

2022

Estimation of causal effects of exposure models and of drug-induced homicide prosecutions on drug overdose deaths

<https://hdl.handle.net/2144/46388>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**ESTIMATION OF CAUSAL EFFECTS OF EXPOSURE
MODELS AND OF DRUG-INDUCED HOMICIDE
PROSECUTIONS ON DRUG OVERDOSE DEATHS**

by

KELLY C. KUNG

M.A., Boston University, 2019

B.A., Wellesley College, 2017

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2022

© 2022 by
KELLY C. KUNG
All rights reserved

Approved by

First Reader

Daniel L. Sussman, Ph.D.
Assistant Professor of Mathematics and Statistics

Second Reader

Judith J. Lok, Ph.D.
Associate Professor of Mathematics and Statistics

Third Reader

Luis Carvalho, Ph.D.
Associate Professor of Mathematics and Statistics

Fourth Reader

Masanao Yajima, Ph.D.
Associate Professor of Statistical Practice

Acknowledgments

I would first like to thank my co-advisors, Dan Sussman and Judith Lok. For the past five years, you both have been so supportive of me on my Ph.D. journey and on my dissertation work. Thank you both for your time and commitment to me and our work. I would not have been able to get through this process without your academic and emotional support. To my committee members, Luis Carvalho and Masanao Yajima, thank you for teaching me and giving me the opportunity to work with you both in MSSP. I have learned a lot from you both in these past five years. I would also like to thank Leo Beletsky and Natasha Martin who allowed me to work on such an interesting and meaningful project. I am deeply grateful for you all.

I would not have been able to survive graduate school without all of my friends. To some of the smartest people I know—Ben Draves, Lily Chou, Nate Josephs, and Zihuan Qiao, I am incredibly lucky to have experienced graduate school with you all, and thank you for sharing your knowledge with me. Thank you to my childhood friends—Hilary Chen, Megan Lieu, Ying Loo, Jack Tam, Johnson Fong, and Michael Lee, who have been there for me since day one and have continued to support me. To my partner, Cyril Caparanga, thank you for being there for me during my ups and my downs and always believing in me. You all reminded me to enjoy life and to enjoy the process.

Lastly, to my family, Mom, Dad, Michelle, and Jason, words cannot describe how much you all mean to me. Thank you for being there for me for the past 26 years. I would be nowhere without your love and support.

I cannot thank all the incredible people who have helped me throughout this time enough. I am indebted to all of you. Thank you, and I love you all.

Dedication

I dedicate my dissertation to my parents. Without all of your sacrifice and support, I would not be where I am today. Thank you both. I love you.

我把我的論文獻給我的父母。沒有你們所有的犧牲和支持，我就不會是今天的我。謝謝你們，我愛你。

**ESTIMATION OF CAUSAL EFFECTS OF EXPOSURE
MODELS AND OF DRUG-INDUCED HOMICIDE
PROSECUTIONS ON DRUG OVERDOSE DEATHS**

KELLY C. KUNG

Boston University, Graduate School of Arts and Sciences, 2022

Major Professors: Daniel L. Sussman, Ph.D.

Assistant Professor of Mathematics and Statistics

Judith J. Lok, Ph.D.

Associate Professor of Mathematics and Statistics

ABSTRACT

Causal inference methods have been applied in various fields where researchers want to establish causal effects between different phenomena. The goal of causal inference is to estimate treatment effects by comparing outcomes had units received treatment versus outcomes had units not received treatment. We focus on estimating treatment effects in three different projects.

We first proposed linear unbiased estimators (LUEs) for general causal effects under the assumption that treatment effects are additive. Under the assumption of additivity, the set of estimands considered grows as contrasts in exposures are now equivalent. Furthermore, we identified a subset of LUEs that forms an affine basis for LUEs, and we characterized LUEs with minimum integrated variance through defining conditions on the support of the estimator.

We also estimated the effect of drug-induced homicide (DIH) prosecutions reported by the media on unintentional drug overdose deaths, which have never been

empirically assessed, using various models. Using a difference-in-differences-like logistic generalized additive model (GAM) with smoothed time effects where we assumed a constant treatment effect, we found that DIH prosecutions reported by the media were associated with a potential harmful effect (risk ratio: 1.064; 95% CI: (1.012, 1.118)) on drug overdose deaths. Upon further research, however, there are potential issues using a constant treatment effect model in a setting where treatment is staggered and treatment effects are heterogeneous. Therefore, we also used a GAM with a linear link function where we assumed that treatment effects may depend on the treatment duration. With this second model, we estimated a risk ratio for having any DIH prosecutions reported by the media of 0.956 (95% CI: (0.824, 1.110)) and a risk ratio of 0.986 (95% CI: (0.973, 0.999)) for the effect of being exposed to DIH prosecutions reported by the media for each additional six months. Despite being statistically significant, the effects were not practically significant. However, the results call for further research on the effect of DIH prosecutions on drug overdose deaths.

Lastly, we shift our focus to Structural Nested Mean Models (SNMMs). We extended SNMMs to a new class of estimators which estimate treatment effects of different treatment regimes in the risk ratio scale—the Structural Nested Risk Ratio Model (SNRRM). We further generalized previous work on SNMMs by estimating treatment effects by modeling a function of treatment, which we choose to be any function that can be modeled by generalized linear models, as opposed to just a model for treatment initiation. We applied SNRRMs to estimate the effect of DIH prosecutions reported by the media on drug overdose deaths.

Contents

1	Introduction	1
1.1	Potential Outcomes Framework	2
1.2	Treatment Effects	4
1.3	Common Assumptions in the Causal Inference Framework	5
1.4	Repeated Outcomes Setting	7
1.5	Overview of Dissertation	9
2	Treatment Effects Under Exposure Models	10
2.1	Introduction	10
2.2	Background	12
2.3	Exposure Models	15
2.4	Linear Unbiased Estimators	19
2.4.1	Unbiased Estimators	21
2.5	Atomic Linear Unbiased Estimators	22
2.5.1	Affine Basis for Linear Unbiased Estimators	24
2.6	Optimal Linear Unbiased Estimators	29
2.6.1	Minimum Integrated Variance Linear Unbiased Estimators (MIV LUE)	29
2.6.2	Characterization of MIV LUEs	32
2.6.3	Example: Six-Term Exposure Set	36
2.7	Simulations	43
2.7.1	Varying Number of Units and Number of Edges	45

2.7.2	Varying Interference Effects and Deviations from Additivity	48
2.7.3	Varying Potential Outcome Distributions	51
2.8	Discussion	53
3	Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths	56
3.1	Introduction	56
3.2	Background	57
3.3	Data	59
3.3.1	Outcome: Unintentional Drug Overdose Deaths	59
3.3.2	DIH Prosecutions and Other Policy Measures	61
3.3.3	Intervention Definition	63
3.4	Constant Treatment Effect	66
3.4.1	Model	66
3.4.2	Sandwich Estimator for Variance of $\hat{\Theta}$	68
3.4.3	Estimating the Number of Attributable Deaths	69
3.4.4	Results	71
3.5	Potential Issues with the Constant Treatment Effect Model in Stag- gered Treatment Settings	74
3.6	Treatment Effects that Depend on Exposure Duration	78
3.6.1	Model	79
3.6.2	Sandwich Estimator for Variance of $\hat{\Theta}$	81
3.6.3	Estimating the Number of Attributable Deaths	82
3.6.4	Sensitivity Analysis	83
3.6.5	Results	84
3.7	Discussion	89

4	Structural Nested Risk Ratio Models	94
4.1	Introduction	94
4.2	Background	95
4.3	Structural Nested Risk Ratio Models	97
4.3.1	Notation	97
4.3.2	Model and Assumptions	98
4.3.3	Estimation, Consistency, and Asymptotic Normality	100
4.4	Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths: Binary Treatment	102
4.4.1	One-parameter Treatment Effect Model	103
4.4.2	Two-parameter Treatment Effect Model	104
4.5	Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths: Count Treatment	105
4.5.1	One-parameter Treatment Effect Model	106
4.6	Results	107
4.7	Discussion	110
A	Treatment Effects Under Exposure Models	113
A.1	Linear Unbiased Constraints	113
A.2	Affine Basis for Set of LUEs \mathcal{U}	114
A.2.1	Construction and Affine Independence of \mathcal{M}	114
A.2.2	Affine Basis for \mathcal{U}	115
A.3	MIV LUEs	117
A.3.1	Optimization Problem	117
A.3.2	Characterizing MIV LUEs	120
A.3.3	Example: Derivation of Weights for Six-Term Exposure	129
A.4	Example: Six-Term Exposure Set	131

A.5	Simulations from an Erdős-rényi Network	141
B	Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths	145
B.1	Intervention Dates	145
B.2	States and their U.S. Regions	148
B.3	Sandwich Estimator	148
B.3.1	Constant Treatment Effect Model	148
B.3.2	Treatment Effect That Depends on Exposure Duration	151
B.4	Estimation of Attributable Deaths	152
B.4.1	Constant Treatment Effect Model	152
B.4.2	Treatment Effect That Depends on Exposure Duration	154
B.5	Results for Constant Treatment Effect	156
B.6	Results of Sensitivity Analyses for Model where Treatment Effect De- pends on Exposure Duration	157
C	Structural Nested Risk Ratio Models	159
C.1	Mimicking Potential Outcomes	159
C.2	Unbiased Estimating Equations	160
C.3	One-parameter Model with Binary Treatment	161
	References	163
	Curriculum Vitae	173

List of Tables

3.1	Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for GAM with linear link function when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years. . .	86
B.1	The intervention date (year-month) of DIH prosecutions reported by the media in each state and the number of DIH prosecutions reported by the media in each state.	146
B.2	The intervention dates (year-month) for the other relevant policy measures in each state: naloxone access law where pharmacists a) can dispense naloxone without prescription (NAL: can dispense) and b) cannot dispense naloxone without prescription (NAL: cannot dispense) c) medical marijuana law (MML) d) recreational marijuana law (RML) e) Prescription Drug Monitoring Program (PDMP) f) 911 Good Samaritan Law (911 GSL) f) Medicaid expansion (Medicaid).	147
B.3	The states considered in the analysis with their U.S. regions.	148
B.4	Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for logistic GAM when assuming constant treatment effects and assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years.	156

B.5	Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for GAM with linear link function when assuming constant treatment effects and assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years.	156
B.6	Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for logistic GAM when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years.	157
B.7	Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for GAM with linear link function when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years. We excluded the last six years of data.	158
B.8	Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for logistic GAM when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years. We excluded the last six years of data.	158

List of Figures

1.1	Trends for ice cream sales and shark attacks in a year. Retrieved from https://www.statology.org/correlation-does-not-imply-causation-examples/ .	2
1.2	Causal diagram for the question: <i>how effective is a new medication on the prevention of cardiovascular disease?</i>	3
2.1	Causal diagram for the question: <i>does showing advertisements for a new game on a user's Facebook Newsfeed increase sales of the new game?</i>	11
2.2	The trends of weight α_1 , α_2 , and α_3 (indicated by line type and color) for $\text{Var}(\theta_{1,m}) = 0.00001$ and different values $\text{Var}(\alpha)$, $\text{Var}(\theta_{1,m_1})$, and $\text{Var}(\theta_{2,j})$ as indicated by the panels. Variances of parameters that are not varying are set to values to maximize α_3 : $\text{Var}(\alpha) = 0.00001$, $\text{Var}(\theta_{1,m_1}) = 100,000$, and $\text{Var}(\theta_{2,j}) = 1$.	39
2.3	The trends of weights $\alpha_1, \alpha_2, \alpha_3$ (indicated by line type and color) for $\text{Var}(\theta_{1,m}) = 0.00001$ and $\text{Var}(\theta_{1,m_1}) = \text{Var}(\alpha) = \text{Var}(\theta_{2,j}) = 1$ and different values of $\text{cor}(\alpha, \theta_{2,j})$, $\text{cor}(\alpha, \theta_{1,m_1})$, $\text{cor}(\theta_{1,m_1}, \theta_{2,j})$ as indicated by the x-axis.	42
2.4	Directed 4-regular network with forty nodes and each unit has in-degree of four. Arrows indicate directions of edges.	46

2-5	IMSE for estimators (indicated by color and line type) when the number of units (indicated by x-axis) increases for a k -regular network for different values of k (indicated by panel) under additivity and when mean interference is zero.	47
2-6	IMSE for estimators (indicated by color and line type) when the interaction (indicated by x-axis) and interference effects (indicated by panel) vary for a 40 node 4-regular graph.	50
2-7	IMSE for estimators (indicated by color and line type) under different potential outcome distributions (indicated by panel) as the interference effect varies (indicated by x-axis) under additivity for a 40-node 4-regular graph.	52
3-1	Causal diagram for a subset of time points for the question: <i>what is the effect of having any drug-induced homicide prosecutions vs. none on drug overdose deaths?</i>	57
3-2	The yearly number of drug-induced homicide charges per year in the U.S. from 1974-2019 collected by the Health in Justice Action Lab.	63
3-3	(a) Yearly number of unintentional drug overdose deaths in the 50 U.S. states from 2000 to 2019; and (b) Cumulative number of states exposed to DIH prosecutions from 2000 to 2019.	71
3-4	Results of fitted models (red and dashed) plotted against observed trend (blue and solid) for all U.S. states in 2000-2019. Vertical black lines indicate the time of the first DIH prosecution reported by the media for each state.	73

3·5	Estimated yearly number of unintentional drug overdose deaths attributable to DIH prosecutions reported by the media for main analysis (red with circles) and secondary analysis (blue with crosses) with 95% confidence intervals (dashed lines for main analysis and dot-dashed lines for secondary analysis) in all 50 U.S. states from 2000 to 2019.	74
3·6	Coefficients and 95% confidence intervals for periods leading up to and following the first date of DIH prosecutions reported by the media. The red dotted vertical line indicates the date of the first DIH prosecution reported by the media. The period right before and 34 periods before the treatment time were excluded.	85
3·7	Estimated risk ratios and 95% confidence intervals for main analysis when assuming treatment effect lasted until at least 2019 and when assuming treatment effect lasted only for two years (indicated by color and shape).	87
3·8	Estimated yearly number of potential lives saved in the 50 U.S. States from 2000 to 2019 due to DIH prosecutions reported by the media 1) when assuming treatment effect lasted until at least 2019 and 2) when assuming treatment effect lasted only for two years (indicated by color and shape).	89
4·1	Number of drug-induced homicide prosecutions reported by media in each state per six-month interval from 2000 to 2019.	95
4·2	Confidence region for ψ_1 and ψ_2 in two-parameter treatment model. The red “×” indicates the estimated $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2)$	108
4·3	The treatment effect given by (a) $\hat{\psi}$ and (b) $\hat{\psi}_1 + \hat{\psi}_2(k - T_s^*)$ by the length of exposure to drug-induced homicide prosecutions reported by the media $k - T_s^*$ with 95% confidence interval when treatment is binary.	109

4.4	Estimated risk ratio of treatment effect for one-parameter treatment model where treatment is number of drug-induced homicide prosecutions and 95% confidence intervals for each $\tilde{a}_t - \tilde{a}_{t-1}$ observed in the data. The size of data point represents the number of observed data-points for the value of $\tilde{a}_t - \tilde{a}_{t-1}$	110
A.1	Directed Erdős-rényi network with 40 nodes and probability of an edge is 0.25.	142
A.2	IMSE for estimators (indicated by color and line type) as the number of units vary (indicated by x-axis) when additivity holds and mean interference effect is zero for ER(0.25) network.	143
A.3	IMSE for estimators (indicated by color and line type) as the interaction effect varies (indicated by x-axis) for different mean interference effects (indicated by the panels) for ER(0.25) network with 40 nodes.	144

List of Abbreviations

911 GSL	911 Good Samaritan Law
95% CI	95% confidence interval
ALUE	atomic linear unbiased estimator
CDC	Centers for Disease Control and Prevention
DIH	drug-induced homicide
ER	Erdős-rényi
GAM	generalized additive model
GLM	generalized linear model
IVAR	integrated variance
LUE	linear unbiased estimator
MIV LUE	minimum integrated variance linear unbiased estimator
MML	medical marijuana law
\mathbb{N}	the set of natural numbers
NAL	Naloxone access law
OLS	Ordinary Least Squares
PDAPs	Prescription Drug Abuse Policy System
PDMP	Prescription Drug Monitoring Program
\mathbb{R}^d	the Real space of dimension d
RML	recreational marijuana law
SNDM	Structural Nested Distribution Model
SNM	Structural Nested Model
SNMM	Structural Nested Mean Model
SNRRM	Structural Nested Risk Ratio Model
SUTVA	Stable Unit Treatment Value Assumption
U.S.	United States of America

Chapter 1

Introduction

It is widely known that *correlation does not imply causation*. Correlation indicates an association between two phenomenon, while in a causal relationship, there is a cause and an effect. One important difference between causal relationships and associations is that the former accounts for confounding, whereas the latter may not. *Confounding* occurs when there are other factors that may affect both phenomenon, and therefore, may affect the relationship between the two phenomenon. For example, Figure 1-1 shows the trends of ice cream sales and number of shark attacks in a year. Since the two trend lines move closely together, one might conclude that ice cream sales causes shark attacks, or vice versa. However, upon further inspection, one may notice that seasonality is a confounding variable that explains the patterns seen—in the summer, more people are buying ice cream and swimming in the ocean where sharks live. Hence, without accounting for potential confounders, one might not be able to accurately determine the relationship between various phenomenon.

More formally, causal inference is the process of analyzing data such that the effect of one phenomenon (*treatment* or *intervention*) on another phenomenon (*response* or *outcome*) can be determined. Causal inference methods have been applied in various fields to answer different questions. For example, some questions discussed in this thesis include: *how effective is a new medication on the prevention of cardiovascular disease? Does showing ads for a new game on a user's Facebook Newsfeed increase sales of the new game? What is the effect of drug-induced homicide prosecutions on*

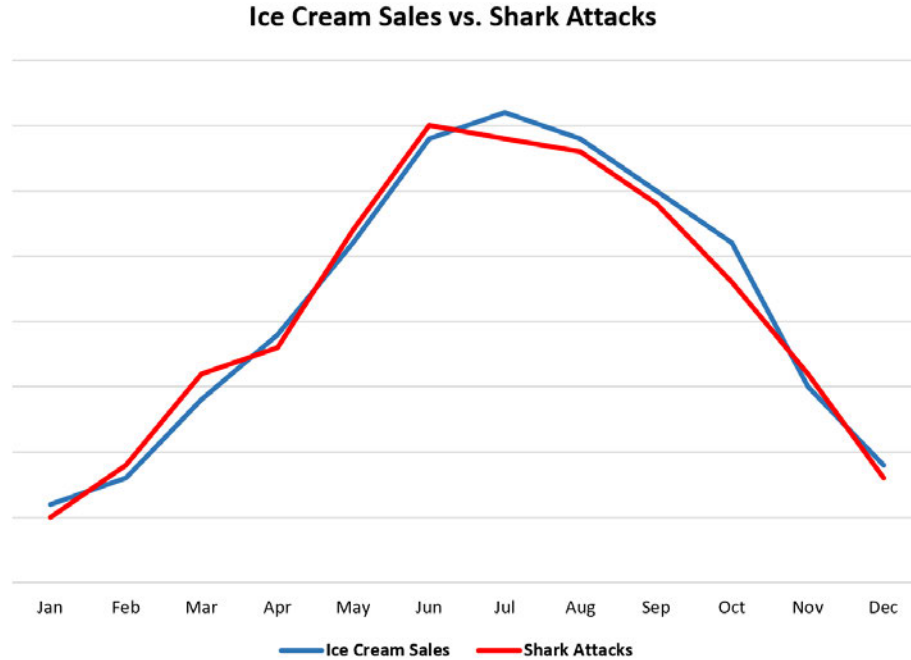


Figure 1.1: Trends for ice cream sales and shark attacks in a year. Retrieved from <https://www.statology.org/correlation-does-not-imply-causation-examples/>.

drug overdose deaths? To answer these questions, causal inference techniques may be used. In particular, the goal of causal inference is to estimate the *treatment effect*, in which outcomes under treatment versus no treatment (*control*) are compared. More generally, we can compare the outcomes under different treatments.

1.1 Potential Outcomes Framework

Treatment assignments for a sample of n units are given by a treatment allocation $\mathbf{z} \in \{0, \dots, m\}^n$. Each $Z_i \in \{0, \dots, m\}$ indicates the treatment assignment for unit i . We will interchangeably denote the treatment assignment of a unit i as Z_i or A_i . For example, we first consider the question of *how effective is a new medication on the prevention of cardiovascular disease?* Figure 1.2 shows a causal diagram containing different components for the example, where circles indicate the different phenomenon and the arrows indicate the directions of the causal relationships. In this example, the

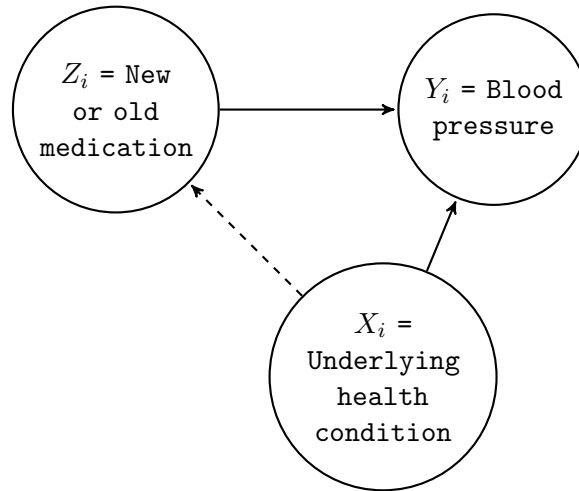


Figure 1.2: Causal diagram for the question: *how effective is a new medication on the prevention of cardiovascular disease?*

units are patients, and the plausible treatment values are whether a patient receives the new medication or the old medication. We denote the control (the old medication in this example) as $Z_i = 0$, and we denote the treatment (the new medication in this example) as $Z_i = 1$. We denote the confounding variables, such as a patient’s underlying health condition in this example, as X_i . Note that the arrow between X_i and Z_i is dashed—whether this arrow is present depends on the study setting. In an experiment, treatment is randomized, and so the treatment is not affected by confounding variables, but in an observational study, treatment is not randomized and may depend on confounding variables. The outcome of interest in this example, denoted as Y_i , is a patient’s blood pressure.

As is common with many research papers, we use the Rubin causal model (Rubin, 1974) or the *potential outcomes framework* to estimate treatment effects. Potential outcomes are outcomes under a given treatment allocation. We denote the potential outcome of patient i under treatment allocation \mathbf{z} as $Y_i(\mathbf{z}) \in \mathbb{R}$. For example, $Y_i(0)$ indicates the blood pressure of patient i had they been given the old medication,

and $Y_i(1)$ indicates the blood pressure of patient i had they been given the new medication. Note, however, in reality, we only observe the treatment allocation \mathbf{z}^{obs} , and so we only observe one potential outcome for unit i , namely $Y_i(\mathbf{z}^{obs})$. We denote the observed outcome of unit i as $Y_i^{obs} = Y_i(\mathbf{z}^{obs})$. This is the Fundamental Problem of Causal Inference (Holland, 1986). Since only one potential outcome is observed for a unit, estimating treatment effects becomes a missing data problem, where we impute missing potential outcomes to estimate treatment effects.

1.2 Treatment Effects

Under the potential outcomes framework, the unit-level *treatment effect* given treatment allocations \mathbf{z} and \mathbf{z}' is commonly defined as the difference between potential outcomes for a unit under \mathbf{z} versus \mathbf{z}' :

$$\tau_i(\mathbf{z}, \mathbf{z}') = Y_i(\mathbf{z}) - Y_i(\mathbf{z}'). \quad (1.1)$$

For example, the treatment effect for patient i when every patient receives the new medication versus when every patient receives the old medication is defined as $\tau_i(\mathbf{1}, \mathbf{0}) = Y_i(\mathbf{1}) - Y_i(\mathbf{0})$, where $\mathbf{0}$ and $\mathbf{1}$ are vectors of all zeros and ones, respectively. Note that treatment effects can also be defined using odds ratios or risk ratios of potential outcomes, to name just a few. In the literature, the *average treatment effect* is more commonly studied than the unit-level treatment effect:

$$\bar{\tau}(\mathbf{z}, \mathbf{z}') = \mathbb{E}(Y_i(\mathbf{z}) - Y_i(\mathbf{z}')), \quad (1.2)$$

where the expectation is taken over the population of units.

1.3 Common Assumptions in the Causal Inference Framework

As with many statistical concepts, there are several core assumptions in the causal inference framework. The first assumption is the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980).

Assumption 1.1 (Stable Unit Treatment Value Assumption). There are two parts to the Stable Unit Treatment Value Assumption.

1. *No interference*: The potential outcomes of every unit do not depend on other units' treatments.
2. *Consistency*: For each unit, there are no different versions of a treatment that could lead to different potential outcomes.

The *no interference* part of SUTVA states that potential outcomes of a unit only depend on the treatment assigned to the unit. The *consistency* part of SUTVA assumes the following:

Assumption 1.2 (Consistency). If $\mathbf{Z}^{obs} = \mathbf{z}$,

$$Y_i(\mathbf{z}) = Y_i^{obs}. \tag{1.3}$$

That is, the potential outcome under treatment assignment \mathbf{z} is equal to the observed outcome if \mathbf{z} is the observed treatment assignment. Under consistency, we can estimate potential outcomes using the corresponding observed outcomes in the data. Under SUTVA, the potential outcomes are given as $Y_i(\mathbf{z}) = Y_i(z_i)$ where z_i is the treatment assignment of unit i given by \mathbf{z} . For example, the blood pressure of a patient only depends on whether they received the old or new medication and does not depend on the type of medication other patients received.

Although we usually assume that SUTVA holds, SUTVA may not be applicable in all settings. For example, in social networks, a unit's outcome may depend not

only on the unit’s treatment but also on the treatment assignments of other units in the network. Hence, when SUTVA does not hold, we have to account for the treatment assignments of other units. In Chapter 2, we propose estimators for causal effects when we relax SUTVA in experimental settings. Results apply to settings when we relax either the first or second part of SUTVA, but we focus on relaxing the no interference assumption. When interference is present, estimation of causal effects becomes harder because we need to account for both the direct effect from the treatment and interference effects from other units. To understand how treatment affects outcomes, Aronow et al. (2017) proposed exposure mappings that map treatment allocations to exposures. Under the assumption that an exposure contains all the necessary information on how the treatment allocations affect the outcomes, we focus on potential outcomes and treatment effects given by the exposures. Given the exposures, we propose linear estimators for general causal effects under the assumption of additivity. Under additivity, contrasts of potential outcomes under different exposures are equivalent, and so the set the of estimators considered grows. We show in Chapter 2 that optimal linear estimators, which are unbiased and have minimum integrated variance (Sussman and Airoidi, 2017), generally leverage information from all observed units, even if the units’ exposures are seemingly unrelated to the estimand of interest.

Assumption 1.3 (Positivity). We assume that for all $\mathbf{z} \in \{0, \dots, m\}^n$, we have $\mathbb{P}(\mathbf{Z}^{obs} = \mathbf{z}) \in (0, 1)$.

Positivity ensures that we may observe units under all possible treatment allocations. Furthermore, treatment is not deterministic, i.e. there are no treatment allocations that always occur or never occur. In addition, we make the no unmeasured confounding assumption:

Assumption 1.4 (No Unmeasured Confounding Assumption). Given potential outcomes $Y_i(\mathbf{z})$ under treatment allocation \mathbf{z} and confounding variables \mathbf{X} , we assume

that

$$Y_i(\mathbf{z}) \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}. \quad (1.4)$$

That is, given confounding variables \mathbf{X} , the treatment allocation \mathbf{Z} does not depend on potential outcome $Y_i(\mathbf{z})$, and vice versa. Note that here, we use potential outcomes $Y_i(\mathbf{z})$ and not the observed outcomes Y_i^{obs} since we do assume that observed outcomes are dependent on the treatment. Assumption 1.4 also says that we have accounted for all confounding variables, leading to accurate treatment effects.

1.4 Repeated Outcomes Setting

In traditional causal inference settings, units receive treatment once, and outcomes are observed after the treatment. However, there are examples where data is collected over time in order to estimate treatment effects, such as estimating the effects of policy measures on drug overdose deaths. We call these settings *repeated outcomes* settings. In studies with repeated outcomes, there are additional complexities which complicate the estimation of treatment effects. First, units may be treated multiple times and at different time points, so the treatment and control groups may differ depending on the time point. Furthermore, time may have an effect on both the treatments and outcomes, and so we take time into account as a potential confounder. In addition, the treatment, potential outcomes, and confounding variables may depend on the past treatments, potential outcomes, and confounding variables. In particular, *confounding by indication* may occur when predictors of potential outcomes may be indicators of treatment. For example, states may enact certain policies in hopes of reducing the number of drug overdose deaths, especially if there was an increase in drug overdose deaths in the past. Because of this indication of the treatment, there may be differences between the treatment and control groups due to selection bias.

Hence, additional confounding variables in repeated outcomes settings need to be taken into account in order for estimated treatment effects to be accurate.

We estimate the effect of exposure to drug-induced homicide (DIH) prosecutions on drug overdose deaths using data collected over time in Chapter 3. Although DIH prosecutions are intended to deter people from delivering drugs that lead to drug overdose deaths, the fear of being prosecuted may lead to people not seeking help for victims and hence, may actually be increasing drug overdose deaths. Despite the conflicting potential results, the effect of DIH prosecutions on drug overdose deaths have never been empirically assessed. Furthermore, we illustrate, through example, a potential issue, which has only recently garnered more interest in the literature (Borusyak and Jaravel, 2017; Borusyak et al., 2021; Sun and Abraham, 2021; Roth et al., 2022), of using a constant treatment effect model in a repeated outcomes setting where the treatment time is staggered and treatment effects are heterogeneous. In particular, the constant treatment effect is a linear combination of individual treatment effects, but where *weights are negative*. This is an issue since there may be conflicting results depending on the model specification.

In Chapter 4, we propose a new class of Structural Nested Mean Models (SNMMs) to estimate the treatment effect on a risk ratio scale through modeling a function of the treatment. SNMMs (Robins et al., 1992; Robins, 1994, 1998; Lok and DeGruttola, 2012) are time-dependent models used to estimate treatment effects in repeated outcomes settings while taking covariates at each time period into account. Prior SNMMs have been developed on the risk differences scale and have been often used to estimate the effect of treatment initiation. We extend SNMMs to the risk ratio scale and call this new class of SNMMs the Structural Nested Risk Ratio Model (SNRRM). We further show that we can estimate treatment effects by modeling a general function, which we choose to be a function that can be modeled by a generalized linear

model, of the treatment. We show that our treatment effect estimators are consistent and asymptotically normal, and we apply SNRRMs to estimate the effect of DIH prosecutions on drug overdose deaths.

1.5 Overview of Dissertation

In Chapter 2, we propose estimators for general causal effects in an experimental setting for exposure models under the additive exposures assumption. We characterize the set of linear unbiased estimators and optimal estimators with minimum integrated variance through the estimator’s support. In Chapter 3, we shift our focus to observational studies where we estimate treatment effects in repeated outcomes settings. We estimate the effect of having any drug-induced homicide (DIH) prosecutions reported by the media versus none on drug overdose deaths using two models: 1) a difference-in-differences-like logistic generalized additive model (GAM) with smoothed time effects where we assume a constant treatment effect and 2) a GAM with a linear link function where we assume that treatment may depend on treatment duration. In Chapter 4, we extend the Structural Nested Mean Model (SNMM) to the risk ratio scale—Structural Nested Risk Ratio Models (SNRRMs). We then apply SNRRMs to conduct an analysis of the effect of DIH prosecutions on drug overdose deaths.

Chapter 2

Treatment Effects Under Exposure Models

2.1 Introduction

Traditionally, one estimates the direct effect of a single treatment under SUTVA Assumption 1.1. However, as we extend causal inference methods to various fields, SUTVA may no longer hold, and so the need for estimation of general treatment effects grows. For example, consider the question: *does showing advertisements for a new game on a user's Facebook Newsfeed increase sales of the new game?* Figure 2-1 shows the causal diagram for this example, where the treatment is whether user i sees an advertisement for the new game, the outcome is whether user i buys the game, and user i 's interests may be an important factor to take into account. We focus on the experiment setting where treatment is randomized in this chapter, and so there is no arrow between X_i and Z_i . Consider users j and k who are connected to user i on Facebook. If users j and/or k see an advertisement for the game, they may tell user i about the game, which may then influence whether user i purchases the game. Hence, whether user i buys the game not only depends on whether he/she sees the advertisement for the game, but also on whether users j and k see the advertisement for the game.

As we stray away from the classical settings of causal inference where SUTVA holds, the estimation of causal effects become more difficult. We have to consider not

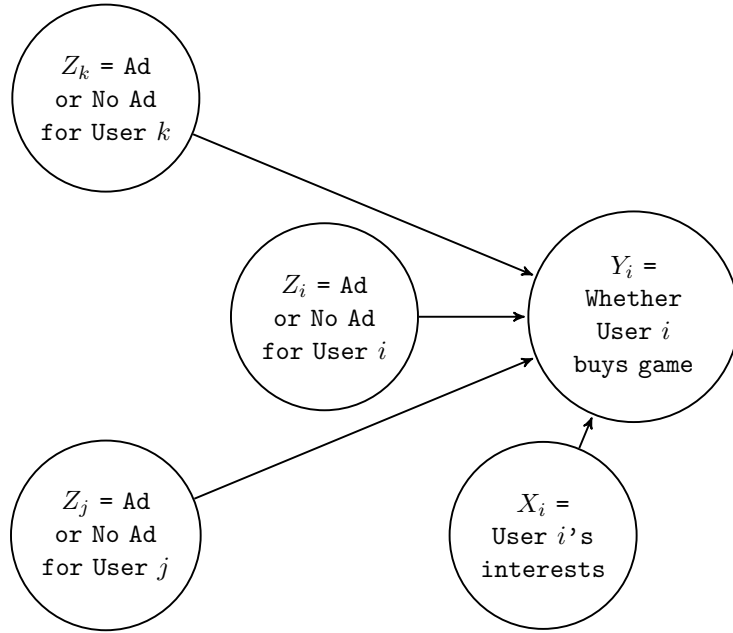


Figure 2.1: Causal diagram for the question: *does showing advertisements for a new game on a user's Facebook Newsfeed increase sales of the new game?*

only how the treatment directly affects the outcome but also how the treatment potentially indirectly affects the outcome. Since there can be nuances in how a treatment allocation affects the potential outcomes, we use an exposure mapping (Aronow et al., 2017) to map the relationship between the treatment allocation and the potential outcomes. Given an exposure mapping, we assume that the potential outcomes depend on the treatment allocation only through the exposures. We then estimate general causal effects for exposure models.

In general, one prefers to make fewer assumptions so that results are generalizable. However, we assume that treatment effects are additive, which provides statistical advantages. Under additivity, contrasts of potential outcomes under different exposures are equivalent, and so we do not have to estimate each estimand separately. Furthermore, when additivity holds, we are no longer limited to only one estimator for the estimand of interest, and so the set of estimators considered grows. In particular, if

units with exposures corresponding to the estimand of interest are not observed, we can still estimate causal effects using units under different exposures corresponding to estimands that are equivalent to the estimand of interest. Even when exposures corresponding to the estimand of interest are observed, we can leverage information from other units, leading to better estimates of treatment effects.

