The Effect of Specialty Mental Health Probation on Public Safety Outcomes

by

Lina Montoya

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Arts

in

Biostatistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Maya Petersen, Chair
Professor Jennifer Skeem
Professor John Marshall

Spring 2017

The Effect of Specialty Mental Health Probation on Public Safety Outcomes

To Ivan and Martha

Si estoy segura de una relación causal es del efecto que el apoyo de ustedes ha tenido en todas las cosas maravillosas que han pasado en mi vida.

Contents

Chapter 1

Introduction

1.1 Background

The United States contains one quarter of the world's prisoners, even though Americans comprise only 5% of the world's population [12]. The US has the highest rate of incarceration in the world: the number of incarcerated people increased from approximately 500,000 to 2.3 million people since 1980 [1, 8]. Recidivism, or the tendency for a prisoner to re-offend, is extremely common: within 3 years of release from prison, two-thirds of released prisoners are re-arrested; within 5 years of release, three-quarters are re-arrested [2]. Such high rates of mass incarceration and re-entry into the criminal justice system lead to overcrowding of prisons and excessive costs, ultimately removing the "correctional" part of "correctional facilities."

People with mental illness are especially susceptible to involvement in the criminal justice system. These people are consistently more prevalent in prison and jail samples (14-17%) than in the general population (although, to be clear: mentally ill persons are *not* more likely to commit violent crimes) [38, 45, 48]. Not only are people with mental illness overrepresented in the criminal justice system, but they also have longer stays in jail and have higher rates of recidivism [45]. Stepping Up, a national initiative whose mission is to reduce the prevalence of people with mental illnesses in jails, notes "Jails spend two to three times more money on adults with mental illnesses that require intervention than on those without those needs, yet often do not see improvements to public safety or these individuals health" [3].

Although these facts are staggering, a hopeful shift has occurred in recent years: the national prison population has stabilized. According to a 2015 report by the US Bureau of Justice Statistics, since 2008 there has been a decline in jail admissions and the incarcerated population is at its lowest since 2004 [26, 15]. The Sentencing Project, a research and advocacy center whose aim is to address US incarceration issues, hypothesizes that this stabilization is partly due to changes in policy and practice learned from evidence-based approaches to public safety [1].

Indeed, one such evidence-based intervention is probation, a court-ordered period of time in which an offender is supervised in the community in lieu of serving time in prison. For example, Petersilia, et al. (1986) found that, in general, prisoners had higher recidivism rates than probationers [35] (for other examples, see [23]). Additionally, probation is cost effective: according to the Washington State Institute for Public Policy's report of the costs and benefits of programs to reduce crime, "when intensive supervision is used as an alternative to prison, there is an immediate, up-front cost savings since supervision in the community is cheaper than incarceration" [4]. Since 2009, at least a dozen US states have updated their community supervision using practices that have been empirically proven to reduce recidivism [21].

Specialty mental health probation (henceforth referred to as "specialty probation") is one type of community-based, balanced supervision that attempts to evenly distribute "care" (social work) with "control" (police work) among probationers with mental illness [25]. It was conceived to address the aforementioned issues, namely to "reduce the disproportionately high rate of community supervision failures among offenders with mental illness" [25]. Specialty probation differs from traditional, surveillance-style probation in five core ways [43, 25]:

1. Specialty probation officers oversee smaller caseloads ($M < 50$ probationers) than traditional officers ($M > 100$), and specialty caseloads are exclusively comprised of offenders with mental illness.

2. Specialty probation officers are trained in handling mental illness issues (e.g., knowledge of diagnoses, how to refer probationers to treatment in the community, etc.)

3. Specialty probation officers aim to integrate probation and treatment, whereas traditional probation officers refer probationers to the community with little direct involvement.

4. Specialty probation officers obtain compliance with probationers via problem solving, whereas traditional officers rely on threats.

5. Specialty officers have, on average, authoritative, as opposed to authoritarian, relationships with probationers.

As of 2006, specialty probation is claimed to have been implemented in 130 agencies; however, not all agencies carry out the program's defining features [43]. This may be due to the limited evidence available of its impact on public safety outcomes. Thus far, only two main studies have addressed specialty probation's efficacy. The first is an unpublished, independent evaluation of the IMPACT program in California in which 800 probationers with mental illness were randomly assigned to one of the following groups: no probation, no probation but receiving treatment, traditional probation, and specialty probation. Results showed that there were no significant differences in recidivism between the four groups (see [44] for a description). The second study reported that specialty probationers had significantly fewer probation violations and fewer jail days 6 months after specialty probation assignment [52]. The results of this quasi-experimental study suggest that specialty probation is effective; however, the follow-up period after probation treatment was short (6 months) and the authors only used baseline demographics and offense type as covariates to correct for the non-experimental nature of their study. Wolff et al. describe other preliminary studies, but most had limitations that made them hard to generalize findings, such as non-randomization, small sample sizes, loss to follow-up, and short follow-up periods.

1.2 Current study

This thesis uses data from a longitudinal, multi-method quasi-experiment designed to rigorously test the effectiveness of specialty probation.

Procedure

Two probation agencies were chosen based on a national survey on probation and mental health ([43], see [25] for more details). First, a prototypical specialty agency was selected in Dallas, Texas. Next, a traditional probation agency in Los Angeles, California was chosen based on its match to the specialty sites characteristics. The two sites were similar in jurisdiction size, urban location, probationer demographic characteristics, and county mental health expenditures. Probationers and their supervising officers were interviewed at baseline, and again 6 and 12 months later. Researchers also collected information about violations from probation, court, and Federal Bureau of Investigation (FBI) records post-placement.

Participants

In order to be eligible for participation in the study, probationers had to be 18-65 years old, English speaking, capable of providing informed consent, living with a probable or known Axis I mental illness (i.e., acute mental illness without personality disorder or mental retardation), have no intellectual disability, on active probation supervision, have at least one year of probation remaining, and have met with their supervising officer at least once since starting probation.

Identification of probationer mental illness differed by site. First, participants were recruited at the specialty site (of 351 recruited, 248 were eligible; of these 248 eligible, 183 were enrolled). Here, probationers were either originally on traditional probation but were subsequently referred to specialty caseload supervision by a probation officer or referred directly from the court.

At the traditional probation site, eligible probationers were identified by a) asking officers which probationers he or she would “refer” to specialty mental health supervision (i.e., clients with a known psychiatric diagnosis, psychotropic medications, and/or hospitalizations) and b) a self-report mental health screen. Then, traditional probation recruits were matched on characteristics observed in the Dallas site based on gender, age, ethnicity, length of time on probation, and type of offense. Eight-hundred and sixty traditional probationers were recruited, of those 504 were eligible. Of 504 eligible participants, 311 matched the enrolled participants in Dallas, and 176 were ultimately enrolled in the study.

At both sites those eligible but not enrolled did not differ on age, gender, or race [25].

1.3 Outline of thesis

This thesis uses the current study to ask: **is specialty probation effective in improving public safety outcomes?**

I will go through the following steps to address this question. First, I will tackle each of the steps of the causal inference roadmap proposed by Petersen & van der Laan [32]. The purpose of using this formal causal framework is to carefully draw conclusions that respect the limits of our knowledge by making explicit any assumptions imposed. Second, I present summary statistics and exploratory data analyses (EDA) to describe the sample, especially to highlight differences (or lack thereof) within covariates by probation type. Third, I look at associations between specialty probation and public safety outcomes using different estimators motivated by the average treatment effect. Fourth, I examine the association between specialty probation and time to re-arrest.

If the aforementioned questions are answered in favor of specialty probation, together these analyses will provide rigorous evidence supporting a practical reform program that improves general public safety. A program that, instead, only removes the "facility" part of "correctional facility."

Chapter 2

Causal Roadmap

2.1 Ideal experiment

In an ideal world, the effect of specialty versus traditional probation on public safety would be studied by way of an experiment, not an observational study. That is, all officers would be trained on specialty and traditional probation tasks and we would randomly assign the probationers to either specialty or traditional probation. Even more hypothetically, instead of randomization, all probationers would receive specialty probation and assess their outcomes, and then go back in time and have all probationers receive traditional probation and assess their outcomes. We could then follow the probationers for up to some fixed amount of time and observe whether they experienced the outcome of interest within that time window, and if so, when. In such an ideal experiment, we would ensure that all probationers were observed until they either experienced the event or for the duration of follow up, whichever happened first. This allows us to prevent any loss to follow up or database closure before that point.

In this alternative universe, the true causal effects of specialty probation on public safety outcomes could be easily assessed. We cannot go back in time (as far as we know), and probation assignment was not random. Therefore, we turn to the causal roadmap.

2.2 Causal questions

The purpose of using the roadmap as a formal causal framework (detailed in [32], [31], [20]) is to let the actual study answer the causal questions of interest as closely as possible. The roadmap allows us to recognize and explicitly state the limitations and assumptions of the current study that keep us from making causal conclusions that could only be made under the ideal experiment. Below, I follow the causal roadmap to answer the following causal questions:

1. What is the effect of specialty probation as compared to traditional probation on the

probability that probationers with mental illness commit a violent act within a year after their baseline interview?

2. What is the effect of specialty probation as compared to traditional probation on the probability that probationers with mental illness are re-arrested within two years after their baseline interview?

3. What is the effect of specialty probation as compared to traditional probation on the probability that probationers with mental illness are re-arrested within t days ($t = 180$, 360, 540, 720, 900, 1080) after their baseline interview, accounting for loss to follow up, for some probationers after 720 days?

The target population for these causal questions are adults (age 18-65), English-speaking probationers with mental illness and no intellectual disability who are on active probation supervision with at least one year remaining on term.

2.3 Step 1: Specify knowledge about the system to be studied using a causal model

Structural causal models (SCM) were used to describe the process that is posited to have given rise to the variables that were and were not observed [29]. The random variables in this SCM, which are called $\mathcal{M}^F$, follow the joint distribution $P_{U,X}$. The following are the variables included in the SCM.

Endogenous variables

Endogenous variables (denoted X) are defined as "factors that are meaningful for the scientific question or about which [there is] some scientific knowledge" [31]. X includes the baseline covariates (W), treatment (A), monitoring at time t (Δ) and outcomes (Y, defined below).

Covariates (W)

The initial data set contained hundreds of covariates from arrest reports, personality and clinical trait exams and various other surveys and interviews measured at baseline. Because assignment to specialty probation was perfectly confounded with location, all variables that might be confounders of assignment to specialty probation and violence/re-arrest were included in the causal model. Therefore, we included 21 core covariates in the multivariate variable W.

Variables predictive of treatment and outcome The model included probationers' demographic characteristics (e.g., age, gender, race, employment status, education level) because they tend to influence probation placement and the outcomes of interest. For example, females and White probationers are historically more likely to get assigned to specialty probation [28]. Certain demographic characteristics are associated with a higher rate of arrest (e.g., young adult, male) and violence (e.g., socio-economic status, education) [47, 53].

Because past criminal behavior could be predictive of both future criminal behavior and mental illness, we considered criminal history covariates such as number of lifetime arrests, most serious crime, index offense, recent violence, and length of time on probation from baseline. It was also necessary to adjust for these covariates because of the uneven distribution among the two populations of probationers (e.g., specialty probationers are likely to have been on probation prior to the baseline interview and the beginning of the study for a longer period of time than traditional probationers, which may decrease the probability of re-arrest/violence throughout the study period).

The model includes mental health status variables, since mental illness is what determines placement into specialty probation [46]. These variables potentially affect public safety outcomes, as well. Three questionnaires were used as measures of substance abuse, externalizing and other psychiatric symptoms. The Colorado Symptom Index (CSI) [41] is a self-report measure which is used to identify mental health status. The Global Assessment of Functioning (GAF) [5] score assesses how capable a subject is at functioning in daily life, due to psychological and social factors. In addition to these two variables, we included several clinical scores from the Personality Assessment Inventory (PAI) [17] in the model. PAI scores are self-report scores which measure clinical traits such as aggression, antisocial personality disorder and drug and alcohol abuse, which are predictive of both treatment and the outcomes [24]. The model included other clinical traits representative of externalizing disorders as theoretical predictors of both treatment and outcome, such as anxiety (physiological), paranoia, mania, and schizophrenia. Additionally, a variable describing the probationer's degree of child abuse seriousness was included, since it is a predictor of future mental illness [50].

For these covariates, missing values were imputed to the median for numeric covariates and re-sampled without replacement for categorical variables.

Missingness variables We also included 19 variables that were binary indicators of missingness for each of the covariates. That is, if the original variable had a missing value, we included an additional indicator variable as a covariate that indicated whether or not an observation was missing for that variable (similar to [14]). For the original covariates, missing values were imputed to the median for numeric covariates and re-sampled without replacement for categorical variables. In total, 41 covariates were included in the analyses.

Treatment (A)

The treatment of interest in this study is type of probation:

$$A = \begin{cases} 1, & \text{Specialty probation} \\ 0, & \text{Traditional probation} \end{cases} \tag{2.1}$$

Probationers were assigned to specialty probation if they had mental illness, and their officers were trained in mental health practices. Compared to traditional probationers, specialty probationers were assigned to smaller caseloads and established higher quality and more direct relationships with officers. Specialty officers used positive compliance strategies, whereas traditional officers relied more on sanction threats (see 5 core features above). More specialty probationers received mental health treatment (91% vs. 60%) and integrated dual diagnosis treatment (34% vs. 15%) than traditional probationers within one year, but they were equally likely to receive substance use treatment (28% vs. 31%) [42, 25]. An in-depth description of the implementation of the probations in the study can be found in Manchak et al. [25].

Outcomes (Y)

We consider two outcomes, an indicator of violence within 2 years and an indicator of re-arrest within t days, for $t = 180, 360, 540, 720, 900, 1080$, we define each below, and in subsequent sections use Y to refer to the general outcome.

