sverre E. Kjeldsen, Thomas Hedner, Kenneth Jamerson, Stevo Julius, William E. Haley, Miguel Zabalgoitia, Amir R. Butt, S. Noor Rahman, and Lennart Hansson. Hypertension optimal treatment (HOT) study. *Hypertension*, 31(4):1014–1020, 1998. 24, 26, 27

[74] Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, RuiJun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, and Patrick B Ryan. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: A systemtatic, multinational, large-scale analysis. *The Lancet*, 394:1816–1826, 11 2019. 34, 48

[75] Peter F. Thall, Christopher Logothetis, Lance C. Pagliaro, Sijin Wen, Melissa A. Brown, Dallas Williams, and Randall E. Millikan. Adaptive Therapy for Androgen-Independent Prostate Cancer: A Randomized Selection Trial of Four Regimens. *JNCI: Journal of the National Cancer Institute*, 99(21):1613–1622, 11 2007. 34

[76] Victoria Lin, Sean McGrath, Zilu Zhang, Lucia C. Petito, Roger W. Logan, Miguel A. Hernán, and Jessica G. Young. gformula: An R package for estimating effects of general time-varying treatment interventions via the parametric g-formula, 2019. 35

[77] Mark van der Laan and Alexander Luedtke. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of Causal Inference*, 3(1):61–95, 2014.

[78] Mohammad Ali Mansournia, Mahyar Etminan, Goodarz Danaei, Jay S Kaufman, and Gary Collins. Handling time varying confounding in observational research. *BMJ*, 359, 08 2017. 35

[79] Yi Zhang, Jessica G Young, Mae Thamer, and Miguel A Hernán. Comparing the Effectiveness of Dynamic Treatment Strategies Using Electronic Health Records: An Application of the Parametric g-Formula to Anemia Management Strategies. *Health Services Research*, 53:1900–1918, June 2018. 35

[80] Bibhas Chakraborty and Susan A. Murphy. Dynamic treatment regimes. *Annual Review of Statistics and Its Application*, 1(1):447–464, 2014. 35

[81] Heather A. Liszka, Arch G. Mainous, Dana E. King, Charles J. Everett, and Brent M. Egan. Prehypertension and Cardiovascular Morbidity. *The Annals of Family Medicine*, 3(4):294–299, 2005. 47

[82] Ramachandran S. Vasan, Martin G. Larson, Eric P. Leip, Jane C. Evans, Christopher J. O'Donnell, William B. Kannel, and Daniel Levy. Impact of high-normal blood pressure on the risk of cardiovascular disease. *New England Journal of Medicine*, 345(18):1291–1297, 2001. 47

[83] Matthew S. Freiberg, Chung-Chou H. Chang, Lewis H. Kuller, Melissa Skanderson, Elliott Lowy, Kevin L. Kraemer, Adeel A. Butt, Matthew Bidwell Goetz, David Leaf, Kris Ann Oursler, David Rimland, Maria Rodriguez Barradas, Sheldon Brown, Cynthia Gibert, Kathy McGinnis, Kristina Crothers, Jason Sico, Heidi Crane, Alberta Warner, Stephen Gottlieb, John Gottdiener, Russell P. Tracy, Matthew Budoff, Courtney Watson, Kaku A. Armah, Donna Doebler, Kendall Bryant, and Amy C. Justice. HIV Infection and the Risk of Acute Myocardial Infarction. *JAMA Internal Medicine*, 173(8):614–622, 04 2013. 48

[84] Madeleine Durand, Sheehy Odile, Jean-Guy Baril, Jacques Lelorier, and Cécile L Tremblay. Association Between HIV Infection, Antiretroviral Therapy, and Risk of Acute Myocardial Infarction: A Cohort and Nested Case–Control Study Using Québec's Public

Health Insurance Database. *Journal of Acquired Immune Deficiency Syndromes*, 57(3): 245–253, July 2011.