In this chapter, we propose linear unbiased estimators for general causal effects under the additive exposure assumption in an experimental setting. We characterize the set of linear unbiased estimators and define an affine basis for the set of linear unbiased estimators. We further characterize a set of optimal estimators with minimum integrated variance. Lastly, we compare the proposed optimal linear unbiased estimators with other linear unbiased estimators through a series of simulations under various settings.

2.2 Background

In early work of causal inference (Neyman, 1923; Rubin, 1974), estimation of treatment effects was commonly done under SUTVA. Since then, there has been a growing interest in estimating treatment effects in settings where SUTVA is relaxed.

Although the simplest setting for an experiment consists of only two treatments, usually treatment and control, there are many instances where one may want to compare the effects of more than two treatments or even different versions of a treatment. For example, one may want to evaluate the effects of different forms of advertisements or different wording of a particular advertisement on the sales of a game. Methods for propensity scores (Rosenbaum and Rubin, 1983) developed for binary treatment were first applied to multiple treatment settings through a series of binary comparisons (SBC) where pairs of treatments are compared as in a binary treatment setting while ignoring units receiving other treatments (Lechner, 2001, 2002). Imbens

(2000) then introduced the generalized propensity score, which was an extension of the propensity score, to settings with multiple treatments in order to adjust for differences in pre-treatment covariates. Using the generalized propensity score, one could then use inverse propensity score weighting or matching to estimate treatment effects. Since then, the generalized propensity score has been applied and extended (Feng et al., 2012; McCaffrey et al., 2013). Matching methods were also initially applied to multiple treatment settings by focusing on pairs of treatments and matching units using the propensity score developed for classical settings (Rassen et al., 2011). Later, matching methods were developed where units with similar generalized propensity scores for each treatment are grouped together (Rassen et al., 2013; Lopez and Gutman, 2017). Similar to prior work, our estimators rely on the probability of exposures. However, we do not use these probabilities for matching. Instead, we focus on estimating treatment effects using inverse probability of exposure weighting.

There have been a growing body of work in relaxing the no interference assumption of SUTVA in which units' outcomes do not depend on other units' treatments. This is likely because there are many settings in which *interference* or *spillover* effects (Cox, 1958; Rubin, 1980) may be present. Early work in estimating treatment effects in the presence of interference began with the assumption that treatment effects may spill over through time, focusing on *residual* effects that may be present from the preceding time point (Grizzle, 1965; Kershner and Federer, 1981). Later, spatial interference attracted attention, where neighboring units or units within the same block may be dependent (Besag and Kempton, 1986; David and Kempton, 1996). Since then, interference has been extended to settings of *partial interference*, where units within a cluster may be dependent but units in different clusters are assumed to be independent (Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen and VanderWeele, 2012).

More recently, with the rise of social media, there has been a growing interest in estimating causal effects in the presence of interference in networks (Ugander et al., 2013; Eckles et al., 2017; Athey et al., 2018; Aronow et al., 2017; Sussman and Airoidi, 2017; Forastiere et al., 2021). This is because networks can be used to represent relationships between units and interference effects may be passed through the connections of the network. While there are papers that estimate treatment effects accounting for units that are up to k connections away for some k through testing exact hypotheses (Athey et al., 2018), we focus on the assumption that interference only occurs for neighboring units. There has also been recent literature on applying methods of estimating causal effects in networks in observational studies (Liu et al., 2019; Barkley et al., 2020; Forastiere et al., 2021). However, we focus on experimental settings, where the treatment is randomized.

Our work extends Aronow et al. (2017), who proposed unbiased estimators for causal effects under general interference, and Sussman and Airoidi (2017), who proposed unbiased estimators for the direct treatment effect with minimum integrated variance under network interference. Under an exposure model (Aronow et al., 2017) in which a treatment allocation is assigned to exposures through an *exposure mapping*, Aronow et al. (2017) proposed two-term unbiased estimators for estimands of interest using Horvitz-Thompson estimators (Horvitz and Thompson, 1952). We propose linear unbiased estimators for unit-level causal effects, but we deviate from Aronow et al. (2017) in that we assume that treatment effects are additive. Additivity provides flexibility such that different estimands under different exposures are equivalent. Assuming additivity, our proposed linear unbiased estimators may place non-zero weights on exposures that are “seemingly unrelated” to the estimand of interest. Furthermore, we deviate from Sussman and Airoidi (2017) in that we estimate general treatment effects, which include both direct and indirect treatment effects.

However, like Sussman and Airoidi (2017), we further characterize an optimal subset of linear unbiased estimators that have minimum integrated variance.

2.3 Exposure Models

Consider a randomized experiment with n units that are together assigned a treatment allocation $\mathbf{z} \in \{0, \dots, m\}^n$ where $z_i \in \{0, \dots, m\}$ represents the treatment that unit i receives. The experimental design of a randomized trial is given by the probability of a treatment allocation, denoted by $p : \{0, \dots, m\}^n \rightarrow [0, 1]$. Since we focus on a randomized experiment setting, we assume that the design is known. The treatment allocation \mathbf{z} provides information, such as unit treatment assignments, number of treated units, etc., which can be used to determine a unit's outcome. While in general $Y_i(\mathbf{z})$ depends on all of \mathbf{z} , we often assume that the outcome only depends on specific aspects of the treatment allocation. For example, under SUTVA, the outcome of a unit only depends on the unit's treatment. That is, under SUTVA, $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$ whenever $z_i = z'_i$. To capture the dependencies of potential outcomes on treatment allocations, Aronow et al. (2017) proposed exposure models as an alternative representation of the potential outcomes that can account for these pathologies while still limiting the complexity of the model.

Exposure models are given by exposure mappings, which are used to capture all the information needed from a treatment allocation to determine a unit's potential outcome:

Definition 2.3.1 (Exposure mapping). Let \mathcal{E} denote the set of exposures. For each unit i , an exposure mapping $f(i, \cdot) : \{0, \dots, m\}^n \rightarrow \mathcal{E}$ maps each treatment allocation to an exposure in the set \mathcal{E} .

Exposure mappings are flexible and can be defined in various ways. However, we assume that this exposure mapping is known and we further assume the following:

Assumption 2.1. For any pair $\mathbf{z}, \mathbf{z}' \in \{0, \dots, m\}^n$, $f(i, \mathbf{z}) = f(i, \mathbf{z}') = \vec{e}$ implies that $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$. That is, we can write

$$Y_i(\mathbf{z}) = Y_i(\vec{e}). \quad (2.1)$$

Assumption 2.1 states that the potential outcome of a unit is determined only by its exposure, and so we assume that potential outcomes are dependent on treatment allocations through the exposures. By using exposures, we reduce the number of possible potential outcomes for unit i from $(m + 1)^n$ to $|\mathcal{E}|$, the size of set \mathcal{E} .

Example 1 (SUTVA). Under SUTVA, a unit's outcome only depends on its own treatment assignment. Here, $\mathcal{E} = \{0, \dots, m\}$, and the exposure mapping is given by $f(i, \mathbf{z}) = z_i \in \{0, \dots, m\}$. Potential outcomes are then given by $Y_i(\mathbf{z}) = Y_i(z_i)$.

Example 2 (Network Interference). Consider a network amongst the n units, given by the $n \times n$ adjacency matrix A . Suppose that a unit's outcome can depend on its own treatment assignment, which is binary, and the treatment assignments of other units in the network. In particular, suppose that the potential outcome of a unit only depends on the number of neighbors that are treated and not necessarily which ones (Sussman and Airoldi, 2017). Note that SUTVA no longer holds since network interference is present. Here, $\mathcal{E} = \{0, \dots, n - 1\} \times \{0, 1\}$, and the exposure mapping is defined as $f(i, \mathbf{z}) = (d_i^{\mathbf{z}}, z_i)$, where $d_i^{\mathbf{z}} = (A^T \mathbf{z})_i$ is the number of treated neighbors or the *treated degree*. Note that for a unit i , the treated degree $d_i^{\mathbf{z}} \in \{0, \dots, d_i\}$ where d_i is the degree of unit i .

Exposures are flexibly defined, but they are often represented with multiple exposure components. For example, exposures in Example 2 are given by two exposure components: $\mathcal{E} = \{0, \dots, n - 1\} \times \{0, 1\}$. We use an *exposure vector* to denote exposures with multiple components:

Definition 2.3.2 (Exposure vector). As the exposure set is finite, without loss of generality, we assume the exposure set has the form $\mathcal{E} = \{0, \dots, m_1\} \times \dots \times \{0, \dots, m_K\}$, where $K \geq 1$ is the number of *exposure components*. Exposures, denoted by $\vec{e} \in \mathcal{E}$, are given by exposure vectors: $\vec{e} = (e_1, \dots, e_K) \in \mathcal{E}$.

We define the vector of all zeros, denoted as $\vec{e} = \vec{0}$, as the *baseline* exposure. We interpret the exposure components as different information given by the exposure mapping. In Example 2, the first exposure component corresponds to the number of treated neighbors for the unit, and the second exposure component corresponds to the treatment assigned to the unit.

Using potential outcomes given by the exposure mappings, we define causal effects given the different exposures. The *causal effect* of $\vec{e} \in \mathcal{E}$ versus $\vec{e}' \in \mathcal{E}$ is defined as the difference between potential outcomes under \vec{e} and \vec{e}' . We focus on estimating causal effects for a single unit i :

$$\tau_i(\vec{e}, \vec{e}') = Y_i(\vec{e}) - Y_i(\vec{e}'). \quad (2.2)$$

Since we focus on unit-level effects, we simplify the notation by dropping the subscript i throughout the rest of the paper. Following from prior work (Aronow et al., 2017), we use unit-level causal effects to estimate the average causal effects through averaging unit-level estimates.

The set of estimands is given by the contrasts in exposures. In general, potential outcomes under an exposure can be decomposed into the baseline, the corresponding direct effects for each exposure component, and interactions between the effects from multiple exposure components. As the number of exposures, and especially the number of exposure components, increase, the number of interaction effects become large. Instead, we assume that additivity holds:

Assumption 2.2 (Additivity). Consider exposure vectors $\vec{e}, \vec{e}' \in \mathcal{E}$. Exposures are additive if, whenever $\vec{e} - \vec{e}' > \vec{0}$,

$$Y(\vec{e}) - Y(\vec{e}') = Y(\vec{e} - \vec{e}') - Y(\vec{0}). \quad (2.3)$$

Under additivity, there are no interaction effects. That is, the difference in potential outcomes given two different exposures only depends on the difference in exposure

components. We can then isolate the effect of the k th exposure component by removing the effects of all other components. To do this, we can add and subtract potential outcomes under different exposures so that the net value of all other exposure components besides the k th exposure component is zero. Additivity provides statistical advantages since certain contrasts are now equivalent, such as

$$Y(m_1, e_2, \dots, e_K) - Y(0, e_2, \dots, e_K) = Y(m_1, e'_2, \dots, e'_K) - Y(0, e'_2, \dots, e'_K)$$

for $e_k \neq e'_k$ for $k \in \{2, \dots, K\}$. Since contrasts in potential outcomes under different exposures are equivalent under additivity, the number of contrasts we consider is then reduced to $\sum_{k=1}^K m_k$.

Since there are no interaction effects under additivity, we can write the potential outcome under exposure $\vec{e} = (e_1, \dots, e_K)$ as:

$$\begin{aligned} Y(e_1, e_2, \dots, e_K) &= Y(0, \dots, 0) \\ &\quad + [Y(e_1, 0, \dots, 0) - Y(0, \dots, 0)] \\ &\quad + [Y(0, e_2, 0, \dots, 0) - Y(0, \dots, 0)] \\ &\quad \dots \\ &\quad + [Y(0, \dots, 0, e_K) - Y(0, \dots, 0)], \end{aligned} \tag{2.4}$$

where the first summand indicates the baseline, and the other summands indicate the various causal effects for the k th exposure component at level $e_k \in \{1, \dots, m_k\}$. We denote the causal effect of the k th exposure component at level j_k as:

$$\theta_{k,j_k} = Y(0, \dots, 0, j_k, 0, \dots, 0) - Y(\vec{0}). \tag{2.5}$$

Let the parameter set, denoted by Θ , contain the baseline parameter $\alpha = Y(0, \dots, 0)$ and parameters θ_{k,j_k} for all $k \in \{1, \dots, K\}$ and $j_k \in \{1, \dots, m_k\}$. Under additivity,

potential outcomes are given as:

$$Y(\vec{e}) = \alpha + \sum_{k=1}^K \sum_{j_k=1}^{m_k} \theta_{k,j_k} \mathbb{I}\{e_k = j_k\}. \quad (2.6)$$

Example 1 (continuing from p.16). Under SUTVA, we define $\vec{e} = z_i \in \{0, \dots, m\}$. The unit-level causal effect for the first (and only) exposure component when treatment is equal to m versus when treatment is equal to zero is given by $\theta_{1,m} = Y(m) - Y(0)$.

Example 2 (continuing from p.16). Under network interference with binary treatment, $\vec{e} = (d_i^z, z_i)$. The causal effect of the first exposure component when unit i has fully treated neighbors versus no treated neighbors is given by $\theta_{1,d_i} = Y(d_i, 0) - Y(0, 0)$. Here, θ_{1,d_i} is the *unit-level interference effect*. Note that we defined θ_{1,d_i} using an estimand with exposures where $z_i = 0$. However, under additivity, contrasts in potential outcomes under different exposures are equivalent, and so $Y(d_i, 1) - Y(0, 1)$ is also an estimand for the unit-level interference effect.

2.4 Linear Unbiased Estimators

In this section, we propose estimators for the unit-level causal effect. Without the loss of generality, for the rest of this chapter, we focus on estimating the effect for a single unit when the first exposure component is m_1 , compared to baseline. No generality is lost since we can remap the exposures to a new exposure set where the k th component is mapped to the first component and the e_k th level is mapped to the maximum.

Linear estimators of the unit-level causal effect of the first exposure component are of the form:

$$\hat{\theta}_{1,m_1} = w(\mathbf{z}^{obs})Y(\vec{e}^{obs}),$$

where $w : \{0, \dots, m\}^n \rightarrow \mathbb{R}$ is a weight function depending only on the treatment allocation \mathbf{z}^{obs} and $Y(\vec{e}^{obs})$ is the observed outcome under observed exposure \vec{e}^{obs} . We

consider only linear estimators with weights that depend only on the unit's exposure, i.e. $w : \mathcal{E} \rightarrow \mathbb{R}$. We denote the support of w , or equivalently of the estimator $\hat{\theta}_{1,m_1}$, as $\text{supp}(\hat{\theta}_{1,m_1}) = \{\vec{e} \in \mathcal{E} : w(\vec{e}) \neq 0\}$. Hence, the linear estimators we consider are of the form:

$$\hat{\theta}_{1,m_1} = w(\vec{e}^{obs})Y(\vec{e}^{obs}). \quad (2.7)$$

Linear estimators under this consideration include Horvitz-Thompson inverse propensity score weighting estimators (Horvitz and Thompson, 1952). The Horvitz-Thompson estimator for a potential outcome is given by:

$$HT_{\vec{e}} = \frac{Y(\vec{e}^{obs})}{p(\vec{e})} \mathbb{I}\{\vec{e}^{obs} = \vec{e}\}, \quad (2.8)$$

where $p(\vec{e}) = \mathbb{P}(\vec{e}^{obs} = \vec{e})$ is the probability of observing the exposure \vec{e} , which is given by the design probabilities. Since the experimental design is known, the probabilities of exposures are also known. Furthermore, $\mathbb{I}\{\vec{e}^{obs} = \vec{e}\}$ indicates whether the exposure \vec{e} is observed. Given exposures $\vec{0}, \vec{e} \in \mathcal{E}$, where $\vec{e} = (m_1, 0, \dots, 0)$, Aronow et al. (2017) proposed estimators for the causal effect using Horvitz-Thompson inverse propensity score weighting estimators:

$$\hat{\theta}_{1,m_1} = HT_{\vec{e}} - HT_{\vec{0}} = Y(\vec{e}^{obs}) \left[\frac{\mathbb{I}\{\vec{e}^{obs} = \vec{e}\}}{p(\vec{e})} - \frac{\mathbb{I}\{\vec{e}^{obs} = \vec{0}\}}{p(\vec{0})} \right].$$

Since the weight only depends on the unit's exposure, the set of linear estimators we consider include these Horvitz-Thompson estimators. On the other hand, the naive difference in means estimator have weights:

$$w(\vec{e}_i) = \frac{\mathbb{I}\{\vec{e}_i^{obs} = \vec{e}\}}{\sum_{j=1}^n \mathbb{I}\{\vec{e}_j^{obs} = \vec{e}\}} - \frac{\mathbb{I}\{\vec{e}_i^{obs} = \vec{0}\}}{\sum_{j=1}^n \mathbb{I}\{\vec{e}_j^{obs} = \vec{0}\}} \quad (2.9)$$

for unit i , where $\vec{e}_i^{obs}, \vec{e}_j^{obs}$ are the observed exposure vectors for units i and j , respec-

tively, for $i, j \in \{1, \dots, n\}$. Equation (2.9) shows that the denominator of the weight for unit i depends on the exposures of other units, and so the linear estimators we consider preclude naive estimators, except under certain highly symmetric designs (e.g. a Completely Randomized Design).

2.4.1 Unbiased Estimators

As a first step to limit the set of linear estimators considered, we further focus on linear estimators that are unbiased for the unit-level causal effect. An estimator is *unbiased* for the unit-level causal effect of the first exposure component if

$$\mathbb{E}(\hat{\theta}_{1,m_1}) = \theta_{1,m_1}. \quad (2.10)$$

Under additivity, linear unbiased estimators (LUEs) exist under certain constraints, which are given by the following proposition.

Proposition 1. Assuming additive exposures, a linear estimator $\hat{\theta}_{1,m_1}$ is unbiased for θ_{1,m_1} if and only if the following constraints hold:

$$\begin{aligned} \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e}) &= 0 && (\alpha \text{ constraints}) \\ \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m_1\} &= 1 && (\theta_{1,m_1} \text{ constraints}) \\ \forall m : m \in \{1, \dots, m_1 - 1\} \quad \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m\} &= 0 && (\theta_{1,m} \text{ constraints}) \\ \forall k, j_k : k \in \{2, \dots, K\}, j_k \in \{1, \dots, m_k\} \quad \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_k = j_k\} &= 0. && (\theta_{k,j_k} \text{ constraints}) \end{aligned}$$

Denote the set of linear unbiased estimators as \mathcal{U} . Given the linear constraints, the size of \mathcal{U} , denoted as $|\mathcal{U}|$, is $|\mathcal{U}| = \prod_{k=1}^K (m_k + 1) - \sum_{k=1}^K m_k - 1$. Here, the product $\prod_{k=1}^K (m_k + 1)$ corresponds to the number of exposures in \mathcal{E} , and $\sum_{k=1}^K m_k + 1$ corresponds to the number of linear constraints. The linear constraints for unbiasedness ensure that when the estimator is averaged across exposures, it leads to a coefficient of 1 in front of the θ_{1,m_1} term, while the coefficients for the other terms are zero.

Hence, when we compute the expected value of $\hat{\theta}_{1,m_1}$, we obtain the parameter of interest θ_{1,m_1} (see Appendix A.1). Example of LUEs include the Horvitz-Thompson inverse probability estimators. Note that unbiasedness holds given the constraints in Proposition 1 only under additivity. Without additivity, we will require more constraints. Hence, under additivity, we consider more estimators that would otherwise be biased.

Example 2 (continuing from p.16). In the network interference example, consider the estimators

$$\begin{aligned}\hat{\theta}_{1,d_i}^{\text{two term},0} &= HT_{(d_i,0)} - HT_{(0,0)} \\ \hat{\theta}_{1,d_i}^{\text{two term},1} &= HT_{(d_i,1)} - HT_{(0,1)} \\ \hat{\theta}_{1,d_i}^{\text{Avg}} &= \frac{1}{2} (HT_{(d_i,0)} - HT_{(0,0)} + HT_{(d_i,1)} - HT_{(0,1)}) \\ \hat{\theta}_{1,d_i}^{\text{four term},2} &= HT_{(d_i,1)} - HT_{(2,1)} + HT_{(2,0)} - HT_{(0,0)}.\end{aligned}$$

Under additivity, all the estimators above are linear unbiased estimators for θ_{1,d_i} . For example, $\hat{\theta}_{1,d_i}^{\text{two term},1}$ introduces the parameter $\theta_{2,1}$ by placing non-zero weight on the $HT_{(d_i,1)}$ term, but $\theta_{2,1}$ is then canceled by the $HT_{(0,1)}$ term. Furthermore, the baseline α is cancelled, and so the parameter that remains is the parameter of interest θ_{1,m_1} . This holds for $\hat{\theta}_{1,d_i}^{\text{Avg}}$, which leverages both estimators $\hat{\theta}_{1,d_i}^{\text{two term},0}$ and $\hat{\theta}_{1,d_i}^{\text{two term},1}$, and $\hat{\theta}_{1,d_i}^{\text{four term},2}$, which leverages “seemingly unrelated” exposures such as $(2,1)$ and $(2,0)$. However, if additivity does not hold, then $\hat{\theta}_{1,d_i}^{\text{two term},0}$ is the only linear unbiased estimator for θ_{1,d_i} . For example, $\hat{\theta}_{1,d_i}^{\text{two term},1}$ is no longer unbiased for θ_{1,d_i} , and the bias is equal to the interference plus the interaction term. Although $\hat{\theta}_{1,d_i}^{\text{two term},1}$ is not unbiased for θ_{1,d_i} when additivity does not hold, it is unbiased for another estimand, namely the estimand that is the sum of the interference and interaction effect.

2.5 Atomic Linear Unbiased Estimators

In the previous section, we defined a class of linear unbiased estimators when additivity holds. Because additivity imposes flexibility in the estimators considered, the class of LUEs can be quite large. However, there are particular subclasses of LUEs that

are of importance. We first focus on a subclass of atomic linear unbiased estimators (ALUEs) which are simpler in terms of its support.

Definition 2.5.1 (Atomic Linear Unbiased Estimators). The estimator $\hat{\theta}_{1,m_1} \in \mathcal{U}$ is atomic within \mathcal{U} if, for all $u \in \mathcal{U}$, if $\text{supp}(u) \subset \text{supp}(\hat{\theta}_{1,m_1})$, then $\text{supp}(u) = \text{supp}(\hat{\theta}_{1,m_1})$.

We denote the set of ALUEs by $\mathcal{A} \subset \mathcal{U}$. The restriction of minimal support reduces the class of linear unbiased estimators considered to estimators whose support cannot be reduced and still be unbiased. Examples of ALUEs include the following two-term and four-term estimators.

Example 1 (continuing from p.16). The treatment effect when SUTVA holds can be estimated using a two-term ALUE:

$$\hat{\theta}_{1,m}^{\text{two term}} = HT_{(m)} - HT_{(0)}.$$

Example 3 (Four Exposure Model). Consider the four exposure model (Aronow et al., 2017), where $z_i \in \{0, 1\}$ and $d_i^z \in \{0, \dots, d_i\}$ is the treated degree of unit i . The exposures are defined as $\vec{e} = (z_i, \mathbb{I}\{d_i^z > 0\}) \in \{0, 1\}^2$. The first exposure component gives the treatment assignment of the unit, and the second exposure component indicates whether network interference is present. Under additivity, we can estimate the direct treatment effect using a two-term ALUE:

$$\hat{\theta}_{1,1}^{\text{two term}} = HT_{(1,1)} - HT_{(0,1)}.$$

Note that $HT_{(1,0)} - HT_{(0,0)}$ is also an ALUE for the direct treatment effect. If we do not assume additivity holds, the only ALUE for the direct treatment effect is $HT_{(1,0)} - HT_{(0,0)}$.

Example 2 (continuing from p.16). We can estimate the network interference effect using a four-term ALUE:

$$\hat{\theta}_{1,d_i}^{\text{four term},d} = HT_{(d_i,1)} - HT_{(d,1)} + HT_{(d,0)} - HT_{(0,0)},$$

where $d \in \{1, \dots, d_i - 1\}$. Note there are also two-term ALUEs for the network interference effect.

In general, the number of Horvitz-Thompson terms in ALUEs can be more than four, but the number of terms in ALUEs is restricted to be even. Generally, the number of terms in ALUEs can be up to $2K$, where K is the number of exposure components. This is because for every exposure component not of interest, whose effects are added by a Horvitz-Thompson term, we need to subtract its effect with another Horvitz-Thompson term so that the estimator is unbiased for θ_{1,m_1} .

2.5.1 Affine Basis for Linear Unbiased Estimators

Atomic linear unbiased estimators are the simplest LUEs in terms of its support. However, we want to be able to generalize properties of ALUEs to the entire class of LUEs. To do this, we relate the class of ALUEs to the rest of the LUEs. We introduce a subclass of ALUEs and show that, with another class of estimators, they form an affine basis for LUEs.

In particular, we focus on a subclass of *monotonic atomic linear unbiased estimators* (MALUEs):

Definition 2.5.2 (Monotonic Atomic Linear Unbiased Estimator). A linear unbiased estimator $\hat{\theta}_{1,m_1} \in \mathcal{A}$ is *monotonic* if the exposures in its support, $\vec{e} \in \text{supp}(\hat{\theta}_{1,m_1})$, can be arranged such that there is a component-wise partial ordering. In particular, each exposure component is simultaneously non-increasing.

Note that all two-term ALUEs are also MALUEs since, by definition, the support only contains exposures (m_1, e_2, \dots, e_K) and $(0, e_2, \dots, e_K)$, where $m_1 > 0$ and all other exposure components are equal. However, ALUEs with more than two terms are not necessarily monotonic.

Example 2 (continuing from p. 16). Consider the following four-term ALUEs for the network interference effect:

$$\begin{aligned}\hat{\theta}_{1,d_i}^{\text{four term},a} &= HT_{(d_i,1)} - HT_{(d,1)} + HT_{(d,0)} - HT_{(0,0)} \\ \hat{\theta}_{1,d_i}^{\text{four term},b} &= HT_{(d_i,0)} - HT_{(d,0)} + HT_{(d,1)} - HT_{(0,1)},\end{aligned}$$

where $d \in \{1, \dots, d_i - 1\}$. Here, $\hat{\theta}_{1,d_i}^{\text{four term},a}$ and $\hat{\theta}_{1,d_i}^{\text{four term},b}$ are both ALUEs, but only $\hat{\theta}_{1,d_i}^{\text{four term},a}$ is also a MALUE. In $\text{supp}(\hat{\theta}_{1,d_i}^{\text{four term},b})$, consider exposures $(d_i, 0)$ and $(d, 1)$, where $d_i > d$ in the first exposure component but $0 < 1$ in the second exposure component. We cannot arrange exposures in $\text{supp}(\hat{\theta}_{1,d_i}^{\text{four term},b})$ according to the component-wise partial order where all exposure components are non-increasing.

We focus on a particular subclass of MALUEs, denoted as $\mathcal{M} \subset \mathcal{A}$:

Lemma 2.1 (\mathcal{M} is affine independent). Consider an ordered set of exposures $\tilde{\mathcal{E}} \subseteq \mathcal{E}$ where

$$\tilde{\mathcal{E}} = \{\vec{e} \in \mathcal{E} : e_1 \in \{1, \dots, m_1 - 1\}, \exists k \in \{2, \dots, K\} \text{ s.t. } e_k \neq 0\} \cup \{\vec{e} \in \mathcal{E} : e_1 = m_1\}$$

such that exposures with $e_1 \in \{1, \dots, m_1 - 1\}$ are first, followed by the exposures with $e_1 = m_1$. Within the subsets of exposures with $e_1 \in \{1, \dots, m_1 - 1\}$ and $e_1 = m_1$, the exposures follow a reverse reflected lexicographic order. Let $\mathcal{M} \subset \mathcal{A}$ contain the following estimators. For each exposure $\vec{e} \in \tilde{\mathcal{E}}$, where $\vec{e} = (e_1, \dots, e_K)$, consider the following:

- If $e_1 \in \{1, \dots, m_1 - 1\}$, add estimator

$$\hat{\theta}_{1,m_1}^{\text{four term}} = HT_{(m_1, e_2, \dots, e_K)} - HT_{(e_1, e_2, \dots, e_K)} + HT_{(e_1, e'_2, \dots, e'_K)} - HT_{(0, e'_2, \dots, e'_K)} \quad (2.11)$$

into \mathcal{M} . Here, $e'_k = 0$ for the first $k \in \{2, \dots, K\}$ such that $e_k \neq 0$ and $e'_{k'} = e_{k'}$ for all other $k' \in \{2, \dots, K\}$ where $k \neq k'$.

- If $e_1 = m_1$, add estimator

$$\hat{\theta}_{1,m_1}^{\text{two term}} = HT_{(m_1, e_2, \dots, e_K)} - HT_{(0, e_2, \dots, e_K)} \quad (2.12)$$

into \mathcal{M} , where $e_k \in \{0, \dots, m_k\}$ for $k \in \{2, \dots, K\}$.

The set \mathcal{M} is affine independent.

By construction, estimators $\hat{\theta}_{1,m_1} \in \mathcal{M}$ have support such that exposures can be ordered such that exposure components are simultaneously non-increasing, and so \mathcal{M} is a subset of MALUEs. Furthermore, note that each estimator $\hat{\theta}_{1,m_1} \in \mathcal{M}$ is uniquely

identifiable by an exposure in $\tilde{\mathcal{E}}$. Namely, the two-term estimators are uniquely identified by exposures where $e_1 = m_1$, and the four term estimators are uniquely identified by exposures where $e_1 \in \{1, \dots, m_1 - 1\}$. To show that the set \mathcal{M} is affine independent, we leverage the fact that the estimators are monotonic and uniquely identifiable (see Appendix A.2.1). Consider estimator $\hat{\theta}$ and let $\hat{\theta} = \sum_{\tilde{\theta} \in \mathcal{M}} g(\tilde{\theta})\tilde{\theta}$. We show that if $\hat{\theta} \in \mathcal{M}$, then

$$g(\tilde{\theta}) = \begin{cases} 1, & \text{if } \tilde{\theta} = \hat{\theta} \\ 0, & \text{otherwise} \end{cases}. \quad (2.13)$$

Since estimators in \mathcal{M} are uniquely identified by the ordered set of exposures $\tilde{\mathcal{E}}$, there is also a natural ordering of the corresponding estimators. Using induction, we iterate through the ordered set of estimators and assign weights $g(\tilde{\theta})$ according to Equation (2.13). At the u th step, if $\tilde{e}^{(u)} \notin \text{supp}(\hat{\theta})$, then $g(\tilde{\theta}^{(u)}) = 0$. Otherwise, since the estimators are ordered according to the estimator's uniquely identifying exposure $\tilde{e}^{(u)} \in \tilde{\mathcal{E}}$, and each estimator is a MALUE, the estimator $\tilde{\theta}^{(u)}$ is the last estimator in \mathcal{M} with $\tilde{e}^{(u)}$ in its support. Hence, if for all $u' < u$, we have $g(\tilde{\theta}^{(u')}) = 0$, $\tilde{e}^{(u)} \in \text{supp}(\hat{\theta})$, and $\hat{\theta} \in \mathcal{M}$, then $\tilde{\theta}^{(u)} = \hat{\theta}$, i.e. $g(\tilde{\theta}^{(u)}) = 1$. If there were at least one $u' < u$ such that $g(\tilde{\theta}^{(u')}) = 1$, then $g(\tilde{\theta}^{(u)}) = 0$ in order for unbiasedness to hold. Since $g(\tilde{\theta}) = 1$ only if $\tilde{\theta} = \hat{\theta}$, then \mathcal{M} is affine independent.

The size of the set of estimators \mathcal{M} , denoted as $|\mathcal{M}|$, is equal to:

$$|\mathcal{M}| = \underbrace{\prod_{k=2}^K (m_k + 1)}_{\text{two term estimators}} + (m_1 - 1) \underbrace{\left[\prod_{k=2}^K (m_k + 1) - 1 \right]}_{\text{four term estimators}}. \quad (2.14)$$

The first term is equal to the number of two-term estimators, which is given by the number of exposures with $e_1 = m_1$. The second term is equal to the number of four-term estimators, where there are $m_1 - 1$ possible values for the first exposure

component, and there are $\prod_{k=2}^K (m_k + 1)$ possible values for e_2, \dots, e_K . We subtract the case when $e_2 = \dots = e_K = 0$; hence the minus one.

Although the estimators in \mathcal{M} are affine independent, there are not enough estimators to span \mathcal{U} . We introduce an additional set of estimators, denoted as \mathcal{Z} :

Definition 2.5.3 (Zero Estimators). Consider a set of estimators \mathcal{Z} , defined as the following:

$$\mathcal{Z} = \{\hat{\theta}_0 : \hat{\theta}_0 = HT_{(0,e_2,\dots,e_K)} - HT_{(0,e_2,0,e_4,\dots,e_K)} - HT_{(0,0,e_3,0,\dots,0)} + HT_{(0,\dots,0)}\}, \quad (2.15)$$

where there are at least two $k, k' \in \{2, \dots, K\}$ such that $e_k \neq 0, e_{k'} \neq 0$, and without the loss of generality, we assumed that $e_2, e_3 \neq 0$.

The size of \mathcal{Z} is:

$$|\mathcal{Z}| = \prod_{k=2}^K (m_k + 1) - 1 - \sum_{k=2}^K m_k. \quad (2.16)$$

The first term is equal to the number of exposures where $e_1 = 0$. Since we require that at least two k, k' are such that $e_k \neq 0, e_{k'} \neq 0$, we subtract the case when $e_1 = \dots = e_K = 0$ and when only one of e_2, \dots, e_K is non-zero. Under additivity, $\mathbb{E}(\hat{\theta}_0) = 0$ (hence we call $\hat{\theta}_0$ a *zero estimator*), which is needed to ensure the unbiased estimation of θ_{1,m_1} . We denote the union of the estimators of \mathcal{M} and the zero estimators as $\hat{\Theta} = \mathcal{M} \cup \mathcal{Z}$.

Theorem 1 (Affine basis for LUE). The set $\hat{\Theta}$ forms an affine basis for the set of linear unbiased estimators \mathcal{U} .