Violence Violence was a binary variable measured one year after the probationer's baseline interview. It was defined as probationer involvement with physical battery, sexual assault, threat with a weapon in hand, or an assaultive act with a weapon [42]. The variable was operationalized using a combination of probationer and officer self-report of violence in the past 6 months, and official probation records of aggressive acts (see Steadman et. al. for details on operationalization) [49]. If any of these sources reported presence of probationer violence, the variable was coded as having occurred:

$$\text{Violence} = \begin{cases} 1, & \text{if probationer was violent within one year after baseline interview} \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

Re-arrest Probationers had different lengths of follow-up for re-arrest reports. Thus, several variables were created that indicated whether FBI reports showed the person was re-arrested while living in the community within t days after their baseline interview for $t = 180, 360, 540, 720, 900, 1080$ days. This excluded days incarcerated or hospitalized (i.e., time not in the community), as there was little or no probability of re-arrest in these settings. The maximum length of time was chosen based on the support available towards the end of the study; that is, enough probationers continued in the study after day 1080, both arrested and not.

$$\text{Re-arrest} = \begin{cases} 1, & \text{if probationer was re-arrested within } t \text{ days after baseline interview} \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

Missingness (Δ)

is an indicator of whether each Y was observed. Specifically:

$$\Delta = \begin{cases} 1, & \text{if } Y = 1 \text{ or if } Y = 0 \text{ and the probationer's follow up time is} \\ & \geq t \text{ for } t = 180, 360, 540, 720, 900, 1080 \\ 0, & \text{otherwise} \end{cases} \tag{2.4}$$

For the first 720 days all probationers were followed; in other words, $\Delta = 1$ for all probationers for the outcome of violence and for the outcomes of re-arrest by $t \leq 720$ days.

Exogenous variables (U)

Exogenous variables are defined as unmeasured variables (or errors) that affect what values the endogenous variables X take, representing all information not known. Because assignment to specialty probation was completely determined by location, there are many unmeasured factors specific to the cities that could affect the endogenous variables. Some possibilities for U variables include gang affiliation (U_W, U_Y, U_Δ), police practices (U_Y, U_Δ), probationers neurobiological factors (U_A, U_W), probationers religion/spiritual affiliation (U_W, U_Δ), community's mental illness beliefs/stigma (U_A), probationer's proximity to data collection site (U_Δ) and socio-cultural factors related to the city (U_W, U_A).

Structural equations

Below are the nodes of the causal model and the structural equations that relate the nodes. There are no exclusion restrictions (i.e., assume all ancestors of X_j affect X_j and thus do not exclude them from $Pa(X_j)$). There are no independence assumptions, meaning the multivariate error U influences all other nodes or assumptions about the functional forms f can take.

$$
\begin{aligned}
W = &\{\text{age, gender, race, index offense, number of lifetime arrests, CSI, GAF,} \\
&\text{most serious crime, employment, education, anxiety [physiological], paranoia,} \\
&\text{mania, schizophrenia, antisocial personality disorder, aggression, alcohol use} \\
&\text{drug use, recent violence, length of time on probation, child abuse seriousness,} \\
&\text{missingness variables}\} \\
A = &\ \{\text{probation type}\} \\
\Delta = &\ \{\text{indicator the otucome is observed}\} \\
Y = &\ \{\text{violence within 1 year, re-arrest within 2 years,} \\
&\text{re-arrest within } t \text{ days: } t = 180, 360, 540, 720, 900, 1080\}
\end{aligned}
$$

The structural equations that relate each of these variables are as follows:

SCM

$$
\begin{aligned}
W &= f_W(U_W) \\
A &= f_A(W, U_A) \\
\Delta &= f_\Delta(W, A, U_\Delta) \\
Y &= f_Y(W, A, \Delta, U_Y)
\end{aligned}
$$

Directed Acyclic Graphs (DAGs)

See Figure 2.1 for the causal graph that show the relationships between variables for the SCM.

2.4 Step 2: Specify the observed data and their link to the causal model

We assume the observed data, O, were generated by sampling n times from a data-generating system contained in (described by) the SCM $\mathcal{M}^F$. There are n independent, identically distributed copies of the random variable O. O contains the observed variables W, A, Y, Δ, and ΔY, which follow the joint distribution P_0 for each of the SCMs. This provides a link between the two SCMs $\mathcal{M}^F$ and their respective statistical models, $\mathcal{M}$.

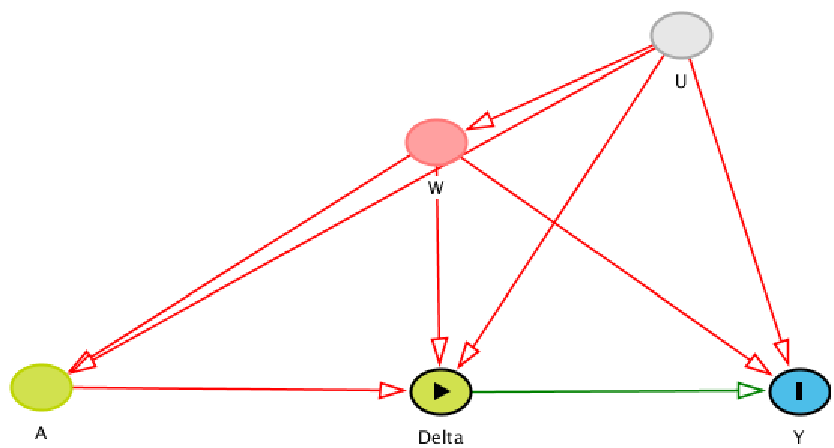

Figure 2.1: DAG visually representing relationships between variables in the SCM.

2.5 Step 3: Specify the target causal quantity

I will specify the target causal parameter (a function of the counterfactual distribution, $P_{U,X}$) and its interpretation for each of the causal questions outline above. Each of the counterfactual outcomes are derived as a solution to the functions f_Y on the post-intervention SCM by setting A equal to the intervention ($A = a$) and having no missing data/loss to follow-up ($\Delta = 1$). The causal parameters are functions of $P_{U,X}$ (denoted ψ^F) and are generally summary measures of the counterfactual outcomes of interest.

Causal questions based on SCM

1. What is the effect of specialty probation as compared to traditional probation on the probability that probationers with mental illness commit a violent act within a year after their baseline interview?

$$\psi^F(P_{U,X}) = E_{U,X}[Y_{a=1,\Delta=1}] - E_{U,X}[Y_{a=0,\Delta=1}]$$

This is the difference in the counterfactual probability of violence during the year following the probationer's baseline interview if all probationers participated in specialty probation ($A = 1$) minus the counterfactual probability of violence if all probationers participated in traditional probation ($A = 0$), with no loss to follow-up. The counterfactual $Y_{a,\Delta=1}$ is the one-year violence status for a probationer if, possibly contrary to fact, the probationer received probation type $A = a$ and his/her violence status was observed ($\Delta = 1$).

2. What is the effect of specialty probation as compared to traditional probation on the probability that probationers with mental illness are re-arrested within two years after their baseline interview?

$$\psi^F(P_{U,X}) = E_{U,X}[Y_{a=1,\Delta=1}] - E_{U,X}[Y_{a=0,\Delta=1}]$$

This is the difference in the counterfactual probability of re-arrest during the two years following the probationer's baseline interview if all probationers participated in specialty probation ($A = 1$) minus the counterfactual probability of re-arrest if all probationers participated in traditional probation ($A = 0$), with no loss to follow-up. The counterfactual $Y_{a,\Delta=1}$ is the two-year re-arrest status for a probationer if, possibly contrary to fact, the probationer received probation type $A = a$ and his/her re-arrest status was observed ($\Delta = 1$).

3. What is the effect of specialty probation as compared to traditional probation on the probability that probationers with mental illness are re-arrested within t days after their baseline interview, under a hypothetical intervention to ensure that the outcome is observed.

$$\psi^F(P_{U,X}) = E_{U,X}[Y_{a=1,\Delta=1}] - E_{U,X}[Y_{a=0,\Delta=1}]$$

This is the difference in the counterfactual probability of re-arrest if all probationers participated in specialty probation ($A = 1$) minus the counterfactual probability of re-arrest if all probationers participated in traditional probation ($A = 0$), with no loss to follow-up. The counterfactual $Y_{a,\Delta=1}$ is the t-day re-arrest status for a probationer if, possibly contrary to fact, the probationer received probation type $A = a$ and his/her re-arrest status was observed ($\Delta = 1$), for $t = 180, 360, 540, 720, 900, 1080$ days.

2.6 Steps 4 & 5: Assess identifiability and commit to a statistical estimand

There are two assumptions necessary for identification so that the causal parameter (a function of our counterfactual distribution) is equal to the statistical estimand (a function of our observed data distribution): 1) positivity assumption and 2) randomization assumption.

Positivity assumption

The relevant positivity assumptions are:

$$\min_{a\in\{0,1\}} P_0(A=a|W=w)>0, \text{ for all } w \text{ for which } P_0(W=w)>0 \tag{2.5}$$

$$P_0(\Delta=1|W=w,A=a)>0, \text{ for all } w \text{ for which } P_0(W=w)>0 \tag{2.6}$$

Equation 2.5 says that for every stratum of baseline covariates, there must exist a positive probability of receiving either probation type. Equation 2.6 states that for every treatment covariate combination, there must exist a non-zero probability of observing a probationer's outcome. Said another way, there are no groups or types of probationers who are systematically censored from the study.

Randomization assumption

When examining the total effect of A on Y, the following independence assumptions must be met:

$$Y_a \perp A|W \tag{2.7}$$

$$Y_{a,\Delta} \perp A,\Delta|W \tag{2.8}$$

In the current SCM, there are several unblocked backdoor pathways that confound the effect of A on Y that is of relevant interest. Thus, identifiability is not achieved. Conditioning on W, the backdoor criteria is satisfied by applying the following convenience assumptions: $U_A \perp U_Y$, $U_A \perp U_W$, $U_\Delta \perp U_Y$, and $U_W \perp U_\Delta$ (see Figure 2.2).

Identification

Under the aforementioned assumptions, the total effects can be identified by the following g-computation formula:

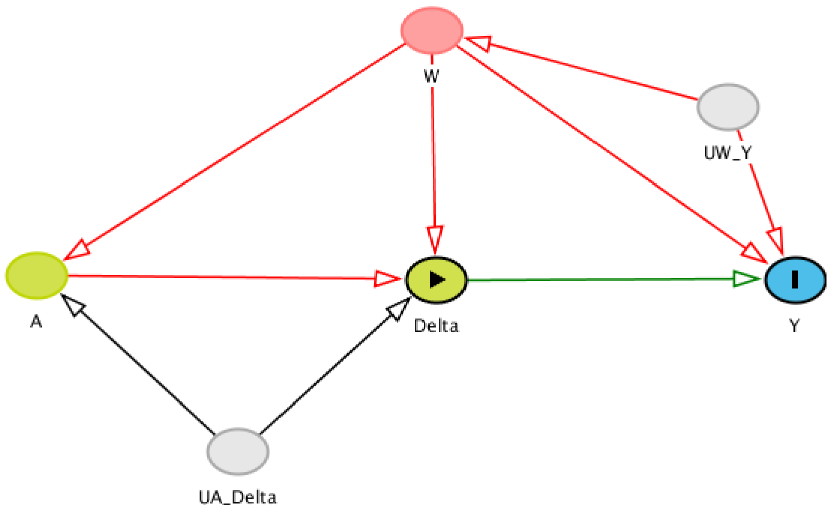

Figure 2.2: $\mathcal{M}^{F*}$, the modified SCM under convenience assumptions, to identify total effects, for the SCM.

$$\begin{aligned}\psi(P_0) &= E_W[E[Y|\Delta = 1, A = 1, W] - E[Y|\Delta = 1, A = 0, W]] \\ &= \sum_W (E[Y|\Delta = 1, A = 1, W = w] - E[Y|\Delta = 1, A = 0, W = w])P(W = w)\end{aligned}$$

2.7 Steps 6 & 7: Estimate and interpret results

The focus of the remainder of this thesis is to estimate the quantities identified above.

Targeted Maximum Likelihood Estimation (TMLE)

TMLE was used to derive the estimates inspired by the aforementioned causal parameters. TMLE is a semi-parametric estimator of treatment effects that, for estimation of the average treatment effect, both incorporates the outcome regression (i.e., the relationship between an outcome versus its treatment and covariates) and the treatment mechanism (i.e., the probability of being assigned to specialty probation given the covariates).

TMLE estimates are consistent (i.e., converge to true parameter value in probability as sample size grows to infinity) if either the outcome regression or the treatment mechanism are consistently estimated; if both are consistently estimated, TMLE has the smallest asymptotic variance among reasonable estimators [20]. This estimation technique is especially suitable for the current study: the estimator does not solely rely on a correctly specified model of treatment assignment process, which is unknown, due to the observational nature of this study [32].

Note: when estimating the treatment mechanism, the bounds on $P_0(A|W)$ and $P_0(\Delta|A,W)$ were set to 0.025 and 0.975 for the lower and upper bounds, respectively.

Influence-curved based confidence intervals were used for inference on the estimates. Influence curves are useful because they give us a measure of the influence of observations on the estimator (i.e., a measure of robustness), as well as an estimate of the variance, allowing us to construct confidence intervals [20].

SuperLearner

Within the TMLE estimation process, $E_0[Y|A,W]$ (the outcome regression, or probability of violence/re-arrest given treatment and background covariates), $P_0(A|W)$ (treatment mechanism, i.e., propensity score, or probability of being assigned to specialty/traditional probation given background covariates), and $P_0(\Delta = 1|A,W)$ (missingness mechanism, or probability of missingness given the treatment and covariates) were estimated using SuperLearner, a data-adaptive method that combines a library of machine learning algorithms and parametric models [19]. Oftentimes prediction estimates are derived by a single parametric model (e.g., gaussian linear regression), which assumes a specific shape and distribution (e.g., linear forms or normality) that, in reality, we do not know to be true, producing biased results (regardless of sample size). SuperLearner does not rely on any model structures or assumptions *a priori*. When estimating via data-adaptive algorithms the only assumptions imposed are that the outcome is a function of the treatment and covariates (or for the treatment mechanism, assume the treatment is a function of baseline covariates). In other words, we assume a non-parametric or large semi-parametric statistical model [20].

This algorithm uses a library of candidate machine learning algorithms and parametric models to find the best weighted combination (i.e., weights that minimize risk over all possible weighted combinations) of individual candidate algorithms that estimate the outcome regression and treatment mechanism. Thus, the advantage of this data-adaptive algorithm is that it performs as well or better asymptotically than any individual algorithm within the library, if the library does not contain a correctly specified parametric model. Additionally, an extra cross-validation step prevents over-fitting of the data, allowing for a more flexible selection of confounder variables [20].