[85] Kaku A. Armah, Chung-Chou H. Chang, Jason V. Baker, Vasan S. Ramachandran, Matthew J. Budoff, Heidi M. Crane, Cynthia L. Gibert, Matthew B. Goetz, David A. Leaf, Kathleen A. McGinnis, Krisann K. Oursler, David Rimland, Maria C. Rodriguez-Barradas, Jason J. Sico, Alberta L. Warner, Priscilla Y. Hsue, Lewis H. Kuller, Amy C. Justice, Matthew S. Freiberg, and for the Veterans Aging Cohort Study (VACS) Project Team. Prehypertension, Hypertension, and the Risk of Acute Myocardial Infarction in HIV-Infected and -Uninfected Veterans. *Clinical Infectious Diseases*, 58(1): 121–129, 09 2013. 48

[86] Judith S Currier. Update on cardiovascular complications in HIV infection. *Topics in HIV medicine : a publication of the International AIDS Society, USA*, 17(3), 2009. 48

[87] Charlotte Lewden, Thierry May, Eric Rosenthal, Christine Burty, Fabrice Bonnet, Dominique Costagliola, Eric Jougla, Caroline Semaille, Philippe Morlat, Dominique Salmon, Patrice Cacoub, Geneviève Chêne, and the ANRS EN19 Mortalité Study Group and Mortavic. Changes in Causes of Death Among Adults Infected by HIV Between 2000 and 2005: The "Mortalité 2000 and 2005" Surveys (ANRS EN19 and Mortavic). *Journal of Acquired Immune Deficiency Syndromes*, 48(5):590–598, August 2008. 48

[88] The Antiretroviral Therapy Cohort Collaboration. Causes of Death in HIV-1—Infected Patients Treated with Antiretroviral Therapy, 1996–2006: Collaborative Analysis of 13 HIV Cohort Studies. *Clinical Infectious Diseases*, 50(10):1387–1396, 05 2010. 48

[89] Stephen R. Cole and Miguel A. Hernán. Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*, 168(6):656–664, 08 2008. 49, 59

[90] Yongling Xiao, Erica E.M. Moodie, and Michal Abrahamowicz. Comparison of approaches to weight truncation for marginal structural cox models. *Epidemiologic Methods*, 2(1):1–20, 2013. 49, 59

[91] Mohammad Ehsanul Karim, Paul Gustafson, John Petkau, Yinshan Zhao, Afsaneh Shirani, Elaine Kingwell, Charity Evans, Mia van der Kop, Joel Oger, and Helen Tremlett. Marginal Structural Cox Models for Estimating the Association Between β-Interferon Exposure and Disease Progression in a Multiple Sclerosis Cohort. *American Journal of Epidemiology*, 180(2):160–171, 06 2014. 49

[92] Haidong Lu, Stephen R Cole, Daniel Westreich, Michael G Hudgens, Adaora A Adimora, Keri N Althoff, Michael J Silverberg, Kate Buchacz, Jun Li, Jessie K Edwards, Peter F Rebeiro, Viviane D Lima, Vincent C Marconi, Timothy R Sterling, Michael A Horberg, M John Gill, Mari M Kitahata, Joseph J Eron, for the North American AIDS Cohort Collaboration on Research Moore, Richard D, and Design of the International Epidemiologic Databases to Evaluate AIDS. Clinical Effectiveness of Integrase Strand Transfer Inhibitor–Based Antiretroviral Regimens Among Adults With Human Immunodeficiency Virus: A Collaboration of Cohort Studies in the United States and Canada. *Clinical Infectious Diseases*, 73(7):e1408–e1414, 08 2020. 49

[93] Ankur Bhargava, Tae Kim, Douglas B. Quine, and Ronald George Hauser. A 20-Year Evaluation of LOINC in the United States' Largest Integrated Health System. *Archives of Pathology & Laboratory Medicine*, 144(4):478–484, 08 2019. 49

[94] Shawn L. Fultz, Melissa Skanderson, Larry A. Mole, Neel Gandhi, Kendall Bryant, Stephen Crystal, and Amy C. Justice. Development and Verification of a "Virtual" Cohort Using the National VA Health Information System. *Medical Care*, 44(8):S25–S30, 2006. 50