The proof for the affine independence of $\hat{\Theta}$ is very similar to the proof of Lemma 2.1 (see Appendix A.2.2). Note now that

$$\begin{aligned} \text{supp}(\hat{\Theta}) &= \{\vec{e} : \vec{e} \in \mathcal{E}, e_1 = m_1\} \\ &\cup \{\vec{e} : \vec{e} \in \mathcal{E}, e_1 \in \{1, \dots, m_1 - 1\}, \exists k \in \{2, \dots, K\} \text{ s.t. } e_k \neq 0\} \end{aligned}$$

$$\cup \{\vec{e} : \vec{e} \in \mathcal{E}, e_1 = 0, \exists k, k' \in \{2, \dots, K\} \text{ s.t. } e_k \neq 0, e_{k'} \neq 0\}.$$

We order the exposures in the support such that the exposures with first exposure component equal to $m \in \{1, \dots, m_1 - 1\}$ are first, the exposures with first exposure component equal to m_1 are next, and the exposures with first exposure component equal to zero are last. Within each subset of exposures, we order the exposures according to the reverse reflected lexicographic order. Similar to the proof of Lemma 2.1, we use induction and rely on the monotonicity and uniquely identifiable estimators to show that $\hat{\Theta}$ is affine independent. Note that each zero estimator is uniquely identified by exposure $(0, e_2, \dots, e_K)$ corresponding to the first Horvitz-Thompson term in the estimator. However, note that the zero estimators are not monotonic in the sense that MALUEs are. Instead, they are monotonic in the sense that the exposures follow a reverse reflected lexicographic order when we arrange them according to the order of the corresponding Horvitz-Thompson terms. For example, for a zero estimator where $e_2, e_3 \neq 0$, the exposures corresponding to the Horvitz-Thompson terms:

$$HT_{(0, e_2, \dots, e_K)} - HT_{(0, e_2, 0, e_4, \dots, e_K)} - HT_{(0, 0, e_3, 0, \dots, 0)} + HT_{(0, \dots, 0)}$$

are ordered (in increasing order) according to the reverse reflected lexicographic order. Since exposures in $\text{supp}(\hat{\Theta})$ are also ordered according to the reverse reflected lexicographic order, then for the u th and $u + 1$ th step, we have $\vec{e}^{(u)} < \vec{e}^{(u+1)}$. Hence, the zero estimator $\hat{\theta}_0^{(u)}$ is the last estimator that contains exposure $\vec{e}^{(u)}$ in its support. We iterate through $\hat{\Theta}$ using induction and show that if an estimator $\hat{\theta} = \sum_{\tilde{\theta} \in \hat{\Theta}} g(\tilde{\theta}) \tilde{\theta}$ such that $\hat{\theta} \in \hat{\Theta}$, then the weights $g(\tilde{\theta})$ are given by Equation (2.13), i.e. $\hat{\Theta}$ is affine independent. Since $\hat{\Theta}$ is affine independent, and the dimension of $\hat{\Theta}$ minus one (since the sum of weights is restricted to equal one for unbiasedness) is equal to the dimen-

sion of \mathcal{U} , then $\text{span}(\hat{\Theta}) = \mathcal{U}$. Hence, $\hat{\Theta}$ forms an affine basis for \mathcal{U} , and properties of the simpler estimators in $\hat{\Theta}$ extend to estimators in \mathcal{U} .

2.6 Optimal Linear Unbiased Estimators

At this point, we have defined a set of estimators $\hat{\Theta} = \mathcal{M} \cup \mathcal{Z}$ which forms an affine basis for the set of LUEs. Although there are fewer estimators in $\hat{\Theta}$ compared to \mathcal{U} , the set of estimators $\hat{\Theta}$ could still be fairly large, especially if the number of components is large. Additionally, thus far, estimators for the same estimand, such as two-term and four-term ALUEs, are equivalent. Hence, a natural question is *which estimator should we use?* In this section, we consider an additional property of variance in order to rank different linear unbiased estimators.

2.6.1 Minimum Integrated Variance Linear Unbiased Estimators (MIV LUE)

We consider a “good” estimator as one that is unbiased and has small variance. Since LUEs depend both on the exposures and the parameters $\Theta = \{\alpha, \theta_{k,j_k}\}$ for $k \in \{1, \dots, K\}$ and $j_k \in \{1, \dots, m_k\}$ corresponding to the given exposures, we would ideally account for the parameters when we compute the variance of LUEs. However, in general, we do not know the true set of parameters Θ . Instead, we use distributions π on Θ which describe the set of parameters. We then focus on minimizing the integrated variance (IVAR), where the variance is computed with respect to distributions π on Θ , i.e. $\text{IVAR} = \int_{\Theta} \text{Var}(\hat{\theta})\pi(d\theta)$. Borrowing from Bayesian statistics, one can view the distributions as “prior” distributions on the parameters. However, note that this is not actually Bayesian since we do not have posterior distributions—instead, we use the prior distributions to inform our choices of the weights for LUEs. These prior distributions act as a weight, where parameters that have a higher likelihood are weighted more when computing the variance of the estimator. Minimum integrated

variance linear unbiased estimators (MIV LUEs) (Sussman and Airoidi, 2017) are then given by weights, which depend on the prior distributions, that minimize the integrated variance. As with linear estimators, MIV LUEs depend only on the prior means and covariances (Hoff, 2009; Bickel and Doksum, 2015; Sussman and Airoidi, 2017).

We seek weights $w(\vec{e})$ that minimize the integrated variance such that the linear constraints in Proposition 1 hold. To simplify the optimization problem, we assume that the parameters are uncorrelated across units, but can be correlated within units. We also assume that the priors have zero mean. However, note that if priors do not have zero mean, the estimator

$$\hat{\theta}_{1,m_1} = w(\vec{e}) (Y(\vec{e}) - \mu_{Y(\vec{e})}) + \mu_{\theta_{1,m_1}}, \quad (2.17)$$

where $\mu_{Y(\vec{e})}$ and $\mu_{\theta_{1,m_1}}$ denote the prior mean of the potential outcome and the prior mean of θ_{1,m_1} , respectively, is unbiased if $w(\vec{e})Y(\vec{e})$ is unbiased for θ_{1,m_1} . If $w(\vec{e})$ minimizes the integrated variance of the estimator when priors have zero mean, then $w(\vec{e})$ also minimizes the integrated variance of the estimator given by Equation (2.17) (Hoff, 2009).

Under these assumptions, the optimization problem is solved by minimizing the following Lagrangian over the weights, $w(\vec{e})$, and lambdas:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \int_{\Theta} \sum_{\vec{e} \in \mathcal{E}} p(\vec{e}) (w(\vec{e})Y(\vec{e}) - \theta_{1,m_1})^2 \pi(\theta') d\theta' + \lambda_1 \left(1 - \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m_1\} \right) \\ & - \sum_{m=1}^{m_1-1} \lambda_{2,m} \left(\sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m\} \right) - \lambda_3 \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e}) \\ & - \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_k = j_k\}, \end{aligned} \quad (2.18)$$

where, by taking the derivative of \mathcal{L} with respect to $w(\vec{e})$ and setting it equal to 0,

the MIV LUE weights $w(\vec{e})$ are defined as:

$$w(\vec{e}) = \frac{\lambda_1 \mathbb{I}\{e_1 = m_1\} + \sum_{m=1}^{m_1-1} \lambda_{2,m} \mathbb{I}\{e_1 = m\} + \lambda_3 + \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} \mathbb{I}\{e_k = j_k\}}{\text{Var}(Y(\vec{e}))}, \quad (2.19)$$

where the variance of the potential outcome, $\text{Var}(Y(\vec{e}))$, is given by the prior variances of the parameters. Note that we added the $\frac{1}{2}$ in the Lagrangian to simplify computations, but this does not change the optimization problem since it is a positive constant.

We can rewrite the optimization problem into a matrix equation. We first define the following matrices. Let \mathbf{W} be a $|\mathcal{E}| \times |\mathcal{E}|$ diagonal matrix where the j th diagonal entry for $j \in \{1, \dots, |\mathcal{E}|\}$ is

$$\mathbf{W}_{j,j} = p(\vec{e}_j) \text{Var}(Y(\vec{e}_j)),$$

where \vec{e}_j is the exposure corresponding to the j th row/column of \mathbf{W} . Let \mathbf{C} be a $|\Theta| \times |\mathcal{E}|$ matrix of linear constraints given by Proposition 1 where the rows correspond to the parameters in Θ (i.e. $k \in \{1, \dots, |\Theta|\}$) and the columns correspond to the exposures (i.e. $j \in \{1, \dots, |\mathcal{E}|\}$). That is, the k, j th entry of matrix \mathbf{C} is equal to

$$\mathbf{C}_{k,j} = p(\vec{e}_j) \mathbb{I}\{\theta_k \in \vec{e}_j\},$$

where we write $\theta_k \in \vec{e}_j$ to mean the k th parameter θ_k contributes to the value of the potential outcome given the j th exposure, $Y(\vec{e}_j)$. The solution vector to the optimization problem, denoted by

$$\mathbf{w} = (w(\vec{e}_1) \quad \dots \quad w(\vec{e}_{|\mathcal{E}|}) \quad \lambda_1 \quad \dots \quad \lambda_{4,K,m_K} \quad \lambda_3)^T,$$

is then the solution to the following matrix equation:

$$\mathbf{P}^{-1}\mathbf{b} = \mathbf{w}, \quad (2.20)$$

where the matrix $\mathbf{P} = \begin{pmatrix} \mathbf{W} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix}$ and \mathbf{b} is a vector of zeros besides at the element corresponding to λ_1 , at which $\mathbf{b}_{\lambda_1} = 1$. The matrix \mathbf{P} is full rank given that the diagonal elements in \mathbf{W} are positive (see Lemma A.1 in Appendix A.3.2), which holds provided that the prior variance for each exposure is positive and the probability of observing each exposure is positive. Equation (2.20) shows that the solution \mathbf{w} depends on the prior variances of parameters and the probabilities of exposures. Hence, not all LUEs are also MIV LUEs—whether LUEs are also MIV LUEs depend on the design probabilities and support of the estimators. We characterize the set of MIV LUEs in the next section through the support of the estimator.

2.6.2 Characterization of MIV LUEs

Before now, we have characterized LUEs through the linear constraints as given in Proposition 1. However, we can also classify LUEs through their support, denoted by $\mathcal{E}' \subseteq \mathcal{E}$. The support \mathcal{E}' of a LUE contains exposures such that there exists weights of exposures, $w(\vec{e})$, where, when multiplied with the vector of indicators for exposures, it solves

$$\mathbf{C}\vec{u} = (0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0)^T, \quad (2.21)$$

where the j th element of $\vec{u} \in \mathbb{R}^{|\mathcal{E}|}$ is $\vec{u}_j = \mathbb{I}\{\vec{e}_j \in \mathcal{E}'\}w(\vec{e}_j)$, and the 1 on the right hand side corresponds to θ_{1,m_1} . Effectively, solving for \vec{u} such that it satisfies Equation (2.21) ensures that the linear unbiased constraints are satisfied.

Example 4. Consider the network interference example where $\vec{e} = (d_i^z, z_i)$ for $d_i^z \in \{0, \dots, d_i\}$, where d_i is the degree of unit i , and $z_i \in \{0, 1\}$. Examples of supports of

LUEs include:

$$\begin{aligned}\mathcal{E}^{\text{two term}, z_i} &= \{(d_i, z_i), (0, z_i)\} \\ \mathcal{E}^{\text{four term}, d} &= \{(d_i, 1), (d, 1), (d, 0), (0, 0)\} \\ \mathcal{E}^{\text{six term}, d} &= \{(d_i, 1), (d_i, 0), (d, 1), (d, 0), (0, 1), (0, 0)\},\end{aligned}$$

where $d \in \{1, \dots, d_i - 1\}$. These sets of exposures satisfy Equation (2.21). For example, the weight vectors $(1, -1)$, $(1, -1, 1, -1)$, and $(\frac{3}{2}, -\frac{1}{2}, -1, 1, -\frac{1}{2}, -\frac{1}{2})$ lead to LUEs with support $\mathcal{E}^{\text{two term}, z_i}$, $\mathcal{E}^{\text{four term}, d}$, and $\mathcal{E}^{\text{six term}, d}$, respectively.

Given a subset of exposures $\mathcal{E}' \subseteq \mathcal{E}$ that is a valid support for LUEs, i.e. it satisfies Equation (2.21), we can divide the set of parameters Θ into the following subsets. Let $\Theta^N \subseteq \Theta$ denote the set of parameters where $\theta^N \in \Theta^N$ are such that $\theta^N \notin \vec{e}'$ for all $\vec{e}' \in \mathcal{E}'$. We further divide the parameters in $\Theta^F = \Theta \setminus \Theta^N$ as $\Theta^F = \Theta^R \cup \Theta^{NR}$. Specifically, Θ^{NR} will be a maximal subset of Θ^F such that the submatrix of \mathbf{C} , with rows given by Θ^{NR} and columns given by \mathcal{E}' , has linearly independent rows. Additionally, we can subdivide matrices \mathbf{W} and \mathbf{C} . Matrix \mathbf{W} is a block diagonal matrix with matrices \mathbf{N} and \mathbf{F} on the diagonal. Matrix \mathbf{N} is a diagonal matrix corresponding to exposures $\vec{e} \in \mathcal{E} \setminus \mathcal{E}'$ and \mathbf{F} is a diagonal matrix with rows corresponding to exposures $\vec{e}' \in \mathcal{E}'$. We denote the constraint submatrices of \mathbf{C} as \mathbf{C}_e^p , where the subscript corresponds to the set of exposures e and superscript corresponds to the set of parameters p . For each $e \in \{N, F\}$ and $p \in \{N, NR, R\}$, we define \mathbf{C}_e^p to contain rows corresponding to constraints of parameters in Θ^p and columns correspond to the exposures in \mathcal{E}^e . Here, $\mathcal{E}^F = \mathcal{E}'$ and $\mathcal{E}^N = \mathcal{E} \setminus \mathcal{E}'$.

Given the subsets of exposures and parameters defined above, we can then characterize MIV LUEs through their support:

Theorem 2. Let $\mathcal{E}' \subseteq \mathcal{E}$ such that $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'} \cap \{\vec{v}_{\vec{e}}\}_{\vec{e} \in \mathcal{E}} = \{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}$ where for $\vec{e} \in \mathcal{E}$, $\vec{v}_{\vec{e}} \in \{0, 1\}^{|\Theta|}$ such that $\vec{v}_{\vec{e}}^T \vec{v} = Y(\vec{e})$ where \vec{v} is the vector of parameters Θ . Furthermore, assume that \mathcal{E}' satisfies Equation (2.21), i.e. there exists an unbiased estimator $\hat{\theta}$ where $\text{supp}(\hat{\theta}) = \mathcal{E}'$. If the design p is such that $p(\vec{e}) > 0$ for all $\vec{e} \in \mathcal{E}$,

then there exists a $\hat{\theta}$ with $\text{supp}(\hat{\theta}) \subseteq \mathcal{E}'$ and $\hat{\theta}$ is a limit of MIV LUEs. Furthermore, if for every exposure $\vec{e}' \in \mathcal{E}'$, we have $\lim_{\eta \rightarrow \infty} \sum_{k=1}^{|\Theta^{NR}|} \text{Adj} \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NRT} \right)_{k,1} \mathbb{I}\{\theta_k \in \vec{e}'^T \vec{v}\} \neq 0$, where Adj is the adjugate, then $\text{supp}(\hat{\theta}) = \mathcal{E}'$.

Theorem 2 states that we can find a limit of MIV LUEs $\hat{\theta}$ whose support is a subset of $\mathcal{E}' \subseteq \mathcal{E}$ as long as \mathcal{E}' is a valid support for LUEs and \mathcal{E}' is such that the corresponding set of vectors of indicators for exposures in \mathcal{E}' , denoted $\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}$, contains all vectors in $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'})$ that correspond to valid exposures of interest. Since $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'})$ is a linear subspace of $\mathbb{R}^{|\Theta|}$, there exists a positive semi-definite matrix Σ such that $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}) = \text{Null}(\Sigma)$. For example, $\Sigma = I - XX^T$, where the columns of X are vectors that form an orthonormal basis for $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'})$. Given prior variance-covariance matrix Σ , MIV LUEs are then given by \mathbf{w} , which we obtain by solving Equation (2.20). However, we require the following lemma:

Lemma 2.2. Let $\Theta = \{\alpha, \theta_{1,1}, \dots, \theta_{K,m_K}\}$ be the set of parameters, and let $\Sigma \in \mathbb{R}^{|\Theta| \times |\Theta|}$ be a variance-covariance matrix for the parameters. Let $\vec{v}_1, \vec{v}_2 \in \{0, 1\}^{|\Theta|}$ be vectors such that $\vec{v}_1^T \Sigma \vec{v}_1 = 0$ and $\vec{v}_2^T \Sigma \vec{v}_2 = a$ where $0 < a < \infty$. There exists a sequence of positive semi-definite matrices $\tilde{\Sigma}_\eta \in \mathbb{R}^{|\Theta| \times |\Theta|}$ such that $\lim_{\eta \rightarrow \infty} \vec{v}_1^T \tilde{\Sigma}_\eta \vec{v}_1 < \infty$ and $\lim_{\eta \rightarrow \infty} \vec{v}_2^T \tilde{\Sigma}_\eta \vec{v}_2 = \infty$.

Specifically, let $\tilde{\Sigma} = \eta \Sigma + B$ for $\eta \in \mathbb{R}$ and $B \in \mathbb{R}^{|\Theta| \times |\Theta|}$ be a positive semi-definite matrix where elements $0 < b_{k,j} < \infty$ are small, where $k, j \in \{1, \dots, |\Theta|\}$. From Theorem 2, since there exists a positive-definite matrix Σ such that $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}) = \text{Null}(\Sigma)$, under Σ , the variances of the potential outcomes corresponding to exposures $\vec{e}' \in \mathcal{E}'$ are zero. Lemma 2.2 then says there exists a sequence of variance-covariance matrices $\tilde{\Sigma}_\eta$ such that the potential outcomes corresponding to exposures $\vec{e}' \in \mathcal{E}'$ have finite limiting variances. On the other hand, potential outcomes given by $\vec{e} \notin \mathcal{E}'$ have infinite limiting variances under $\tilde{\Sigma}_\eta$ since $\vec{v}_{\vec{e}} \notin \text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}) = \text{Null}(\Sigma)$. Denote \mathbf{P}_η as the matrix \mathbf{P} , where submatrix \mathbf{W}_η depends on variances given by $\tilde{\Sigma}_\eta$. We also denote submatrices of \mathbf{W}_η with the subscript η , i.e. \mathbf{F}_η and \mathbf{N}_η are matrices given by

η and corresponds to diagonal matrices \mathbf{F} and \mathbf{N} , respectively. Together with Theorem 2, we then see that potential outcomes with finite limiting variances potentially have non-zero weights, while potential outcomes with infinite limiting variances have weights of zero. Note that this is supported by Equation (2.19), where the variance of the potential outcome is inversely related to the MIV LUE weights. We can interpret this as we put more weight on exposures that we are more confident about, i.e. potential outcomes with smaller prior variances, while we put less weight on exposures that we are not as informed about, i.e. potential outcomes with larger prior variances.

To ensure that weights of the potential outcomes corresponding to exposures in \mathcal{E}' are non-zero, we further require that, for every exposure in \mathcal{E}' , the limit of the sum of the entries of the adjugate of $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NRT}$ in the column corresponding to parameter θ_{1,m_1} as $\eta \rightarrow \infty$ is non-zero. Although it is possible for the weights of exposures in \mathcal{E}' to be zero, we show through an example of a six-term exposure set (see Appendix A.3.3) that “typical” choices of design p will lead to non-zero weights if we assume finite limiting prior variances. Hence, under most designs p , we have $\text{supp}(\hat{\theta}) = \mathcal{E}'$. In general, if limiting prior variances of all parameters are finite, $\hat{\theta}$ is a MIV LUE with non-zero weight on all exposures, and $\hat{\theta}$ is an affine combination of estimators in $\hat{\Theta}$. Note that formally, $\hat{\theta}$ is a solution to the matrix equation in Equation (2.20) while taking the limit \mathbf{P}_η as $\eta \rightarrow \infty$. Since the matrix \mathbf{P}_η may contain infinite values in the limit, it may not be a well-defined problem, and so $\hat{\theta}$ lies on the boundary of MIV LUEs. However, for convenience, we say that a limit of MIV LUEs is also MIV LUE. Hence $\hat{\theta}$ is a MIV LUE.

Example 4 (continuing from p.32). We considered three examples of supports for LUEs in the context of network interference:

$$\begin{aligned} \mathcal{E}^{\text{two term}, z_i} &= \{(d_i, z_i), (0, z_i)\} \\ \mathcal{E}^{\text{four term}, d} &= \{(d_i, 1), (d, 1), (d, 0), (0, 0)\} \\ \mathcal{E}^{\text{six term}, d} &= \{(d_i, 1), (d_i, 0), (d, 1), (d, 0), (0, 1), (0, 0)\}. \end{aligned}$$

Consider $\mathcal{E}^{\text{two term}, z_i} = \{(0, z_i), (d_i, z_i)\}$. Note that $\text{span}(\{\vec{v}_{0, z_i}, \vec{v}_{d_i, z_i}\}) \cap \{\vec{v}_{\vec{e}}\}_{\vec{e} \in \mathcal{E}} = \{\vec{v}_{0, z_i}, \vec{v}_{d_i, z_i}\}$ for $z_i \in \{0, 1\}$. By Theorem 2, there exists weights $w(\vec{e})$ under a given prior such that $\hat{\theta}^{\text{two term}, z_i} = \sum_{\vec{e} \in \mathcal{E}^{\text{two term}, z_i}} w(\vec{e}) Y(\vec{e})$ is a MIV LUE with support $\mathcal{E}^{\text{two term}, z_i}$. Specifically, examples of priors include the following, depending on whether $z_i = 0$ or $z_i = 1$. First consider $z_i = 0$, i.e. $\mathcal{E}^{\text{two term}, 0} = \{(0, 0), (d_i, 0)\}$. Let $\Sigma^{\text{two term}, 0}$ be defined such that parameters $\text{Var}(\alpha) = \text{Var}(\theta_{1, d_i}) = 0$ and variances of all other parameters are positive. Now consider $z_i = 1$, i.e. $\mathcal{E}^{\text{two term}, 1} = \{(0, 1), (d_i, 1)\}$. Let $\Sigma^{\text{two term}, 1}$ be such that $\text{Var}(\theta_{1, d_i}) = 0$, $\alpha = -\theta_{2, 1}$, $\text{Var}(\alpha) = \text{Var}(\theta_{2, 1}) > 0$, $\text{cov}(\alpha, \theta_{2, 1}) = -\text{Var}(\alpha)$, and variances of all other parameters are positive while covariances are non-negative. The prior variance matrices $\Sigma^{\text{two term}, 0}$ and $\Sigma^{\text{two term}, 1}$ inform the MIV LUE weights. In particular, the MIV LUE weights given by priors $\Sigma^{\text{two term}, 0}$ and $\Sigma^{\text{two term}, 1}$ are equal to the weights of the two-term ALUEs $\hat{\theta}_{1, d_i}^{\text{two term}, 0}$ and $\hat{\theta}_{1, d_i}^{\text{two term}, 1}$, respectively. That is, the two-term ALUEs are also MIV LUEs for some prior.

We now consider $\mathcal{E}^{\text{four term}, d} = \{(0, 0), (d, 0), (d, 1), (d_i, 1)\}$, where $d \in \{1, \dots, d_i - 1\}$. We consider the span of $\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}^{\text{four term}, d}}$. In particular, the vector $\vec{v}_{d_i, 0} \in \text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}^{\text{four term}, d}}) \cap \{\vec{v}_{\vec{e}}\}_{\vec{e} \in \mathcal{E}}$, where $\vec{v}_{d_i, 0} = \vec{v}_{d_i, 1} - \vec{v}_{d, 1} + \vec{v}_{d, 0}$. However, $\vec{v}_{d_i, 0} \notin \{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}^{\text{four term}, d}}$. Then by Theorem 2, there do not exist MIV LUEs for any prior under our formulation with support $\mathcal{E}^{\text{four term}, d}$.

The set of exposures $\mathcal{E}^{\text{six term}, d}$ is a support for a six-term MIV LUE. We focus on a generalized example of a six-term exposure set in the next section.

2.6.3 Example: Six-Term Exposure Set

For notational simplicity, we focus on exposures with two exposure components, but the results generalize to cases with more than two exposure components where all other exposure components are the same for all six exposures. Let $\mathcal{E}^{\text{six term}, m} = \{(0, 0), (0, j), (m, 0), (m, j), (m_1, 0), (m_1, j)\}$, where $j \in \{1, \dots, m_2\}$ and $m \in \{1, \dots, m_1 - 1\}$. By Theorem 2, since $\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}^{\text{six term}, m}} = \text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}^{\text{six term}, m}}) \cap \{\vec{v}_{\vec{e} \in \mathcal{E}}\}$, there exists a MIV LUE $\hat{\theta}_{1, m_1}^{\text{six term}}$ such that $\text{supp}(\hat{\theta}_{1, m_1}^{\text{six term}}) = \mathcal{E}^{\text{six term}, m}$ for a given prior. By Theorem 1, since $\hat{\theta}_{1, m_1}^{\text{six term}}$ is a LUE, we can write

$$\hat{\theta}_{1, m_1}^{\text{six term}} = \alpha_1 (HT_{(m_1, 0)} - HT_{(0, 0)}) + \alpha_2 (HT_{(m_1, j)} - HT_{(0, j)})$$

$$+ \alpha_3 (HT_{(m_1,j)} - HT_{(m,j)} + HT_{(m,0)} - HT_{(0,0)}), \quad (2.22)$$

where the three estimators are ALUEs and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Furthermore, by a similar argument as in the proof of Theorem 1, the set of ALUEs

$$\{\hat{\theta}_{1,m_1}^{\text{two term},0}, \hat{\theta}_{1,m_1}^{\text{two term},j}, \hat{\theta}_{1,m_1}^{\text{four term},m}\} = \{HT_{(m_1,0)} - HT_{(0,0)}, HT_{(m_1,j)} - HT_{(0,j)}, \\ HT_{(m_1,j)} - HT_{(m,j)} + HT_{(m,0)} - HT_{(0,0)}\}$$

forms a basis for estimators with exposure set $\mathcal{E}^{\text{six term},m}$. Hence, we only need to focus on the three weights $\alpha_1, \alpha_2, \alpha_3$ as opposed to the six weights on the different exposures. Recall that in the previous section, we showed that the two-term ALUEs, $\hat{\theta}_{1,m_1}^{\text{two term},0}$ and $\hat{\theta}_{1,m_1}^{\text{two term},j}$, are also MIV LUEs for some prior, and so it is possible that $\alpha_1 = 1, \alpha_2 = 0, \alpha_3 = 0$ or $\alpha_1 = 0, \alpha_2 = 1, \alpha_3 = 0$. However, since the four-term estimator $\hat{\theta}_{1,m_1}^{\text{four term},m}$ is not a MIV LUE for any prior, then $\alpha_3 \neq 1$. Although four-term ALUEs are not MIV LUEs, exposures in the supports of four-term ALUEs may still contribute to MIV LUEs. Through the weights $\alpha_1, \alpha_2, \alpha_3$, we investigate how much emphasis might be put on exposures that are “seemingly unrelated” to the estimand of interest, such as exposures (m, j) and $(m, 0)$.

Solving for the MIV LUE weights given by the MIV LUE problem in Equation (2.20) given exposure set $\mathcal{E}^{\text{six term},m}$ and some prior $\Sigma^{\text{six term}}$ (see Appendix A.4), we determine that

$$\alpha_3 = \frac{r(m,0)r(m,j) \left\{ r(m_1,j)r(0,0) - r(m_1,0)r(0,j) \right\}}{D} \quad (2.23)$$

where $r(\vec{e}) = \frac{p(\vec{e})}{\text{Var}(Y(\vec{e}))}$ and

$$D = r(m_1,0) \left[r(0,0)r(m,0)r(m_1,j) + r(0,0)r(m,j)r(m_1,j) \right]$$

$$\begin{aligned}
& + r(m_1, j) \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \\
& + \left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0)r(m, 0)r(0, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& \left. + r(0, 0)r(m, j)r(0, j) + r(0, j)r(m, 0)r(m, j) \right]. \tag{2.24}
\end{aligned}$$

Hence, the weight α_3 is determined by the prior variance-covariance matrix $\Sigma^{\text{six term}}$ and design probabilities $p(\vec{e}) > 0$ for $\vec{e} \in \mathcal{E}$. Since we assume that the design is fixed, we focus on how α_3 changes as we vary the different prior variances.

We first assume that the parameters are independent, i.e. covariances of parameters are zero. Rearrange Equation (2.23) such that $\text{Var}(\theta_{1,m})$ appears only in the denominator of α_3 . Hence $\text{Var}(\theta_{1,m})$ is inversely related to α_3 , and the weight α_3 is maximized as $\text{Var}(\theta_{1,m}) \rightarrow 0$. This makes sense since exposures with $e_1 = m$ contribute the most in estimating θ_{1,m_1} when we are certain about $\theta_{1,m}$, and $\hat{\theta}_{1,m_1}^{\text{four term},m}$ is the only estimator in Equation (2.22) whose support contains exposures with $e_1 = m$. If we are not as certain about $\theta_{1,m}$ relative to the other parameters, we put more weight on the two-term estimators.

Figure 2.2 shows the trajectories of α_1 , α_2 , and α_3 as $\text{Var}(\alpha)$, $\text{Var}(\theta_{1,m_1})$, and $\text{Var}(\theta_{2,j})$ vary when the probability of a unit being treated follows the Bernoulli distribution with probability 0.5, $m_1 = 3$, and $m_2 = 1$. In each of the panels, the variances of parameters that are not varying are fixed to values aimed to maximize α_3 (see Appendix A.4 for details). That is, we set $\text{Var}(\theta_{1,m}) = 0.00001$, $\text{Var}(\alpha) = 0.00001$, $\text{Var}(\theta_{1,m_1}) = 100,000$, and $\text{Var}(\theta_{2,j}) = 1$. In general, the weights depend on the fraction $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\theta_{1,m_1})}$. As $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\theta_{1,m_1})}$ increases, the weight α_1 is generally non-decreasing while α_2 is generally non-increasing. This is because $\hat{\theta}_{1,m_1}^{\text{two term},0}$, which corresponds to α_1 , does not contain the parameter $\theta_{2,j}$, but $\hat{\theta}_{1,m_1}^{\text{two term},j}$, which corresponds to α_2 , contains the parameter $\theta_{2,j}$. When we are less certain about $\theta_{2,j}$ relative to θ_{1,m_1} , i.e. when the variance of $\theta_{2,j}$ is relatively larger than the variance of θ_{1,m_1} , the exposures

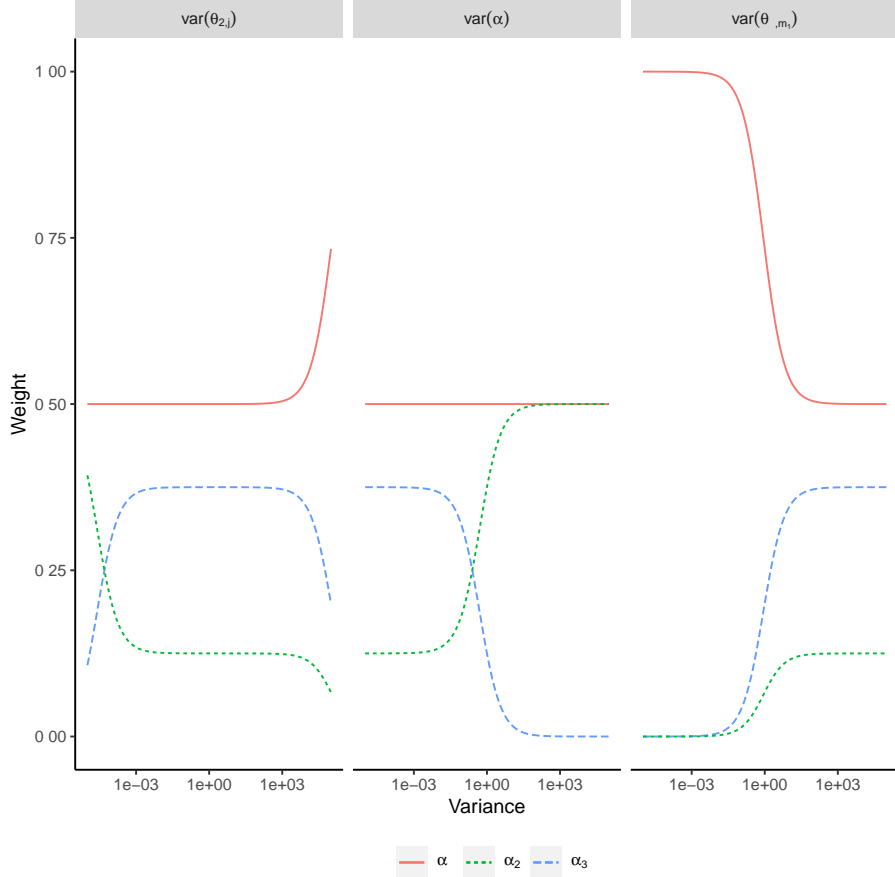


Figure 2.2: The trends of weight α_1 , α_2 , and α_3 (indicated by line type and color) for $\text{Var}(\theta_{1,m}) = 0.00001$ and different values $\text{Var}(\alpha)$, $\text{Var}(\theta_{1,m_1})$, and $\text{Var}(\theta_{2,j})$ as indicated by the panels. Variances of parameters that are not varying are set to values to maximize α_3 : $\text{Var}(\alpha) = 0.00001$, $\text{Var}(\theta_{1,m_1}) = 100,000$, and $\text{Var}(\theta_{2,j}) = 1$.

of $\hat{\theta}_{1,m_1}^{\text{two term},j}$ contribute less to the estimation of θ_{1,m_1} . When we are more certain about $\theta_{2,j}$ relative to θ_{1,m_1} , i.e. when the variance of $\theta_{2,j}$ is relatively smaller than the variance of θ_{1,m_1} , the exposures of $\hat{\theta}_{1,m_1}^{\text{two term},0}$ contribute less to the estimation of θ_{1,m_1} . Recall that the support of the four-term estimator contains exposures that are found in both of the supports of the two-term estimators. Hence, the weight α_3 is not necessarily monotonic as $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\theta_{1,m_1})}$ changes. The weight α_3 increases in general as $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\theta_{1,m_1})}$ approaches the interval $(2 \times 10^{-6}, 2 \times 10^{-4})$. As $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\theta_{1,m_1})}$ falls outside of $(2 \times 10^{-6}, 2 \times 10^{-4})$, then α_3 decreases. Hence, when the probability for a unit to

be treated follows a Bernoulli distribution with probability 0.5, $m_1 = 3$, and $m_2 = 1$, α_3 is maximized if $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\theta_{1,m_1})}$ is inside the range $(2 \times 10^{-6}, 2 \times 10^{-4})$.

The weights α_2 and α_3 also depend on $\text{Var}(\alpha)$, specifically on the ratio $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\alpha)}$. As $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\alpha)}$ increases, the weight α_3 is non-decreasing, while the weight α_2 is non-increasing. This is possibly explained because $\hat{\theta}_{1,m_1}^{\text{four term},m}$, which corresponds to α_3 , also has exposures with $e_2 = 0$ in its support, which only depends on parameters α and either $\theta_{1,m}$ or θ_{1,m_1} . On the other hand, $\hat{\theta}_{1,m_1}^{\text{two term},j}$ only has exposures with $e_2 = j$ in its support. When the variance of α is relatively higher than the variance of $\theta_{2,j}$, we prioritize the exposures in the support of $\hat{\theta}_{1,m_1}^{\text{two term},j}$. When the variance of α is relatively lower than the variance of $\theta_{2,j}$, we prioritize the exposures in the support of the four-term ALUE compared to exposures in the support of $\hat{\theta}_{1,m_1}^{\text{two term},j}$. When the other variances are fixed to values to maximize α_3 , we see that the weight α_1 does not depend on $\frac{\text{Var}(\theta_{2,j})}{\text{Var}(\alpha)}$.