The following algorithms (and their abbreviations) were used in the SuperLearner library to estimate the treatment mechanism, outcome regression, and missingness mechanism: generalized linear models (glm), bayesian linear models (bayesglm), multivariate adaptive regression splines (earth), classification and regression training (caret), generalized boosted regression models (gbm), lasso and elastic-net regularized generalized linear models (glmnet) with tuning parameters (alpha) set to 0, 0.25, 0.75, 0.5, 0.1, improved predictors (ipred), k-nearest neighbors (knn), mean, neural networks (nnet), multivariate adaptive polynomial spline regression (polymars), pruned recursive partitioning (rpartPrune), backward and forward stepwise AIC selection (stepAIC).

Kaplan-Meier Estimator

The Kaplan-Meier (KM) estimator was used as an unadjusted (i.e., no controlling of baseline covariates) measure of probability of "survival" in the community (i.e., no re-arrest) for each probation type. The KM estimator is non-parametric, making it robust to varying interval durations between events with long duration times (as opposed to, for example, estimation using the average days to re-arrest) [37]. Days until the event or study dropout, or `time`, whether or not the probationer was re-arrested, or `arrest`, and probation type were the three variables used to estimate the survival function. Specifically, for each probation type, the procedure for generating this estimates was as follows:

1. From the smallest to largest `time`, count the number of probationers who have not yet been re-arrest, or who are "at risk" in that time interval. For example, for traditional probationers, the smallest `time` is 1 day, and for that first time point, no one has been re-arrested yet, so the number at risk is 173 (i.e., the total number of probationers in traditional probation). The next time point is 5. We remove the probationer from `time` = 1, so now the number at risk is 172. Repeat this for every `time`. This is the *denominator* for the interval.

2. From the smallest to largest `time`, count the number of probationers who survive after the event (this is just one less than the number at risk). This is the *numerator* for the interval.

3. For every timepoint in which there was a re-arrest, calculate the interval survival rate after the event: divide the *numerator* by the *denominator*. For example, at `time` = 1, there was a re-arrest, so the survival probability for that interval is $172/173 = 0.994$.

4. Calculate the cumulative survival rate by multiplying the interval survival probabilities up until that `time`.

By calculating the cumulative probability of no re-arrest at each timepoint in which there was a re-arrest event we can generate an estimate of the survival (or no re-arrest) curve.

Software

Software: Analyses were performed using R version 3.3.0. TMLE analyses were done using the “tmle” package [13]. Estimation of the outcome regression/treatment mechanism were done using “SuperLearner” [11].

Chapter 3

Descriptive Analyses

3.1 Methods

Descriptive analyses (i.e., non-inference based statistics) were done to learn about the sample. Specifically, cross-tabulations and percentiles were used to compare raw (unadjusted) characteristics between the two exposure groups before missing value imputation.

To provide a visual display of the distribution of the covariates, stripcharts are (separately) shown for categorical and continuous variables [9]. Histograms showed the distribution of missing variables for each covariate set by probation type. In this and all proceeding exploratory data analysis (EDA) graphics, specialty probation is depicted in blue, whereas traditional probation is depicted in red.

In this section a variety of clustering techniques are presented that examine the distribution of the observed covariates. Given the sampling scheme, it is clear that some degree of confounding by location of residence will yield covariates that are not uniformly distributed between probation type, as we would see in a randomized controlled trial (e.g., specialty probationers have more of some type of covariate, say, females, compared to traditional probationers). To examine visually the degree to which variables differ by probation type, following are Principal Components Analysis (PCA) scatterplots, K-means clustering, and dendrograms.

3.2 Results

Descriptive statistics

The full dataset consisted of 359 observations (probationers): 183 received specialty probation and 176 received traditional probation. Table 3.1 shows summary statistics for each covariate broken down by each probation type. To assess which characteristics significantly differed between probation groups, two-sample t-tests were run on continuous variables to assess differences in means and χ^2 tests were run on categorical variables to assess independence between probation type and levels of that variable. Significance was assessed at the $\alpha = 0.05$ level. See Table 3.1 for reference on which variables significantly differed by group.

Table 3.1: Comparing Specialty and Traditional Probationer Characteristics

		Specialty ($N = 183$)	**Traditional** ($N = 176$)
Baseline Covariates		% or mean (SD)	% or mean (SD)
Demographics/SES			
Age (years)		36.13 (10.16)	37.62 (10.96)
Male		54.10%	60%
White (vs. non-White)		37.70%	38.60%
Employment status			
	Full-time	12.70%	14.90%
	Part-time	15.50%	12.10%
	Unemployed	71.80%	73.00%
Educational attainment			
	One year or less of college	84.70%	76.60%
	Over one year college-BS/BA	14.20%	21.70%
	Some graduate post-graduate	1.10%	1.70%
Criminal/Childhood Abuse History			
Index offense ***			
	Person arrest	30.20%	41.30%
	Property arrest	32.80%	19.60%
	Drug arrest	21.00%	36.70%
	Minor/other arrest	16.40%	2.50%
Number of lifetime arrests *			
	One time	13.90%	5.10%
	Two times	11.10%	12.00%
	$\geq$ 3 times	75.00%	82.90%
Most serious crime ***			
	Person	41.10%	66.10%
	Property	24.40%	12.10%
	Drug	28.30%	20.70%
	Minor	6.10%	1.10%
Violence, prior 6 months		38.50%	30.70%
Time on probation (months) **		15.27 (14.86)	11.35 (9.98)
Child abuse seriousness ***			
	None	17.50%	34.10%
	Bare hand only (no physical injury)	1.60%	4.00%
	With an object (no physical injury)	62.30%	42.00%

Continuation of Table 3.1			
		Specialty ($N = 183$)	**Traditional** ($N = 176$)
	Resulting in physical injury	18.60%	19.90%
Symptoms			
PAI subscales:			
	Anxiety ***	37.24 (13.47)	29.54 (12.90)
	Paranoia	33.91 (9.39)	33.19 (11.58)
	Mania	32.61 (11.55)	32.20 (11.35)
	Schizophrenia **	30.60 (12.43)	26.27 (12.41)
	Antisocial	26.53 (10.75)	26.93 (11.00)
	Aggression	24.16 (11.04)	23.64 (10.37)
	Alcohol	9.30 (8.11)	10.66 (8.24)
	Drug	14.08 (8.18)	15.23 (8.47)
Psychiatric symptoms (CSI Total) **		44.94 (12.14)	49.25 (12.75)
Psychiatric functioning (GAF) ***		45.26 (11.97)	54.91 (15.14)

*$p < .05$, **$p < .01$, ***$p < .001$

Distribution of covariates

Figure 3.1 and Figure 3.2 are stripcharts presenting the distribution of the data, at an individual level, for continuous and categorical (including binary) covariates, respectively, after scaling and centering. Dark regions on the stripchart correspond to values of high density and light regions correspond to values of low density, for a given covariate. On the whole, the distribution (e.g., the center and spread) of the covariates are relatively similar.

Distribution of missing values for covariates

The average number of missing observations across all covariates was 9.63 observations per variable, or 2.68% missing observations per variable. Figure 3.3 displays the per-variable frequency of missingness by treatment. There was a relatively large amount of missingness for traditional probationers (i.e., Los Angeles residents) with respect to length of probation ($N = 10$, or 5.68% of traditional probationers) and index offense ($N = 18$, or 10.23% of traditional probationers). For specialty probationers (i.e., Dallas residents), there was a high amount missingness with respect to index offense ($N = 67$, or 36.61% of specialty probationers) and GAF score ($N = 8$, 4.37% of specialty probationers). Despite this, there was a high degree of missingness concordance for most variables across both cities of residence, and less than 20% missingness was found for all but the four aforementioned variables.

Clustering results

PCA

Figure 3.4 depicts a scatterplot of the 359 probationers arranged along the first 2 principle components, which explain 29.38% of the covariates' variability (see Figure 3.5). The range and spread along both axes for the blue points (specialty probation) is about the same as that for the red (traditional probation) points. The geometric center of the red looks slightly higher than the blue, but on the whole, the point pattern of the two colors is not all that different. These two lines of evidence suggest that probation type/city of residence is not a substantial driver of the variability seen in the non-treatment explanatory variables.

K-Means

Figure 3.6 displays the data using k-means clustering with k=2, 3, and 4 clusters. The k=2 plot shows the first cluster (the red cluster) with slightly more blue circles than red circles, and the second cluster with slightly more red triangles than blue triangles. However, these differences are quite small, leading us to believe that there is substantial heterogeneity by city of residence in each k-means cluster. Similarly, the k=3 and k=4 cluster plots also show substantial heterogeneity in each of their respective clusters. Table 3.2, Table 3.3, and Table 3.4 each summarize the number of individuals by city of residence within each cluster.

	Traditional	Specialty	Proportion of Homogeneity
Cluster 1	108.00	77.00	0.58
Cluster 2	75.00	99.00	0.57

Table 3.2: Composition of the clusters using k-means with 2 means. The first two columns displays the number of individuals from Los Angeles and Dallas in each cluster. The third column displays the proportion of the cluster that is comprised of individuals from the dominant group.

	Los Angeles	Dallas
Cluster 1	61	46
Cluster 2	58	95
Cluster 3	64	35

Table 3.3: Composition of the clusters using k-means with 3 means. The columns display the number of individuals from Los Angeles and Dallas in each cluster.

Agglomerative Hierarchical Clustering

Figure 3.7 gives an agglomerative hierarchical clustering of the data by probation type. This dendrogram was produced under complete linkage to avoid the chaining phenomenon

	Los Angeles	Dallas
Cluster 1	38	24
Cluster 2	49	38
Cluster 3	41	78
Cluster 4	55	36

Table 3.4: Composition of the clusters using k-means with 4 means. The columns display the number of individuals from Los Angeles and Dallas in each cluster.

sometimes seen in dendrograms drawn under single linkage. Again, there is a lot of heterogeneity with respect to city of residence across the branches of the dendrogram. The initial branch in the dendrogram does not cleanly separate study subjects by probation type. Table 3.5 shows the proportion of subjects in each city of residence broken down by the initial branching event of the dendrogram, and a 39/61 split is observed in the table.

	Traditional	Specialty
Top Branch	0.48	0.52
Bottom Branch	0.50	0.50

Table 3.5: Proportion of subjects by city of residence in coarsest level of hierarchical branching.

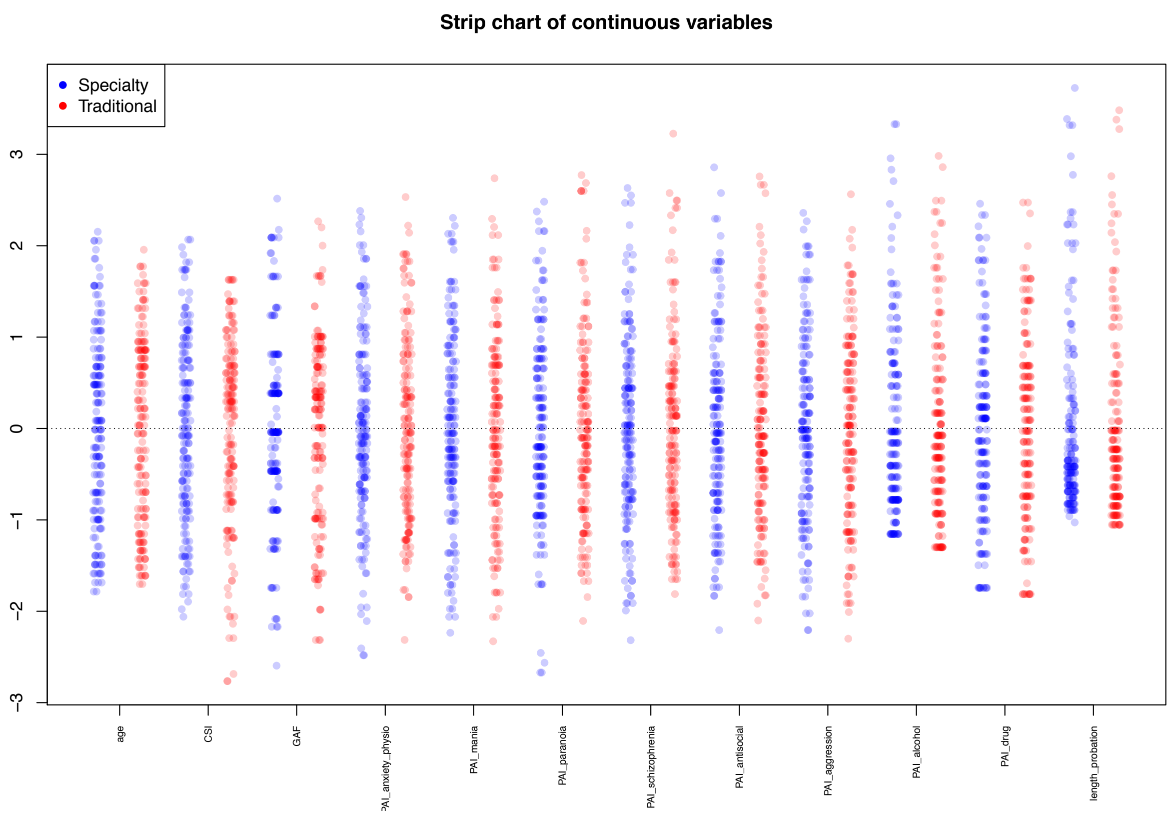

Figure 3.1: Strip chart showing side-by-side distributions of the continuous covariates for each city/probation type. Higher color density translates to a greater number of participants with a similar value for a given variable.