[95] Amy C. Justice, Kirsha S. Gordon, Melissa Skanderson, Eva Jennifer Edelman, Kathleen M. Akgun, Cynthia L Gibert, Vincent Lo Re, David Rimland, Julie A Womack, Christina M. Wyatt, Janet P. Tate, and VACS Project Group. Nonantiretroviral polypharmacy and adverse health outcomes among HIV-infected and uninfected individuals. *AIDS*, 32:739–749. 52

[96] Diane C. Cowper, Joseph D. Kubal, Charles Maynard, and Denise M. Hynes. A Primer and Comparative Review of Major U.S. Mortality Databases. *Annals of Epidemiology*, 12(7):462–468, 2002. ISSN 1047-2797. 52

[97] Amy C. Justice, Sharada P. Modur, Janet P. Tate, Keri N. Althoff, Lisa P. Jacobson, Kelly A. Gebo, Mari M. Kitahata, Michael A. Horberg, John T. Brooks, Kate Buchacz, Sean B. Bourke, Anita Rachlis, Sonia Napravnik, Joseph Eron, James H. Willig, Richard Moore, Gregory D. Kirk, Ronald Bosch, Benigno Rodriguez, Robert S. Hogg, Jennifer Thorne, James J. Goedert, Marina Klein, John Gill, Steven Deeks, Timothy R. Sterling, Kathryn Anastos, Stephen J. Gange, NA-ACCORD, and VACS Project Teams. Predictive accuracy of the Veterans Aging Cohort Study index for mortality with HIV infection: a North American cross cohort analysis. *Journal of Acquired Immune Deficiency Syndrome*, 62(2):149–163, February 2013. 54

[98] Janet P. Tate, Amy C. Justice, Michael D. Hughes, Fabrice Bonnet, Peter Reiss, Amanda Mocroft, Jacob Nattermann, Fiona C Lampe, Heiner C. Bucher, Timothy R Sterling, Heidi M. Crane, Mari M. Kitahata, Margaret T May, and Jonathan A. C. Sterne. An internationally generalizable risk index for mortality after one year of antiretroviral therapy. *AIDS*, 27:563–572, 2013. 54

[99] LA Barakat, M Juthani-Mehta, H Allore, M Trentalange, J Tate, D Rimland, M Pisani, KM Akgun, MB Goetz, AA Butt, M Rodriguez-Barradas, M Duggal, K Crothers, AC Justice, and VJ Quagliarellao. Comparing clinical outcomes in HIV-infected and uninfected older men hospitalized with community-acquired pneumonia. *HIV Medicine*, 16(7):421–430, August 2015. 54

[100] Lloyd D. Fisher and D. Y. Lin. Time-dependent covariates in the Cox proportional hazards regression model. *Annual Review of Public Health*, 20(1):145–157, 1999. 57

[101] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011. URL `https://www.jstatsoft.org/v45/i03/`. 59

[102] Ian R. White, Patrick Royston, and Angela M. Wood. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399, 2011. 59

[103] Terry M Therneau. *A Package for Survival Analysis in R*, 2021. URL `https://CRAN.R-project.org/package=survival`. R package version 3.2-13. 59

[104] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000. ISBN 0-387-98784-3. 59

[105] Ayeshi Ali, M Adnan Ali, and Zhe Wei. On computing standard errors for marginal structural Cox models. *Lifetime Data Analysis*, 20:106–131, 2014. 59

[106] Andrea Marshall, Douglas G Altman, Roger L Holder, and Patrick Royston. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*, 9(57), 2009. 66

[107] Miguel Hernán. The Hazards of Hazard Ratios. *Epidemiology*, 21(1):13–15, January 2010. 72

[108] Michele Santacatterina and Matteo Bottai. Optimal Probability Weights for Inference With Constrained Precision. *Journal of the American Statistical Association*, 113(523): 983–991, 2018. 72

[109] Brian K. Lee, Justin Lessler, and Elizabeth A. Stuart. Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3):337–346, 2010. 72