The weight α_3 approaches zero when we take the limit of the variances of α and θ_{1,m_1} , specifically $\text{Var}(\alpha) \rightarrow \infty$ and $\text{Var}(\theta_{1,m_1}) \rightarrow 0$. However, α_3 remains non-zero even in the limits of $\text{Var}(\theta_{2,j})$, both towards zero and towards infinity. Hence, even if one is highly uncertain about $\theta_{2,j}$, exposures in the support of the four-term ALUE may still contribute to the estimation of θ_{1,m_1} . Taking the limits of the variances of the parameters when the parameters are uncorrelated, the weight α_3 is maximized at the following:

Corollary 3. Consider a set of exposures

$$\mathcal{E}^{\text{six term},m} = \{(0, 0), (0, j), (m, 0), (m, j), (m_1, 0), (m_1, j)\},$$

where $j \in \{1, \dots, m_2\}$ and $m \in \{1, \dots, m_1 - 1\}$. Under the assumption that all covariances are zero, the weight α_3 is maximized when we take $\text{Var}(\alpha), \text{Var}(\theta_{1,m}) \rightarrow 0$, $\text{Var}(\theta_{1,m_1}) \rightarrow \infty$, and $\text{Var}(\theta_{2,j}) < \infty$. Given a design $p(\vec{e})$ for $\vec{e} \in \mathcal{E}$, the maximum of

α_3 is constant and depends only on the design:

$$\max_{\substack{\text{Var}(\alpha), \text{Var}(\theta_{1,m}) \\ \text{Var}(\theta_{1,m_1}), \text{Var}(\theta_{2,j})}} \{\alpha_3\} = \frac{p(m, j)p(m_1, j)}{[p(0, j) + p(m, j)][p(m_1, 0) + p(m_1, j)]}. \quad (2.25)$$

The maximum contribution of potential outcomes corresponding to exposures with $e_1 = m$ depends on the design. The choice of the design is out of the scope for this paper, and future work may be done on this topic. Under the conditions when α_3 is maximized, we require $\text{Var}(\theta_{1,m_1}) \rightarrow \infty$, $\text{Var}(\theta_{1,m}) \rightarrow 0$, and $\text{Var}(\alpha) \rightarrow 0$. Thus, estimators with a weight α_3 equal to Equation (2.25) formally lies on the boundary of the set of MIV LUEs. However, since we considered the set of MIV LUEs to be closed for convenience, estimators with α_3 equal to Equation (2.25) are MIV LUEs.

When covariances between parameters are non-zero, similar deductions can be made—the weights $\alpha_1, \alpha_2, \alpha_3$ depend on the overall variances of the potential outcomes of the corresponding estimators. Figure 2.3 shows the trends of the weights α_1, α_2 , and α_3 as the correlation between pairs of parameters: $\text{cor}(\alpha, \theta_{2,j})$, $\text{cor}(\alpha, \theta_{1,m_1})$, and $\text{cor}(\theta_{2,j}, \theta_{1,m_1})$ changes when the probability for a unit to be treated follows a Bernoulli distribution with probability 0.5, $m_1 = 3$, and $m_2 = 1$. Note that $\alpha_1, \alpha_2, \alpha_3$ do not depend on the covariances when the variances of parameters are taken to maximize α_3 . Hence, we consider when $\text{Var}(\theta_{1,m}) = 0.00001$ and $\text{Var}(\theta_{1,m_1}) = \text{Var}(\alpha) = \text{Var}(\theta_{2,j}) = 1$. Since the variances are equal to 1, the correlations here are equivalent to the covariances between the parameters. The sign of $\text{cor}(\alpha, \theta_{2,j})$ indicates whether the variance of potential outcomes with exposures where $e_2 = j$ increases or decreases since all potential outcomes contain the baseline effect α —a negative correlation indicates a decrease in variance while a positive correlation indicates an increase in variance. As $\text{cor}(\alpha, \theta_{2,j})$ increases, the weight α_1 increases while the weight α_2 decreases—we are more certain about the parameters corresponding to exposures in $\hat{\theta}_{1,m_1}^{\text{two term},0}$ than the parameters corresponding to exposures in $\hat{\theta}_{1,m_1}^{\text{two term},j}$. The tra-

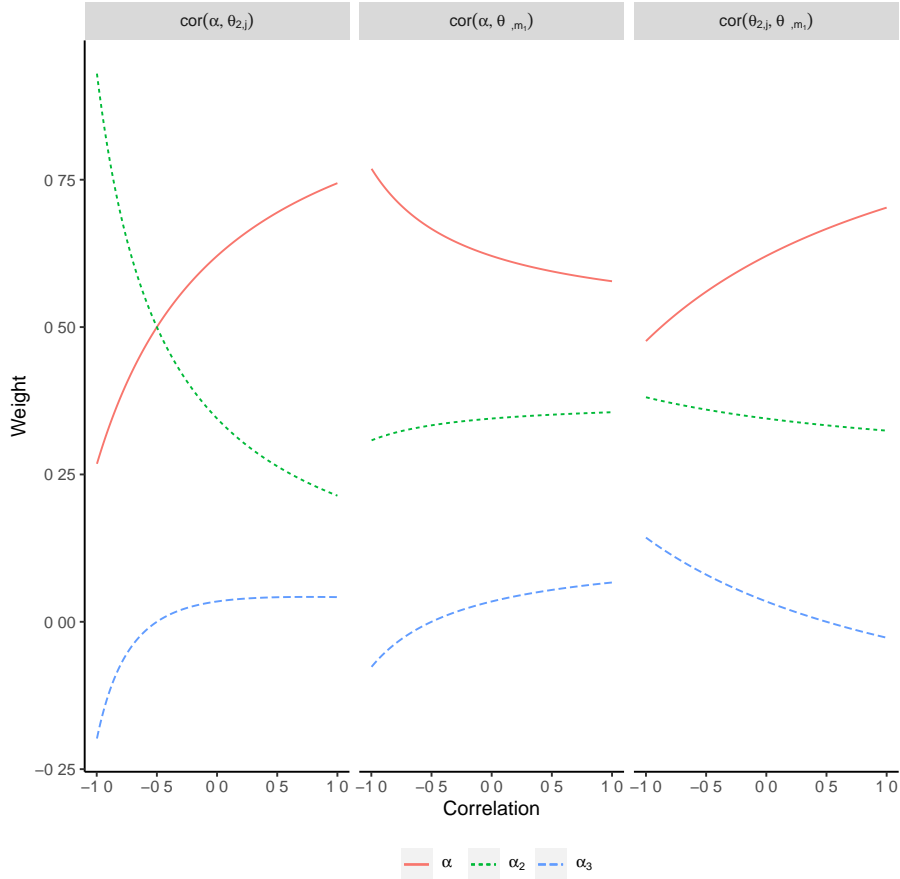


Figure 2.3: The trends of weights $\alpha_1, \alpha_2, \alpha_3$ (indicated by line type and color) for $\text{Var}(\theta_{1,m}) = 0.00001$ and $\text{Var}(\theta_{1,m_1}) = \text{Var}(\alpha) = \text{Var}(\theta_{2,j}) = 1$ and different values of $\text{cor}(\alpha, \theta_{2,j}), \text{cor}(\alpha, \theta_{1,m_1}), \text{cor}(\theta_{1,m_1}, \theta_{2,j})$ as indicated by the x-axis.

jectory of weight α_3 follows a similar pattern of the trajectory of α_1 as $\text{cor}(\alpha, \theta_{2,j})$ varies, but at a smaller magnitude. By a similar argument, as $\text{cor}(\alpha, \theta_{1,m_1})$ increases, the weight α_1 decreases as the weight α_2 increases. However, the trajectory of weight α_3 now follows a similar pattern of the trajectory of α_2 . Since α_2 and α_3 correspond to estimators whose supports contain exposure (m_1, j) , we also see that α_2 and α_3 decrease as $\text{cor}(\theta_{2,j}, \theta_{1,m_1})$ increases and α_1 increases with $\text{cor}(\theta_{2,j}, \theta_{1,m_1})$. Overall, when covariances are non-zero, the weight α_3 is generally smaller than when covariances are zero. However, even when covariances are non-zero, we see that exposures with $e_1 = m$ may contribute to the estimation of θ_{1,m_1} .

2.7 Simulations

In the previous chapter, we characterized MIV LUEs through their supports which vary depending on the prior distribution. Here, we evaluate the performance of the MIV LUEs presented in Chapter 2.6 through simulations to estimate network interference effects as described in Example 2. Recall that we assume that the potential outcomes of a unit depend on the unit's treatment and the treatment of the unit's neighbors. In particular, we assume a binary treatment and assume that the potential outcome of a unit depends on the number of treated neighbors and not necessarily which units are treated. The set of exposures is given by $\mathcal{E} = \{\vec{e} : \vec{e} = (d_i^z, z_i)\}$, where $d_i^z \in \{0, \dots, d_i\}$ is the treated degree, or the number of treated neighbors. Furthermore, d_i is the degree of unit i , and $z_i \in \{0, 1\}$ is the treatment assignment of unit i . Under additivity, the potential outcome of unit i , given exposure \vec{e}_i , is given by:

$$Y_i(\vec{e}_i) = \alpha^{(i)} + \theta_{2,1}^{(i)} z_i + \sum_{d=1}^{d_i} \theta_{1,d}^{(i)} \mathbb{I}\{d_i^z = d\}. \quad (2.26)$$

Here, we include i as a superscript and subscript to indicate different parameters, exposures, treatment assignments, and treated degree for different units. We focus on directed networks here, but results can be applied to undirected graphs. In the case of directed networks, d_i is the in-degree of unit i or the number of edges pointing at unit i .

The parameter of interest is $\theta_{1,d_i}^{(i)}$, which is the interference effect when the treated degree is equal to the degree of the unit versus when the treated degree is zero. More specifically, we are interested in the average interference effect when all of the neighbors of a unit versus none are treated, $\bar{\theta}_{1,d_i} = \frac{1}{n} \sum_{i=1}^n \theta_{1,d_i}^{(i)}$. We compare the performance of various linear estimators with inverse probability of exposure weighting:

Two-term Horvitz-Thompson for untreated units: Horvitz-Thompson inverse

probability weighting estimators where $w_i(\vec{e}_i) = \frac{\mathbb{I}\{\vec{e}_i=(d_i,0)\}}{np(d_i,0)} - \frac{\mathbb{I}\{\vec{e}_i=(0,0)\}}{np(0,0)}$. The two-term Horvitz-Thompson estimator for untreated units is unbiased even when additivity does not hold. We denote this estimator as HT_0 .

Two-term Horvitz-Thompson for treated units: Horvitz-Thompson inverse probability weighting estimators where $w_i(\vec{e}_i) = \frac{\mathbb{I}\{\vec{e}_i=(d_i,1)\}}{np(d_i,1)} - \frac{\mathbb{I}\{\vec{e}_i=(0,1)\}}{np(0,1)}$. We denote this estimator as HT_1 .

Average Horvitz-Thompson: Horvitz-Thompson inverse probability weighting estimators where $w_i(\vec{e}_i) = \frac{1}{2} \left[\frac{\mathbb{I}\{\vec{e}_i=(d_i,1)\}}{np(d_i,1)} - \frac{\mathbb{I}\{\vec{e}_i=(0,1)\}}{np(0,1)} \right] + \frac{1}{2} \left[\frac{\mathbb{I}\{\vec{e}_i=(d_i,0)\}}{np(d_i,0)} - \frac{\mathbb{I}\{\vec{e}_i=(0,0)\}}{np(0,0)} \right]$. We denote this estimator as HT_{Avg} .

MIV LUE with Independent Priors: LUE where weights are given by solving the MIV LUE problem with prior distributions: $\alpha^{(i)}, \theta_{2,1}^{(i)}, \theta_{1,1}^{(i)}, \theta_{1,d}^{(i)} \sim \mathcal{N}(0, 1)$ for all $d \in \{1, \dots, d_i\}$. We assume that priors are uncorrelated between units and independent between parameters. We denote this estimator as M_{Ind} .

MIV LUE with Dilated Priors: LUE where weights are given by solving the MIV LUE problem with prior distributions: $\alpha^{(i)} \sim \mathcal{N}(0, 1)$, $\theta_{2,1}^{(i)} = \alpha^{(i)}$, and $\theta_{1,d}^{(i)} = \frac{d}{d_i} \eta_1 \times \alpha^{(i)}$ for all $d \in \{1, \dots, d_i\}$ for a fixed value of η_1 . Note that the prior variances are: $\text{Var}(\theta_{2,1}^{(i)}) = 1$ and $\text{Var}(\theta_{1,d}^{(i)}) = \left(\frac{d}{d_i} \eta_1\right)^2$. We assume that priors are uncorrelated between units. However, there are covariances between the parameters. We let $\eta_1 = 1$ for our simulations, and we denote this estimator as M_{Dil} .

Recall that Aronow et al. (2017) proposed the linear unbiased estimator HT_0 to estimate the network interference effect. Under additivity, the other estimators: HT_1 , HT_{Avg} , M_{Ind} , and M_{Dil} are also linear unbiased estimators. However, the supports of HT_0 and HT_1 are of size two, whereas HT_{Avg} , M_{Ind} , and M_{Dil} puts non-zero weights on more than two exposures. Specifically, the support of HT_{Avg} is equal to the union of the supports of HT_0 and HT_1 , while the supports of M_{Ind} and M_{Dil} may be equal to the entire set of exposures.

We fix the design to be a Bernoulli design where the probability of being treated is 0.5. The probability of a given exposure is then given by:

$$\mathbb{P}(\vec{e} = (d, z)) = \binom{d_i}{d} 0.5^{d_i+1}. \quad (2.27)$$

For each simulation, we sampled 1000 sets of parameters for each unit to generate the potential outcomes. Unless otherwise specified, we generated the parameters for the potential outcomes as follows: $\alpha^{(i)}, \theta_{2,1}^{(i)}, \theta_{1,d}^{(i)} \sim \mathcal{N}(0, 1)$ for all $d \in \{1, \dots, d_i\}$. Hence, the sampling distribution for the parameters for the potential outcomes may be different from the prior distributions used for the estimators of interest.

We compare the estimators using the integrated mean squared error (IMSE), which is integrated over the parameters as in the integrated variance. Under additivity, the estimators considered are unbiased, and so IMSE is largely driven by the integrated variance. We evaluate the performance of the estimators under settings of: varying the number of units and the number of edges, varying the level of additivity and interference, and varying the potential outcome distributions.

2.7.1 Varying Number of Units and Number of Edges

We first investigate how the IMSEs of the estimators change as we vary the size of the network. In particular, we vary the number of units and the number of edges in a network. For each $n = 10, 20, \dots, 50$, we generated k -regular directed networks, where each unit has in-degree k , for $k = 2, 4, 6, 8$. For example, Figure 2-4 shows a 4-regular directed (as indicated by the arrows) network with 40 nodes. We fixed the networks while we sampled different sets of potential outcome parameters and iterated through the treatment allocations. Since each unit in a k -regular network has the same in-degree, each unit contributes equally to the estimation of the average interference effect. For networks with $n = 10$, for each sampled set of parameters,

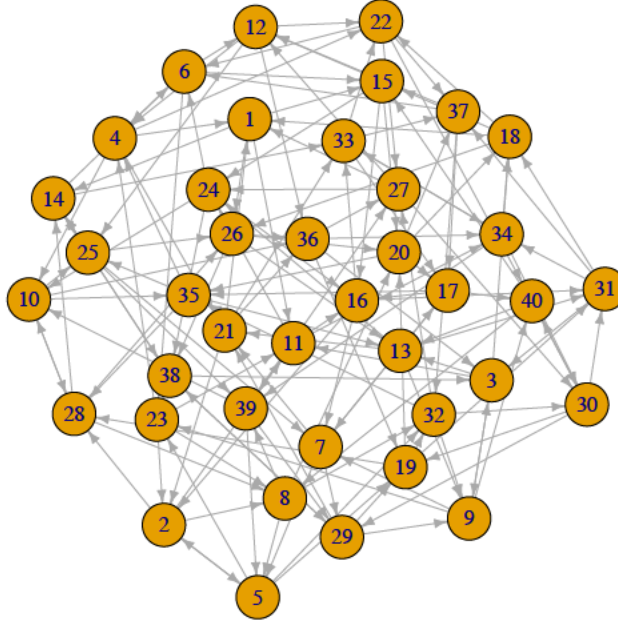


Figure 2-4: Directed 4-regular network with forty nodes and each unit has in-degree of four. Arrows indicate directions of edges.

we computed the IMSE using all 2^{10} possible treatment allocations. For networks with $n = 20, 30, 40, 50$, we computed the IMSE over a sample of 1500 treatment allocations. Hence, for $n = 10$, we computed the exact integrated bias and variance whereas we estimated these for $n = 20, 30, 40, 50$. Potential outcomes were simulated under additivity, i.e. there were no interaction effects. Hence, in this simulation setting, we expect HT_{Ind} to perform the best since the prior of HT_{Ind} matches the distributions of the parameters for the potential outcomes.

Figure 2-5 shows the IMSE for the estimators as the number of units (indicated by x-axis) increases for different values of k (indicated by panels). Overall, the IMSE

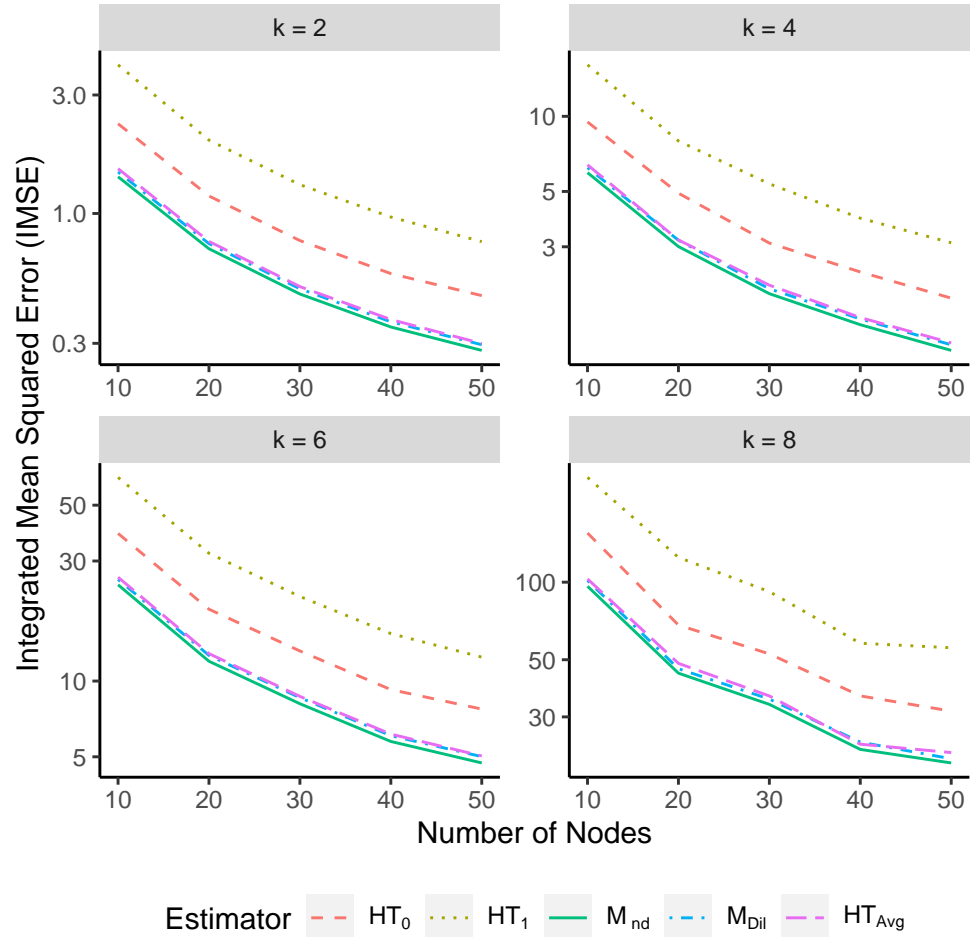


Figure 2.5: IMSE for estimators (indicated by color and line type) when the number of units (indicated by x-axis) increases for a k -regular network for different values of k (indicated by panel) under additivity and when mean interference is zero.

decreases as the number of units increase. Since all units have the same in-degree and hence the same exposure distribution, increasing the number of units leads to a decrease in the IMSE. On the other hand, as the number of edges (or in-degree) k increases, the IMSE increases for all estimators. This is possibly explained by the fact that as the number of edges increases, the probability of a unit having treated degree zero or treated degree d_i decreases. Weights on exposures with treated degree equal to zero or d_i then increase with k since weights are inversely related to the

probabilities of exposures. On the other hand, weights on other exposures are either zero (for two-term HT estimators and HT_{Avg}) or are relatively smaller (for M_{Ind} and M_{Dil}) since the probability of exposures with $e_1 \in \{1, \dots, d_i - 1\}$ increase with k , leading to a greater IMSE.

The red, dashed line indicates the IMSE for HT_0 . Under additivity, HT_1 is also a linear unbiased estimator of $\bar{\theta}_{1,d_i}$, but HT_1 has higher IMSE than HT_0 . This is likely due to the extra variance introduced by the treated units. However, there is a significant reduction in IMSE across the different panels when we average both HT_0 and HT_1 . Indeed, the IMSE of HT_{Avg} , given by the purple, long-dashed line, is lower than the IMSEs of HT_0 and HT_1 . There is an additional reduction in IMSE when we take the integrated variance into account and compute weights to minimize the integrated variance. Although the IMSEs of M_{Ind} and M_{Dil} are only slightly lower than the IMSE of HT_{Avg} , we still see the benefit of using optimal weights. Furthermore, M_{Ind} performs the best as expected. The performance of the estimators suggest that there is an advantage in leveraging information from all data available as opposed to just using a subset of units.

2.7.2 Varying Interference Effects and Deviations from Additivity

Throughout this chapter, we derived linear unbiased estimators under the assumption that causal effects are additive. In this section, we examine the robustness of the MIV LUEs when the additivity assumption is violated. We represent varying levels of additivity through an interaction effect between the direct effect and the interference effect. Potential outcomes in this section were simulated according to the parameterization:

$$Y_i(\vec{e}_i) = \alpha^{(i)} + \theta_{2,1}^{(i)} z_i + \sum_{d=1}^{d_i} \theta_{1,d}^{(i)} \mathbb{I}\{d_i^z = d\} + \sum_{d=1}^{d_i} \Delta_d^{(i)} z_i \mathbb{I}\{d_i^z = d\}, \quad (2.28)$$

where $\alpha^{(i)}, \theta_{2,1}^{(i)} \sim \mathcal{N}(0, 1)$, and $\theta_{1,d}^{(i)} \sim \mathcal{N}(\frac{d}{d_i}\mu_1, 1)$ and $\Delta_d^{(i)} \sim \mathcal{N}(\frac{d}{d_i}\delta_1, \mathbb{I}\{\delta_{1,d_i} > 0\})$ are the interference effects and interaction effects, respectively. When $\Delta_d^{(i)} = 0$ for all $d \in \{1, \dots, d_i\}$, additivity holds. We simulated potential outcomes under $\mu_1 \in \{0, 10, 50\}$ and $\delta_1 \in \{0, 2, 4, 6\}$. Note when $\delta_1 = 0$, we set $\text{Var}(\Delta_d^{(i)}) = 0$ so that $\Delta_d^{(i)} = 0$ to ensure that additivity holds. Even though interference effects were not necessarily mean zero, we maintained zero-mean priors to evaluate the performance of our estimators when the priors do not match the potential outcome distributions. In particular, we estimated average network interference effects on a 4-regular graph when $n = 40$.

Figure 2.6 shows the IMSE of the estimators as the interaction effect increases when the mean interference effect is 0, 10, and 50 (indicated by the panels). As the mean interference effect increases across the three panels, the IMSE increases for all estimators. Since we used zero-mean priors to derive the MIV LUEs, it is reasonable that when the potential outcome distributions stray further away from the prior distribution, all estimators do not perform as well. When the true mean interference effect is zero and additivity holds, M_{Ind} outperforms the other estimators, as expected. However, as the mean interference effect increases, M_{Dil} actually outperforms M_{Ind} , despite the fact that the potential outcome parameters are independent. Hence, there may be slight concerns when using an estimator with weights obtained from a prior distribution different from the potential outcome distribution. Even though M_{Ind} does not perform the best when the mean interference is non-zero, in general, the multi-term MIV LUEs outperform the other estimators.

As the interaction effect (indicated by x-axis) increases, the IMSEs of estimators increase in general. However, since HT_0 puts non-zero weight on untreated units, it is invariant to the interaction effect. Furthermore, it is the only estimator considered that is unbiased even when additivity does not hold. When the interaction effect is not

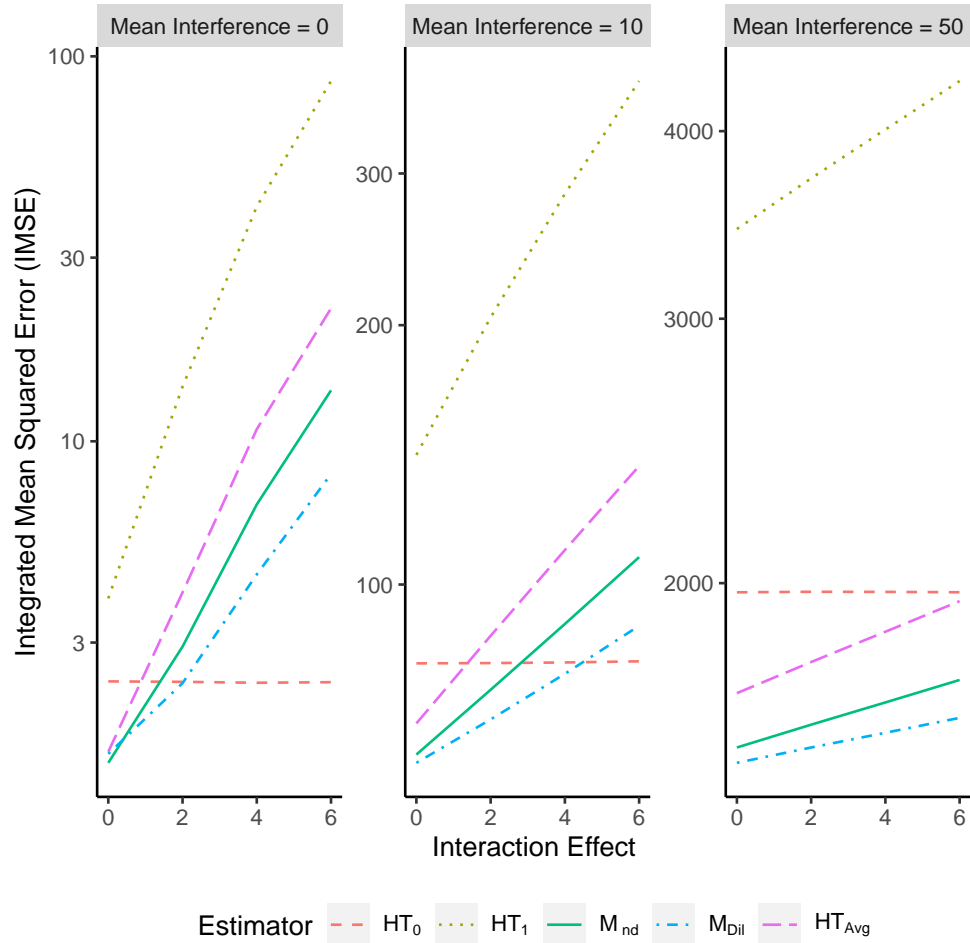


Figure 2-6: IMSE for estimators (indicated by color and line type) when the interaction (indicated by x-axis) and interference effects (indicated by panel) vary for a 40 node 4-regular graph.

zero, the other estimators are biased, which partially explains the increase in IMSE as the interaction effect increases. In particular, HT_1 performs the worst as it only puts non-zero weights on treated exposures, and so the interaction effect is always present. However, even when the interaction effect is non-zero, i.e. when additivity does not hold, we see that there are instances when HT_{Avg} , M_{Ind} , and M_{Dil} outperform HT_0 . This is especially seen as the mean interference effect increases. Indeed, when the mean interference effect is equal to 50, the three estimators outperform HT_0 for all

of the values of interaction effects considered. This suggests that the estimators are fairly robust to violations of the additivity assumption, especially when the mean interference effect is large. Furthermore, as the interaction effect increases, there is a bigger distinction between the IMSE of HT_{Avg} and the IMSEs of M_{Ind} and M_{Dil} , which was not seen in the previous section when additivity holds. Hence, there is a benefit in using M_{Ind} and M_{Dil} , over HT_{Avg} , especially when additivity does not hold.

2.7.3 Varying Potential Outcome Distributions

Lastly, we compare estimators in settings with different potential outcome parameter distributions. In the previous sections, potential outcomes were sampled such that units and parameters were independent. When additivity holds and the true mean interference effect is zero, M_{Ind} outperforms the other estimators. In this section, in addition to the independent parameters, we also simulated potential outcome parameters under a dilated distribution where parameters are correlated. That is, $\alpha^{(i)} \sim \mathcal{N}(0, 1)$, $\theta_{2,1}^{(i)} = \alpha^{(i)}$, and $\theta_{1,d}^{(i)} = \frac{d}{d_i} \eta_1 \alpha^{(i)}$ for $\eta_1 = 0, 1, 5, 10, 50$. Under this setting, we expect M_{Dil} to perform the best. We compared results for a 4-regular graph with forty nodes.

Figure 2-7 shows the IMSEs of the different estimators under the independent and dilated potential outcome distributions (indicated by panel) as we vary η_1 or μ_1 (indicated by x-axis) for the dilated and independent distributions, respectively, and assuming that additivity holds. The IMSEs for estimators under the two different potential outcome distributions are fairly similar, with estimators using potential outcome parameters sampled from independent Normal distributions having slightly higher IMSEs. As seen in Chapter 2.7.2 when we varied the magnitude of the interference effect, the IMSEs of the estimators increase as the interference effect increases.

The performance of the estimators under different potential outcome distributions

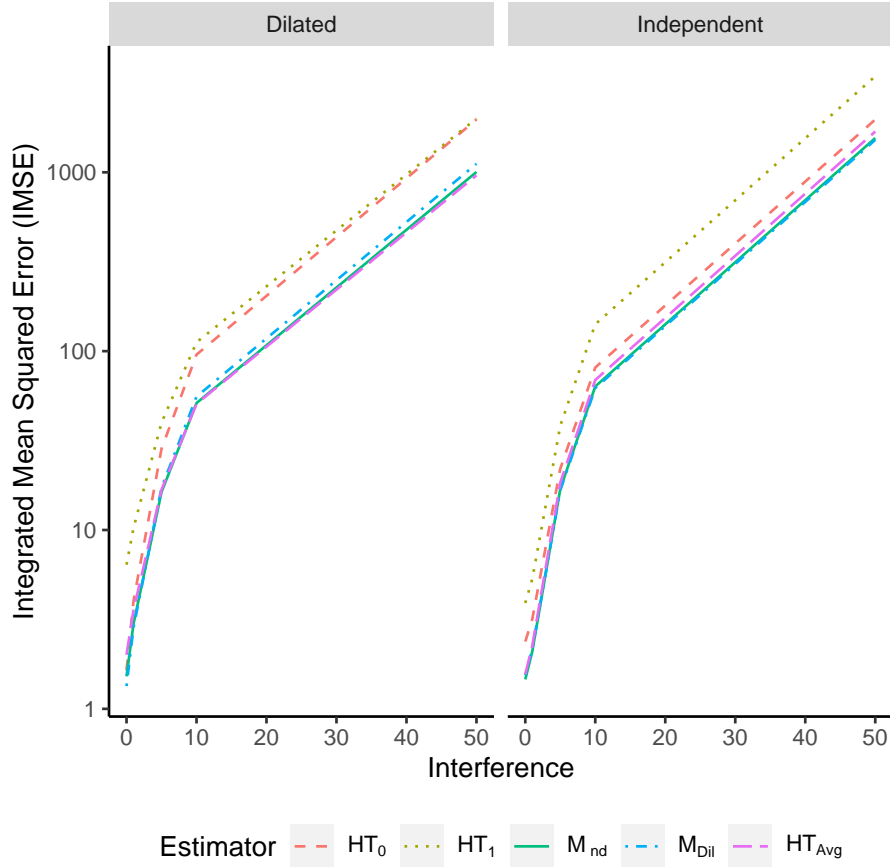


Figure 2.7: IMSE for estimators (indicated by color and line type) under different potential outcome distributions (indicated by panel) as the interference effect varies (indicated by x-axis) under additivity for a 40-node 4-regular graph.

reflected the results seen in Chapter 2.7.1. HT_1 performs worse than HT_0 , but HT_{Avg} , M_{Ind} , and M_{Dil} outperform HT_0 , with M_{Ind} and M_{Dil} generally performing the best. The multi-term MIV LUE whose prior distribution matches the distribution of the potential outcomes performs the best when the true mean interference effect was low, as expected. However, when μ_1 and η_1 increase, the multi-term MIV LUE whose prior distribution matches the distribution of the potential outcomes does not perform as well. Even so, the IMSEs of the two multi-term MIV LUEs are comparable. Hence, even if we use a prior distribution that does not match that of the potential outcomes, there is benefit in the multi-term MIV LUEs since they outperform the

other estimators.

2.8 Discussion

We proposed linear unbiased estimators for general causal effects as specified by exposure mappings under the assumption of additivity across exposure components. Under this assumption, the space of linear unbiased estimators becomes much larger, and exposures that are “seemingly unrelated” to the estimand of interest can contribute to the estimation. We can then leverage the information from units under other exposures not in the estimand of interest. Given the set of exposures, we defined linear constraints for when these LUEs exist, and we introduced a class of atomic estimators which, when combined with some unbiased estimators for zero, forms an affine basis for the set of LUEs. Additionally, we characterized an optimal subset of LUEs with minimum integrated variance.

In general, there is benefit to adding non-zero weight to more exposures. Even if we just take the average of the two-term Horvitz-Thompson estimators for untreated and treated units (hence putting non-zero weight on four exposures), we saw a significant reduction in IMSE compared to the IMSEs of each of the two-term estimators separately. If we further compute optimal weights for a LUE given a prior distribution, there is an additional reduction in the IMSE. However, these multi-term estimators are only LUEs under additivity. Under additivity, these multi-term estimators perform well in practice. Although we require additivity for theoretical results, the multi-term estimators are fairly robust to violations of additivity in practice. In fact, these multi-term estimators outperform two-term estimators for low levels of interaction effects and large interference effects.