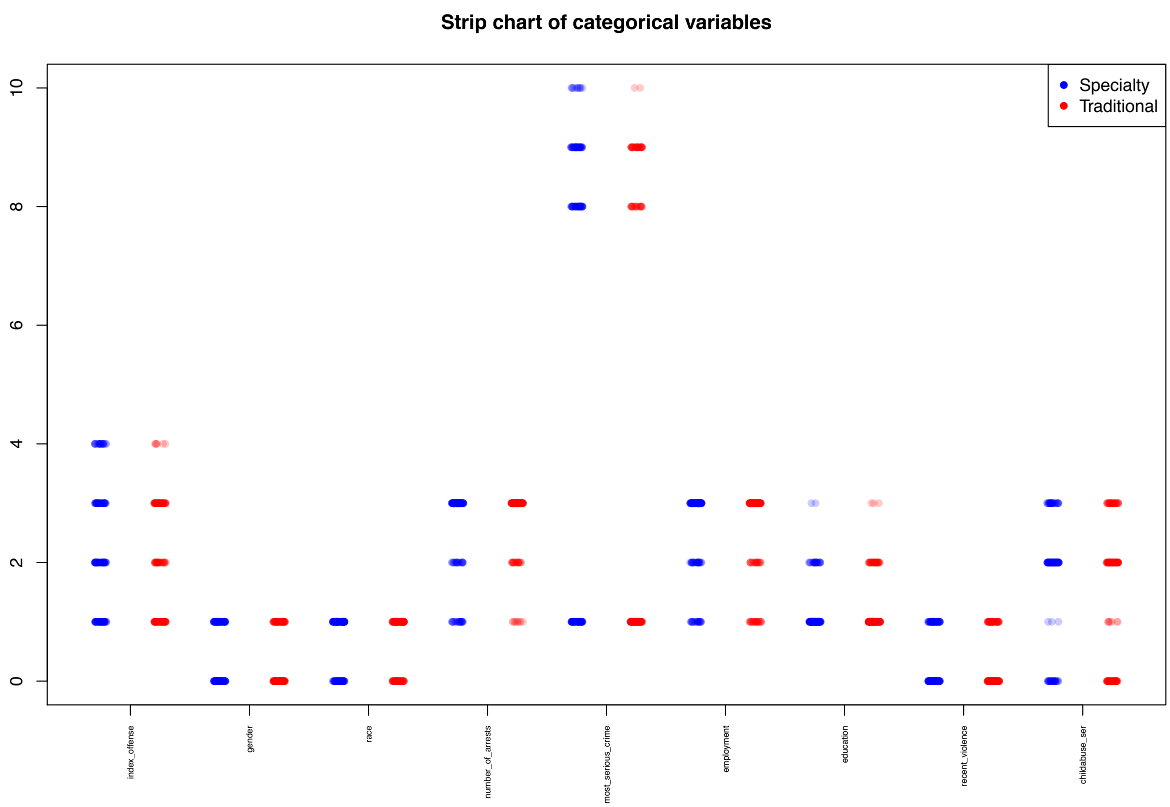

Figure 3.2: Strip chart showing side-by-side distributions of the categorical covariates for each city/probation type. Higher color density translates to a greater number of participants with a similar value for a given variable.

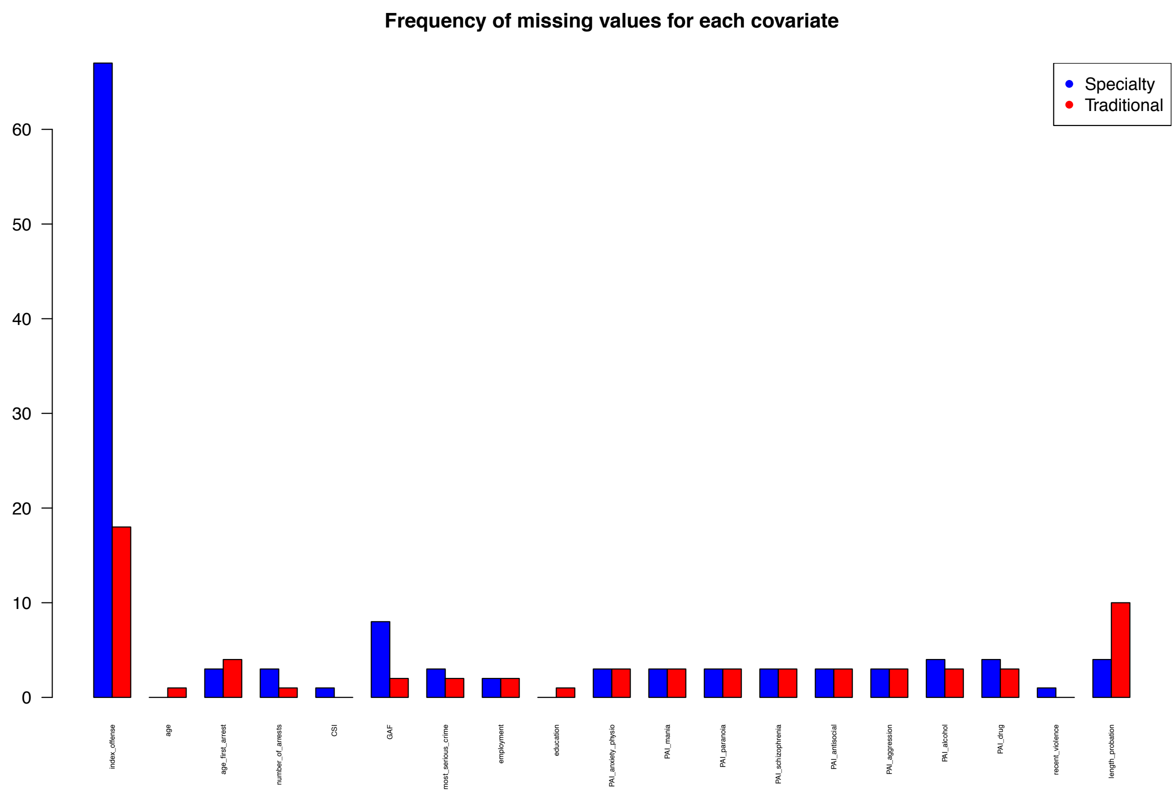

Figure 3.3: Bar plot with proportion of missing data values for each covariate for each city of residence/probation type.

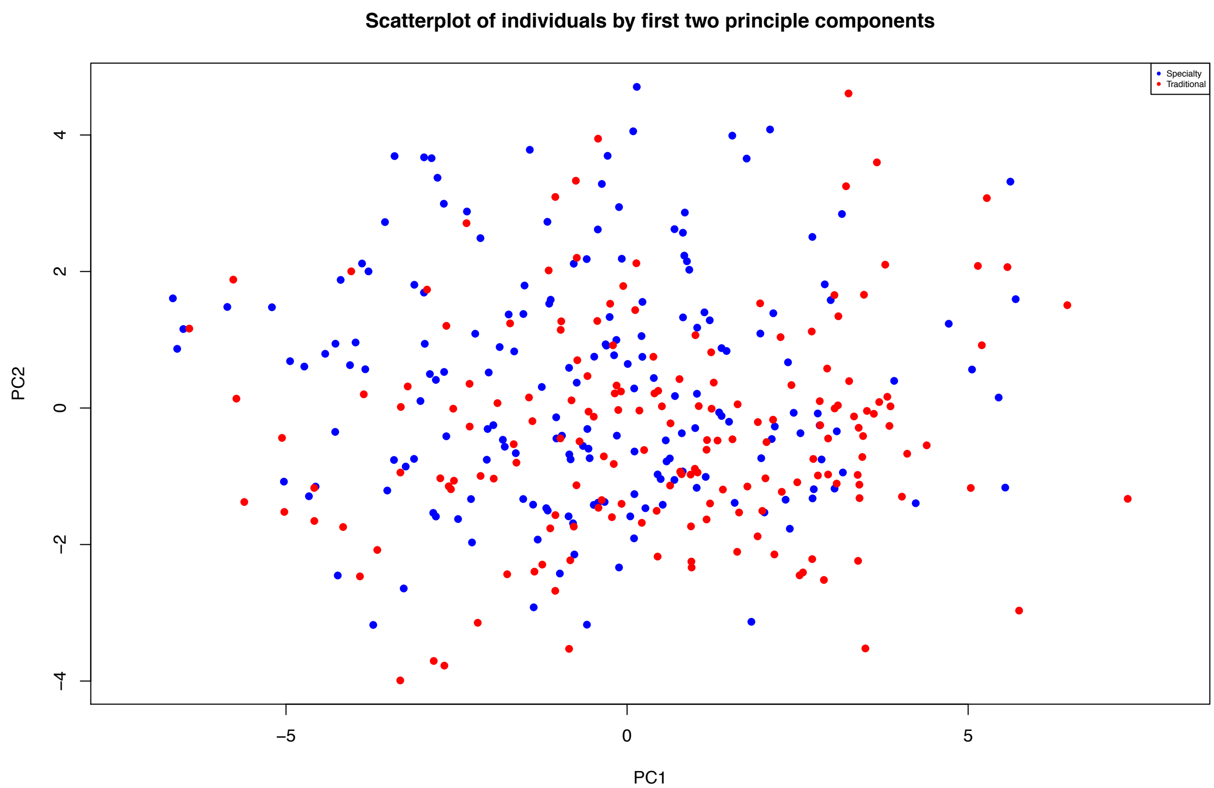

Figure 3.4: A plot of the first two principle components as determined by PCA on confounders (W).

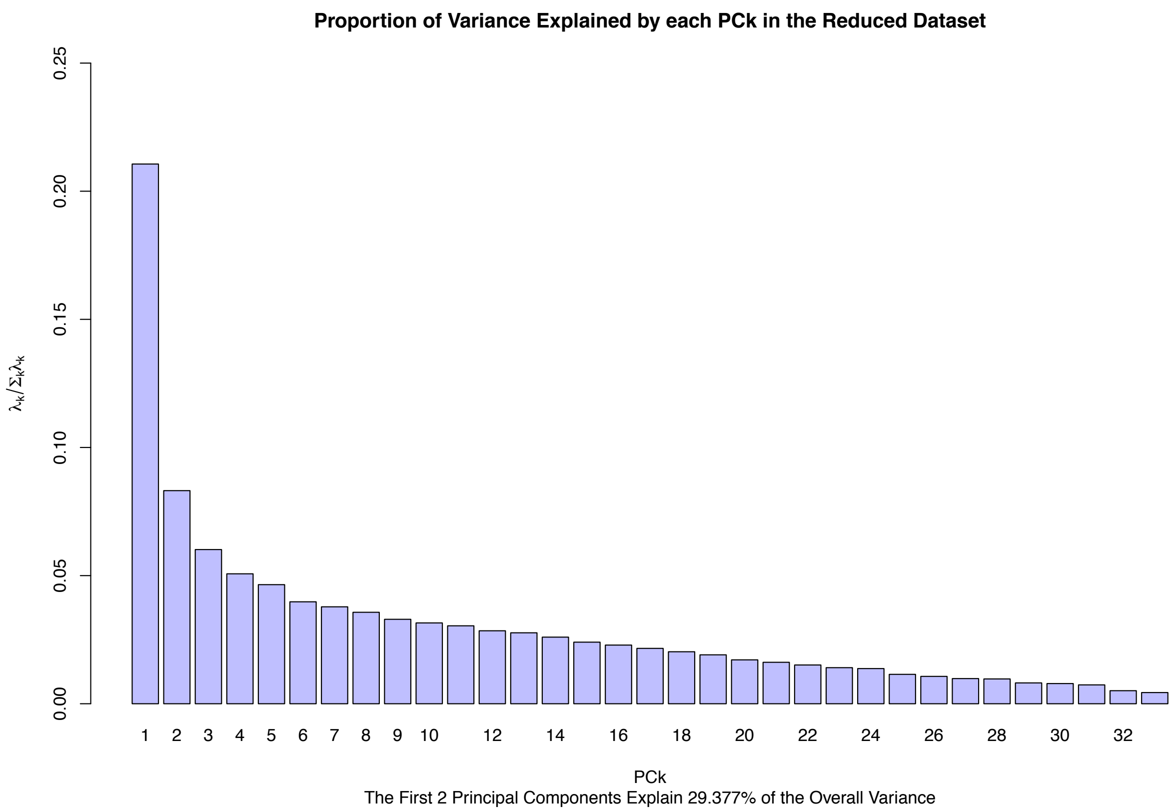

Figure 3.5: Proportion of variance explained by each of the principal components as determined by PCA on confounders (W)

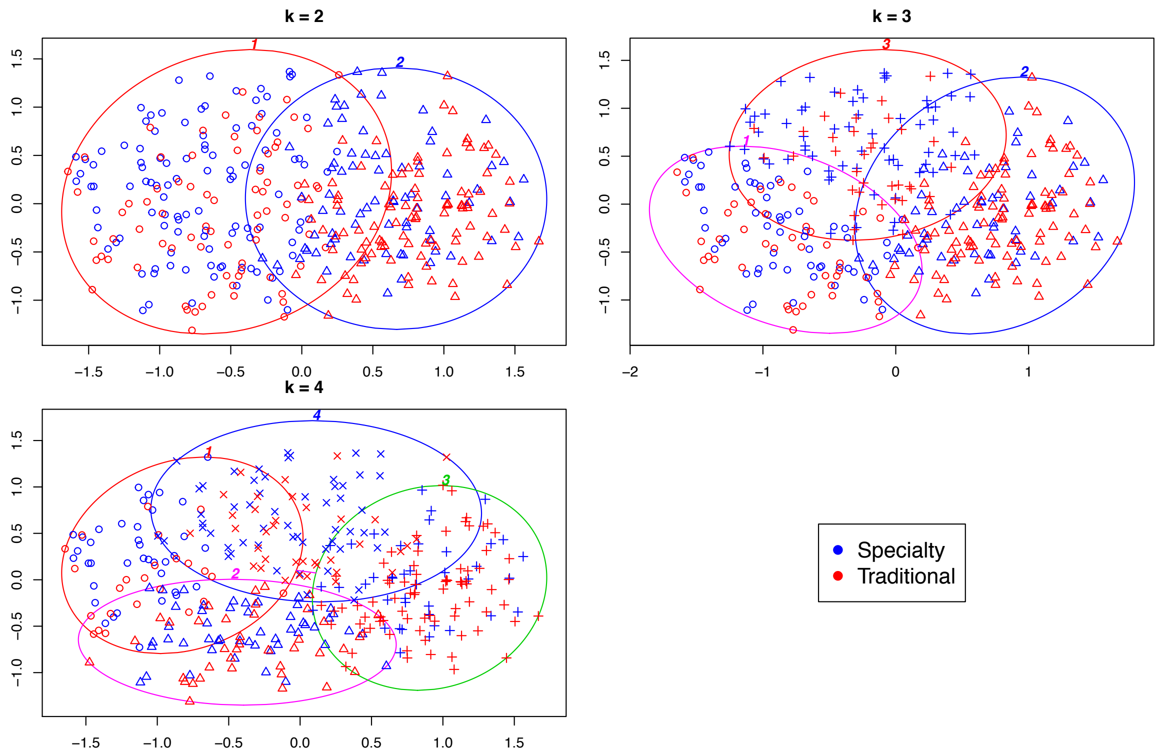

Figure 3.6: Plots corresponding to k-means clustering when k = 2 (top-left), k = 3 (top-right), k = 4 (bottom-left). The 1-correlation distance matrix was used to produce k-means results.