[110] Yeying Zhu, Debashis Ghosh, Nandita Mitra, and Bhramar Mukherjee. A data-adaptive strategy for inverse weighted estimation of causal effects. *Health Services and Outcomes Research Methodology*, pages 69–41, 2014. 72

[111] Zunyou Wu and Jennifer M McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72,314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, 323 (13):1239–1242, February 2020. 90

[112] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, Lulu Guan, Yuan Wei, Hui Li, Xudong Wu, Jiuyang Xu, Shengjin Tu, Yi Zhang, Hua Chen, and Bin Cao. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229):1054–1062, March 2020. 90, 94

[113] Giacomo Grasselli, Alberto Zangrillo, Alberto Zanella, Massimo Antonelli, Luca Cabrini, Antonio Castelli, Danilo Cereda, Antonio Coluccello, Giuseppe Foti, Roberto Fumagalli, Giorgio Iotti, Nicola Latronico, Luca Lorini, Stefano Merler, Giuseppe Natalini, Alessandra Piatti, Marco Vito Ranieri, Anna Mara Scandroglio, Enrico Storti, Maurizio Cecconi, Antonio Pesenti, and for the COVID-19 Lombardy ICU Network. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*, April 2020. 90

[114] Josefa Velasquez, Ann Choi, and Yoav Gonen. Hospitals nearing ICU bed limits as COVID-19 surges in NYC. *The City*, March 2020. 90

[115] Miguel Marquez and Sonia Moghe. Inside a Brooklyn hospital that is overwhelmed with Covid-19 patients and deaths. *CNN*, March 2020. 90

[116] Bill Parry. City's public health system to triple ICU capacity and expand personnel as COVID-19 surge continues. *QNS*, April 2020. 90

[117] Lisa Schencker and David Heinzmann. As COVID-19 cases increase, intensive care beds are filling up at Chicago-area hospitals. *Chicago Tribune*, April 2020. 90

[118] Nick Iannelli. Hospitals scramble to increase capacity in Prince George's Co. *WTOP News*, April 2020.

[119] Healthcare organizations take steps to increase patient capacity during COVID-19. *MEDITECH*, April 2020. 90

[120] Monica Tapia. Governor orders hospitals to expand capacity. *AZ BEX*, March 2020. 90

[121] Stephen Fowler. Georgia expands COVID-19 testing criteria, hospital capacity as peak approaches. *GBP Radio News*, April 2020.