Besides from additivity, we assumed that priors were uncorrelated between units which allowed for easier computation of the variances of estimators. By assuming

independent priors between units, we only had to account for the prior variances for unit i when computing the LUEs for the unit-level effect. Estimators may be derived to account for the covariance between units when computing the integrated variance. However, when priors are correlated between units, we are likely not able to derive a closed form solution. Furthermore, we did not discuss estimators for the variance in this chapter. To derive estimators for the variance, we may leverage the work done in Aronow et al. (2017), who derived estimators for variances for two-term estimators. However, in our work, estimators may have more than two terms. Since we have to account for covariances between the various Horvitz-Thompson terms, estimators for the variance could be quite complicated.

Although we focused on experimental settings, we would like to extend our multi-term MIV LUEs to observational studies as a next step. In the context of observational studies, we would likely have to account for noise in the exposure mapping and noise in the probability of exposures. In our current work, we did not make any assumptions about the treatment effects, but we did assume that the exposure mapping was known. If the exposure mapping used is not the true underlying exposure mapping, which could happen in both experiments and observational studies, then results may not be accurate. Aronow et al. (2017) showed that in the case when an exposure mapping maps two treatment allocations to the same exposure, but the potential outcomes under the two treatment allocations are different, the two-term estimator HT_0 is unbiased for a weighted average of the potential outcomes under the different treatment allocations. In our case, we could possibly account for the various types of noise in the exposure mapping. Furthermore, we assumed that the probability of exposures were known, which is typically not true in observational studies. In observational studies, we would have to estimate the probability of exposures using a model given covariate variables. Therefore, we would like an estimator that is doubly

robust (Robins et al., 1994; Li et al., 2021). However, unlike the typical doubly robust models, where one can make misspecifications in the outcome model or the treatment model, we would ideally want an estimator that is robust to misspecifications in the exposure mapping and/or the probability of exposure model. Lastly, inclusion of covariates was not discussed in this work, but we would likely benefit from including information from covariates when estimating treatment effects and can better quantify treatment effect heterogeneity. Again, we could possibly leverage the work of Aronow et al. (2017) who proposed linear unbiased estimators for treatment effects using models that account for the covariates.

In summary, in this chapter, we characterized the set of linear unbiased estimators under the assumption of additive exposures. We further specified conditions of supports of estimators that lead to MIV LUEs with non-zero weights on all exposures in the support. Using these proposed MIV LUEs, we saw an added benefit of incorporating information from all units as opposed to two-term LUEs which only place non-zero weight on units with exposures in the estimand of interest.

Chapter 3

Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths

3.1 Introduction

We now transition to an observational study where we estimate the treatment effect of drug-induced homicide (DIH) prosecutions reported by the media on drug overdose deaths. In particular, we focus on the question: *what is the effect of having any drug-induced homicide prosecutions versus none on drug overdose deaths?* We estimate the treatment effects using data that is collected over time. First, units can be treated at different time points, and so at each time point, the treatment and control groups can be different. In addition, treatments, confounding variables, and outcomes were collected at every time point—a repeated outcomes setting. With the additional consideration of time, we have to account for additional complexities. In particular, in a repeated outcomes setting, *confounding by indication* may occur. Confounding by indication occurs when there are differences between the treatment and control groups due to selection bias based on the predictors of the outcome. For example, states with a bigger drug overdose problem may be more likely to prosecute more people in hopes of potentially reducing the number of drug overdose deaths. If pre-treatment factors are not taken into account, then estimated treatment effects may be biased. Hence, there may be dependencies of treatments, outcomes, and confounding variables on

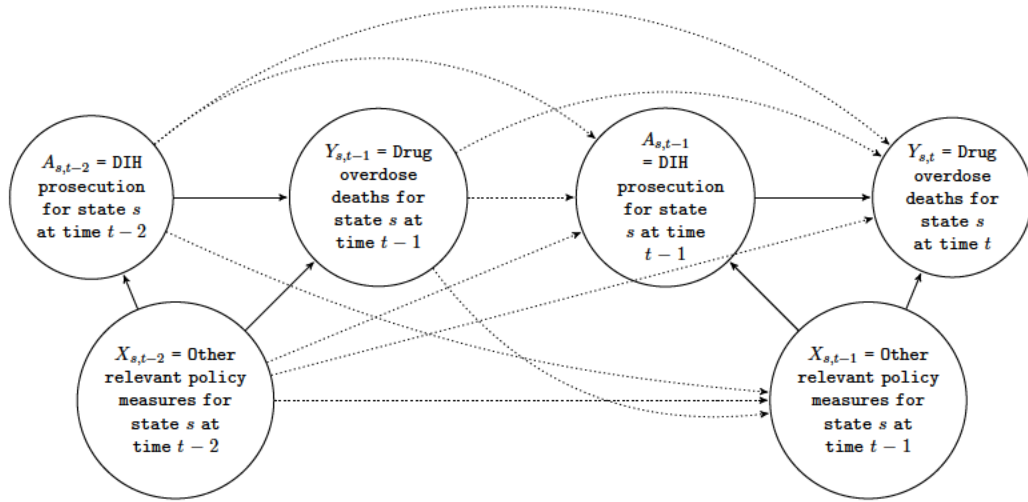


Figure 3-1: Causal diagram for a subset of time points for the question: *what is the effect of having any drug-induced homicide prosecutions vs. none on drug overdose deaths?*

the past treatments, outcomes, and confounding variables. Figure 3-1 shows the causal diagram for multiple time points. The dotted lines represent the potential dependencies between the different time points. Note that the causal diagram shown only represents a snapshot in time, and the full causal diagram would contain variables and dependencies for all time periods. Hence, the estimation of treatment effects of drug-induced homicide prosecutions on drug overdose deaths is more complicated than in traditional causal inference settings.

This work was done jointly with Leo Beletsky, J.D. (Northeastern University) and Natasha Martin, DPhil (University of California, San Diego).

3.2 Background

Since 2000, the drug overdose death rate tripled, from 7.0 deaths per 100,000 people in 2000 to 22.7 deaths per 100,000 people in 2019 (Xu et al., 2021). In just one year, from 2018 to 2019, the death rate for drug-induced causes increased by 4.6%

(Xu et al., 2021). As the world struggles with the COVID-19 pandemic, there are additional concerns that the coronavirus emergency may compound the overdose risk in the United States through exacerbating social isolation and despair (which are contributing factors to opioid addiction) and perturbing drug markets, further fueling the ongoing overdose crisis (Friedman et al., 2020a; Volkow, 2020; Wakeman et al., 2020). As the number of drug overdose deaths continued to grow through the 20th and 21st centuries, governments on all levels passed a number of policy measures. Although some of the policy measures passed followed an evidence-based public health approach (such as 911 Good Samaritan laws), there are also policy responses that are carceral and punitive. We focus on one of these policy responses: drug-induced homicide (DIH) prosecutions.

In 1986, the U.S. Congress passed the Anti-Drug Abuse Act, which includes a law in which anyone who knowingly or intentionally distributes controlled substances resulting in death is subject to a sentencing enhancement of a 20-year mandatory minimum in prison (21 U.S. Code §841, 1986). A number of states also passed similar DIH laws to punish those who distributed drugs resulting in death. However, until 2000, these laws were almost never invoked in drug law enforcement (Health in Justice Action Lab, 2021). As the drug overdose crisis grew, these DIH laws became an increasingly popular tool because of a theory that the perceived threat of harsh legal sanctions deters illicit drug activity, thus preventing overdose deaths. As of January 2019, the Prescription Drug Abuse Surveillance System (PDAPs) suggests that 23 states have specific DIH laws. However, many states without specific DIH laws have also used generic manslaughter or felony murder laws to charge drug-induced homicide.

States adopt DIH laws as deterrents in order to send a message by targeting high-level drug dealers with steep penalties, thereby deterring drug trafficking, reducing

illicit drug supply, and preventing overdoses as a result (New Jersey Revised Statutes §2C:35-1.1, 1987). Per the wording of DIH laws, though, anyone who distributes controlled substances, including friends and family members of victims, can be prosecuted (People v. Boand, 838 N.E.2d, 2005). Thus, the message intended for high-level drug dealers may be reaching friends and family members of victims and low-level drug dealers instead, as these are the individuals who are modally prosecuted (LaSalle, 2017; Walker, 2017).

As a result, due to legal liability concerns, DIH prosecutions may be discouraging people from seeking help during a drug overdose (Carroll et al., 2021) and may in fact be aggravating the overdose risk. Despite these conflicting notions, the effect of DIH prosecutions on overdose death risk has never been empirically assessed. Heterogeneity in state-level adoption of DIH prosecutions in response to the overdose crisis provides an opportunity for quantitative modeling of the effect of DIH prosecutions on fatal overdose patterns. Here, we estimate the effect of DIH prosecutions reported by the media on drug overdose deaths.

3.3 Data

3.3.1 Outcome: Unintentional Drug Overdose Deaths

To assess the effect of DIH prosecutions reported by the media on drug overdose deaths, we obtained monthly outcome data on unintentional drug overdose deaths for people who were at least 18 years old for all 50 U.S. states from 1999 to 2019 from the Centers for Disease Control and Prevention (CDC), using International Classification of Diseases, Tenth Revision codes X40-X44. However, the CDC suppresses the number of drug overdose deaths if it is below ten, leading to missing monthly data. Even though the monthly data were sometimes missing, we were able to obtain most of the yearly data since almost all of the number of unintentional drug overdose deaths

surpassed ten when aggregated yearly. If the yearly total number of drug overdose deaths were missing in a state, we imputed the missing yearly totals by dividing the total number of unaccounted drug overdose deaths in that state from 2000 to 2019 equally by the number of years for which the yearly number of drug overdose deaths were missing.

Using the yearly number of drug overdose deaths, combined with linear interpolation, we imputed the missing monthly number of unintentional drug overdose deaths in 2000-2019 as follows. For state s , we first linearly interpolated the number of unintentional drug overdose deaths for month m in year u with missing outcomes (indicated by $C_{s,u,m} = 1$). Denote the linearly interpolated number of unintentional drug overdose deaths in state s for month m in year u by $y'_{s,u,m}$. To ensure that the imputed monthly drug overdose death data were consistent with the yearly CDC data, we weighted the interpolated values as follows. First, we computed the number of yearly overdose deaths that were not accounted for by the non-missing monthly data using the observed number of yearly drug overdose deaths ($y_{s,u}$) and the observed, unsuppressed, number of monthly drug overdose deaths ($y_{s,u,m}$):

$$\# \text{ of unaccounted OD deaths in state } s, \text{ year } u = y_{s,u} - \sum_{\tilde{m}=1}^{12} y_{s,u,\tilde{m}} (1 - C_{s,u,\tilde{m}}).$$

Using the total number of drug overdose deaths that were not accounted for by the monthly data, we scaled the linearly interpolated outcomes, $y'_{s,u,m}$, by a factor so that the total number of imputed drug overdose deaths in a year is equal to the total number of drug overdose deaths that actually needed to be imputed for that year for that state. Thus, the number of imputed unintentional drug overdose deaths for state s for year u at month m was:

$$\tilde{y}_{s,u,m} = \frac{y_{s,u} - \sum_{\tilde{m}=1}^{12} y_{s,u,\tilde{m}} (1 - C_{s,u,\tilde{m}})}{\sum_{\tilde{m}=1}^{12} y'_{s,u,\tilde{m}} C_{s,u,\tilde{m}}} y'_{s,u,m}. \quad (3.1)$$

If there were no suppressed values in state s in year u , we did not need to calculate this imputed value. If we could not linearly interpolate the values for a specific state-year-month combination (e.g. when there were a sequence of suppressed monthly drug overdose deaths), we imputed the number of overdose deaths by dividing the total number of unaccounted overdose deaths in state s in year u evenly amongst the months for which the outcome data was suppressed:

$$\tilde{y}_{s,u,m} = \frac{y_{s,u} - \sum_{\tilde{m}=1}^{12} y_{s,u,\tilde{m}} (1 - C_{s,u,\tilde{m}})}{\sum_{\tilde{m}=1}^{12} C_{s,u,\tilde{m}}}. \quad (3.2)$$

We then aggregated the number of unintentional drug overdose deaths into six-month intervals (from January to June and July to December) per year and analyzed the data as such. Since, in general, we observed the yearly number of drug overdose deaths, the imputation process of the missing monthly unintentional drug overdose deaths essentially just allocated the observed yearly number of overdose deaths to the two six-month periods.

3.3.2 DIH Prosecutions and Other Policy Measures

In the absence of centralized criminal justice data tracking DIH prosecutions, we used data detailing mass media coverage of DIH prosecutions in 2000-2019 collected by the Health in Justice Action Lab (Health in Justice Action Lab, 2021). Using Media Cloud, a web-scraping media analysis tool, web articles published between 1978 and April 2021 that were related to drug-induced homicide cases were identified. We then filtered the data to exclude DIH prosecutions reported by the media where the victims were known to be younger than 18 years old. The resulting dataset was then analyzed to check for relevance and duplication. We focused on DIH prosecution media reports because media reports are a good medium to spread information about DIH prosecutions since they are readily available and widely accessible. Furthermore,

media reports captured the intended effect of DIH prosecutions—to deter people from using and distributing drugs by instilling fear of being prosecuted for homicide.

The dataset used for the analysis consisted of information such as the date and the state where each individual was charged. For each state, the *intervention date* for DIH prosecutions reported by the media was the first charge date that we found for DIH prosecutions reported by the media in the state in 2000-2019. We denote the intervention date for state s as T_s^* . The intervention dates for the intervention of interest, DIH prosecutions reported by the media, for each state can be found in Table B.1 in Appendix B.1. We focused on the years 2000-2019 because 1) the CDC data on unintentional drug overdose deaths date back only to 1999 and 2) prior to 2000, states rarely prosecuted people for drug-induced homicide (Figure 3·2).

To address the growing drug overdose risk during this time period, states also adopted other policies. To account for potential confounding due to other policy measures, we obtained data on the following additional policies that have been seen in prior research to potentially impact drug overdose deaths (Abouk et al., 2019; McClellan et al., 2018). From the Prescription Drug Abuse Policy System (PDAPS), we collected data on the presence of naloxone access laws— one law where pharmacists can dispense naloxone without a prescription (NAL: can dispense) and one law where pharmacists cannot dispense naloxone without a prescription (NAL: cannot dispense), the presence of legalized medical marijuana laws (MML) and recreational marijuana laws (RML), the presence of 911 Good Samaritan laws (911 GSL), and the presence of Prescription Drug Monitoring Programs (PDMP) in each state. Data on whether a state expanded Medicaid as part of the Affordable Care Act were also collected from the Henry J. Kaiser Family Foundation. Intervention dates for the relevant policy measures for each state can be found in Table B.2 in Appendix B.1.

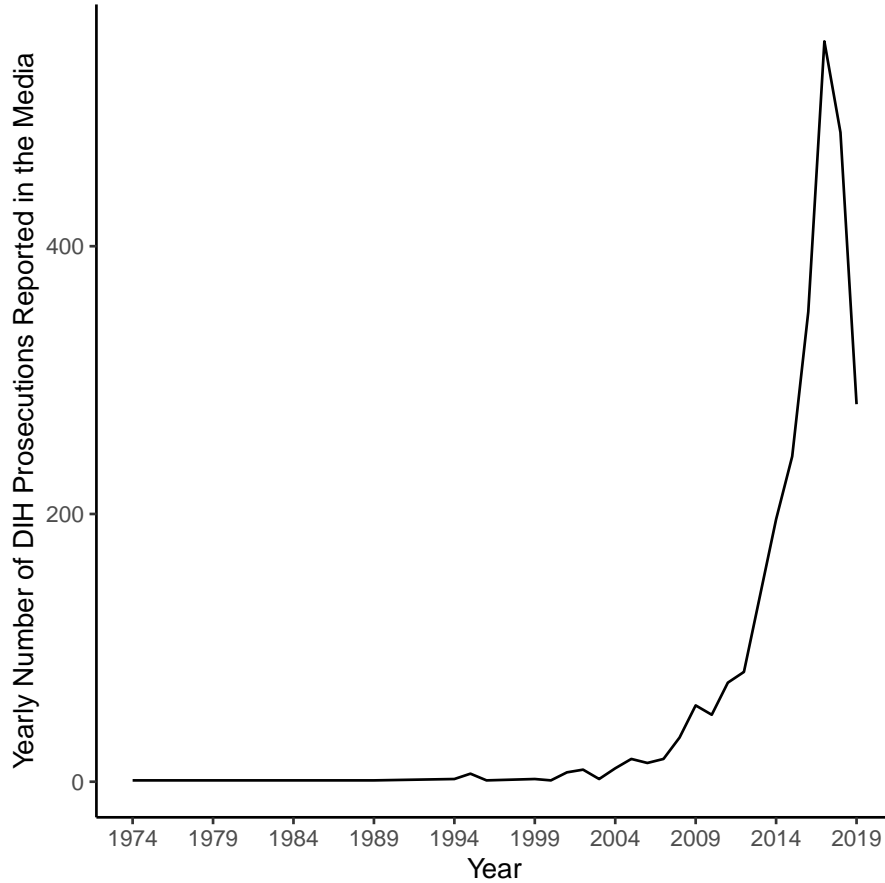


Figure 3.2: The yearly number of drug-induced homicide charges per year in the U.S. from 1974-2019 collected by the Health in Justice Action Lab.

3.3.3 Intervention Definition

Our primary focus was to estimate the effect of having at least one DIH prosecution media report versus never having any DIH prosecutions reported by the media on unintentional drug overdose deaths. We defined the intervention variables for the other relevant policy measures and the DIH prosecutions reported by the media in the same way, and so we describe the intervention variable in the context of DIH prosecutions reported by the media. Denote the six-month time interval t as $\mathcal{I}_t = (D_{t,1}, D_{t,L})$ where $D_{t,1}$ and $D_{t,L}$ indicate the first and last dates of time interval t , respectively. For each time interval t , the intervention variable was defined as the

proportion of days in time interval t in which the intervention was in effect:

$$A_{s,t}^{\text{primary}} = \begin{cases} 0 & \text{if } D_{t,L} < T_s^* \\ 1 & \text{if } D_{t,1} > T_s^* \\ \frac{\# \text{ of days exposed to intervention in } \mathcal{I}_t}{\text{number of days in } \mathcal{I}_t} & \text{if } T_s^* \in \mathcal{I}_t \end{cases} \quad (3.3)$$

That is, if by the end of time interval t , there had not been any DIH prosecutions reported by the media in state s , then $A_{s,t}^{\text{primary}} = 0$. If by the beginning of time interval t , there had been at least one DIH prosecution reported by the media in state s , then $A_{s,t}^{\text{primary}} = 1$. Otherwise, if the intervention date for state s occurred in \mathcal{I}_t , $A_{s,t}^{\text{primary}}$ is equal to the proportion of days in the time interval in which the state was exposed to the intervention.

As a secondary analysis, we assumed that the treatment effect of DIH prosecutions reported by the media only lasted for two years. In this secondary analysis, the intervention variables for the other policy measures were defined as in the primary analysis. However, the intervention variable for DIH prosecutions reported by the media was defined slightly different. We now have to account for all the DIH prosecutions reported by the media in state s , as opposed to just the first one. Let $k_{s,1}(t)$ and $k_{s,L}(t)$ be the indices of the first and last DIH prosecutions reported by the media in state s at time interval t , respectively. If there were no DIH prosecutions reported by the media in state s at time interval t , then $k_{s,1}(t) = k_{s,L}(t) = 0$. Similar to the primary analysis, for each state s and time interval t , we computed the proportion of

time that a state was exposed to DIH prosecutions reported by the media:

$$A_{s,t}^{\text{sec}} = \begin{cases} 1 & \text{if } k_{s,1}(t-3), k_{s,1}(t-2), \\ & \text{or } k_{s,1}(t-1) > 0 \\ \frac{\# \text{ of days exposed to intervention } k_{s,L}(t-4) \text{ in } \mathcal{I}_t}{\text{number of days in } \mathcal{I}_t} & \text{if } k_{s,L}(t-4) > 0, k_{s,1}(t) = 0 \\ \frac{\# \text{ of days exposed to intervention } k_{s,1}(t) \text{ in } \mathcal{I}_t}{\text{number of days in } \mathcal{I}_t} & \text{if } k_{s,L}(t-4) = 0, k_{s,1}(t) > 0 \\ \frac{\# \text{ of days exposed to interventions } k_{s,L}(t-4) \text{ or } k_{s,1}(t) \text{ in } \mathcal{I}_t}{\text{number of days in } \mathcal{I}_t} & \text{if } k_{s,L}(t-4) > 0, k_{s,1}(t) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.4)$$

Recall that each time interval t consists of a six-month time period. Therefore, $t - M$ is the indicator for a six-month time interval occurring $\frac{M}{2}$ years before time interval t , e.g. \mathcal{I}_{t-4} indicates a time interval that occurred two years before time interval t . If there were any previous DIH prosecutions reported by the media in state s within the time intervals \mathcal{I}_{t-3} , \mathcal{I}_{t-2} , or \mathcal{I}_{t-1} , then $A_{s,t}^{\text{sec}} = 1$. Otherwise, if there were any previous DIH prosecutions reported by the media in state s two years prior to time interval t , i.e. $k_{s,L}(t-4) > 0$, but no DIH prosecutions reported by the media in time interval t , then $A_{s,t}^{\text{sec}}$ is equal to the proportion of days in time interval t exposed to the last intervention from two years ago. Otherwise, if there were no previous DIH prosecutions reported by the media in state s two years prior to time interval t , but there were DIH prosecutions reported by the media in time interval t , then $A_{s,t}^{\text{sec}}$ is equal to the proportion of days in time interval t exposed to interventions in time interval t . Otherwise, if there were DIH prosecutions reported by the media in state s in \mathcal{I}_{t-4} and \mathcal{I}_t , then $A_{s,t}^{\text{sec}}$ is equal to the proportion of days in time interval t exposed to DIH prosecutions reported by the media either from two years ago and/or DIH prosecutions reported by the media in time interval t . Otherwise, $A_{s,t}^{\text{sec}} = 0$.

3.4 Constant Treatment Effect

3.4.1 Model

We first assumed a constant treatment effect for DIH prosecutions reported by the media on unintentional drug overdose deaths. To analyze the effect of DIH prosecutions reported by the media on the unintentional drug overdose death risk when assuming a constant treatment effect, we used a difference-in-differences-like generalized additive model (GAM). The difference-in-differences model (Snow, 1856; Bertrand et al., 2004) is often used to estimate policy effects by comparing outcomes before and after policies were enacted (Powell et al., 2018; Abouk et al., 2019). The difference-in-differences model relies on the *parallel trends assumption*, where in the absence of the treatment, the difference in outcomes in the treatment and control groups is assumed to be constant. We can then estimate potential outcomes for the treatment group, had the treatment not occurred, for post-treatment time periods. The treatment effect in the treated is the difference between the observed outcome for the treatment group and the potential outcome had the treatment group not received treatment. In a classic difference-in-differences setting, there are only two time periods (before and after treatment) and two groups of units (treated units and control units). In our analysis, we applied the differences-in-differences model to a setting with forty time periods and fifty states which can be treated at different time periods (Bertrand et al., 2004). Hence, we also had to account for state and time effects, which we estimated using a semi-parametric generalized additive model.

The generalized additive model (GAM) (Hastie and Tibshirani, 1990) assumes that a function of the mean outcome is given by the summation of functions of $m < \infty$ predictors:

$$g(\mathbb{E}(Y)) = \sum_m f_m(X_m), \quad (3.5)$$

where g is a link function and f_m can be parametric or non-parametric functions such as regression splines. For example, the ordinary least squares (OLS) model is a specific example of GAMs, in which g is the identity link function and $f_m(X_m) = \beta_m X_m$. By using a GAM, we can estimate treatment effects using both non-parametric and parametric functions. We used a GAM to estimate smoothed time effects over different U.S. regions using cubic regression splines. The list of states and their U.S. regions were obtained from the U.S. Census (U.S. Census, 2018) (see Table B.3 in the Appendix B.2).

We estimated the effects of the various interventions on a risk ratio scale as opposed to the risk difference scale (Powell et al., 2018; Abouk et al., 2019), since the intervention effect could be proportionally higher in states with a higher overdose death risk. Since the probability of a drug overdose death for a random person was small (with a maximum probability of approximately 0.0003 in any state in a six-month time interval), the odds ratio was essentially equivalent to the risk ratio. Therefore, we fit the following logistic GAM to predict the overdose death risk:

$$\text{logit}(\mathbb{E}(Y_{s,t} \mid s, t, N_t, X_{s,t}, A_{s,t})) = \alpha_s + \gamma_{r(s)}(t) + N_t\psi + X_{s,t}\delta + A_{s,t}\beta, \quad (3.6)$$

where $Y_{s,t}$ is the risk of unintentional drug overdose deaths in a state s at time interval t , α_s indicates the state fixed effect of state s , $\gamma_{r(s)}(t)$ indicates the smoothed time effects, which could differ by the U.S. regions of the states (denoted by $r(s)$) and estimated using cubic regression splines, N_t indicates the number of states exposed to DIH prosecutions reported by the media by time t , $X_{s,t}$ indicates the various policy measures (given by Equation (3.3)), and $A_{s,t}$ indicates the intervention of having any DIH prosecutions reported by the media versus none given by Equations (3.3) and (3.4). Note we included N_t as a variable to measure additional time effects to avoid treatment effects picking up time effects.

3.4.2 Sandwich Estimator for Variance of $\hat{\Theta}$

Under the logistic GAM, we typically assume independence between units. However, this assumption may not hold in our setting of drug overdose deaths since there may be dependencies between neighboring states and also within states between the different time intervals. Hence, we could not use the standard errors estimated by standard software. Instead, we derived a sandwich estimator (Van der Vaart, 1998) for the standard errors of the logistic GAM (see Appendix B.3.1 for details). Denote the vector of state fixed effects, time effects, other relevant policy measures, and the DIH prosecutions reported by the media for state s at time interval t by

$$\vec{Z}_{s,t} = \begin{pmatrix} \mathbb{I}\{\text{state} = s\} \\ \mathbb{I}\{\text{time} = t\} \\ N_t \\ X_{s,t} \\ A_{s,t} \end{pmatrix}. \quad (3.7)$$

To derive the sandwich estimator, we made the following two assumptions.

Assumption 3.1. Consider states s and s' , where if $s = s'$, let $t' < t$, and if $s \neq s'$, let $t' \leq t$. Given $\vec{Z}_{s,t}$ for state s and $\vec{Z}_{s',t'}$ and $Y_{s',t'}$ for state s' at time interval t' ,

$$Y_{s,t} \perp\!\!\!\perp \left(\vec{Z}_{s',t'}, Y_{s',t'} \right) \mid \vec{Z}_{s,t}. \quad (3.8)$$

Assumption 3.1 assumes that the dependence of unintentional drug overdose death rates in state s at time interval t on the state's (and on other states') past unintentional drug overdose death rates, DIH prosecutions reported by the media, and relevant policy measures was through the DIH prosecutions reported by the media and relevant policy measures of state s at time interval t . Furthermore, we assumed that given $\vec{Z}_{s,t}$, we have a correctly specified model for $p_{s,t}$.

Assumption 3.2. Given the state fixed effects, time effects, other relevant policy measures, and the DIH prosecutions reported by the media for state s at time interval

t , denoted as $\vec{Z}_{s,t}$,

$$\text{logit} \left(\mathbb{E}(Y_{s,t} \mid \vec{Z}_{s,t} = \vec{z}_{s,t}) \right) = p_{s,t,\Theta^*}, \quad (3.9)$$

where p_{s,t,Θ^*} , given by the true parameter Θ^* , is a correctly specified model for $p_{s,t}$.

Under Assumptions 3.1 and 3.2, unintentional drug overdose deaths in a state may depend on different states and on past time intervals, but the unintentional drug overdose deaths in a state are uncorrelated with the outcomes and predictors from different states and from past time intervals (see Appendix B.3.1 for a proof). The sandwich estimator for the variance-covariance matrix equals:

$$\hat{\Sigma} = \frac{N}{N-d} C^{-1} \left(\sum_{s,t} (Y_{s,t} - p_{s,t,\hat{\Theta}})^2 \vec{Z}_{s,t} \vec{Z}_{s,t}^T \right) (C^{-1})^T, \quad (3.10)$$

where

$$C = \sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T p_{s,t,\hat{\Theta}} (1 - p_{s,t,\hat{\Theta}}),$$

where we estimated the true parameter Θ^* by $\hat{\Theta}$. We included a factor of $\frac{N}{N-d}$ in Equation (3.10) for bias correction, where N equals the number of state-time combinations and d equals the number of parameters (Li and Redden, 2015).

3.4.3 Estimating the Number of Attributable Deaths

We also estimated the number of excess drug overdose deaths attributable to DIH prosecutions reported by the media. To estimate the number of deaths in a state that is attributable to drug-induced homicide prosecutions reported by the media, we derived an estimator that is similar to the etiologic fraction (Miettinen, 1974). Let $n_{s,t,\text{attributable deaths}}$ be the number of deaths attributable to DIH prosecutions reported by the media in state s for time intervals t where there were at least one DIH prosecution reported by the media. The number of observed deaths in state s

at time interval t is denoted by $n_{s,t,\text{observed overdose deaths}}$, and $n_{s,t,\text{overdose deaths}}^{(a_s,t=0)}$ denotes the number of unintentional drug overdose deaths that would have occurred had the intervention not occurred in state s at time interval t . The number of drug overdose deaths attributable to DIH prosecutions reported by the media was estimated by computing the difference between the observed number of drug overdose deaths and the estimated number of deaths had there not been any DIH prosecutions reported by the media:

$$\hat{n}_{s,t,\text{attributable deaths}} = n_{s,t,\text{observed overdose deaths}} - \hat{n}_{s,t,\text{overdose deaths}}^{(a_s,t=0)}, \quad (3.11)$$

where

$$\hat{n}_{s,t,\text{overdose deaths}}^{(a_s,t=0)} = n_{s,t,\text{population}} \times \hat{p}_{s,t}^{(a_s,t=0)}. \quad (3.12)$$

That is, we estimated the number of drug overdose deaths had there not been any DIH prosecutions reported by the media by multiplying the population in state s at time interval t by the estimated probability of drug overdose deaths had there not been any DIH prosecutions reported by the media for state s at time interval t , denoted as $\hat{p}_{s,t}^{(a_s,t=0)}$. Using the estimated effect of DIH prosecutions reported by the media, we estimated $\hat{p}_{s,t}^{(a_s,t=0)}$ using (see Appendix B.4.1 for more details):

$$\hat{p}_{s,t}^{(a_s,t=0)} = \text{expit} \left(\text{logit} \left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}} \right) - A_{s,t} \hat{\beta} \right), \quad (3.13)$$

where $\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}$ is the observed risk of unintentional drug overdose deaths under the actual intervention. To find a 95% confidence interval for the number of attributable deaths, we substituted $\hat{\beta}$ by the 95% confidence interval upper and lower limits of β .

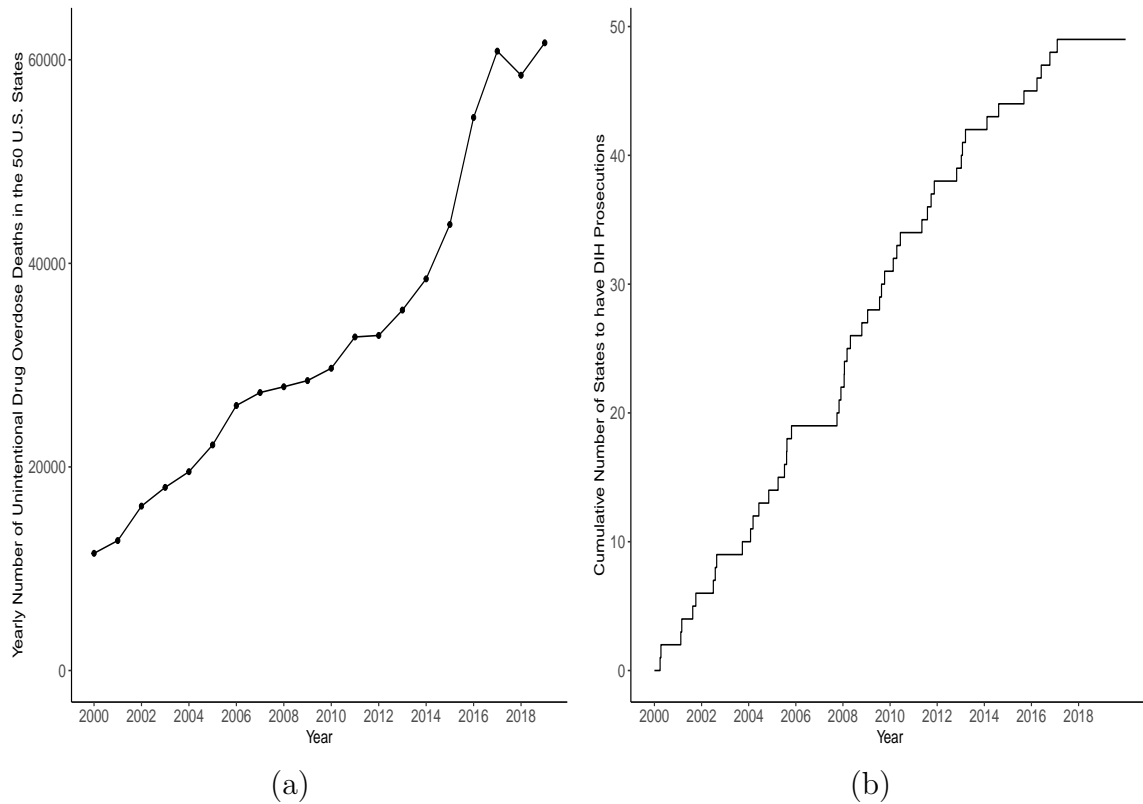


Figure 3-3: (a) Yearly number of unintentional drug overdose deaths in the 50 U.S. states from 2000 to 2019; and (b) Cumulative number of states exposed to DIH prosecutions from 2000 to 2019.

3.4.4 Results

Figure 3-3(a) presents the yearly total number of unintentional drug overdose deaths in the 50 U.S. states in 2000-2019 for victims who were at least 18 years old. The total number of unintentional drug overdose deaths in the 50 U.S. states for those at least 18 years of age ranged from 11,514 deaths in 2000 to 61,665 deaths in 2019. In total, there were approximately 658,215 unintentional drug overdose deaths for those at least 18 years of age in the U.S. from 2000 to 2019. Blue, solid lines in Figure 3-4 show the observed unintentional drug overdose death rates per 100,000 for each state. Red, dashed lines in Figure 3-4 show the fitted values for the unintentional drug overdose death rates per 100,000 people under the assumption of constant treatment

effect. Figure 3-3(b) shows the cumulative number of states that had at least one DIH prosecution media reported by the media from 2000 to 2019. The black vertical lines in Figure 3-4 indicate the first intervention date of a DIH prosecution reported by the media in each state. Hawaii was the only state without any DIH prosecutions reported by the media by the end of 2019.