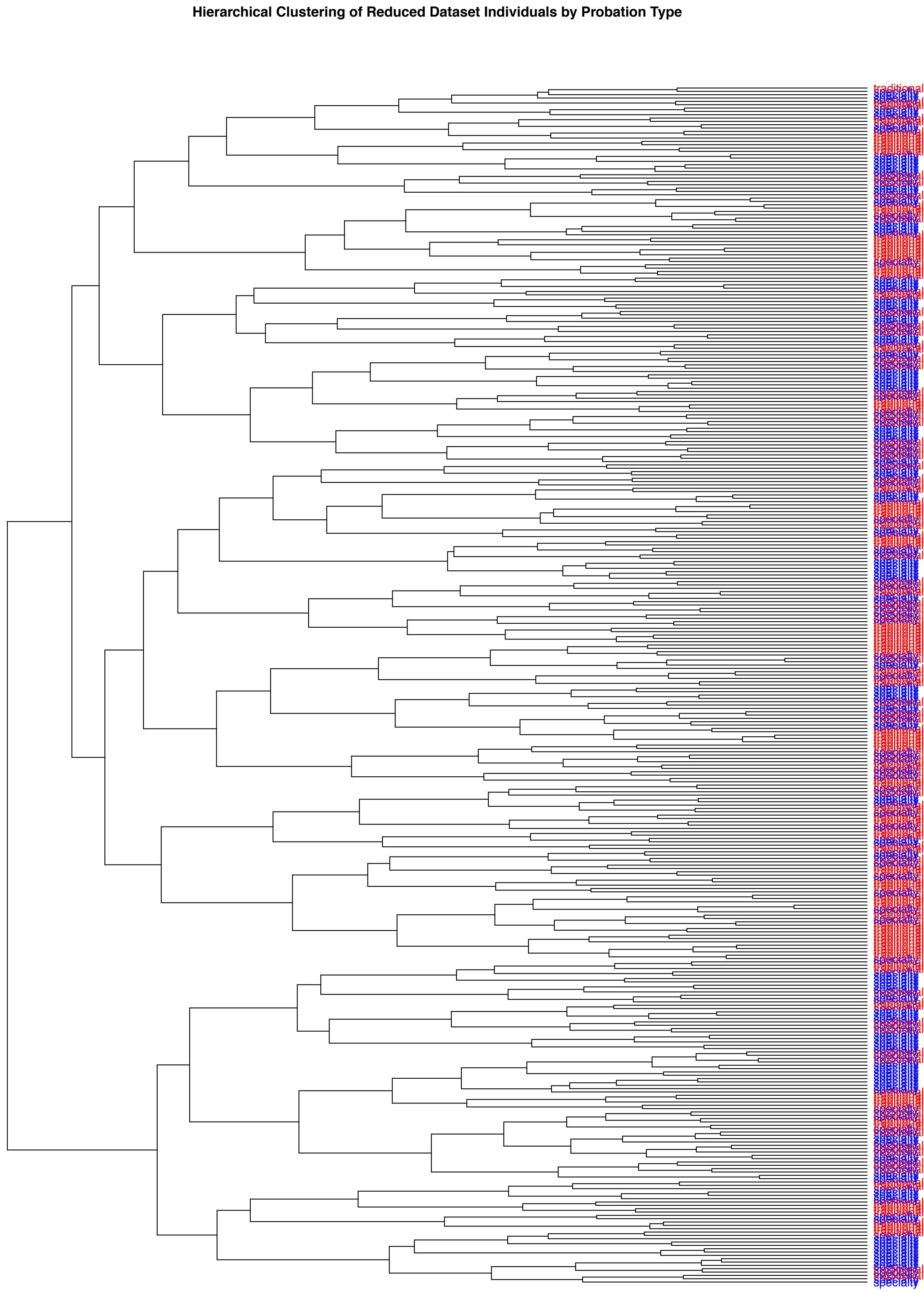

Figure 3.7: Hierarchical clustering of individuals by city/probation type.

Chapter 4

Estimation of Causal Parameters

4.1 Diagnostics and performance

Algorithm performance

	Coef.g	Risk.g	Coef.Qarrest	Risk.Qarrest	Coef.Qvio	Risk.Qvio
glm	0.00	0.20	0.00	0.25	0.25	0.21
bayesglm	0.25	0.19	0.14	0.24	0.00	0.20
earth	0.21	0.21	0.02	0.29	0.02	0.25
caret	0.08	0.19	0.00	0.24	0.00	0.20
gbm	0.27	0.18	0.12	0.23	0.06	0.20
glmnet.0	0.00	0.19	0.00	0.23	0.00	0.20
glmnet.0.25	0.00	0.19	0.00	0.23	0.00	0.19
glmnet.0.75	0.00	0.19	0.00	0.23	0.45	0.19
glmnet.0.5	0.00	0.19	0.00	0.23	0.00	0.19
glmnet.1	0.00	0.19	0.42	0.23	0.00	0.19
ipredbagg	0.00	0.19	0.00	0.24	0.00	0.20
knn	0.00	0.22	0.15	0.25	0.00	0.25
mean	0.00	0.25	0.00	0.25	0.00	0.23
nnet	0.00	0.24	0.00	0.27	0.00	0.25
polymars	0.03	0.20	0.00	0.24	0.00	0.20
rpartPrune	0.16	0.20	0.00	0.25	0.01	0.23
stepAIC	0.00	0.21	0.15	0.23	0.20	0.20

Table 4.1: Algorithm performance as measured by coefficients (Coef) and risk (Risk) for SuperLearner estimates on $E[Y|A,W]$ (Q) and $P(A|W)$ (g), where Y = violence within one year after baseline (vio) and re-arrest within 2 years after baseline (arrest). For example, the last column, "Risk.Qvio" shows the risk for each SuperLearner algorithm when estimating the outcome regression using violence as the outcome.

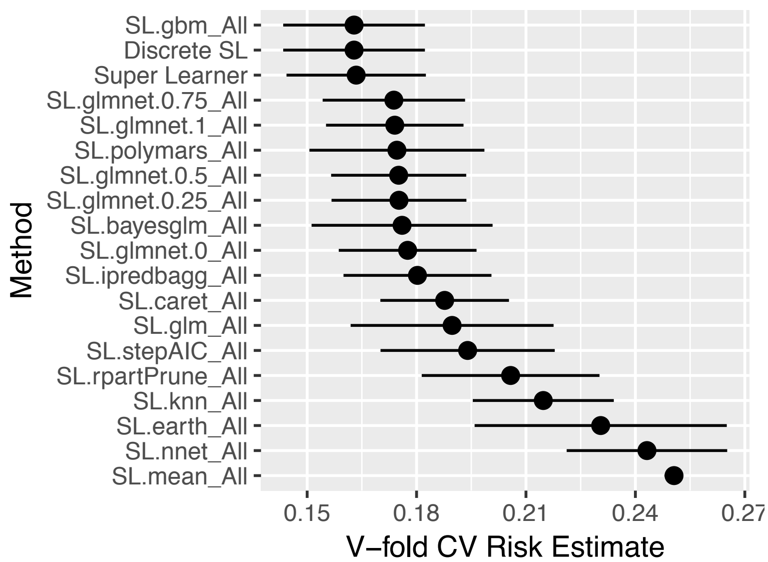

Figure 4.1: SL performance on estimation of treatment mechanism.

	Q180	Q360	Q540	Q720	Q900	gDelta900	Q1080	gDelta1080
gam	0.01	0.42	0.08	0.00	0.28	0.00	0.02	0.00
glm	0.29	0.00	0.00	0.20	0.15	0.00	0.00	0.03
glmnet	0.31	0.36	0.45	0.05	0.00	1.00	0.56	0.00
ipredbagg	0.00	0.00	0.22	0.61	0.19	0.00	0.25	0.01
mean	0.31	0.16	0.00	0.14	0.13	0.00	0.00	0.96
rpartPrune	0.00	0.00	0.00	0.00	0.26	0.00	0.00	0.00
step	0.07	0.07	0.25	0.00	0.00	0.00	0.17	0.00

Table 4.2: Algorithm performance as measured by coefficients for SuperLearner estimates on $E[Y|A,W]$ (Qt) and $P(\Delta = 1, |A, W)$ (gDelta$t$), where $Y =$ re-arrest within t days after baseline for $t = 180, 360, 540, 720, 900, 1080$. For example, the last column, "gDelta1080" shows the coefficient for each SuperLearner algorithm when estimating the missingness mechanism for $t = 1080$.

SuperLearner coefficients are assigned to each algorithm to let us know how much "weight" that estimator had in estimating the regression of interest. Risk (in this case, mean squared error) is a measure of the algorithm's accuracy.

As shown in Table 4.1, the highest weighted algorithms chosen by SuperLearner for estimating the fixed timepoint outcome regressions and the propensity scores included: Bayesian GLM, Multivariate Adaptive Regression Splines (earth), Generalized Boosted Regression Modeling (gbm), regularized regression with various tuning parameters (glmnet), and stepwise model selection by AIC. These algorithms also have lower risk estimates (of note, low coefficients do not necessarily translate to low risk estimates).

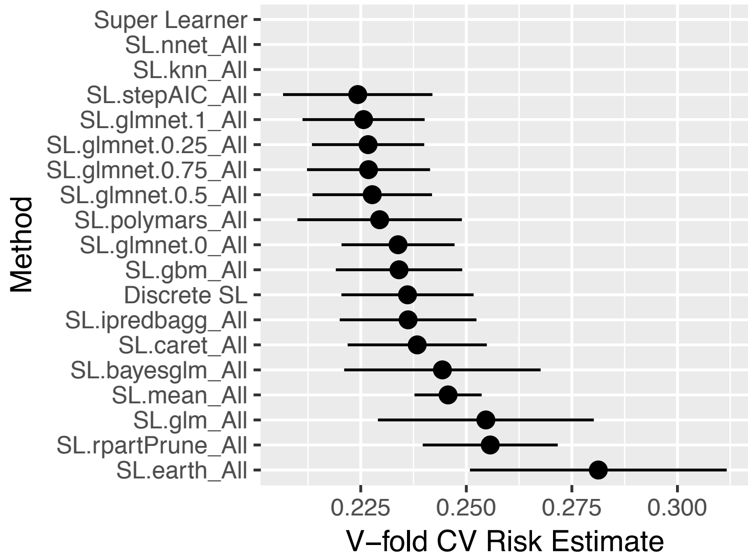

Figure 4.2: SL performance on estimation of outcome regression mechanism - re-arrest after 2 years.

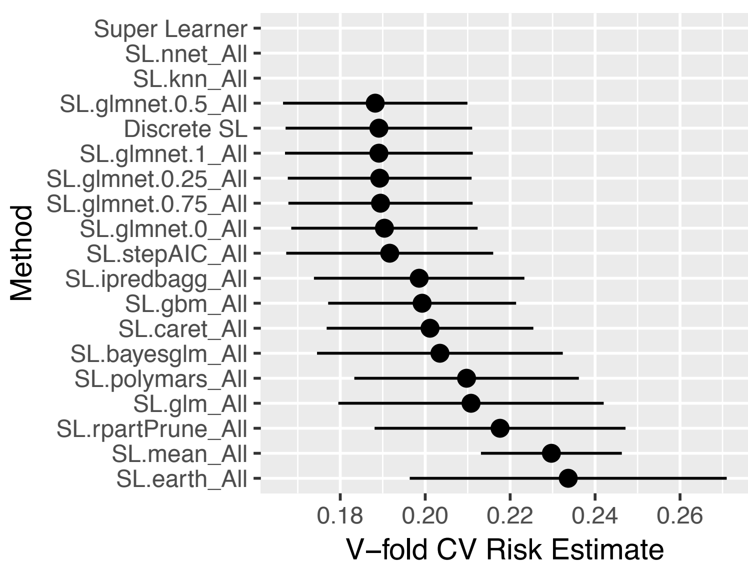

Figure 4.3: SL performance on estimation of treatment mechanism - violence after 1 year.

Table 4.2 shows that the algorithms contributing most to the estimates of outcome regression for various t included: regularized regression, the mean, generalized additive models, and bagging classification/regression/survival trees. Estimation of the missignmenss mechanism heavily relied on the mean as its primary algorithm.

SuperLearner performance

Based on 4.1, we see that SuperLearner did not have the "best" performance in estimating the treatment mechanism. In this case, "best" means lowest estimated risk, or mean squared error. SuperLearner is performing almost the same as the "best" algrithm, which is gbm.

Distribution of estimated treatment mechanism

Part of the TMLE estimation process of the average treatment effect requires an estimator of the treatment mechanism, or $P_n(A|W)$. If there is lack of data support for certain treatment/covariate combinations, these probabilities will be either close to 1 or 0. This is also called the practical positivity violation [31]. If positivity is violated, consistent estimation of TMLE ends up exclusively relying on consistent estimation of the outcome regression, removing its beneficial double robust properties [34]. Instead of looking directly at the distribution of $P_n(A|W)$, I will look at the inverse propensity score weights, or $1/P_n(A|W)$ (thus, if there are large weights, this should be cause for concern).

The minimum of the weights was 1.026 and the maximum was 5.328 (M = 1.457, SD = 0.481). Based on Figure 4.4, we see that most of the weights are distributed between 1 and 2. According to Cole and Hernan, positivity should be a concern if the mean of the weights is very far from one or there are extreme weights [7]. Neither seems to be the case here, so we will assume nonpositivity.

Distribution of estimated missingness mechanism

For TMLE estimates when missingness on the outcome exists (i.e., when looking at time to re-arrest), it is necessary to examine practical positivity violations on $P_n(\Delta|A, W)$. Thus, we will address the distribution of $1/P_n(\Delta = 1|A = 0, W)$ and $1/P_n(\Delta = 1|A = 1, W)$. We will only do this for $t = 1080$, as there is no missingness for $t = 180, 360, 540$ and 720, and the SuperLearner estimate for $t = 900$ only used the mean as its algorithm, generating one estimated weight for all probationers.

Figure 4.5 and 4.6 both the distribution of the weights for $1/P_n(\Delta = 1|A = a, W)$ for $a \in \{0, 1\}$. The range for the weights are [1.255, 1.284] and [1.254, 1.282] and the averages are 1.265 and 1.264 for $1/P_n(\Delta = 1|A = 0, W)$ and $1/P_n(\Delta = 1|A = 1, W)$, respectively. Thus, we will again assume nonpositivity.