[122] Jon Campbell. Andrew Cuomo's order to hospitals: Expand capacity or face state takeover. *Democrat & Chronicle*, April 2020.
[123] Governor Newsom Announces Progress in Expanding Hospital Capacity to Fight Coming Surge in COVID-19 Cases. *Office of Governor Gavin Newsom, CA.gov*, April 2020. 90, 91
[124] Adam Blumenberg, Matt Noble, and Robert G. Hendrickson. COVID-19 and surge capacity in U.S. hospitals. *The Hospitalist*, April 2020. 91
[125] Anil M. Makam and David C. Grabowski. How can we ramp up hospital capacity to handle the surge of COVID-19 patients? Long-term acute care hospitals can play a critical role. *Health Affairs*, April 2020. 91
[126] Yale converts gym space for Covid-19 . *The CT Mirror*, March 2020. 91
[127] David Robinson, Ashley Biviano, and Steve Orr. Coronavirus in New York: Hospitals race to expand bed capacity 50% as COVID-19 cases surge. *lohud.*, April 2020. 91
[128] Lisa Rosenbaum. Facing Covid-19 in Italy — Ethics, Logistics, and Therapeutics on the Epidemic's Front Line. *New England Journal of Medicine*, 382:1873–1875, March 2020. 91
[129] Robert D. Truog, Christine Mitchell, and George Q. Daley. The toughest triage — allocating ventilators in a pandemic. *New England Journal of Medicine*, 382:1973–1975, May 2020.
[130] Mattia Ferraresi. A coronavirus cautionary tale from Italy: Don't do what we did. *The Boston Globe*, March 2020. 94
[131] J Horowitz and David D Kirkpatrick. Dip in Italy's cases does not come fast enough for swamped hospitals. *New York Times*, March 2020. 91
[132] Matthew Schwartz. U.S. Adds 184,000 Coronavirus Cases in 1 Day, With No End In Sight. *National Public Radio*, November 2020. 91
[133] Mitch Smith. Covid-19: Pandemic Shatters More Records in U.S., as States and Cities Tighten Restrictions. *The New York Times*, November 2020. 91
[134] Steven Elbow. Hospitals scramble for staff, ICU beds as Wisconsin passes a quarter million COVID-19 cases. *The Cap Times*, November 2020. 91
[135] Meredith Deliso. Over 900 Mayo Clinic staff in Midwest have contracted COVID-19 in past 2 weeks. *ABC News*, November 2020. 91
[136] Vicente Arenas. COVID-19 hospitalizations fill ICU beds to capacity at Denver Health. *Fox News*, November 2020.
[137] Julia Jacobo and Arielle Mitropoulos. Hospital ICUs running out of space due to COVID-19 surges across the country. *ABC News*, November 2020. 91
[138] Joanne Silberner. Covid-19: North Dakota and Belgium have let infected health staff work on wards. *BMJ*, 371, November 2020. 91
[139] Charlie Smart. Covid hospitalizations hit crisis levels in Southern ICUs. *The New York Times*, September 2021. 91
[140] Centers for Disease Control and Prevention. Delta variant. August 2021. 91
[141] Shadman Aziz, Yaseen M. Arabi, Waleed Alhazzani, Laura Evans, Giuseppe Citerio, Katherine Fischkoff, Jorge Salluh, Geert Meyfroidt, Fayez Alshamsi, Simon Oczkowski, Elie Azoulay, Amy Price, Lisa Burry, Amy Szierba, Andrew Benintende, Jill Morgan, Giacomo Grasselli, Andrew Rhodes, Morten H. Møller, Larry Chu, Shelly Schwedhelm,

John J. Low, Du Bin, and Michael D. christian. Managing ICU surge during the COVID-19 crisis: Rapid guidelines. *Intensive Care Medicine*, 46:1303–1325, June 2020. 92, 108

[142] Yue-Nan Ni, Jian Luo, He Yu, Dan Liu, Bin-Miao Liang, and Zong-An Liang. The effect of high-flow nasal cannula in reducing the mortality and the rate of endotracheal intubation when used before mechanical ventilation compared with conventional oxygen therapy and noninvasive positive pressure ventilation: A systematic review and meta-analysis. July 2017. 92

[143] Xiaofeng Ou, Yusi Hua, Jin Liu, Cansheng Gong, and Wenling Zhao. Effect of high-flow nasal cannula oxygen therapy in adults with acute hypoxemic respiratory failure: a meta-analysis of randomized controlled trials. *CMAJ*, 189(7):E260–E267, 2017.

[144] B. Rochwerg, D. Granton, D.X. Wang, Y. Helviz, S. Einav, J.P. Frat, A. Mekontso-Dessap, A. Schreiber, Azoulay E., A. Mercat, A. Demoule, V. Lemiale, A. Pesenti, E.D. Riviello, T. Mauri, J. Mancebo, L. Brochard, and K. Burns. High flow nasal cannula compared with conventional oxygen therapy for acute hypoxemic respiratory failure: A systematic review and meta-analysis. *Intensive Care Medicine*, 45:563–572, March 2019. 92

[145] Sharon Einav, John L. Hick, Dan Hanfling, Brian L. Erstad, Eric S. Toner, Richard D. Branson, Robert K. Kanter, Niranjan Kissoon, Jeffrey R. Dichter, Asha V. Devereaux, and Michael D. Christian. Surge Capacity Logistics: Care of the Critically Ill and Injured During Pandemics and Disasters: CHEST Consensus Statement. *Chest*, 146:e17S – e43S, October 2014. 92

[146] Chris Carter and Joy Notter. COVID-19 disease: a critical care perspective. *Clinics in Integrated Care*, 1:100003, 2020. 92

[147] CovidActNow. `https://covidactnow.org/`, 2020. Accessed: 2021-06-04. 92

[148] Nicholas B Noll, Ivan Aksamentov, Valentin Druelle, Abrie Badenhorst, Bruno Ronzani, Gavin Jefferies, Jan Albert, and Richard Neher. COVID-19 Scenarios: An interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2. *medRxiv*, 2020.