Exposure to DIH prosecutions reported by the media was associated with a 6.4% increase in unintentional drug overdose deaths (risk ratio: 1.064; 95% CI: (1.012, 1.118)). Table B.4 in Appendix B.5 presents the risk ratios and 95% confidence intervals for the exposure to DIH prosecution reported by the media and the other relevant policy measures. We estimated a total of 33,115 (95% CI: (12,684, 52,776)) unintentional drug overdose deaths were attributable to DIH prosecutions reported by the media in all states from 2000 to 2019 (Figure 3-5), which made up approximately 5.03% (95% CI: (1.93%, 8.02%)) of the total number of unintentional drug overdose deaths from 2000 to 2019. When we assumed that the effect of DIH prosecutions reported by the media only lasted for two years, exposure to DIH prosecutions reported by the media was associated with a 5.9% increase in unintentional drug overdose deaths (risk ratio: 1.059; 95% CI: (1.013, 1.107)). Further, we estimated a total of 28,145 (95% CI: (6,511, 48,850)) unintentional drug overdose deaths attributable to DIH prosecutions reported by the media which made up approximately 4.28% (95% CI: (1.92%, 6.55%)) of the total number of unintentional drug overdose deaths from 2000 to 2019. Hence, our results suggest that, when assuming a constant treatment effect and controlling for other factors, exposure to DIH prosecutions reported by the media may not decrease the number of drug overdose deaths. Instead, exposure to DIH prosecution media reports is associated with an increase in drug overdose deaths.

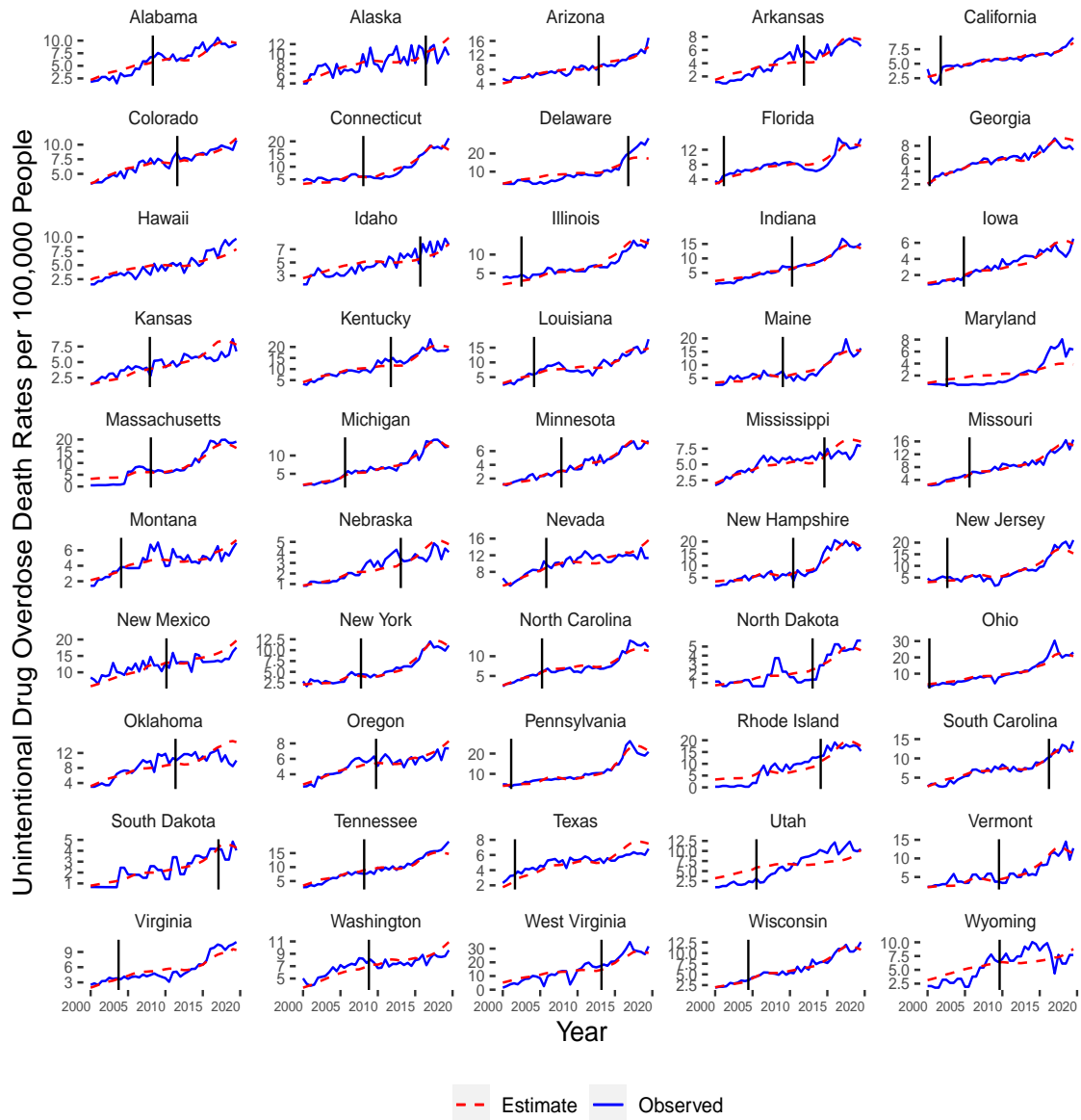


Figure 3.4: Results of fitted models (red and dashed) plotted against observed trend (blue and solid) for all U.S. states in 2000-2019. Vertical black lines indicate the time of the first DIH prosecution reported by the media for each state.

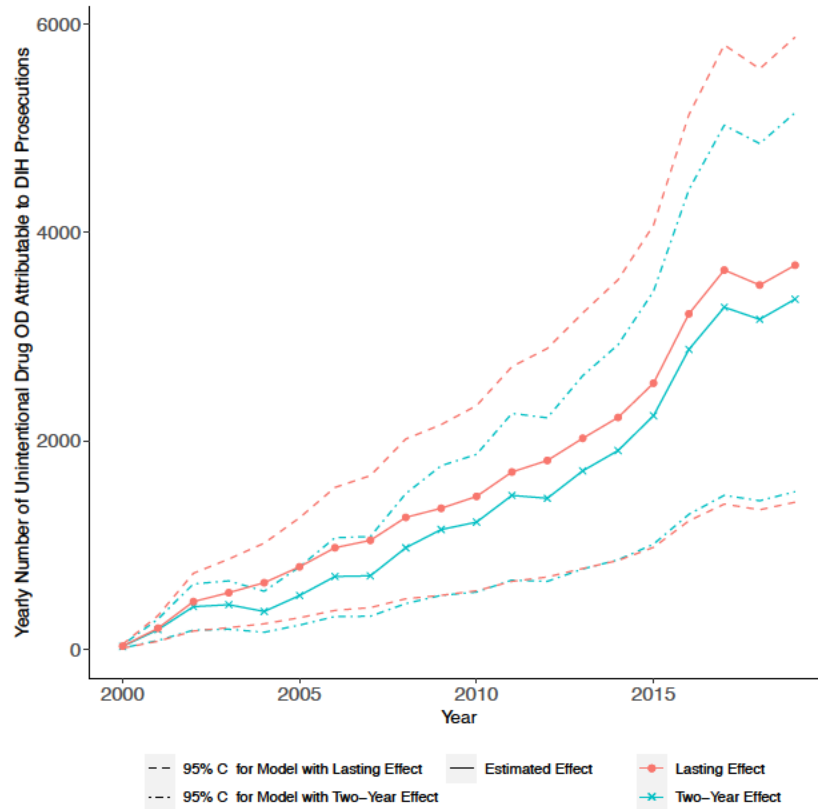


Figure 3-5: Estimated yearly number of unintentional drug overdose deaths attributable to DIH prosecutions reported by the media for main analysis (red with circles) and secondary analysis (blue with crosses) with 95% confidence intervals (dashed lines for main analysis and dot-dashed lines for secondary analysis) in all 50 U.S. states from 2000 to 2019.

3.5 Potential Issues with the Constant Treatment Effect Model in Staggered Treatment Settings

When we assumed that the treatment effect of DIH prosecutions reported by the media on drug overdose deaths was constant, we concluded that DIH prosecutions reported by the media may have a harmful effect on drug overdose deaths. However, there may be issues with the constant treatment effect model given by Equation (3.6).

First, for models which are linear in parameters, coefficients are consistent when predictors are uncorrelated with the errors and the inner product of the design matrix

is full rank (Wooldridge, 2010, p. 53, Theorem 4.1). Hence, even if there are not enough data points to accurately estimate fixed effects, the estimated treatment effect will still be consistent under the linearity in parameters. However, consistency of parameters in non-linear models such as logistic regression models requires stronger assumptions, and so consistency of parameters in non-linear models may not hold (Wooldridge, 2010, p. 348, Theorem 12.2). If we do not have enough data points to accurately estimate the fixed effects, the estimated treatment effect may also not be accurate. Although we have forty data points for each state, we conducted a sensitivity analysis using a GAM with a linear link function to compare our results to a logistic GAM. We fit the same model as in Equation (3.6), but using a GAM with a linear link function as opposed to a logit link function. For the GAM with a linear link function, we transformed the risk of unintentional drug overdose deaths using the log transformation. Because of the linearity in parameters in the GAM with a linear link function, coefficient estimates are consistent. The risk ratios of the other relevant policy measures with their 95% confidence intervals are found in Table B.5 in Appendix B.5. Using a linear link function, we estimated that having any DIH prosecutions reported by the media when assuming that the treatment effect lasts until at least 2019 is associated with a risk ratio of 0.975 (95% CI: (0.931, 1.021)), but it is not statistically significant at level $\alpha = 0.05$. When we assumed that treatment effect only lasts for two years, we estimated a risk ratio of 1.014 (95% CI: (0.974, 1.056)). Hence, under a different model specification, the effect of DIH prosecutions reported by the media was no longer statistically significant at level $\alpha = 0.05$, and we even estimated a potential protective effect.

Another potential issue with the difference-in-differences model given by Equation (3.6) is that we assumed that the treatment effects are constant. Consider a general OLS difference-in-differences model with a constant treatment effect in a set-

ting with more than two time periods:

$$Y_{it} = \alpha_i + \gamma_t + \beta A_{it} + \epsilon_{it}, \quad (3.14)$$

where Y_{it} indicates the outcome for unit i at time t , α_i and γ_t indicate the unit and time fixed effects, respectively, A_{it} is a treatment indicator where once a unit is treated, it remains treated, and ϵ_{it} is a zero-mean error term. Following Borusyak et al. (2021), by the Frisch-Waugh-Lovell theorem (Frisch and Waugh, 1933), we can estimate β from Equation (3.14) using the following model:

$$\tilde{Y}_{it} = \tilde{A}_{it}\beta + \tilde{\epsilon}_{it}, \quad (3.15)$$

where \tilde{Y}_{it} and \tilde{A}_{it} are the residuals for the regressions regressing Y_{it} and A_{it} on the fixed unit and time effects, respectively (Borusyak et al., 2021). Furthermore, we can model Y_{it} using the individual treatment effects: $\mathbb{E}(Y_{it}|A_{it}) = \alpha_i + \gamma_t + \beta_{it}A_{it}$ (Borusyak et al., 2021). The coefficient β is given by:

$$\beta = \frac{\text{cov}(\tilde{Y}_{it}, \tilde{A}_{it})}{\text{var}(\tilde{A}_{it})} = \frac{\text{cov}(Y_{it} - \hat{\alpha}_i^{(Y_{it})} - \hat{\gamma}_t^{(Y_{it})}, \tilde{A}_{it})}{\text{var}(\tilde{A}_{it})} = \frac{\text{cov}(Y_{it}, \tilde{A}_{it})}{\text{var}(\tilde{A}_{it})},$$

where the last equality holds since the residuals \tilde{A}_{it} are orthogonal to the projection of Y_{it} on the space of unit and time fixed effects given by $\hat{\alpha}_i^{(Y_{it})} + \hat{\gamma}_t^{(Y_{it})}$. We can then estimate β as follows:

$$\hat{\beta} = \frac{\sum_{it} Y_{it} \tilde{A}_{it}}{\sum_{it} \tilde{A}_{it}^2} = \frac{\sum_{it} (\hat{\alpha}_i + \hat{\gamma}_t + \hat{\beta}_{it} A_{it}) \tilde{A}_{it}}{\sum_{it} \tilde{A}_{it}^2} = \frac{\sum_{it} \hat{\beta}_{it} A_{it} \tilde{A}_{it}}{\sum_{it} \tilde{A}_{it}^2},$$

where the last equality holds since the residuals are orthogonal to $\hat{\alpha}_i$ and $\hat{\gamma}_t$. Letting

$w_{it} = \frac{\tilde{A}_{it}}{\sum_{it} \tilde{A}_{it}^2}$, we have

$$\hat{\beta} = \sum_{it} w_{it} \hat{\beta}_{it} A_{it} = \sum_{it: A_{it}=1} w_{it} \hat{\beta}_{it}. \quad (3.16)$$

Furthermore, note that $\sum_{it: A_{it}=1} w_{it} = 1$ since

$$\sum_{it: A_{it}=1} w_{it} = \sum_{it: A_{it}=1} \frac{\tilde{A}_{it}}{\sum_{it} \tilde{A}_{it}^2} = \frac{\sum_{it} \tilde{A}_{it} A_{it}}{\sum_{it} \tilde{A}_{it}^2} = \frac{\sum_{it} \tilde{A}_{it} (\tilde{A}_{it} + \hat{\alpha}_i^{(A_{it})} + \hat{\gamma}_t^{(A_{it})})}{\sum_{it} \tilde{A}_{it}^2} = \frac{\sum_{it} \tilde{A}_{it}^2}{\sum_{it} \tilde{A}_{it}^2} = 1,$$

where the second equality holds because A_{it} is a binary treatment indicator, the third equality holds because $\hat{\alpha}_i^{(A_{it})}$ and $\hat{\gamma}_t^{(A_{it})}$ are the coefficients for the unit and time fixed effects using A_{it} as the response variable, and the second to last equality holds because $\sum_{it} \hat{\alpha}_i^{(A_{it})} \tilde{A}_{it} = 0$ and $\sum_{it} \hat{\gamma}_t^{(A_{it})} \tilde{A}_{it} = 0$ since residuals are orthogonal to $\hat{\alpha}_i^{(A_{it})}$ and $\hat{\gamma}_t^{(A_{it})}$. Hence, the constant treatment effect β is a weighted average of the individual treatment effects β_{it} . However, the weights w_{it} may not be “proper” in the sense that some w_{it} may be negative in settings where treatment is staggered and treatment effects are heterogeneous (Borusyak et al., 2021).

Typically in difference-in-differences, one compares outcomes of the treated group after treatment occurred with the outcomes of the reference group where units are untreated. However, when treatment is staggered, i.e. when units are treated at different times, the reference group used could be made up of treated units where, throughout the entire time period considered, the treatment variable does not change. OLS uses these treated units to estimate the time fixed effects; a “forbidden comparison” (Borusyak et al., 2021). When the true treatment effects are heterogeneous, but one assumes a constant treatment effect, these forbidden comparisons lead to biased estimates of the treatment effect (Borusyak et al., 2021). In particular, in settings with staggered treatment, there may be an overweighting of short-term effects and negative weighting (or at least underweighting) of long-term effects (De Chaisemartin

and d’Haultfoeuille, 2018; De Chaisemartin and d’Haultfoeuille, 2020; Borusyak et al., 2021; Roth et al., 2022). Whether the weights are negative depend on \tilde{A}_{it} , i.e. the residual for the regression where one regresses A_{it} on the fixed unit and time effects. Unit fixed effects tend to be higher for those who were treated earlier in the analysis and time effects tend to be higher in later time periods of the analysis. Hence, units who were treated earlier are more likely to have negative weights at later time periods. This results in potential negative weighting of long-term effects. As a result of the negative weights, it is possible that β is positive even though the β_{it} are negative.

Because of the potential issues with a constant treatment effect model given by the difference-in-differences model, we also estimated the effect of DIH prosecutions reported by the media using a model where treatment effects depend on treatment duration.

3.6 Treatment Effects that Depend on Exposure Duration

Since there may be issues with the constant treatment effect model, we also considered an approach resembling an event study, which allows for treatment effects to vary depending on treatment duration. We estimated the effect of having at least one DIH prosecution reported by the media on drug overdose deaths, assuming that treatment effects may depend on the amount of time since the state was first exposed to a DIH prosecution reported by the media.

Event study analyses are typically used in settings where treatment is staggered (Sun and Abraham, 2021; Borusyak et al., 2021; Roth et al., 2022). Event study models differ from the constant treatment effect model in that instead of a single variable that indicates whether a unit is treated, event study models contain lead and lag indicators which estimate the effects leading up to and following treatment

initiation, respectively. In general, under an event study model, we assume that

$$\mathbb{E}(Y_{it}) = \alpha_i + \gamma_t + \delta X_{it} + \sum_{k=B}^C \beta_k \mathbb{I}\{K_{it} = k\}, \quad (3.17)$$

where α_i and γ_t are unit and time fixed effects, respectively, X_{it} is a vector of other potential confounding variables, and K_{it} is the time relative to the treatment time. Values of $k < 0$ indicate periods before the treatment, and values of $k \geq 0$ indicate periods at or after treatment. Under a *fully specified* model, all leads and lags are included besides one (commonly $k = -1$). Another common specification of the event study model include $B = 0, C > 0$. Under this specification of the event study model, one assumes that there are no pre-treatment effects.

We focus on the latter model specification, where $B = 0$, i.e. we assumed there were no pre-treatment effects. To verify the assumption that there were no pre-treatment effects, we checked if coefficients β_k given by Equation (3.17) for $k < 0$ are zero. Typically, the values of β_k are checked both visually and statistically by hypothesis tests. If there were no pre-treatment effects, i.e. $\beta_k = 0$ for $k < 0$, then we could focus only on the estimation of post-treatment effects. Although in practice, one usually excludes β_{-1} as the reference group, we excluded both β_{-K} and β_{-1} , where $-K$ is the maximum number of leads (Borusyak and Jaravel, 2017). This is because in absence of a control group (in our case, only Hawaii did not have any DIH prosecutions reported by the media by the end of 2019), the leads and lags may be collinear with the state and time fixed effects. By assuming that another lead is zero, e.g. $\beta_{-K} = 0$, we avoid the issue of multicollinearity (Borusyak and Jaravel, 2017).

3.6.1 Model

Variables for the outcome, intervention of interest, and other policy measures were defined the same as in Chapter 3.4. Again, we conducted 1) a primary analysis

where we assumed that treatment effects lasted until at least the end of 2019 and 2) a secondary analysis where we assumed that treatment effects lasted only for two years. Similar to the constant treatment effect model, we estimated the treatment effect using a GAM where we estimated smoothed time effects over the U.S. regions using a cubic regression splines model. Instead of a logistic regression model, our primary focus was a GAM with a linear link function. Under the GAM with a linear link, we transformed the risk of unintentional drug overdose deaths using a log transformation.

We used a model that resembles the event study model to estimate the effect of DIH prosecutions reported by the media on unintentional drug overdose deaths. In addition to the state and time fixed effects, we included an indicator for having at least one DIH prosecution reported by the media in state s at time interval t and a linear effect for the treatment duration. We also included this indicator for the various policy measures, which we now index by l . The GAM model for the log of the risk of unintentional drug overdose deaths was given by:

$$\begin{aligned} \mathbb{E}(\log(Y_{s,t})|s, t, X_{l,s,t}, K_{l,s,t}, A_{s,t}, K_{s,t}) &= \alpha_s + \gamma_{r(s)}(t) + X_{l,s,t}\delta_{0,l} + K_{l,s,t}\delta_{1,l} \\ &+ A_{s,t}\beta_0 + K_{s,t}\beta_1, \end{aligned} \quad (3.18)$$

where $Y_{s,t}$ indicates the risk of unintentional drug overdose deaths in state s at time interval t , α_s indicates state fixed effects, $\gamma_{r(s)}(t)$ indicates the smoothed time effects which may differ by U.S. regions indicated by $r(s)$, $X_{l,s,t}$ and $A_{s,t}$ indicate exposure to the l th (out of L) policy measure and DIH prosecutions reported by the media, respectively, as defined in Section 3.3.2, and $K_{l,s,t}$ and $K_{s,t}$ indicate the treatment duration of the l th policy measure and DIH prosecutions reported by the media, respectively. Since we assumed there were no pre-treatment effects, for time intervals t that occurred before the intervention time or if the intervention occurred in time

interval t , we let the corresponding $K_{l,s,t}$ or $K_{s,t}$ be zero.

3.6.2 Sandwich Estimator for Variance of $\hat{\Theta}$

Similar to the constant treatment effect model, we derived a sandwich estimator for the standard errors. Note now,

$$\vec{Z}_{s,t} = \begin{pmatrix} \mathbb{I}\{\text{state} = s\} \\ \mathbb{I}\{\text{time} = t\} \\ X_{l,s,t} \\ K_{l,s,t} \\ A_{s,t} \\ K_{s,t} \end{pmatrix}. \quad (3.19)$$

As in Section 3.4.2, we assumed the conditional independence of the unintentional drug overdose deaths in state s at time interval t and the state's (and other states') past policy measures, DIH prosecutions reported by the media, and unintentional drug overdose deaths, given $\vec{Z}_{s,t}$. In addition, we assumed the following:

Assumption 3.3. Given the past history of state fixed effects, time effects, other relevant policy measures, and DIH prosecutions reported by the media for state s at time interval t , denoted as $\vec{Z}_{s,t}$ we have

$$\mathbb{E}(\log(Y_{s,t}) \mid \vec{Z}_{s,t} = \vec{z}_{s,t}) = \log(p_{s,t,\Theta^*}), \quad (3.20)$$

where p_{s,t,Θ^*} is the underlying probability of unintentional drug overdose deaths given by the true parameter Θ^* . That is, Assumption 3.3 states that we have a correctly specified model for $\log p_{s,t}$ given the vector of predictors $\vec{Z}_{s,t}$. Under Assumptions 3.1 and 3.3, we derived the Sandwich estimator (see Appendix B.3.2 for details):

$$\hat{\Sigma} = \frac{N}{N-d} C^{-1} \left(\sum_{s,t} \left(\log(Y_{s,t}) - \vec{Z}_{s,t}^T \hat{\Theta} \right)^2 \vec{Z}_{s,t} \vec{Z}_{s,t}^T \right) (C^{-1})^T, \quad (3.21)$$

where now $C = \sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T$ and where we estimated the true parameter Θ^* by $\hat{\Theta}$.

3.6.3 Estimating the Number of Attributable Deaths

We also estimated the number of excess drug overdose deaths attributable to DIH prosecutions reported by the media (see Appendix B.4.2 for details). The probability of an unintentional drug overdose death had intervention not occurred is:

$$\hat{p}_{s,t}^{(a_{s,t}=0)} = \frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}} \exp\left(-A_{s,t}\hat{\beta}_0 - K_{s,t}\hat{\beta}_1\right), \quad (3.22)$$

where $\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}$ is the observed risk of unintentional drug overdose deaths. As in Section 3.4.3, we estimated the number of attributable deaths by:

$$\hat{n}_{s,t,\text{attributable deaths}} = n_{s,t,\text{observed overdose deaths}} - \hat{p}_{s,t}^{(a_{s,t}=0)} \times n_{s,t,\text{population}}.$$

Note that since there are two parameters for the treatment effects, namely β_0 and β_1 , we cannot simply substitute $\hat{\beta}_0$ and $\hat{\beta}_1$ by their 95% confidence interval upper and lower limits because we have to account for any covariances between the parameters. Hence, we used the Delta Method to estimate the variance of the number of attributable deaths in a time interval t . By the theory of unbiased estimating equations:

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow \mathcal{N}(\vec{0}, \Sigma_{\beta}),$$

where $\beta = (\beta_0 \ \beta_1)^T$ and $\hat{\beta}$ estimates β . Furthermore, Σ_{β} indicates the variance-covariance matrix for β . For a time interval t , we let

$$g(\beta) = \sum_s n_{\text{population},s,t} (p_{s,t} - p_{s,t} \exp(-A_{s,t}\beta_0 - K_{s,t}\beta_1)), \quad (3.23)$$

where $g(\beta)$ indicates the number of attributable deaths in time interval t , and we

estimated $p_{s,t}$ using $\hat{p}_{s,t}^{\text{obs}} = \frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}$. Using the Delta Method,

$$\sqrt{n} \left(g(\hat{\beta}) - g(\beta) \right) \rightarrow \mathcal{N} \left(\vec{0}, J_{\beta,t} \Sigma_{\beta} J_{\beta,t}^T \right),$$

where $J_{\beta,t}$ is the Jacobian matrix:

$$\begin{aligned} J_{\beta,t} &= \begin{pmatrix} \frac{\partial g(\beta)}{\partial \beta_0} & \frac{\partial g(\beta)}{\partial \beta_1} \end{pmatrix} \\ &= \begin{pmatrix} \sum_s n_{\text{population},s,t} p_{s,t} \exp(-A_{s,t}\beta_0 - K_{s,t}\beta_1) A_{s,t}, \\ \sum_s n_{\text{population},s,t} p_{s,t} \exp(-A_{s,t}\beta_0 - K_{s,t}\beta_1) K_{s,t} \end{pmatrix}. \end{aligned} \quad (3.24)$$

Hence, the lower and upper bounds of the 95% confidence interval of the attributable deaths at time interval t were given by:

$$\sum_s n_{\text{attributable deaths},s,t} \pm \frac{1.96}{\sqrt{n}} \sqrt{J_{\beta,t} \Sigma_{\beta} J_{\beta,t}^T}. \quad (3.25)$$

3.6.4 Sensitivity Analysis

As described in Chapter 3.5, OLS potentially makes “forbidden comparisons” in which treated units are also used to estimate fixed unit and time effects. Because of the invalid extrapolation of data points to estimate fixed effects, treatment effects may not be separable from the time effects when all (or most) units are treated by the end of the study (Borusyak et al., 2021). This potentially leads to an inaccurate estimation of treatment effects, especially for long-run treatment effects. Borusyak et al. (2021) suggested using the difference between the earliest and the latest treatment times as an upper bound on the number of periods post-treatment. The first state that had at least one DIH prosecution reported by the media was Ohio in March of 2000, and the last state in the treatment group that had at least one DIH prosecution reported by the media was South Dakota in February of 2017. We took extra precaution and

conducted a sensitivity analysis where we removed the last six years; i.e. the analysis period ends at the end of 2013. With the removal of the last six years, the control group consisted of eight states (Alaska, Delaware, Hawaii, Idaho, Mississippi, Rhode Island, South Carolina, and South Dakota).

In addition to a GAM with linear link, we modeled treatment effects using a logistic GAM to compare results. Recall that since the probability of a drug overdose death for a random person was small, odds ratios were essentially equivalent to risk ratios, so both the logistic GAM and GAM with linear link were essentially evaluated on the risk ratio scale. The logistic GAM for unintentional drug overdose deaths was given by

$$\begin{aligned} \text{logit}(\mathbb{E}(Y_{s,t} \mid s, t, X_{l,s,t}, K_{l,s,t}, A_{s,t}, K_{s,t})) &= \alpha_s + \gamma_{r(s)}(t) + X_{l,s,t}\delta_{0,l} + K_{l,s,t}\delta_{1,l} \\ &+ A_{s,t}\beta_0 + K_{s,t}\beta_1, \end{aligned} \quad (3.26)$$

where variables were defined as in Equation (3.18).

3.6.5 Results

We first checked for any pre-treatment trends. Figure 3-6 shows the coefficients and 95% confidence intervals for the periods leading up to and following the first charge date of DIH prosecutions reported by the media. The red dotted vertical line indicates the intervention time of DIH prosecutions reported by the media. Periods right before the treatment time and 34 periods before treatment were excluded. From the coefficients and the 95% confidence intervals, we concluded that there were likely no pre-treatment trends—all coefficients prior to the treatment were not statistically significantly different from zero. Hence, we focused on estimating post-treatment effects. Furthermore, there was a decreasing trend for treatment effects for periods after treatment, suggesting that the DIH prosecutions reported by the media might

actually have a protective effect in the long run.

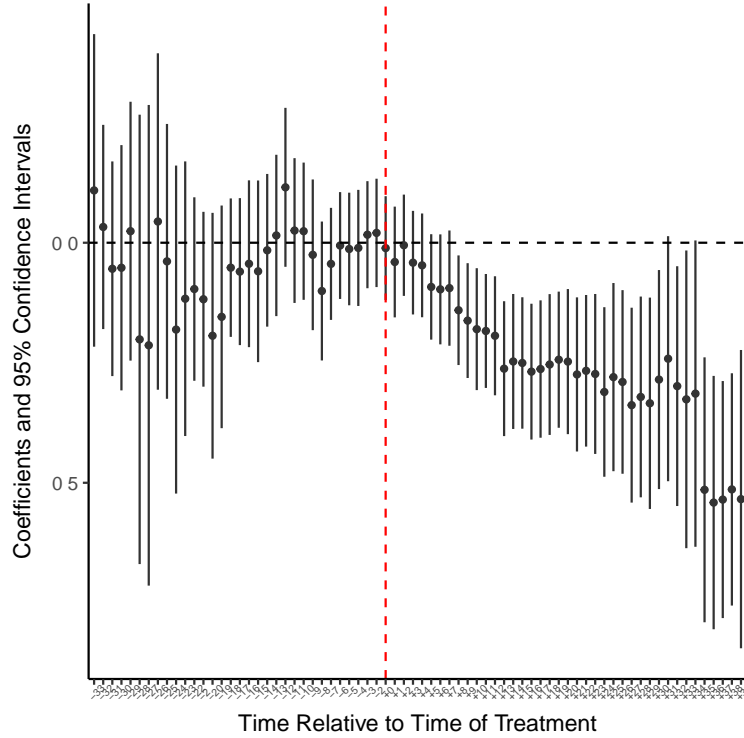


Figure 3-6: Coefficients and 95% confidence intervals for periods leading up to and following the first date of DIH prosecutions reported by the media. The red dotted vertical line indicates the date of the first DIH prosecution reported by the media. The period right before and 34 periods before the treatment time were excluded.

Table 3.1 shows the estimated risk ratios and 95% confidence intervals for effects of the different relevant policy measures and the DIH prosecutions reported by the media. We first considered the case where we assumed that the treatment effect lasted until at least 2019. There was not a statistically significant effect when states were first exposed DIH prosecutions reported by the media (0.9561, 95% CI: (0.8235, 1.1100)). However, DIH prosecutions reported by the media potentially have a protective effect as the exposure duration of DIH prosecutions reported by the media increases (0.9863, 95% CI: (0.9734, 0.9994)). Results did not differ much when we assumed that the effect of DIH prosecutions reported by the media only lasted for

two years. In particular, the risk ratio of the treatment duration of DIH prosecutions reported by the media when assuming the effect only lasts two years was 0.9884 (95% CI: (0.9781, 0.9989)). However, when we assumed that treatment effects only lasts for two years, there was a stronger and potentially harmful immediate effect of exposure to DIH prosecutions reported by the media: 1.0476 (95% CI: (0.9263, 1.1849)).

Table 3.1: Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for GAM with linear link function when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years.

	GAM with linear link with Lasting Effect (95% Confidence Interval)	GAM with linear link with Two Year Effect (95% Confidence Interval)
Exposure to NAL: can dispense	0.9599 (0.8188, 1.1253)	0.9459 (0.8046, 1.1119)
Linear effect of NAL: can dispense	0.9734 (0.9475, 0.9999)	0.9718 (0.9450, 0.9994)
Exposure to NAL: cannot dispense	1.1384 (0.9098, 1.4245)	1.1298 (0.8806, 1.4495)
Linear effect of NAL: cannot dispense	0.9775 (0.9590, 0.9963)	0.9781 (0.9583, 0.9983)
Exposure to MML	1.2120 (1.0112, 1.4526)	1.2203 (0.9898, 1.5046)
Linear effect of MML	0.9915 (0.9784, 1.0049)	0.9917 (0.9768, 1.0067)
Exposure to RML	0.9841 (0.7995, 1.2114)	0.9657 (0.7583, 1.2297)
Linear effect of RML	0.9916 (0.9538, 1.0309)	0.9998 (0.9687, 1.0319)
Exposure to 911 GSL	1.0741 (0.9397, 1.2278)	1.0885 (0.9486, 1.2491)
Linear effect of 911 GSL	1.0106 (0.9898, 1.0318)	1.0103 (0.9903, 1.0307)
Exposure to PDMP	0.8790 (0.7515, 1.0282)	0.8845 (0.7454, 1.0495)
Linear effect of PDMP	1.0063 (0.9894, 1.0234)	1.0073 (0.9881, 1.0269)
Exposure to Medicaid expansion	1.0800 (0.9273, 1.2578)	1.0828 (0.9324, 1.2575)
Linear effect of Medicaid expansion	1.0130 (0.9883, 1.0383)	1.0155 (0.9928, 1.0386)
Exposure to DIH prosecutions	0.9561 (0.8235, 1.1100)	1.0476 (0.9263, 1.1849)
Linear effect of DIH prosecutions	0.9863 (0.9734, 0.9994)	0.9884 (0.9781, 0.9989)

Figure 3-7 shows the estimated risk ratios and 95% confidence intervals of having any DIH prosecutions reported by the media versus not having any, at each time interval after being exposed to DIH prosecutions reported by the media. Figure 3-7 shows there was initially an increase in unintentional drug overdose deaths when we assumed that treatment effects only lasted for two years. However, when we assumed that treatment effects lasts until at least 2019, there was a decrease in unintentional drug overdose deaths when the state was first exposed to DIH prosecutions reported by the media. In both cases, though, there was a protective effect of DIH prosecutions

reported by the media on unintentional drug overdose deaths as the exposure duration increased. However, the treatment effects at each time interval post-treatment were not statistically significant at level $\alpha = 0.05$.

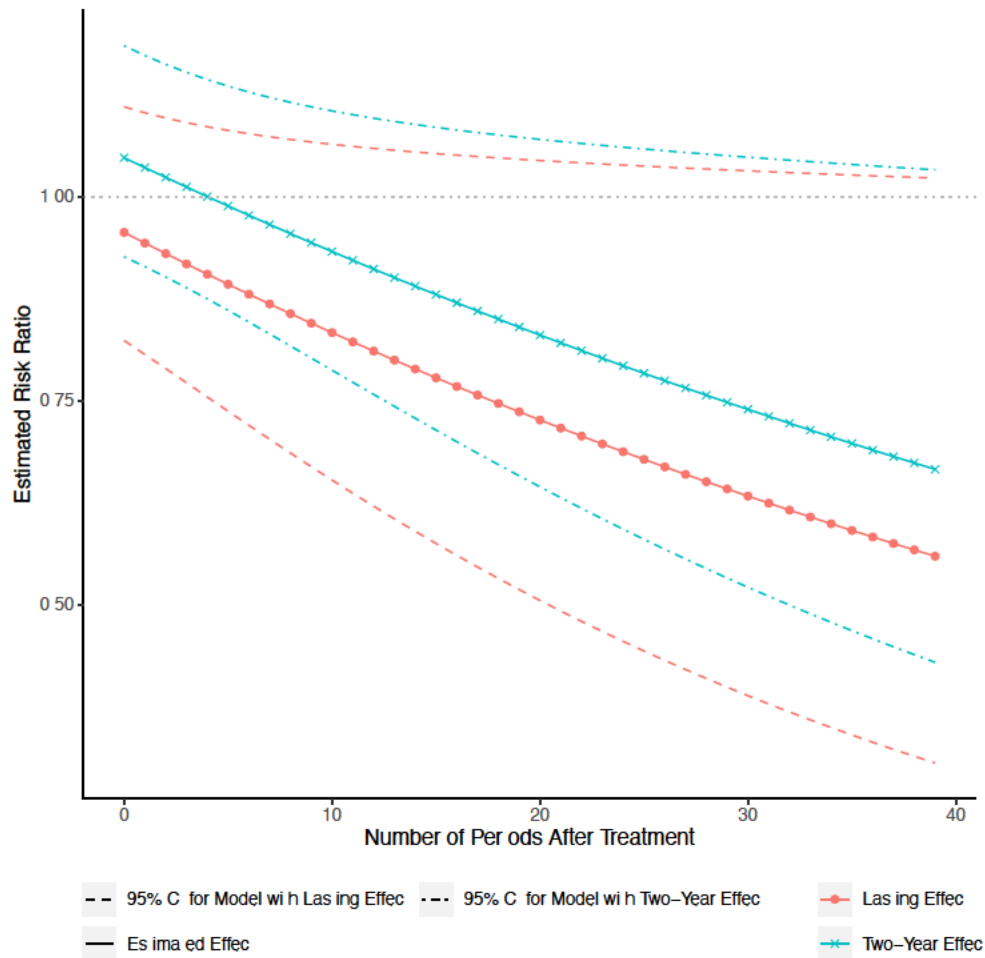


Figure 3-7: Estimated risk ratios and 95% confidence intervals for main analysis when assuming treatment effect lasted until at least 2019 and when assuming treatment effect lasted only for two years (indicated by color and shape).