4.2 Average Treatment Effect on Fixed Time Outcomes

TMLE with SuperLearner based propensity score/outcome regression estimates were used to estimate the average treatment effect of specialty probation on violence and re-arrest within a fixed period of time.

Estimate of the effect of specialty probation on violence 1 year after baseline interview

During the one-year follow-up, 35.1% of probationers for whom there was violence outcome data ($N = 291$) were involved in violence. Naïve, unadjusted differences between groups indicate a 2.4% increase ($p = 0.663$) in violence for specialty (36.2%) versus traditional (33.8%) probationers.

Results of TMLE estimation (adjusting for confounding) and inference indicate that specialty supervision has no significant effect on violence. The probability of violence within one year is similar for traditional and specialty probationers (risk difference estimate = 2.5%, CI = [-4.3%, 9.3%]; p =0.4779). TMLE estimates the probability of violence to be 32.7% for specialty probation and 30.2% for traditional probation.

Estimate of the effect of specialty probation on re-arrest 2 years after baseline interview

Two years after their baseline interview, 42.37% of probationers for whom there was re-arrest data were arrested ($N = 354$; charge type: 15.1% person, 9.7% property, 14.5% drug, and 4.7% minor). Unadjusted differences between groups show a 22.3% significant decrease in re-arrest ($p < 0.0001$) for specialty (31.5%) versus traditional (53.8%) probationers.

TMLE estimation shows that specialty probation is associated with a significant reduction in re-arrest. As shown in Figure 4.7, the probability of re-arrest within two years is 30.6% (CI = [23.3%, 37.9%]; $p < 0.0001$) higher for probationers in traditional (55.7%) than specialty probation (25%).

4.3 Time to Re-arrest

KM and TMLE with SuperLearner based propensity score/outcome regression/missingness mechanism estimates were used to estimate the probability of no re-arrest in the community and difference in these probabilities for several timepoints, t.

Unadjusted estimates of survival probabilities

The average follow-up period across all probationers was 738.76 days (N = 354; SD = 448.44), excluding days incarcerated or hospitalized. KM estimates of unadjusted survival probabilities plotted in Figure 4.8 show that specialty probationers (top red curve) have longer times until the first arrest. At about three years, the probability of having "survived" in the community without a re-arrest is 61.8% and 35.4% for specialty and traditional probationers, respectively.

Estimate of the effect of specialty probation on re-arrest t days after baseline interview

	P(Y|A = 1, W)	P(Y|A = 0, W)	Risk difference	CI
180	0.30	0.06	-0.24	[-0.181, -0.302]
360	0.44	0.18	-0.26	[-0.189, -0.331]
540	0.48	0.26	-0.23	[-0.153, -0.299]
720	0.54	0.28	-0.26	[-0.185, -0.325]
900	0.60	0.34	-0.26	[-0.182, -0.336]
1080	0.73	0.42	-0.30	[-0.218, -0.392]

Table 4.3: TMLE estimates and inference of the difference in probability of re-arrest for each probation type within time = t, accounting for missingness in the outcome. Each row is a different timpoint, $t = 180, 360, 540, 720, 900, 1080$

Figure 4.9 and Table 4.3 show the probability of re-arrest until time t for specialty and traditional probationers, accounting for missingness in the re-arrest outcome. Figure 4.10 shows the difference in the probabilities between the two groups. All differences in probabilities and their confidence intervals are below 0, suggesting that the estimated risk difference between these two groups is significantly different than 0. In other words, at all timepoints, the probability of re-arrest is always significantly lower for probationers, taking account loss to follow-up before observing the probationer's outcome. Additionally, as also seen in the KM plot, the probability of re-arrest increases with time (i.e., probability of no re-arrest decreases with time in the KM plot), which makes sense: the longer a probationer is in the community, the more opportunity the probationer has of getting re-arrested.

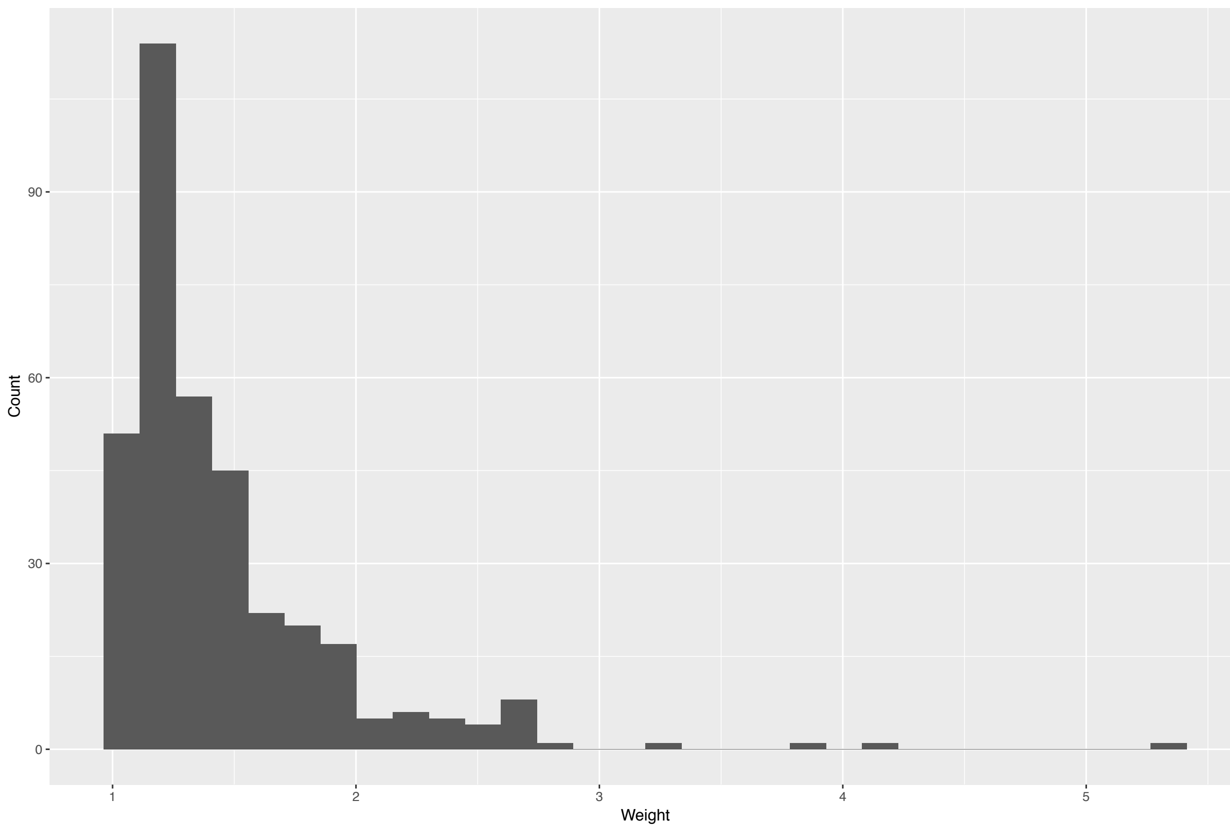

Figure 4.4: Distribution of weights for estimated treatment mechanism $1/P_n(A|W)$

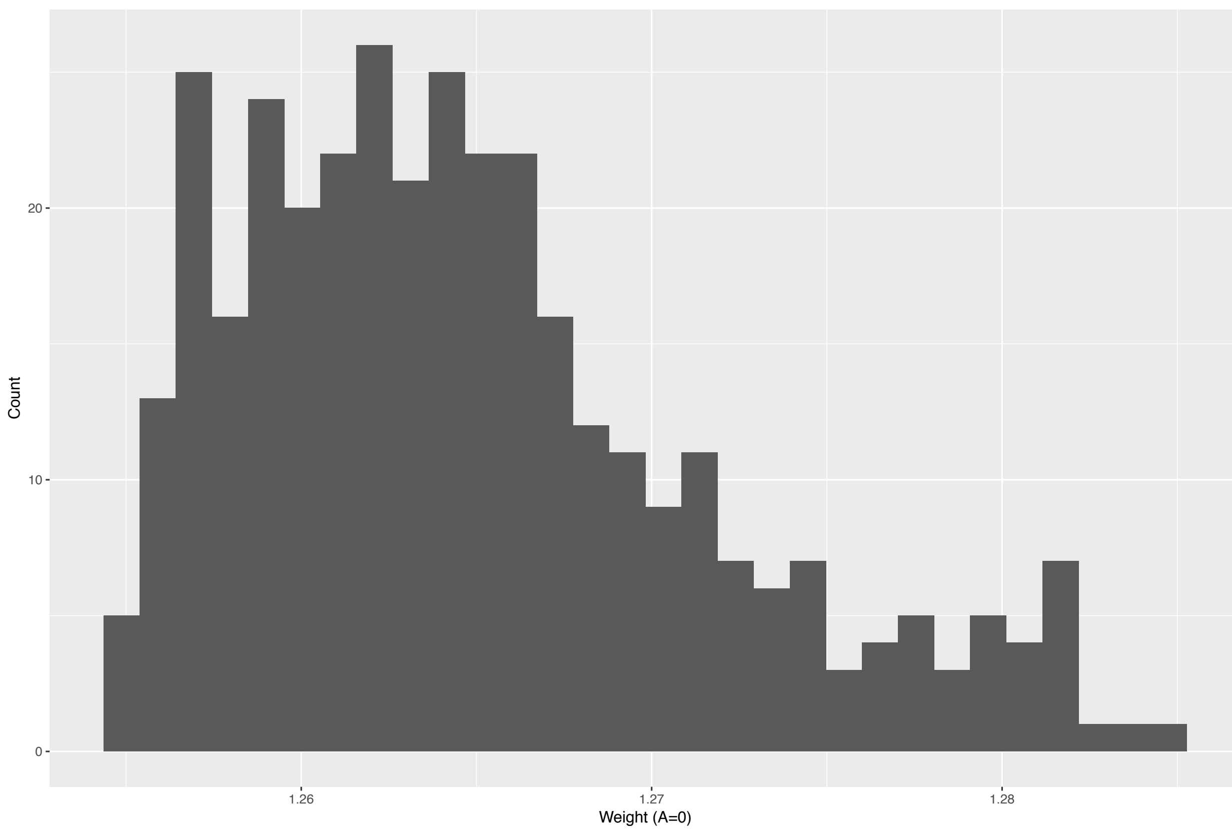

Figure 4.5: Distribution of weights for estimated missingness mechanism $1/P_n(\Delta = 1|A = 0, W)$ for $t = 1080$

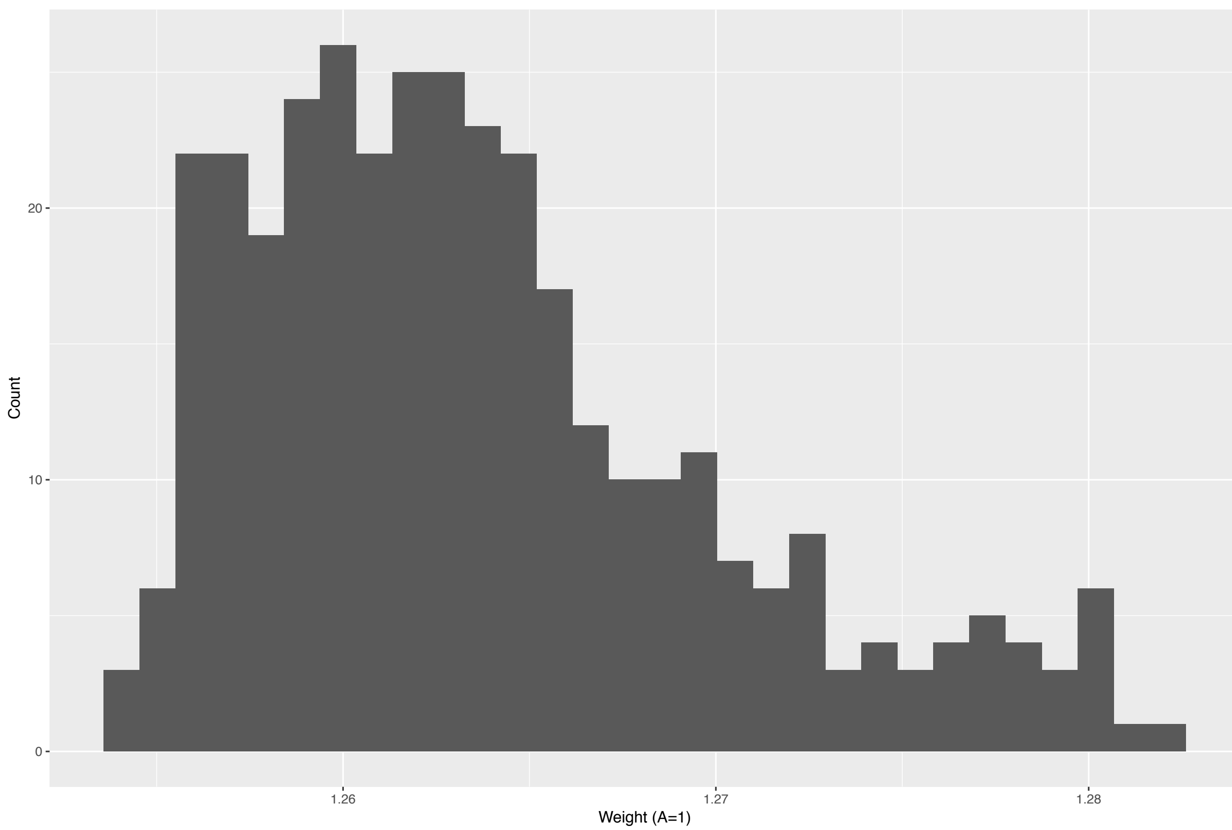

Figure 4.6: Distribution of weights for estimated missingness mechanism $1/P_n(\Delta = 1|A = 1, W)$ for $t = 1080$