[149] Covid-19 scenarios. `https://covid19-scenarios.org/`, 2020. Accessed: 2021-06-04.

[150] Ihme covid-19 model. `https://covid19.healthdata.org/`, 2020. Accessed: 2021-06-04.

[151] Christopher JL Murray and IHME COVID-19 health service utilization forecasting team. Forecasting the impact of the first wave of the covid-19 pandemic on hospital demand and deaths for the usa and european economic area countries. *medRxiv*, April 2020.

[152] Covid-19 simulator. `https://www.covid19sim.org/`, 2020. Accessed: 2021-06-04.

[153] Mack Grenfell. US COVID19 Forecaster. `https://mackgrenfell.com/forecaster/covid19`. Accessed: 2021-06-04.

[154] Gabriel Goh. Epidemic Calculator. `https://gabgoh.github.io/COVID/index.html`, 2020. Accessed: 2021-06-04.

[155] Ken O'Day, Dylan Mezzio, and Kellie Meyer. The Potential Impact of Pharmaceutical and Non-Pharmaceutical Interventions to Mitigate the COVID-19 Crisis in the United States: A Model-Based Analysis. *Xcenda (Amerisource Bergen)*, March 2020.

[156] GLEAM COVID-19 Modeling. https://covid19.gleamproject.org/, 2020. Accessed: 2021-06-04.

[157] COVID-19 Hospital Impact Model for Epidemics (CHIME). https://penn-chime.phl.io/, 2020. Accessed: 2021-06-04.

[158] Alison Hill. Modeling COVID-19 Spread vs. Healthcare Capacity. https://alhill.shinyapps.io/COVID19seir/, 2020. Accessed: 2021-06-04.

[159] BB Adhikari, LS Fischer, B Greening, S Jeon, EB Kahn, GJ Kang, G Rainisch, MI Meltzer, and ML Washington. COVID19Surge: a manual to assist state and local public health officials and hospital administrators in estimating the impact of a novel coronavirus pandemic on hospital surge capacity (Beta test version). *Centers for Disease Control and Prevention, U.S. Department of Health and Human Services*, April 2020.

[160] Seyed M Moghadas, Affan Shoukat, Meagan C Fitzpatrick, Chad R Wells, Pratha Sah, Abhishek Pandey, Jeffrey D Sachs, Zheng Wang, Lauren A Meyers, Burton H Singer, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences*, 117(16):9122–9126, March 2020.

[161] A model to estimate bed demand for COVID-19 related hospitalization, author=Zhang, Teng and McFarlane, Kelly and Vallon, Jacqueline and Yang, Linying and Xie, Jin and Blanchet, Jose and Glynn, Peter and Staudenmayer, Kristan and Schulman, Kevin and Scheinker, David. *medRxiv*, March 2020.

[162] Johannes Opsahl Ferstad, Angela Jessica Gu, Raymond Ye Lee, Isha Thapa, Andrew Y Shin, Joshua A Salomon, Peter Glynn, Nigam H Shah, Arnold Milstein, Kevin Schulman, et al. A model to forecast regional demand for COVID-19 related hospital beds. *medRxiv*, April 2020. 92

[163] Vincent Chin, Noelle I. Samia, Roman Marchant, Ori Rosen, John P.A. Ioannidis, Martin A. Tanner, and Sally Cripps. A case study in model failure? COVID-19 daily deaths and Icu bed utilisation predictions in New York State. *European Journal of Epidemiology*, 35:733–742, August 2020. 92