Figure 3-8 shows the estimated yearly number of drug overdose deaths attributable to DIH prosecutions reported by the media. Since we found protective effects of the DIH prosecutions reported by the media, the “deaths attributable” equate to “lives saved”. From 2000 to 2019, we estimated a total of 197,298 (95% CI: (-73,712,

468,307)) lives saved due to DIH prosecutions reported by the media when we assumed that treatment effects lasted until at least 2019. This makes up approximately 29.97% (95% CI: (-11.20%, 71.15%)) of the total unintentional drug overdose deaths in the U.S. from 2000 to 2019. When we assumed that treatment effects lasted only for two years, we estimated that there were 67,364 (95% CI: (-55,714, 190,441)) lives saved due to DIH prosecutions reported by the media, which makes up approximately 10.23% (95% CI: (-8.46%, 28.93%)) of the total unintentional drug overdose deaths in the U.S. from 2000 to 2019. Note that a negative number of “lives saved” suggests that there were also deaths that were attributable to DIH prosecutions reported by the media. Hence, we cannot definitively say that DIH prosecutions reported by the media reduced the number of unintentional drug overdose deaths.

Tables with the estimated risk ratios and 95% confidence intervals for the relevant policy measures and DIH prosecutions reported by the media when we used a logistic GAM and when we excluded the last six years are included in Appendix B.6. The risk ratio estimates from the logistic GAM were similar to the risk ratio estimates from the GAM with a linear link function. However, all risk ratio estimates were not statistically significant at level $\alpha = 0.05$. When we excluded the last six years of the analysis, the risk ratios for the exposure to DIH prosecutions reported by the media and the linear effect of DIH prosecutions reported by the media using a GAM with linear link were similar to the corresponding effects in the main analysis. The risk ratios for the exposure to DIH prosecutions reported by the media and linear effect of DIH prosecutions reported by the media using a logistic GAM when excluding the last six years of the analysis were slightly different from the risk ratios from a logistic GAM without excluding the last six years. In general, results suggest a small harmful immediate effect of exposure to DIH prosecutions reported by the media, but a potential longer-term protective effect of DIH prosecutions reported by the media

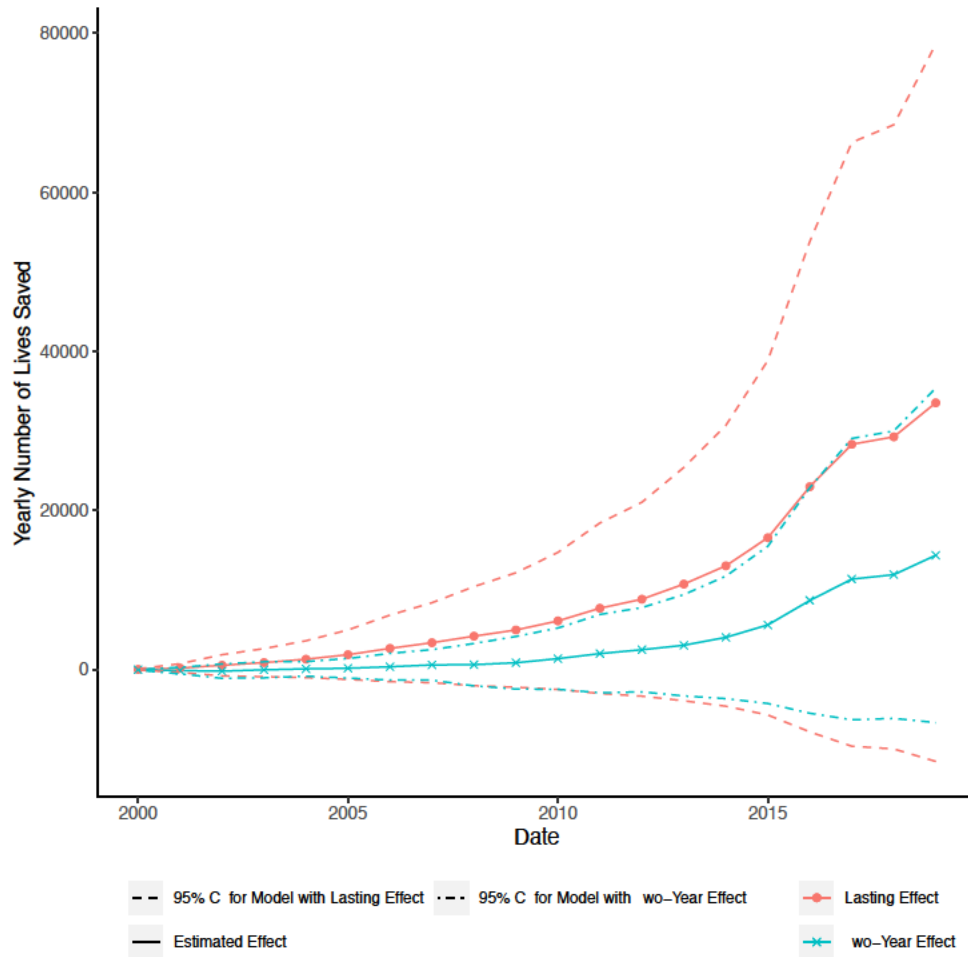


Figure 3·8: Estimated yearly number of potential lives saved in the 50 U.S. States from 2000 to 2019 due to DIH prosecutions reported by the media 1) when assuming treatment effect lasted until at least 2019 and 2) when assuming treatment effect lasted only for two years (indicated by color and shape).

on unintentional drug overdose deaths.

3.7 Discussion

We found that under the different model specifications, there were different conclusions for the effect of having any DIH prosecutions reported by the media versus having none on the unintentional drug overdose deaths. Under the assumption of constant treatment effects, results suggest that DIH prosecutions reported by the

media may have harmful effects on unintentional drug overdose deaths. However, under the assumption that treatment effects may depend on treatment duration, results suggest that DIH prosecutions reported by the media may have protective effects on unintentional drug overdose deaths.

The conflicting results support the recent literature which have found potential issues when applying difference-in-differences methods with a constant treatment effect to settings where units are treated at different times and when treatment effects are heterogeneous (Borusyak and Jaravel, 2017; Sun and Abraham, 2021; Roth et al., 2022). Our results suggest that the estimated treatment effect in a model assuming constant treatment effect is a weighted average of the treatment effects in a model assuming that the treatment effect depends on treatment duration where the weights are negative. A possible explanation for the negative weights in our case is that negative weights tend to occur later in the analysis for units who were treated early in the analysis period (Borusyak and Jaravel, 2017; Roth et al., 2022). Figure 3-3 shows that by 2010, approximately half of the states had at least one DIH prosecution reported by the media, which suggests that approximately half of states had been treated at least nine years by the end of 2019. Since many states had at least one DIH prosecution “early” in the analysis period, the conflicting results from the two treatment effects models are likely due to a negative weighting of treatment effects from the model where treatment effects depended on treatment duration. We saw similar results even in the analyses in which we excluded the last six years.

Our results shed light on potential issues in applying difference-in-differences to settings with staggered adoption of treatment when treatment effects are heterogeneous. To our knowledge, previous work done in analyzing effects of interventions on drug overdose deaths conducted both difference-in-differences and event study analyses, but none have encountered conflicting results such as the ones we found

(Pacula et al., 2015; Powell et al., 2018). It was not until upon further investigation that we found recent work (dating within the past few years) that highlighted these potential issues with models assuming constant treatment effects such as difference-in-differences models. The difference-in-differences model has been a very common method in estimating treatment effects. However, our work and the recent literature suggest that there are settings where traditional difference-in-differences methods are not appropriate. We hope that our work serves as an example of what could happen when one does not carefully consider all aspects of the constant treatment effect difference-in-differences model before applying the model to estimate treatment effects.

Although we found statistically significant results for the linear effect of the treatment duration of DIH prosecutions reported by the media when using a GAM with a linear link function, we cannot definitively conclude that DIH prosecutions reported by the media have a protective effect. This is because the risk ratios for the treatment duration were 0.9863 with 95% confidence interval equal to (0.9734, 0.9994) and 0.9884 with 95% confidence interval equal to (0.9781, 0.9989) when assuming treatment effects lasts until at least the end of 2019 and when assuming treatment effects lasts for two years, respectively. Since the upper bounds of the 95% confidence intervals are very close to one, these risk ratios are barely statistically significant. Furthermore, even though the effect of the treatment duration of DIH prosecutions reported by the media are statistically significant at level $\alpha = 0.05$, the treatment effects at each time interval post-treatment (shown in Figure 3-7) are not statistically significant. Figures 3-7 and 3-8 further show that the confidence intervals are quite wide, suggesting that results are not practically significant. Figure 3-8 also shows that negative “lives saved” were included in the 95% confidence intervals, suggesting that DIH prosecutions reported by the media could also potentially lead to drug overdose

deaths.

In addition to making conclusions based on the results from our analysis, we may also want to consider other potential side effects of DIH prosecutions. First, DIH prosecutions, and their coverage in the mass media, may actually discourage people from seeking emergency medical help. In two previous separate studies, participants stated that police involvement was the main reason people did not make a 911 call or had a delay in doing so (Baca and Grant, 2007; Pollini et al., 2006). In addition, awareness of 911 Good Samaritan laws is broadly lacking, and criminal justice measures such as DIH prosecutions may hurt efforts to increase public understanding and the lifesaving spirit of 911 Good Samaritan Laws (Schneider et al., 2020; Carroll et al., 2021). Furthermore, there are additional concerns about the impact of criminal justice contact on the health risk of drug users. A recent systematic literature review of global research found robust associations between police contact and HIV infection and risk behaviors (Baker et al., 2019). Baker et al. (2019) also highlighted the structural impact of law enforcement as a barrier to protective measures, including impact on Opioid Agonist Therapy (OAT) nonattendance, Site Engagement Program (SEP) avoidance, and healthcare avoidance. Lastly, there are also ethical concerns about law enforcement activities producing and reinforcing disparities by race and vulnerability. The perceived risks of legal repercussions for seeking help during overdose events are likely affected by existing racial disparities in contact with the criminal legal system (Beletsky et al., 2011). A recent study explored the relationship between structural vulnerability and police abuse and harassment among people who inject drugs, and they found that a certain subgroup: men who were experiencing homelessness, were from rural areas, had traded sex, and dropped out of high school were most likely to experience police abuse and harassment (Friedman et al., 2020b).

There were limitations to our analysis. First, it is possible that there was a mis-

classification in the exposure to DIH prosecutions reported by the media. Specifically, it is possible that there was an underreporting of the number of states being exposed to DIH prosecutions reported by the media at any given time, since the exposure indicator depended on the media reports that we collected. If we missed a media report, then the state would be considered as not being exposed to DIH prosecutions reported by the media. Previous research showed that in scenarios of underreporting (here, it is likely that we have differential misclassification), the estimate of the treatment effect is biased towards the null (Ferrão and Goldstein, 2014; Kesmodel, 2018; Moradzadeh et al., 2018). Second, we only had information about the charge date of the DIH prosecution and not about the date of the media report, which may lead to misspecification of the intervention date. Third, we did not account for interference effects from neighboring states, but the unintentional drug overdose deaths in a state may depend on whether a neighboring state is prosecuting people for drug-induced homicide. Lastly, we only considered the exposure to DIH prosecutions reported by the media and not the number of DIH prosecutions reported by the media in each time interval. In Chapter 4, we explore another model where we can take the number of DIH prosecutions reported by the media into account when estimating the effect of DIH prosecutions reported by the media.

Given the results from our analyses and the potential side effects of implementing DIH prosecutions, we cannot make concrete conclusions or suggestions regarding DIH prosecutions. That is, we cannot conclude whether DIH prosecutions have a harmful or protective effect. However, our findings suggest that the effect of DIH prosecutions on drug overdose deaths need to be at least studied further.

Chapter 4

Structural Nested Risk Ratio Models

4.1 Introduction

In the previous chapter, we estimated the effect of DIH prosecutions reported by the media on unintentional drug overdose deaths. In particular, we estimated the effect of exposure to DIH prosecutions reported by the media, i.e. a state is considered treated at time t if there were any DIH prosecution reported by the media by time t . Hence, we did not take the number of DIH prosecutions into account. Figure 4-1 shows the number of DIH prosecutions reported by the media in each state per six-month interval from 2000 to 2019. While most states continued to have DIH prosecutions reported by the media after the first media report, the number of DIH prosecutions reported by the media in each state varied greatly—of those with at least one DIH prosecution reported by the media, Pennsylvania had a total of 652 DIH prosecutions reported by the media between 2000 and 2019, while Mississippi and Nebraska only had one DIH prosecution reported by the media between 2000 and 2019.

In this chapter, we take the number of DIH prosecutions reported by the media into account and estimate the effect of an increase in DIH prosecutions. Recall that we have a repeated outcomes setting where the data collected to estimate the effect of DIH prosecutions on unintentional drug overdose deaths, as described in Chapter 3.3, consisted of data for each state, for each six-month time interval from 2000 to 2019. We focus on a different model in this chapter: Structural Nested Models (SNMs) (Robins, 1992, 1994; Lok and DeGruttola, 2012). SNMs have been used to estimate

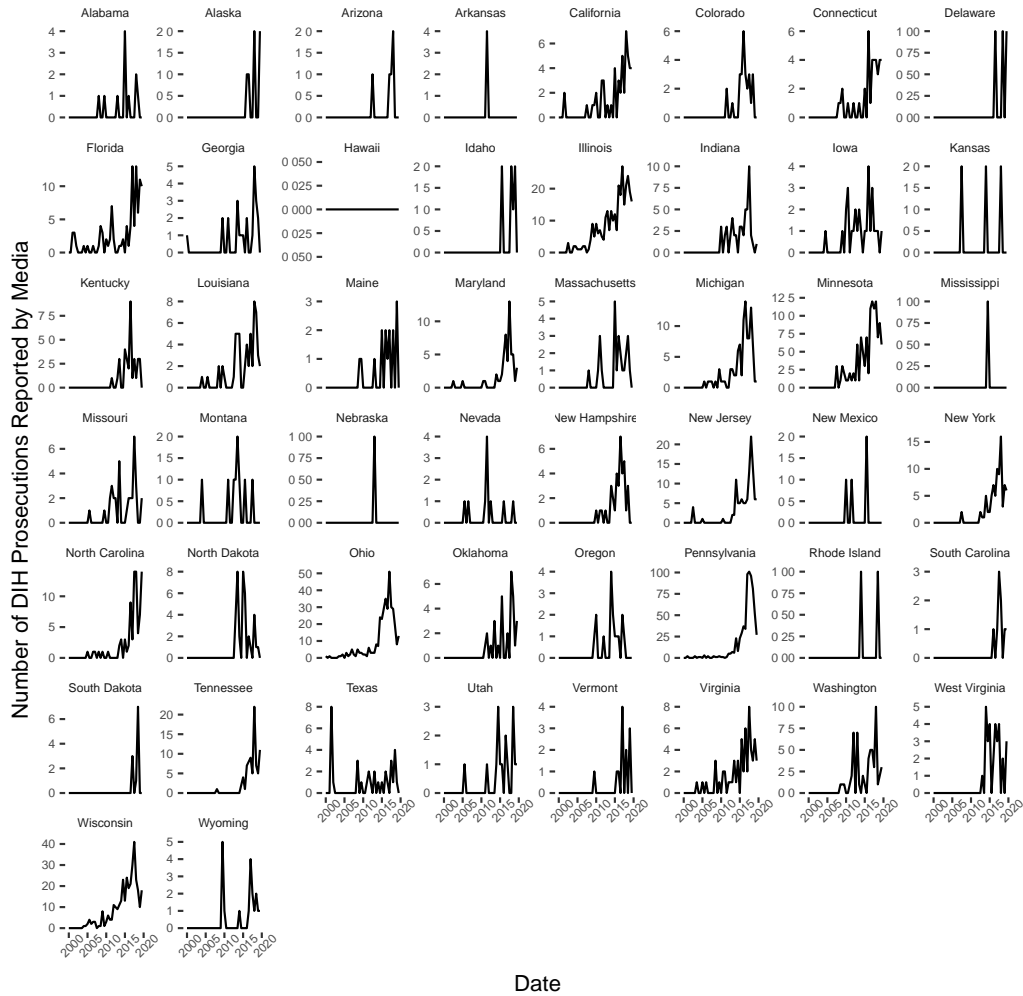


Figure 4-1: Number of drug-induced homicide prosecutions reported by media in each state per six-month interval from 2000 to 2019.

treatment effects in settings with repeated outcomes. We propose a new class of structural nested models in which we estimate the treatment effect on the risk ratio scale by modeling a function of the treatment.

4.2 Background

There have been several models that have been proposed to estimate treatment effects in settings with repeated outcomes. Some examples include the Structural

Nested Models (SNMs) (Robins, 1994, 1998) and Marginal Structural Models (MSMs) (Robins et al., 2000; Robins, 1999). MSMs are used to estimate treatment effects that depend on baseline covariates, whereas SNMs use a g-estimation method to estimate treatment effects depending on time-dependent covariates—covariates which could change depending on the prior outcomes and prior treatments. Furthermore, these time-dependent covariates may affect the future outcomes and treatments. We focus on Structural Nested Models. There are two main classes of SNMs: 1) Structural Nested Mean Models (SNMMs) (Robins et al., 1992; Robins, 1994, 1998; Lok and DeGruttola, 2012) where one estimates treatment effects on the mean of the outcome and 2) Structural Nested Distribution Models (SNDMs) (Vansteelandt and Joffe, 2014) where one estimates treatment effects on the distribution of the outcome. We focus on SNMMs.

In general, SNMMs are used to estimate treatment effects under two different treatment regimes. For example, early SNMMs were used to estimate the mean difference between outcomes as a result of stopping treatment at time t versus stopping treatment at time $t - 1$, given covariates at time $t - 1$ (Robins et al., 1992). Since the proposal of SNMMs, there have been newer classes of SNMMs that have been proposed. The coarse SNMM (Robins, 1994; Lok and DeGruttola, 2012; Yang and Lok, 2018) is a subclass of SNMMs where treatment occurs at time m and outcomes are measured at a later time k . Coarse SNMMs are used to estimate the effect of treatment initiation, conditional on a unit's prior covariates up to time m . For example, coarse SNMMs have been used to study the effect of the initiation time of antiretroviral treatment in HIV infected patients (Lok and DeGruttola, 2012). We extend SNMMs in this chapter to a model where, instead of comparing outcomes on the risk difference scale, we compare outcomes on the risk ratio scale. In addition, our work differs from previous work on SNMMs in that we estimate the treatment effect

by modeling a function of the treatment. Hence, we are not restricted to estimating the effect of treatment initiation, which was the focus on many previous work. In particular, we choose a function that can be modeled using a generalized linear model. We call this model the Structural Nested Risk Ratio Model.

4.3 Structural Nested Risk Ratio Models

4.3.1 Notation

To formalize Structural Nested Risk Ratio Models (SNRRMs), we first introduce notation in the context of DIH prosecution media reports. Let $Y_{s,t}$ be the observed risk of drug overdose deaths in state s at time interval t for $t \in \{0, \dots, K\}$. In our analysis, $t = 0$ corresponds to the first six-month period, i.e. January to June of 2000, and $t = K$ corresponds to the last six-month period, i.e. July to December of 2019. Let $\tilde{A}_{s,t}$ be the number of DIH prosecutions reported by the media in state s at time interval t , and let $g(\tilde{A}_{s,t})$ be a function of the number of DIH prosecutions reported by the media in state s at time interval t . We choose g as a function that can be modeled using a generalized linear model (GLM). For example, $g(\tilde{A}_{s,t}) = \tilde{A}_{s,t}$ is the identity function, where $\tilde{A}_{s,t}$ can be modeled using a Poisson GLM, and $g(\tilde{A}_{s,t}) = \mathbb{I}\{\tilde{A}_{s,t} \geq 0\}$ indicates whether there were any DIH prosecutions reported by the media in state s at time interval t , which can be modeled using a logistic GLM. Let $X_{s,t}$ represent the other relevant policy measures in state s at time interval t . These other policy measures are: naloxone access laws where pharmacists can and cannot dispense naloxone without a prescription, the legalization of medical and recreational marijuana, prescription drug monitoring programs, the 911 Good Samaritan law, and the expansion of Medicaid.

We denote the history of a variable using the bar notation. Let $\bar{Y}_{s,t}$ be the history of the risk of drug overdose deaths in state s up to time interval t , i.e. $\bar{Y}_{s,t} =$

$(Y_{s,0}, \dots, Y_{s,t})$. Let $\bar{A}_{s,t} = (\bar{A}_{s,0}, \dots, \bar{A}_{s,t})$ and $g(\bar{A}_{s,t}) = (g(\bar{A}_{s,0}), \dots, g(\bar{A}_{s,t}))$ be the history of the number of DIH prosecutions reported by the media in a state s up to time interval t and the history of a function of number of DIH prosecutions reported by the media in a state s up to time interval t , respectively. Let $\bar{X}_{s,t} = (X_{s,0}, \dots, X_{s,t})$ be the history of the other relevant policy measures in a state s up to time interval t . Given a treatment regime $\bar{A}_{s,K} = \bar{a}_K$ up to time interval K for a state s , we denote the policy measures that naturally progress under $g(\bar{a}_K)$ as $X_{s,t}^{(g(\bar{a}_K))}$. Then, for a given treatment regime \bar{a}_K and $X_{s,t}^{(g(\bar{a}_K))}$, we denote the potential outcome in state s at time t as $Y_{s,t}^{(g(\bar{a}_K))}$. We use the bar notation again to denote the history of the policy measures and the potential outcomes in state s up to time interval t under $g(\bar{a}_K)$ as $\bar{X}_{s,t}^{(g(\bar{a}_K))}$ and $\bar{Y}_{s,t}^{(g(\bar{a}_K))}$, respectively.

4.3.2 Model and Assumptions

Contrary to previous work (Robins, 1994; Lok and DeGruttola, 2012) in which risk differences were studied, we will work in the risk ratio scale.

Definition 1. For $k = 1, \dots, K$, for $t < k$, we define the treatment effect of having treatment regime $(g(\bar{a}_{t-1}), g(\bar{a}_t), \dots, g(\bar{a}_t))$ versus $(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))$ as:

$$\gamma_k^t(\bar{x}_{s,t}, \bar{a}_t) = \frac{\mathbb{E}(Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_t), \dots, g(\bar{a}_t))} \mid \bar{X}_{s,t} = \bar{x}_{s,t}, \bar{A}_{s,t} = \bar{a}_t)}{\mathbb{E}(Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))} \mid \bar{X}_{s,t} = \bar{x}_{s,t}, \bar{A}_{s,t} = \bar{a}_t)}. \quad (4.1)$$

The treatment effect γ_k^t from Equation (4.1) is the ratio of the expected risks of drug overdose deaths observed at time k , under different fixed treatment regimes, $(g(\bar{a}_{t-1}), g(\bar{a}_t), \dots, g(\bar{a}_t))$ versus $(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))$, while the other policy measures naturally progress. Hence, we are interested in the treatment effect when DIH prosecutions reported by the media at times $t, t+1, \dots, K-1$ are $g(\bar{a}_t)$ versus $g(\bar{a}_{t-1})$, while the other policy measures naturally progress. If $t \geq k$, there is no difference between $Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_t), \dots, g(\bar{a}_t))}$ and $Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))}$, and so when $t \geq k$,

$$\gamma_k^t(\bar{X}_{s,t}, \bar{A}_{s,t}) = 1.$$

Assumption 4.1. For $k = 1, \dots, K$, for $t < k$, $\gamma_{k,\psi^*}^t(\bar{X}_{s,t}, \bar{A}_{s,t})$ is a correctly specified model for γ_k^t , where $\psi^* \in \mathbb{R}^d$ (for $d \geq 1$) indicates the true parameter.

Similar to previous work on SNMs (Robins et al., 1992; Robins, 1994; Lok and DeGruttola, 2012), we require the Assumption of No Unmeasured Confounding for time-dependent models:

Assumption 4.2 (No Unmeasured Confounding). Let $\tilde{A}_{s,t}$ be the number of DIH prosecutions reported by the media in state s at time interval t . Let $\bar{A}_{s,t-1} = \bar{a}_{t-1}$ and $\bar{X}_{s,t}$ be the history of DIH prosecutions for state s up to time interval $t - 1$ and the history of other policy measures in state s up to time interval t , respectively. Given $\bar{A}_{s,t-1}$ and $\bar{X}_{s,t}$,

$$\tilde{A}_{s,t} \perp\!\!\!\perp Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))} \mid \bar{A}_{s,t-1} = \bar{a}_{t-1}, \bar{X}_{s,t}. \quad (4.2)$$

Assumption 4.2 states that conditional on the treatment history $\bar{A}_{s,t-1} = \bar{a}_{t-1}$ and the history of other policy measures $\bar{X}_{s,t}$, the potential outcomes $Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))}$ do not depend on and do not predict the treatment at time t . Hence, under No Unmeasured Confounding, we assume that we have all the information on factors that potentially affects both the treatment and the potential outcomes. In addition to No Unmeasured Confounding, we also require the consistency assumption for time-dependent models:

Assumption 4.3 (Consistency). For time intervals $T > k - 1$, given $\bar{A}_{s,T} = \bar{a}_T$, $Y_{s,k}^{(g(\bar{a}_{k-1}), g(\bar{a}_k), \dots, g(\bar{a}_T))} = Y_{s,k}^{(g(\bar{a}_{k-1}))} = Y_{s,k}$. Additionally, for the history of other policy measures, $\bar{X}_{s,k}^{(g(\bar{a}_{k-1}), g(\bar{a}_k), \dots, g(\bar{a}_T))} = \bar{X}_{s,k}^{(g(\bar{a}_{k-1}))} = \bar{X}_{s,k}$.

Assumption 4.3 states that potential outcomes and history of other policy measures at time k do not depend on treatments at time T where $T > k - 1$. Furthermore, if $\bar{A}_{s,k-1} = \bar{a}_{k-1}$, then the potential outcome and history of other policy measures under $g(\bar{a}_{k-1})$ are given by the observed potential outcomes and histories of other policy measures at time interval k .

4.3.3 Estimation, Consistency, and Asymptotic Normality

To estimate the treatment effect of DIH prosecutions reported by the media on unintentional drug overdose deaths, we first estimate potential outcomes under treatment regime $(g(\bar{a}_{t-1}), g(\tilde{a}_{t-1}), \dots, g(\tilde{a}_{t-1}))$. We use a function that “blips off” the treatment (Robins, 1994; Lok et al., 2004) to mimic potential outcomes under treatment regime $(g(\bar{a}_{t-1}), g(\tilde{a}_{t-1}), \dots, g(\tilde{a}_{t-1}))$.

Definition 2. For $k = 1, \dots, K$, for $t \geq k$, we define $H_{s,k}(t) = Y_{s,k}$. We define recursively for $t = k - 1, \dots, 0$:

$$H_{s,k}(t) = \frac{H_{s,k}(t+1)}{\gamma_k^t(\bar{X}_{s,t}, \bar{A}_{s,t})}, \quad (4.3)$$

where γ_k^t is given by Equation (4.1).

Note that for $t \geq k$, we defined $H_{s,k}(t) = Y_{s,k}$ since we assumed that the outcome at time interval k is not affected by a treatment that occurs after time interval $k - 1$. $H_{s,k}(t)$ mimics the potential outcome at time interval k had the intervention at time interval t remained \tilde{a}_{t-1} for time intervals $t, t + 1, \dots, K$:

Theorem 4 (Mimicking Potential Outcomes). Under Consistency Assumption 4.3, for $k = 1, \dots, K$ and $t < k$, $H_{s,k}(t)$ mimics the potential outcome at time k had the treatment at time t been blipped off. That is,

$$\mathbb{E}(H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t) = \mathbb{E}(Y_{s,k}^{(g(\bar{a}_{t-1}), g(\tilde{a}_{t-1}), \dots, g(\tilde{a}_{t-1}))} \mid \bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t). \quad (4.4)$$

We prove Theorem 4 using backwards induction (see Appendix C.1). Let $k \in \{1, \dots, K\}$. Given $\bar{A}_{s,k-1} = \bar{a}_{k-1}$, $H_{s,k}(k) = Y_{s,k} = Y_{s,k}^{(g(\bar{a}_{k-1}), g(\bar{a}_{k-1}))}$ by definition of $H_{s,k}(k)$ and Consistency Assumption 4.3. Starting at time $t = k - 1$, by the definition of $H_{s,k}(k - 1)$ given by Equation (4.3) and the Law of Iterated Expectations, we divide $\mathbb{E}(Y_{s,k}^{(g(\bar{a}_{s,k-1}), g(\bar{a}_{k-1}))} \mid \bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1})$ by $\gamma_k^{k-1}(\bar{X}_{s,k-1}, \bar{A}_{s,k-1})$. Using the definition of $\gamma_k^{k-1}(\bar{X}_{s,k-1}, \bar{A}_{s,k-1})$, the quotient is equal to $\mathbb{E}(Y_{s,k}^{(g(\bar{a}_{k-2}), g(\bar{a}_{k-2}), g(\bar{a}_{k-2}))} \mid$

$\bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1}$), and so Equation (4.4) holds. Assuming that Equation (4.4) holds for $t = k - 2, k - 3, \dots, m + 1$, we show that Equation (4.4) holds for $t = m$ (in a similar way as in the case of $t = k - 1$) by the Law of Iterated Expectations and the definition of the treatment effect γ_k^m in Equation (4.1). Hence, $H_{s,k}(t)$ mimics the potential outcome under treatment regime $(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))$ and can be used to estimate treatment effects.

Using No Unmeasured Confounding Assumption 4.2, we estimate the true parameter ψ^* for the treatment effect model $\gamma_{k,\psi}^t$ by adding $\alpha(\psi)H_{s,k}(t)$ to the model to predict a function of the treatment, $g(\bar{A}_{s,t})$, given the past treatment history and history of the other policy measures (Robins, 1992). We estimate ψ^* by the ψ for which the coefficient $\hat{\alpha}(\psi) = 0$ (Robins, 1992), and we can do this because of the following theorem:

Theorem 5 (Unbiased Estimating Equations). Assume that No Unmeasured Confounding Assumption 4.2 and Consistency Assumption 4.3 hold. Let $\lambda_{s,t} = \mathbb{E}(g(\bar{A}_{s,t}) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1})$. For any measurable and bounded function $\bar{q}_k^t : (\bar{X}_{s,t}, \bar{A}_{s,t-1}) \rightarrow \mathbb{R}^d$:

$$\mathbb{E} \left(\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) H_{s,k}(t) (g(\bar{A}_{s,t}) - \lambda_{s,t}) \right) = \vec{0}. \quad (4.5)$$

If λ_{s,t,θ^*} and γ_{k,ψ^*}^t are correctly specified models for $\lambda_{s,t}$ and γ_k^t , respectively, then

$$P_n \left(\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) H_{s,k}(t) (g(\bar{A}_{s,t}) - \lambda_{s,t}) \right) = \vec{0}, \quad (4.6)$$

are the unbiased estimating equations for ψ^* and θ^* when stacked with other unbiased estimating equations for θ , where P_n is an empirical measure where $P_n X = \frac{1}{n} \sum_{i=1}^n X_i$. We prove Theorem 5 using the Law of Iterated Expectations by first conditioning on $\bar{X}_{s,t}$ and $\bar{A}_{s,t}$ (see Appendix C.2). By No Unmeasured Confounding Assumption 4.2, $\mathbb{E} \left(H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t} \right) = \mathbb{E} \left(H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right)$. We then condition on $\bar{X}_{s,t}$ and

$\bar{A}_{s,t-1}$ by the Law of Iterated Expectations, which gives:

$$\mathbb{E} \left(\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) \mathbb{E} \left(H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right) \mathbb{E} \left(g(\tilde{A}_{s,t}) - \lambda_{s,t} \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right) \right),$$

where if λ_{s,t,θ^*} is a correctly specified model for $\lambda_{s,t}$, then Equation (4.5) is equal to zero. Hence, by finding the ψ such that $\hat{\alpha}(\psi) = 0$, we solve the unbiased estimating equations. By the theory of unbiased estimating equations, under regularity conditions, the estimator $\hat{\psi}$ is consistent and asymptotically normal (Van der Vaart, 1998, Theorems 5.9 and 5.21).

4.4 Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths: Binary Treatment

We applied SNRRMs to estimate the effect of DIH prosecutions reported by the media on unintentional drug overdose deaths. For simplicity, we chose $\bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) = 1$. To compare the SNRRM to the models described in Chapter 3, we first considered the effect of having any DIH prosecutions reported by the media versus none on unintentional drug overdose deaths. Let $g(\tilde{A}_{s,t}) = \mathbb{I}\{\tilde{A}_{s,t} > 0\} = A_{s,t}$, and denote the intervention time for state s by T_s^* . Denote the history of treatment indicators for state s up to time interval t as $\bar{A}_{s,t}$. For time interval t in which state s first had at least one DIH prosecution reported in the media, the treatment indicator $A_{s,t} = 1$. For time intervals $t' > t$, the treatment indicators $A_{s,t'}$ remain as 1. We considered the following treatment initiation model:

$$\text{logit} \left(\mathbb{P}(A_{s,t} = 1 \mid \bar{A}_{s,t-1} = \bar{0}, \bar{X}_{s,t}) \right) = \beta_0 + \beta_1 \sum_{l=1}^L X_{l,s,t-1} + \beta_2 t + \beta_3 Y_{s,t-1}, \quad (4.7)$$

where $\sum_{l=1}^L X_{l,s,t}$ is the number of policy measures, out of L policy measures, that have been enacted by time interval t , and we accounted for linear time effects t .