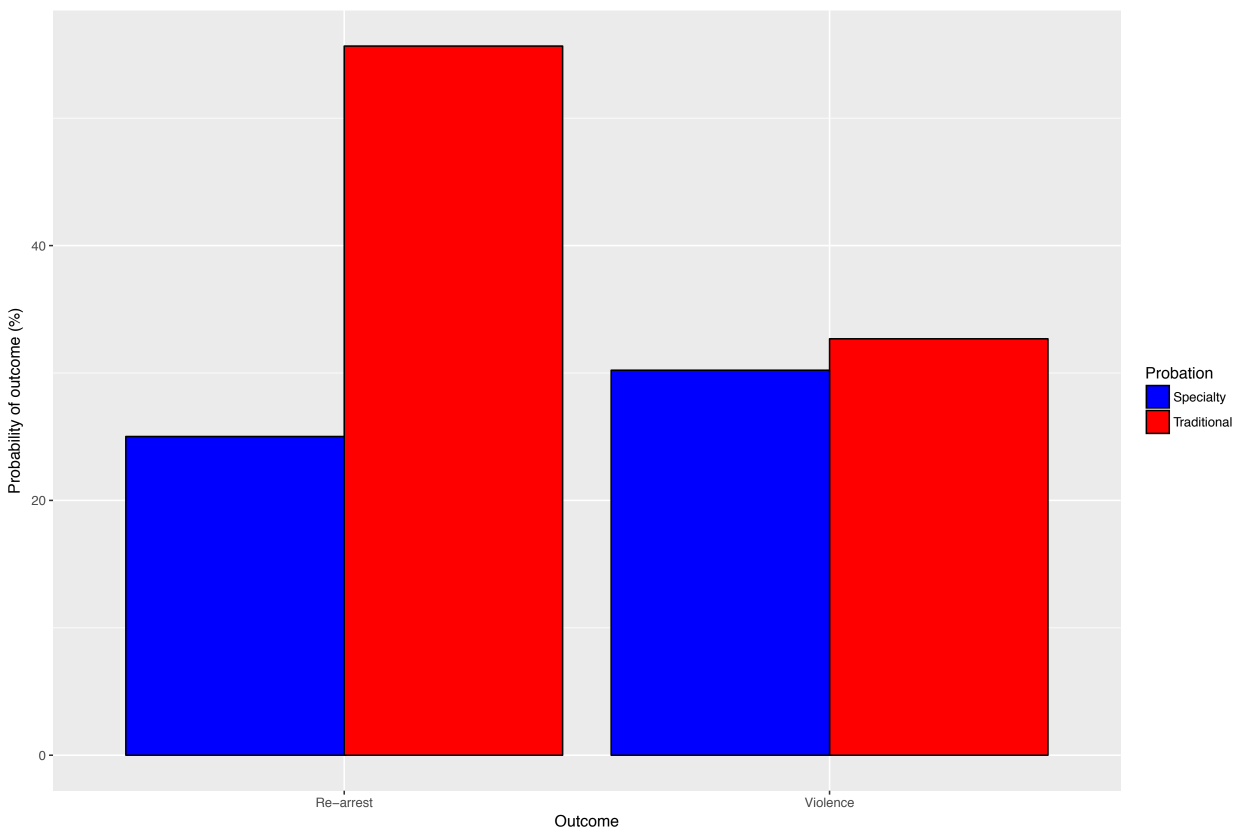

Figure 4.7: Estimated rates of violence (one year) and re-arrest (two years) for specialty and traditional probationers

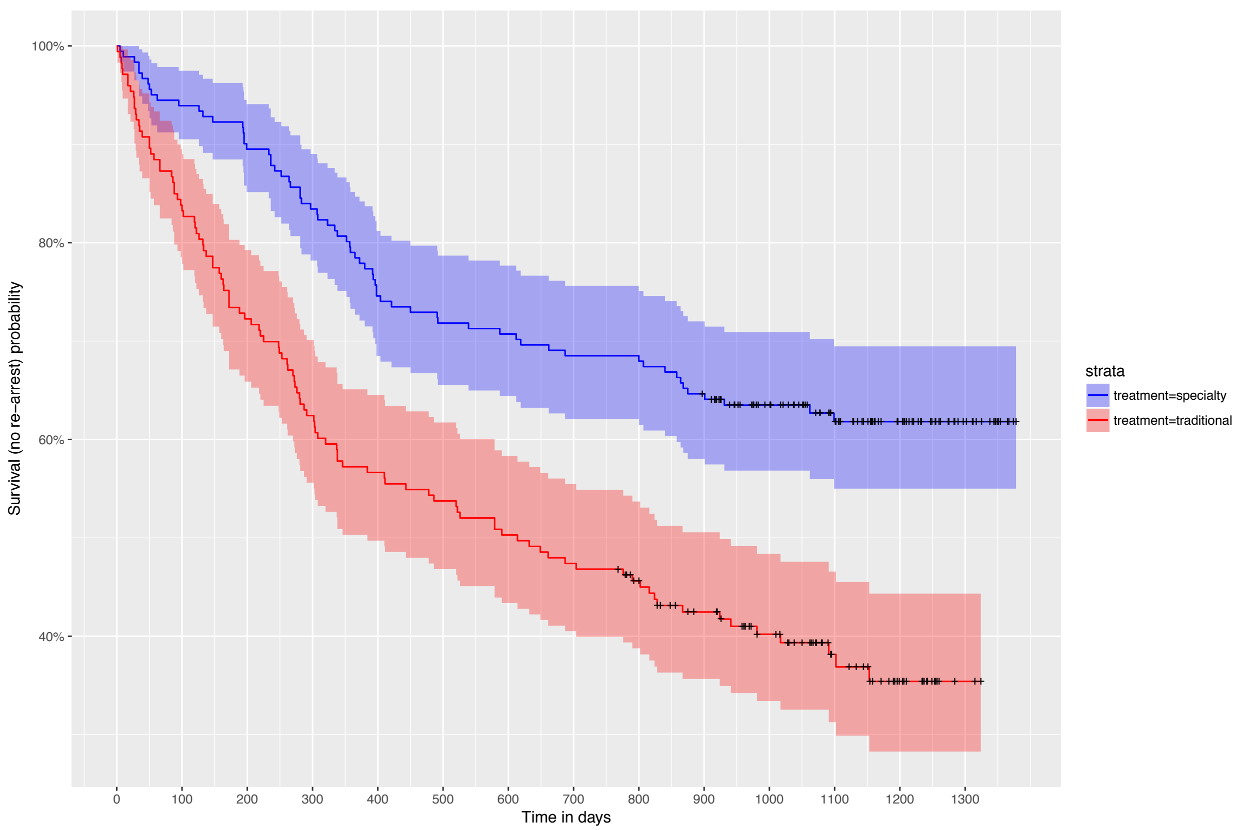

Figure 4.8: Estimates of the probability of survival (i.e., no re-arrest) for specialty and traditional probationers

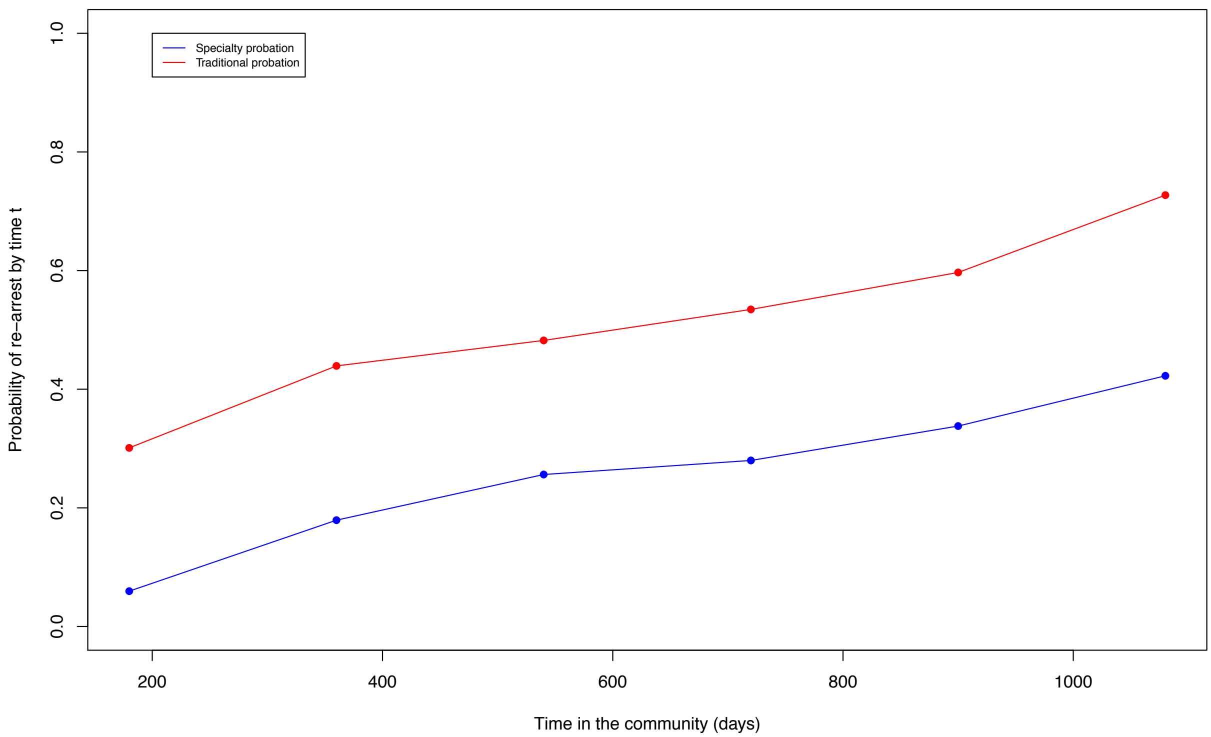

Figure 4.9: Estimates of the probability of re-arrest for different t accounting for missingness in outcome variable

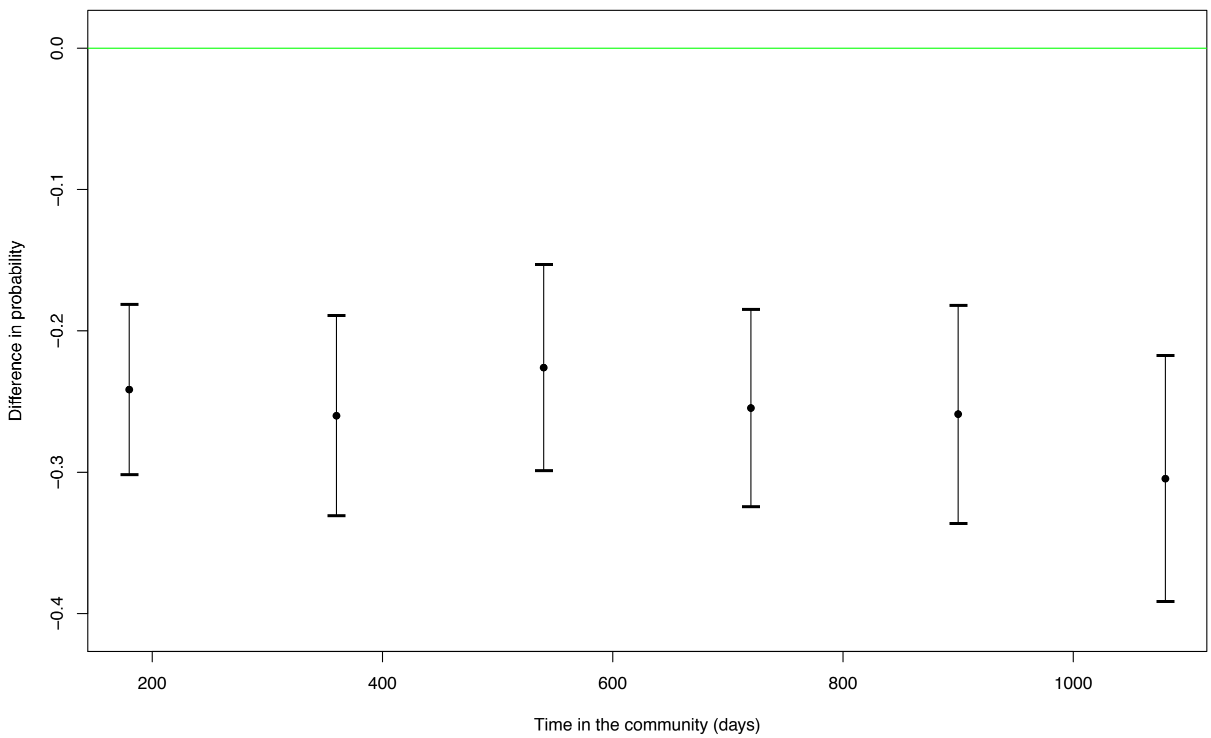

Figure 4.10: Estimates of the difference in probability of re-arrest by time t between specialty versus traditional probationers. The green line denotes a risk difference estimate of 0 (the null).

Chapter 5

Discussion

Specialty mental health probation is a type of probation aimed at improving outcomes for people with mental illness. Estimates inspired by the causal effect show that the difference in the probability of re-arrest among subjects with specialty versus traditonal probabiion, after adjsuting for a wide range of potential confounders including demographics, mental health indicators, and city characteristics, was significantly lower than for traditional probation (30.6%, CI = [23.3%, 37.9%]; $p < 0.0001$). This is true not just for outcomes measured at a fixed time point after the probationer's baseline interview, but also further out in time when not all probationer's outcomes were observed due to study dropout. We also found that specialty probation was not significantly associated with violence (2.5%, CI = [-4.3%, 9.3%]; $p = 0.4779$). Under the backdoor criteria and relevant positivity assumptions, these associations are equivalent to the causal effect of specialty probation on public safety outcomes. The plausibility of these assumptions is described below.

5.1 Study limitations

The main study limitation here is that probation type is completely determined by whether a probationer lives in Dallas or Los Angeles. Thus, many variables pertinent to the probationer's city of residence could be the actual driver of differences between the two probation groups, confounding the true effect of specialty probation on our outcomes of interest.

We tried to address this limitation in as many places as we could within the study design and analysis process. For example, probationers were matched based on similar background characteristics, and we ensured strong study implementation and precise measurements. In the analysis step, we included a large number of diverse covariates that we believe were related to probation assignment and the outcomes in the hopes of controlling for confounding. We also used a double-robust estimator that allowed us two chances to generate a consistent estimate. Additionally, sensitivity analyses published elsewhere [42] using the same estimators on different covariate sets yielded similar results as seen here. Our efforts to address the main study limitation point to promising conclusions on specialty probation's effects; however, the study must be replicated in a randomized controlled trial that more closely resembles the ideal experiment.

5.2 Plausibility of assumptions

Positivity assumption

The assumption described in 2.5 is theoretically violated if we include city of residence in the adjustment sest because the probability of receiving specialty probation for probationers who live in Los Angeles is 0. We therefore excluded city from our adjustment set. This implies, however, that other measured characterstics of individuals must be sufficient to control for any differences in the risk of public safety outcomes between probationers in the two cities. Subjects who receive specialty probation may only possess background characteristics unique to Los Angeles. However, traditional probationers were matched at the sample level to specialty probationers, yielding similar background characteristics for both sites. Therefore, it is more unlikely that there are no unique background characteristics pertaining to one city that make it impossible for a type of probation to occur.

Randomization assumptions

Plausibility of the randomization assumption 2.8 is of concern here. This states that the outcome is independent of the treatment and missingness mechanism, given covariates. Δ as defined here inherently depends on the outcome: a subject who is lost to follow-up before time t will have $\Delta = 1$ only if he or she experienced re-arrest before lost to follow-up. We note however, that this only applies to the re-arrest outcome at the two time points after 720 days; outcomes prior to this are not subject to missingness (Δ is deterministcally equal to 1). Below we describe extensions to better handle this missingness.

5.3 Implications of study

In this analysis and in a previous sensitivity analysis [42], we did not find an association between specialty probation and reduced violence. The reason for this null finding may be because mental illness is not strongly related to violent recidivism [6]. In this study, assignment to specialty probation is mostly based on probationers' mental illness status, and there is a large emphasis on mental health treatment while they are on probation. Because there is a weak relationship between mental illness and violent recidivism, the factors that drive assignment to treatment and the actual treatment do not relate to the outcome. Additionally, in general, probation is not designed to specifically reduce violence.

We did observe a reduction in re-arrest for specialty probationers, which is in line with the desirable goal of improving public safety. These results were not only significant, but also lasting for years after probationers' baseline interview (see Figures 4.8 and 4.9) — an extension of results found in previous studies examining specialty probation [52].

Specialty probation is reportedly implemented in 130 agencies, but not all agencies carry out the five core defining features described previously. The current findings are presented in a context in which the core features of specialty probation are preserved. Moving forward, from a policy standpoint, it is important that the core features of specialty probation are maintained; otherwise, it will risk losing its efficacy on public safety. Although specialty probation requires more specialized supervision and careful allocation of resources, in future analyses we show that it is ultimately more cost effective than traditional probation.

5.4 Future directions

Longitudinal TMLE

In this thesis, we estimated the probability of re-arrest for both probation groups based on prior re-arrest history, not adjusting for baseline confounders (i.e., KM estimation). Additionally, we estimated the difference probability of re-arrest between the two probation groups as point treatments for various timepoints (using TMLE). The next step in estimation would be making of use of both adjustment of baseline confounders and prior drop-out/re-arrest by longitudinal TMLE (LTMLE) [39]. Because LTMLE takes into account previous covariate/censoring/outcome history, it is a more unbiased and efficient estimator than KM or point treatment TMLE.

Two-community intervention

The current analysis treats all probationers as n i.i.d. observations drawn from one random variable O. However, another way to look at this study design is to treat the two probation sites as two communities that each receive a level of the intervention, and variables are drawn independently from two populations, instead of one target population. The motivation for looking at the problem in this way is that the two cities have the same environmental factors and intervention, making it difficult to distinguish the true cause of the outcome as the intervention or other environmental factors. Petersen and van der Laan [18] go through the steps of the causal roadmap to estimate the causal effect of an intervention for a community based intervention with two communities. In this case, the ideal experiment is as follows: the specialty probation is assigned to both Dallas and Los Angeles and individual outcomes are measured in the combined population, and then we go back in time and traditional probation is assigned to both cities and individual outcomes are measured in the combined population. With this setup, the marginal treatment effect can be estimated using TMLE at the community level.

Mediation

Now that we have examined whether specialty probation has an effect on public safety outcomes, it will be important to see how specialty probation reduces re-arrest. In future analyses we hope to examine pathways from probation type to 1) better correctional practices and 2) symptom control to re-arrest. Specifically, we hope to estimate the natural direct effect via these two mediators.

5.5 Concluding remarks

The results of this study could have significant implications for probation practices within the criminal justice system. Our main result is that specialty probation reduces re-arrest relative to traditional probation. By carefully going through the steps of the causal roadmap, we shed light on the assumptions under which we can declare causal inference. If those assumptions hold and specialty probation does indeed protect against re-arrests, implementing this program in a systematic way could mean an additional alternative to reduce mass incarceration.

Bibliography

[1] URL: http://www.sentencingproject.org/criminal-justice-facts/.

[2] URL: https://www.nij.gov/topics/corrections/recidivism/Pages/welcome.aspx.

[3] URL: https://stepuptogether.org/the-problem.

[4] Steve Aos et al. "The Comparative Costs and Benefits of Programs To Reduce Crime. Version 4.0." In: (2001).

[5] American Psychiatric Association, American Psychiatric Association, et al. "DSM-IV-TR: Diagnostic and statistical manual of mental disorders, text revision". In: *Washington, DC: American Psychiatric Association* 75 (2000).

[6] James Bonta, Julie Blais, and Holly A Wilson. "A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders". In: *Aggression and violent behavior* 19.3 (2014), pp. 278–287.

[7] Stephen R Cole and Miguel A Hernán. "Constructing inverse probability weights for marginal structural models". In: *American journal of epidemiology* 168.6 (2008), pp. 656–664.

[8] *Criminal Justice Fact Sheet*. URL: http://www.naacp.org/criminal-justice-fact-sheet/.

[9] Sandrine Dudoit. *Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine I*. http://www.stat.berkeley.edu/ sandrine/Teaching/PH240C.F16/. 2016.

[10] *Effect of Specialty Mental Health Probation on Public Safety Outcomes*.

[11] Chris Kennedy Mark van der Laan Eric Polley Erin LeDell. *SuperLearner*. R Package. 2016.

[12] Daniela Golinelli and Ann E. Carson. *Prisoners in 2012: Advance Counts*. Bulletin. Bureau of Justice Statistics, 2013.

[13] Susan Gruber and Mark J. van der Laan. *tmle: An R Package for Targeted Maximum Likelihood Estimation*. 13. 2012, pp. 1–35. URL: http://www.jstatsoft.org/v51/i13/.

[14] Alan Hubbard et al. "Time-Dependent Prediction and Evaluation of Variable Importance Using SuperLearning in High Dimensional Clinical Data". In: *The journal of trauma and acute care surgery* 75.1 0 1 (2013), S53.

[15] Danielle Kaeble and Lauren Glaze. *Correctional Populations in the United States*. Tech. rep. Bureau of Justice Statistics, 2016.

[16] Patrick J Kennealy et al. "Firm, fair, and caring officer-offender relationships protect against supervision failure." In: *Law and human behavior* 36.6 (2012), p. 496.

[17] YH Kim et al. "Personality assessment inventory professional manual". In: *Seoul: Harkgisa* (2001).

[18] Mark J van der Laan, Maya Petersen, Wenjing Zheng, et al. "Estimating the effect of a community-based intervention with two communities". In: *Journal of causal inference* 1.1 (2013), pp. 83–106.

[19] Mark J Van der Laan, Eric C Polley, and Alan E Hubbard. "Super learner". In: *Statistical applications in genetics and molecular biology* 6.1 (2007).

[20] Mark J Van der Laan and Sherri Rose. "Targeted learning: causal inference for observational and experimental data. 00216". In: *Springer Science & Business Media. Accessed October* 5 (2011), p. 2015.

[21] Alison Lawrence. "Trends in Sentencing and Corrections: State Legislation". In: *National Conference of State Legislatures*. Vol. 15. 2013.

[22] Samuel D Lendle, Meenakshi S Subbaraman, and Mark J van der Laan. "Identification and efficient estimation of the natural direct effect among the untreated". In: *Biometrics* 69.2 (2013), pp. 310–317.

[23] Mark W Lipsey and Francis T Cullen. "The effectiveness of correctional rehabilitation: A review of systematic reviews". In: *Annu. Rev. Law Soc. Sci.* 3 (2007), pp. 297–320.

[24] Mary Lee Luskin and Bradley Ray. "Selection Into Mental Health Court: Distinguishing Among Eligible Defendants". In: *Criminal justice and behavior* 42.11 (2015), pp. 1145–1158.

[25] Sarah M Manchak et al. "High-fidelity specialty mental health probation improves officer practices, treatment access, and rule compliance". In: *Law and human behavior* 38.5 (2014), pp. 450–461.

[26] Todd D. Minton and Zhen Zeng. *Jail Inmates in 2015*. Tech. rep. Bureau of Justice Statistics, 2016.

[27] John Monahan et al. *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. Oxford University Press, 2001.

[28] Michelle Naples, Laura S Morris, and Henry J Steadman. "Factors in disproportionate representation among persons recommended by programs and accepted by courts for jail diversion". In: *Psychiatric Services* 58.8 (2007), pp. 1095–1101.

[29] Judea Pearl. *Causality*. Cambridge university press, 2009.

[30] Judea Pearl. "Interpretation and identification of causal mediation." In: *Psychological Methods* 19.4 (2014), p. 459.

[31] Maya Petersen and Laura Balzer. *Introduction to Causal Inference*. http://www.ucbbiostat.com. 2014.

[32] Maya L Petersen and Mark J van der Laan. "Causal models and learning from data: integrating causal modeling and statistical estimation". In: *Epidemiology (Cambridge, Mass.)* 25.3 (2014), p. 418.

[33] Maya L Petersen, Sandra E Sinisi, and Mark J van der Laan. "Estimation of direct causal effects". In: *Epidemiology* 17.3 (2006), pp. 276–284.

[34] Maya L Petersen et al. "Diagnosing and responding to violations in the positivity assumption". In: *Statistical methods in medical research* 21.1 (2012), pp. 31–54.

[35] Joan Petersilia and Susan Turner. *Prison versus Probation in California*. Tech. rep. National Institute of Justice, U.S. Department of Justice, 1986.

[36] Eric Polley et al. *SuperLearner: Super Learner Prediction*. R package version 2.0-22. 2016. URL: `https://github.com/ecpolley/SuperLearner`.

[37] Jason T Rich et al. "A practical guide to understanding Kaplan-Meier curves". In: *Otolaryngology-Head and Neck Surgery* 143.3 (2010), pp. 331–336.

[38] Teresa L. Scheid and Tony N. Brown. *A Handbook for the Study of Mental Health*. Second Edition. 2010.

[39] J Schwab et al. "ltmle: Longitudinal targeted maximum likelihood estimation". In: *R package version 0.9* 3 (2014).

[40] Joshua Schwab et al. *ltmle: Longitudinal Targeted Maximum Likelihood Estimation*. R package version 0.9-9. 2016. URL: `https://CRAN.R-project.org/package=ltmle`.

[41] David L Shern et al. "Client outcomes II: Longitudinal client data from the Colorado treatment outcome study". In: *The Milbank Quarterly* (1994), pp. 123–148.

[42] Jennifer Skeem, Sarah Manchak, and Lina Montoya. "Effect of Specialty Mental Health Probation on Public Safety Outcomes". In: *Journal of the American Medical Association - Psychiatry* ((in progress)).

[43] Jennifer L Skeem, Paula Emke-Francis, and Jennifer Eno Louden. "Probation, mental health, and mandated treatment: A national survey". In: *Criminal Justice and Behavior* 33.2 (2006), pp. 158–184.

[44] Jennifer L Skeem and Jennifer Eno Louden. "Toward evidence-based practice for probationers and parolees mandated to mental health treatment". In: *Psychiatric Services* 57.3 (2006), pp. 333–342.

[45] Jennifer L Skeem, Sarah Manchak, and Jillian K Peterson. “Correctional policy for offenders with mental illness: Creating a new paradigm for recidivism reduction”. In: *Law and human behavior* 35.2 (2011), pp. 110–126.

[46] Jennifer L Skeem and John Petrila. “Problem-solving supervision: Specialty probation for individuals with mental illness”. In: *Court Review* 40.4 (2004), p. 8.

[47] Howard N. Snyder. *Arrest in the United States, 1990-2010.* Tech. rep. Bureau of Justice Statistics, 2012.

[48] Henry J Steadman et al. “Prevalence of serious mental illness among jail inmates”. In: *Psychiatric services* 60.6 (2009), pp. 761–765.

[49] Henry J Steadman et al. “Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods”. In: *Archives of general psychiatry* 55.5 (1998), pp. 393–401.

[50] Martin H Teicher et al. “The neurobiological consequences of early stress and childhood maltreatment”. In: *Neuroscience & Biobehavioral Reviews* 27.1 (2003), pp. 33–44.

[51] Johannes Textor, Juliane Hardt, and Sven Knüppel. “DAGitty: a graphical tool for analyzing causal diagrams”. In: *Epidemiology* 22.5 (2011), p. 745.

[52] Nancy Wolff et al. “Mental health specialized probation caseloads: Are they effective?” In: *International journal of law and psychiatry* 37.5 (2014), pp. 464–472.

[53] *World report on violence and health: summary.* Tech. rep. World Health Organization, 2002.

Acknowledgments

First and foremost I would like to thank my advisers for their mentorship and guidance when writing this thesis and throughout my Masters degree. Dr. Jennifer Skeem opened the door for me to work on a project that I find most meaningful, and she has been immensely supportive of me trying out methodologies that may be regarded as unconventional in other disciplines. I am truly grateful and proud to be a part of her research team. I thank Dr. Maya Petersen for her advising throughout my Masters and her role as my teacher. She has changed the way I think about causality, and because of her tutelage I am passionate about the framework and methodologies. I greatly admire not only her knowledge on the subject, but also her ability to clearly translate complex concepts to colleagues across many disciplines. Thank you to Dr. John Marshall for his support, especially for giving me freedom in teaching his classes, which I have no doubt helped me grow as an instructor. I am truly grateful for his trust, collaboration, and teamwork.

I am grateful to my peers who worked on analyzing these data for various group projects, especially Chris Kennedy, Courtney Schiffman, Jonathan Larson, Fausto Bustos, Dana Goin, and Suzanne Dufault. Thank you to Dr. Alan Hubbard, Dr. Sandrine Dudoit, Lucia Petito, Jonathan Levy, and Kelly Street for their help and feedback on these group projects. Additionally, thank you to Dr. David Benkeser and Dr. Caleb Miles for taking time out of their busy schedules to check if I was doing things right :)

Last but not least, I sincerely thank my friends and family for their support. A warm thanks to Caitlin Lynch and Juan Manuel Ramirez for their revisions to the thesis, among many, many other things. Ivan and Martha, I owe everything to you, and words cannot come close to expressing my immense gratitude for everything you do for me on a daily basis. I hope someday you will learn it is pronounced “causal” instead of “casual.”