Furthermore, we took the state's risk of unintentional drug overdose deaths at time interval $t - 1$, denoted as $Y_{s,t-1}$, into account. Given a model for the treatment effect $\gamma_{k,\psi}^t$, we wanted to find the true parameter ψ^* that solves the unbiased estimating equation given by Equation (4.5). We considered two models for the treatment effect.

4.4.1 One-parameter Treatment Effect Model

We first considered a one-parameter model for the treatment effect:

$$\gamma_k^t(\bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t) = \exp(\psi(a_t - a_{t-1})). \quad (4.8)$$

Since the treatment variables are binary indicators, $a_t - a_{t-1} \geq 0$. If $\psi > 0$, there is a harmful effect of DIH prosecutions reported by the media on drug overdose deaths. If $\psi < 0$, there is a protective effect of DIH prosecutions reported by the media on drug overdose deaths. When $a_t = a_{t-1}$, the treatment effect is 1. Under this one-parameter treatment effect model, the function $H_{s,k}(t)$ for $k = 1, \dots, K$ and $t < k$ was defined as follows:

$$H_{s,k}(t) = \begin{cases} Y_{s,k} & \text{if } k \leq T_s^* \\ \frac{Y_{s,k}}{\exp(\psi)} & \text{if } k > T_s^* \end{cases}. \quad (4.9)$$

Since the one-parameter model for the treatment effect γ_k^t is linear in ψ , we solved the unbiased estimating equations in Equation (4.5) for a closed form solution of ψ (see Appendix C.3):

$$\hat{\psi} = -\log \left(-\frac{\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k \leq T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t})}{\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k > T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t})} \right), \quad (4.10)$$

where we estimated $p_{s,t}$ using a weighted logistic regression model given by Equation (4.7). Here, the weights were deterministic and were equal to the number of time intervals from time interval t until time interval K . Furthermore, Equation (4.10)

includes only datapoints for states s at time intervals t in which $\bar{A}_{s,t-1} = \bar{0}$, i.e. data points that occurred at or before treatment time for state s . The numerator of the fraction within the log of Equation (4.10) contains terms corresponding to states s at time intervals k that occurred at or before treatment time T_s^* , i.e. untreated outcomes. Similarly, the denominator contains terms corresponding to states s but at time intervals k that occurred after treatment time T_s^* , i.e. treated outcomes. We estimated the 95% confidence intervals using a grid-search: for a grid of potential ψ , we computed $H_{s,k}(t)$ using $\gamma_{k,\psi}^t$ and added $\alpha(\psi)H_{s,k}(t)$ to the treatment initiation model in Equation (4.7) (Robins, 1992, 1998). The 95% confidence intervals included values of ψ such that the p-value of a Wald test for null hypothesis $H_0 : \alpha(\psi) = 0$ was greater than 0.05 (Robins, 1992, 1998; Lok, 2007).

4.4.2 Two-parameter Treatment Effect Model

We also considered a two-parameter treatment effect model:

$$\gamma_t^k(\bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t) = \begin{cases} 1 & \text{if } a_t = 0 \text{ or } a_{t-1} = 1 \\ \exp(\psi_1 + \psi_2(k - T_s^*)) & \text{if } a_t = 1 \text{ and } a_{t-1} = 0 \end{cases}, \quad (4.11)$$

where we accounted for the treatment duration of DIH prosecutions reported in the media, denoted as $k - T_s^*$. Since $a_t - a_{t-1} \geq 0$, $\psi_1 > 0$ means a harmful effect of exposure to DIH prosecutions reported by the media. If $\psi_2 > 0$, then for a longer treatment duration, $k - T_s^*$, there is a more harmful effect. Alternatively, if $\psi_1 < 0$, there is a protective effect of being exposed to DIH prosecutions reported by the media, and if $\psi_2 < 0$, then for a longer treatment duration, there is a more protective effect of drug overdose deaths.

Under the two-parameter model, we have

$$H_{s,k}(t) = \begin{cases} Y_{s,k} & \text{if } k \leq T_s^* \\ \frac{Y_{s,k}}{\exp(\psi_1 + \psi_2^*(k - T_s^*))} & \text{if } k > T_s^* \end{cases}. \quad (4.12)$$

Unlike the one-parameter model, we were not able to obtain closed form solutions for both ψ_1 and ψ_2 when solving the unbiased estimating equations for a two-parameter model for the treatment effect γ_k^t . Instead, we used a grid-search approach to solve for $\psi^* = (\psi_1^*, \psi_2^*)$, where for a grid of values for $\psi = (\psi_1, \psi_2)$, we computed $\gamma_{k,\psi}^t$ and found $\hat{\psi}$ such that $\alpha(\hat{\psi})$ was closest to zero. The 95% confidence region for ψ_1^* and ψ_2^* consisted of values of $\psi = (\psi_1, \psi_2)$ in which the coefficients for the corresponding treatment variables were not jointly statistically significant at level $\alpha = 0.05$.

4.5 Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths: Count Treatment

In this section, we consider a treatment given by $\tilde{A}_{s,t}$, the number of DIH prosecutions reported in the media in state s at time interval t . Here, $g(\tilde{A}_{s,t}) = \tilde{A}_{s,t}$, which can be modeled using a Poisson regression model with a log link function. To determine how to incorporate the history of DIH prosecutions reported by the media into the model for $\tilde{A}_{s,t}$, we compared the log likelihood ratios of the various models of consideration and a baseline model without accounting for past DIH prosecutions reported by the media. In particular, the models of consideration included a weighted Poisson regression model with a predictor of the total number of DIH prosecutions reported by the media in state s before time interval t and a weighted Poisson regression model with a predictor of just the number of DIH prosecutions reported by the media in state s at time interval $t - 1$. We used deterministic weights given by the number

of time intervals from time interval t until time interval K . The log likelihood ratio for the model with the number of DIH prosecutions reported by the media at time interval $t - 1$ and the baseline model was higher than the log likelihood ratio for the model accounting for all of the past DIH prosecutions reported by the media and the baseline model, suggesting that former model was more likely given the data. Hence, the model for the mean number of DIH prosecutions reported by the media for state s at time interval t was given by:

$$\log \left(\mathbb{E} \left(\tilde{A}_{s,t} \mid \bar{A}_{s,t-1}, \bar{X}_{s,t} \right) \right) = \beta_0 + \beta_1 \sum_{l=1}^L X_{l,s,t-1} + \beta_2 t + \beta_3 Y_{s,t-1} + \beta_4 \tilde{A}_{s,t-1}, \quad (4.13)$$

where the predictors are the same as the predictors in Equation (4.7) but with the addition of $\tilde{A}_{s,t-1}$, the number of DIH prosecutions reported by the media in state s at time interval $t - 1$.

4.5.1 One-parameter Treatment Effect Model

We considered a one-parameter model for the treatment effect:

$$\gamma_k^t(\bar{X}_{s,t}, \bar{A}_{s,t} = \tilde{a}_t) = \exp(\psi(\tilde{a}_t - \tilde{a}_{t-1})). \quad (4.14)$$

Under this model, we no longer restrict $\tilde{a}_t - \tilde{a}_{t-1} \geq 0$ since there may be more or less DIH prosecutions reported by the media at time interval $t - 1$ versus time interval t . Then, if $\psi > 0$, there is a positive relationship between the change in the number of DIH prosecutions reported by the media and the change in drug overdose deaths from time interval $t - 1$ to time interval t . That is, an increase in DIH prosecutions reported by the media means a harmful effect on drug overdose deaths. If $\psi < 0$, there is a negative relationship between the change in the number of DIH prosecutions reported by the media and the change in drug overdose deaths from time interval $t - 1$ to time interval t . That is, an increase in DIH prosecutions reported by the media leads to a

protective effect on drug overdose deaths.

Recall that we defined $H_{s,k}(t) = H_{s,k}(t+1)\gamma_t^k(\bar{X}_{s,t}, \bar{A}_{s,t})^{-1}$. In the binary case, since the treatment only changed once, we only have to “blip” off the treatment once. However, when the treatment is not binary, we compute $H_{s,k}(t)$ recursively, and so we take $H_{s,k}(t+1)$ into account. Hence,

$$H_{s,k}(t) = \begin{cases} Y_{s,k} & \text{if } k \leq T_s^* \\ H_{s,k}(t+1)\gamma_t^k(\bar{X}_{s,t}, \bar{A}_{s,t})^{-1} & \text{if } k > T_s^* \end{cases}. \quad (4.15)$$

Since we could not find a closed form solution for ψ using the unbiased estimating equations, we used a grid search to find ψ such that when $\alpha(\psi)H_{s,k}(t)$ was added to the model in Equation (4.13), we have $\alpha(\psi) = 0$. The 95% confidence intervals were estimated as in the one-parameter model for binary treatment.

4.6 Results

First, we considered when treatment was binary, i.e. the effect of having any DIH prosecutions reported by the media versus having none on unintentional drug overdose deaths. Solving Equation (4.10) for ψ for the one-parameter treatment model, $\hat{\psi} = -0.107$ with 95% CI (-.217, -.011) and $\exp(-\hat{\psi}) = 0.899$ (95% CI: (0.805, 0.989)) on the risk ratio scale. Results suggest a protective effect of exposure to DIH prosecutions reported by the media on unintentional drug overdose deaths. When taking the treatment duration into account in the two-parameter treatment model, we estimated parameters $\hat{\psi}_1 = -0.132$ (risk ratio: 0.876) and $\hat{\psi}_2 = 0.001$ (risk ratio: 1.001). Figure 4-2 shows the 95% confidence region for (ψ_1, ψ_2) , where the red “×” indicates the estimated $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2)$. The 95% confidence region for (ψ_1, ψ_2) contains (0, 0), suggesting that treatment effects were not statistically significant at $\alpha = 0.05$ level.

Figure 4-3 shows the estimated risk ratios of treatment effects at each post-

treatment time interval for (a) the one-parameter model, i.e. $\hat{\psi}$, and (b) two-parameter model, i.e. $\hat{\psi}_1 + \hat{\psi}_2(k - T_s^*)$. Since the one-parameter treatment model did not depend on the treatment duration of DIH prosecutions reported by the media, there was a constant, statistically significant protective effect. For the two-parameter model, there was a potential protective effect when a state was first exposed to DIH prosecutions reported by the media. As the exposure duration to DIH prosecutions increased, there was a small, potentially harmful effect. However, for all possible values of $k - T_s^*$, $\hat{\psi}_1 + \hat{\psi}_2(k - T_s^*) < 0$, suggesting an overall protective effect. The treatment effect at each post-treatment time interval was not statistically significant.

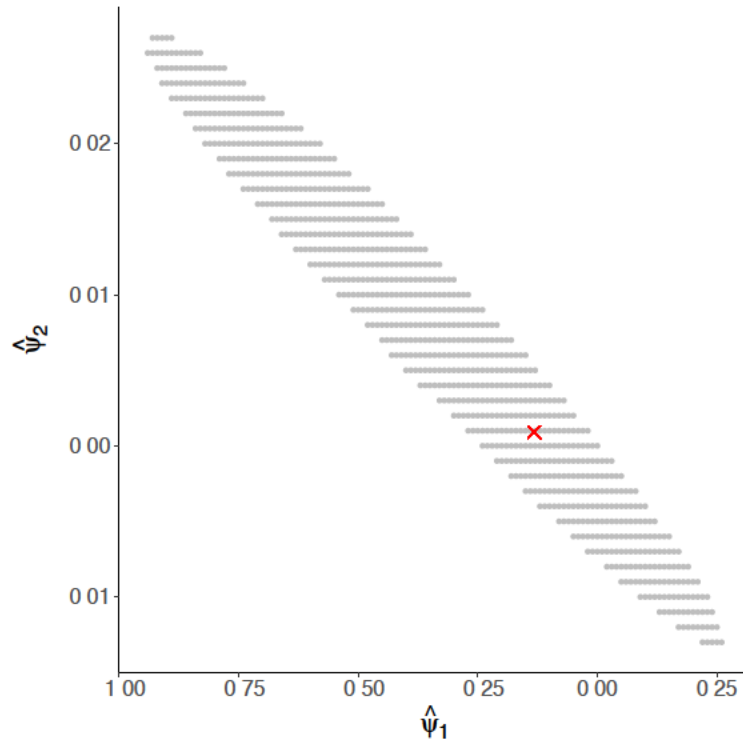


Figure 4.2: Confidence region for ψ_1 and ψ_2 in two-parameter treatment model. The red “x” indicates the estimated $\hat{\psi} = (\hat{\psi}_1, \hat{\psi}_2)$.

When we estimate the effect of an increase in the number of DIH prosecutions reported by the media, we estimated a $\hat{\psi}$ of 0.00749 (95% CI: (0.00703, 0.00798)). On the risk ratio scale, $\exp(\hat{\psi}) = 1.0075$ (95% CI: (1.0071, 1.0080)). Since $\hat{\psi} > 0$,

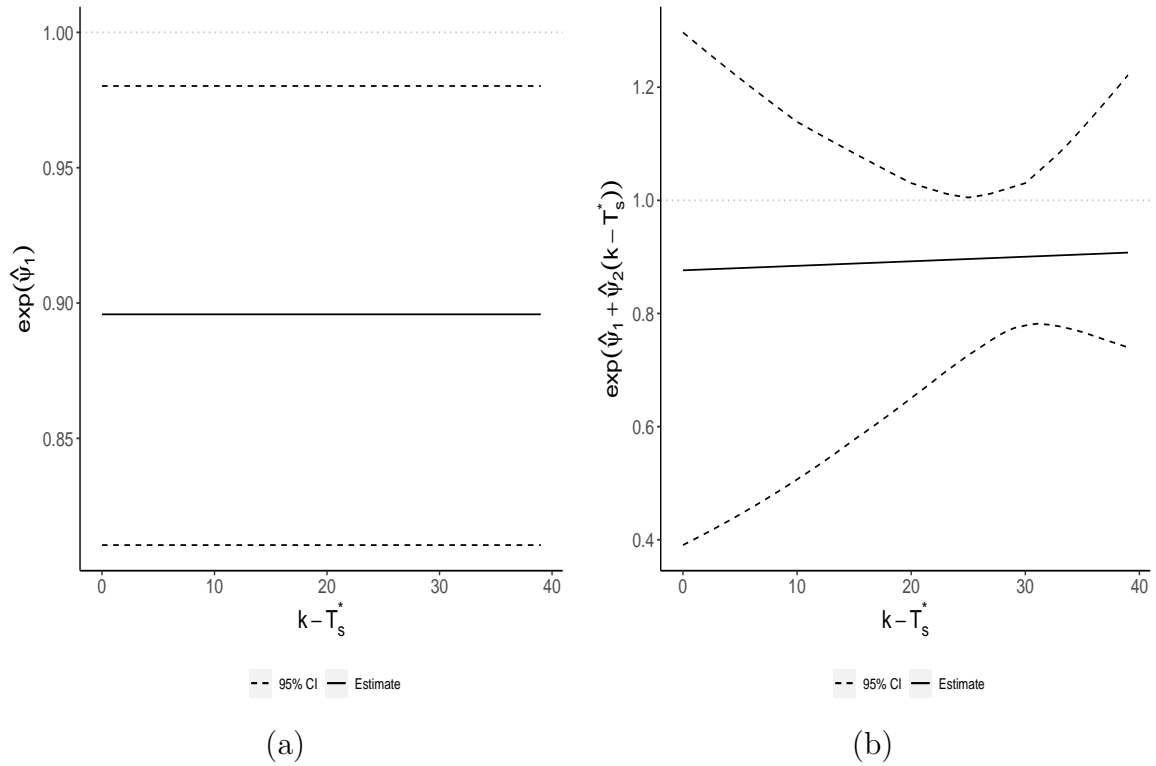


Figure 4-3: The treatment effect given by (a) $\hat{\psi}$ and (b) $\hat{\psi}_1 + \hat{\psi}_2(k - T_s^*)$ by the length of exposure to drug-induced homicide prosecutions reported by the media $k - T_s^*$ with 95% confidence interval when treatment is binary.

results suggest that the risk of unintentional drug overdose deaths move in the same direction as the change in the number of DIH prosecutions from time interval $t - 1$ to time interval t . That is, an increase in DIH prosecutions reported by the media led to an increase in unintentional drug overdose deaths. Figure 4-4 shows the estimated risk ratio of treatment effect given $\tilde{A}_t - \tilde{A}_{t-1} = \tilde{a}_t - \tilde{a}_{t-1}$. The observed change in the number of DIH prosecutions reported by the media from time interval $t - 1$ to time interval t , given by $\tilde{a}_t - \tilde{a}_{t-1}$, in the data ranged from -29 to 64. However, there were often no changes in the number of DIH prosecutions from time interval $t - 1$ to time interval t in the data: the median $\tilde{a}_t - \tilde{a}_{t-1}$ was -1 with interquartile range 17. Hence, even though estimated risk ratios of treatment effects, given the data, ranged from 0.805 to 1.615, most of the estimated risk ratios fell between 0.931 and 1.058

(computed from the first and third quartiles).

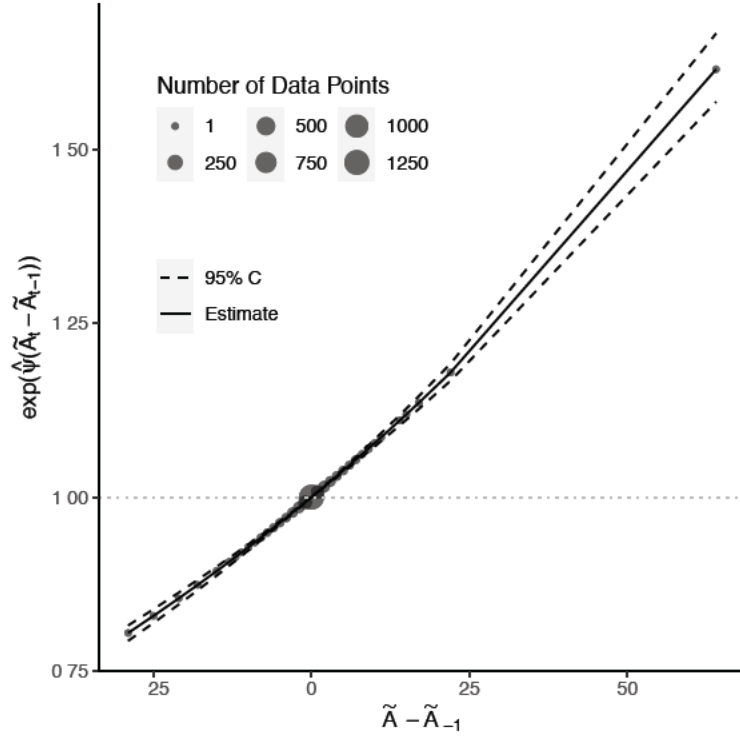


Figure 4.4: Estimated risk ratio of treatment effect for one-parameter treatment model where treatment is number of drug-induced homicide prosecutions and 95% confidence intervals for each $\tilde{a}_t - \tilde{a}_{t-1}$ observed in the data. The size of data point represents the number of observed datapoints for the value of $\tilde{a}_t - \tilde{a}_{t-1}$.

4.7 Discussion

The effect of DIH prosecutions reported by the media on unintentional drug overdose deaths depended on whether we took the exposure duration and number of DIH prosecutions reported by the media into account. Exposure to DIH prosecutions reported by the media assuming a constant treatment effect model had a potential protective effect on unintentional drug overdose deaths. This result differed from the results in Chapter 3.4.4 when we considered a constant treatment effect model. In Chapter 3.4.4, we estimated a potential harmful effect of exposure to DIH prosecutions

reported by the media on unintentional drug overdose deaths. When we took the exposure duration to DIH prosecutions reported by the media into account in the SNRRM, we estimated a potentially protective effect of exposure to DIH prosecutions while we estimated a potentially harmful effect as the treatment duration increases. Again, this differed from results in Chapter 3.6.5, where, when we took the exposure duration into account as a linear effect in the GAM, we estimated a potentially harmful effect at the time of exposure but a potentially protective effect as the length of exposure increases. However, overall, the treatment effect in the one and two-parameter models at each post-treatment time interval remained negative, suggesting an overall protective effect.

When we took the number of DIH prosecutions into account, we estimated a potential harmful effect as the number of DIH prosecutions reported by the media increases. For each additional DIH prosecution reported by the media from one six-month interval to the next, we estimated 0.75% (95% CI: (0.71%, 0.80%)) more unintentional drug overdose deaths. However, since there could many more DIH prosecutions reported by the media from one time interval to the next, there could be a substantial increase in the number of unintentional drug overdose deaths due to DIH prosecutions reported by the media.

There were limitations to the analysis. Of note, the treatment effect models were simplified, e.g. we only took the number of other relevant policy measures that have been enacted in the past into account as opposed to focusing on each of the relevant policy measures separately. We also only considered one and two-parameter treatment effect models. However, the estimation of the parameters becomes harder with the addition of more parameters in the treatment effect model. Furthermore, we chose $\bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t}) = 1$ for simplicity, but there are other choices of $\bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t})$ that can be used for the estimation of treatment effects (Lok et al., 2021).

In the next steps, we would likely benefit from finding an optimal function $\bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t})$ and a treatment effect model with more parameters (Lok et al., 2021). Furthermore, we would like to estimate the number of attributable deaths as we did in the previous chapter. We would also like to estimate the effect of DIH prosecutions reported by the media using the model using the number of DIH prosecutions reported by the media as treatment and accounting for exposure duration. This would allow us to determine the effect of the intensity and treatment duration of DIH prosecutions reported by the media on drug overdose deaths.

Overall, results suggest that there might be an overall protective effect on unintentional drug overdose deaths of being exposed to DIH prosecutions reported by the media. However, when we account for the intensity of DIH prosecutions reported by the media, there is a harmful effect of DIH prosecutions reported by the media on unintentional drug overdose deaths. However, similar to conclusions in the previous chapter, more work needs to be done to actually assess the impact of DIH prosecutions reported by the media.

Appendix A

Treatment Effects Under Exposure Models

A.1 Linear Unbiased Constraints

Proof of Proposition 1. Let $\hat{\theta}_{1,m_1}$ be a linear estimator for the unit-level causal effect for a unit i whose weights only depend on the unit's exposure, i.e. $\hat{\theta}_{1,m_1} = w(\vec{e})Y(\vec{e})$, where $\vec{e}_i^{obs} = \vec{e}$. Without the loss of generality, we assume that the exposure component of interest is the first one. The parameter of interest for the unit-level causal effect of the first exposure component being m_1 versus zero is denoted as θ_{1,m_1} . The expected value of the estimator under additivity is given as follows:

$$\begin{aligned} \mathbb{E}(\hat{\theta}_{1,m_1}) &= \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e})Y(\vec{e}) \\ &= \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e}) \left[\alpha + \theta_{1,m_1} \mathbb{I}\{e_1 = m_1\} \right. \\ &\quad \left. + \sum_{m=1}^{m_1-1} \theta_{1,m} \mathbb{I}\{e_1 = m\} + \sum_{k=2}^K \sum_{j_k=1}^{m_k} \theta_{k,j_k} \mathbb{I}\{e_k = j_k\} \right]. \end{aligned}$$

In order for $\mathbb{E}(\hat{\theta}_{1,m_1}) = \theta_{1,m_1}$, we need:

$$\begin{aligned} \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e}) &= 0 \\ \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e}) \mathbb{I}\{e_1 = m_1\} &= 1 \\ \forall m : m \in \{1, \dots, m_1 - 1\} \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e}) \mathbb{I}\{e_1 = m\} &= 0 \\ \forall k, j_k : k \in \{2, \dots, K\}, j_k \in \{1, \dots, m_k\} \sum_{\vec{e} \in \mathcal{E}} p(\vec{e})w(\vec{e}) \mathbb{I}\{e_k = j_k\} &= 0. \end{aligned}$$

These give us the constraints needed for unbiasedness. \square

A.2 Affine Basis for Set of LUEs \mathcal{U}

We first define notation for the weight of exposures in an estimator.

Definition A.2.1 (Weight of Exposure in Estimator in \mathcal{M}). Let $\hat{\theta} \in \mathcal{M}$. We denote the weight on the Horvitz Thompson term associated with exposure $\vec{e} \in \mathcal{E}$ as $f_{\hat{\theta}}(\vec{e}) : \mathcal{E} \rightarrow \{-1, 0, 1\}$.

A.2.1 Construction and Affine Independence of \mathcal{M}

Proof of Lemma 2.1. Let \mathcal{M} be the set of estimators described in Theorem 2.1. Note that all estimators in \mathcal{M} are MALUEs. Let $\hat{\theta} = \sum_{\tilde{\theta} \in \mathcal{M}} g(\tilde{\theta})\tilde{\theta}$ where $\sum_{\tilde{\theta} \in \mathcal{M}} g(\tilde{\theta}) = 1$. Assume $\hat{\theta} \in \mathcal{M}$. If \mathcal{M} is affine independent, then it implies that

$$g(\tilde{\theta}) = \begin{cases} 1, & \text{if } \tilde{\theta} = \hat{\theta} \\ 0, & \text{otherwise} \end{cases}.$$

Since $\tilde{\mathcal{E}}$ is ordered and each MALUE in \mathcal{M} is uniquely identified by the exposures in $\tilde{\mathcal{E}}$, there is a natural ordering of the corresponding MALUEs. We prove that \mathcal{M} is affine independent using induction. Consider each of the MALUEs $\tilde{\theta} \in \mathcal{M}$.

Base Case: $j = 1$

Consider the first MALUE $\tilde{\theta}^{(1)} \in \mathcal{M}$. In particular, we have the MALUE:

$$\tilde{\theta}^{(1)} = HT_{(m_1, m_2, \dots, m_K)} - HT_{(m_1-1, m_2, \dots, m_K)} + HT_{(m_1-1, 0, m_3, \dots, m_K)} - HT_{(0, 0, m_3, \dots, m_K)}$$

Based on the construction of \mathcal{M} , $\tilde{\theta}^{(1)}$ is a unique MALUE for which $\vec{e}^{(1)} \in \text{supp}(\tilde{\theta}^{(1)})$ where $\vec{e}^{(1)} = (m_1 - 1, m_2, \dots, m_K)$. Thus, if $\vec{e}^{(1)} \in \text{supp}(\hat{\theta})$, then $\tilde{\theta} = \hat{\theta}$, i.e. $g(\tilde{\theta}) = 1$. Otherwise, $g(\tilde{\theta}) = 0$.

Induction Hypothesis: Now assume that for $j \in \{2, \dots, u\}$, the weight g for the j th MALUE in \mathcal{M} is given by Equation (2.13).

Case: $j = u + 1$

Now consider the $(u+1)$ th estimator $\tilde{\theta}^{(u+1)} \in \mathcal{M}$. From the definition of the estimators in \mathcal{M} , $\tilde{\theta}^{(u+1)}$ is uniquely identified by an exposure $\vec{e}^{(u+1)} \in \tilde{\mathcal{E}}$ where either $e_1^{(u+1)} \in \{1, \dots, m_1 - 1\}$ or $e_1^{(u+1)} = m_1$, depending on whether $\tilde{\theta}^{(u+1)}$ is a four-term or two-

term MALUE, respectively. If $\vec{e}^{(u+1)} \notin \text{supp}(\hat{\theta})$, then $g(\tilde{\theta}^{(u+1)}) = 0$. Otherwise, we consider the different cases.

If for all j' 'th MALUEs $\tilde{\theta}^{(j')} \in \mathcal{M}$ for $j' < j$ we have $g(\tilde{\theta}^{(j')}) = 0$ and $\vec{e}^{(u+1)} \in \text{supp}(\hat{\theta})$, then $g(\tilde{\theta}^{(u+1)}) = 1$. This is because \mathcal{M} is ordered and since $\tilde{\theta}^{(u+1)}$ is a MALUE, the exposure components in the exposures corresponding to each of the HT terms are simultaneously decreasing. Hence, $\tilde{\theta}^{(u+1)}$ is the last MALUE in \mathcal{M} for which \vec{e} is in its support. Since all previous MALUEs $\tilde{\theta}^{(j')}$ have weight 0, then $\tilde{\theta}^{(u+1)}$ must be equal to $\hat{\theta}$, i.e. $g(\tilde{\theta}^{(u+1)}) = 1$.

If there exists a j' 'th MALUE $\tilde{\theta}^{(j')} \in \mathcal{M}$ for $j' < j$ such that $g(\tilde{\theta}^{(j')}) = 1$, it means that $\tilde{\theta}^{(j')} = \hat{\theta}$. Because of the induction hypothesis, there can be at most one estimator before the $(u+1)$ th estimator that has non-zero weight (namely the estimator equal to $\hat{\theta}$). If $\vec{e}^{(u+1)} \in \text{supp}(\hat{\theta})$, then $\vec{e}^{(u+1)} \in \text{supp}(\tilde{\theta}^{(j')})$. Furthermore, $f_{\hat{\theta}}(\vec{e}^{(u+1)}) = f_{\tilde{\theta}^{(j')}}(\vec{e}^{(u+1)})$. If $g(\tilde{\theta}^{(u+1)}) \neq 0$, the weight $f_{\tilde{\theta}^{(j')} + g(\tilde{\theta}^{(u+1)})\tilde{\theta}^{(u+1)}}(\vec{e}^{(u+1)})$ is given by $f_{\tilde{\theta}^{(j')}}(\vec{e}^{(u+1)}) + g(\tilde{\theta}^{(u+1)})f_{\tilde{\theta}^{(u+1)}}(\vec{e}^{(u+1)})$. Since $\tilde{\theta}^{(u+1)}$ is the last MALUE in \mathcal{M} with exposure $\vec{e}^{(u+1)}$ in its support, there are no other MALUEs for which we cancel the extra weight on the Horvitz Thompson term with exposure $\vec{e}^{(u+1)}$. Thus, if $g(\tilde{\theta}^{(u+1)}) \neq 0$, we have $f_{\tilde{\theta}^{(j')}}(\vec{e}^{(u+1)}) + g(\tilde{\theta}^{(u+1)})f_{\tilde{\theta}^{(u+1)}}(\vec{e}^{(u+1)}) \neq f_{\tilde{\theta}^{(j')}}(\vec{e}^{(u+1)}) = f_{\hat{\theta}}(\vec{e}^{(u+1)})$. This leads to biasedness. Hence, $g(\tilde{\theta}^{(u+1)})$ must be 0.

For all estimators in $\tilde{\theta} \in \mathcal{M}$, the weights $g(\tilde{\theta})$ are defined as in Equation (2.13). Thus, if $\hat{\theta} \in \mathcal{M}$, then $\hat{\theta}$ cannot be written as an affine combination of estimators $\tilde{\theta}$ in \mathcal{M} , i.e. the set of estimators \mathcal{M} are affine independent. □

A.2.2 Affine Basis for \mathcal{U}

Proof of Theorem 1. First, we want to show that $\hat{\Theta}$ is affine independent. Consider the support of $\hat{\Theta}$:

$$\begin{aligned} \text{supp}(\hat{\Theta}) &= \{\vec{e} : \vec{e} \in \mathcal{E}, e_1 = m_1\} \\ &\cup \{\vec{e} : \vec{e} \in \mathcal{E}, e_1 \in \{1, \dots, m_1 - 1\}, \exists k \in \{2, \dots, K\} \text{ s.t. } e_k \neq 0\} \\ &\cup \{\vec{e} : \vec{e} \in \mathcal{E}, e_1 = 0, \exists k, k' \in \{2, \dots, K\} \text{ s.t. } e_k \neq 0, e_{k'} \neq 0\}. \end{aligned} \quad (\text{A.1})$$

Note that each estimator in $\hat{\Theta}$ can be uniquely identified by an exposure. In particular, each two-term estimator in \mathcal{M} can be uniquely identified by an exposure where $e_1 = m_1$. Each four term estimator in \mathcal{M} can be uniquely identified by an exposure where $e_1 \in \{1, \dots, m_1 - 1\}$ and there is at least another non-zero exposure. Each estimator

in \mathcal{Z} can be uniquely identified by an exposure where $e_1 = 0$ and there are at least two other non-zero exposure components.

Order the set of exposures such that the set of exposures with $e_1 \in \{1, \dots, m_1 - 1\}$ are first, then the exposures with $e_1 = m_1$, and finally $e_1 = 0$. Within each subset of exposures in $\text{supp}(\hat{\Theta})$, order the exposures in a reverse reflected lexicographic manner by uniquely identifying exposure. Recall that each estimator in \mathcal{M} is a MALUE, and so the exposures, and their corresponding Horvitz-Thompson terms, can be arranged such that the exposure components are simultaneously non-increasing. Furthermore, exposures, and their corresponding Horvitz-Thompson terms, in the support of estimators in \mathcal{Z} can also be ordered. In particular, exposures in the support of an estimator in \mathcal{Z} can be arranged, in increasing order, according to the reverse reflected lexicographic order. Hence, the zero estimators are also monotonic. Because of the ordering of $\text{supp}(\hat{\Theta})$ and the monotonicity of the estimators, each estimator $\tilde{\theta} \in \hat{\Theta}$ is the last estimator in $\hat{\Theta}$ for which the uniquely identifying exposure is in its support. For example, the two term estimator $HT_{(m_1, e_2, \dots, e_K)} - HT_{(0, e_2, \dots, e_K)}$, where $e_k \in \{0, \dots, m_k\}$, is the last estimator for which (m_1, e_2, \dots, e_K) is in its support.

Now, let $\hat{\theta} = \sum_{\tilde{\theta} \in \hat{\Theta}} g(\tilde{\theta})\tilde{\theta}$, where $\sum_{\tilde{\theta} \in \hat{\Theta}} g(\tilde{\theta}) = 1$. Suppose that $\hat{\theta} \in \hat{\Theta}$. Based on the ordering of the estimators, we can extend the proof of Theorem 2.1 to here. By doing so, we argue that g is given by:

$$g(\tilde{\theta}) = \begin{cases} 1, & \text{if } \tilde{\theta} = \hat{\theta} \\ 0, & \text{otherwise} \end{cases}.$$

Hence, $\hat{\Theta}$ is affine independent.

Next, we show that $\hat{\Theta}$ spans the set of LUE. To do so, we determine the dimension of $\hat{\Theta}$ and the dimension of \mathcal{U} . The number of estimators in $\tilde{\theta} \in \hat{\Theta}$ is equal to the number of uniquely identifying exposures. That is,

$$|\hat{\Theta}| = \underbrace{\left(\prod_{k=2}^K (m_k + 1) \right)}_{\text{two term estimators}} + \underbrace{\left((m_1 - 1) \left[\prod_{k=2}^K (m_k + 1) - 1 \right] \right)}_{\text{four term estimators}}$$

$$\begin{aligned}
& + \left(\underbrace{\prod_{k=2}^K (m_k + 1) - 1 - \sum_{k=2}^K m_k}_{\text{zero estimators}} \right) \\
& = \prod_{k=1}^K (m_k + 1) - \sum_{k=1}^K m_k. \tag{A.2}
\end{aligned}$$

However, since the weights of the estimators must sum to 1, there is one less free dimension. Thus, the dimension of the affine space is

$$\prod_{k=1}^K (m_k + 1) - \sum_{k=1}^K m_k - 1.$$

The dimension of the LUEs is determined by the number of exposures minus the number of constraints. Hence, we have,

$$\begin{aligned}
|\mathcal{U}| & = \left(\underbrace{\prod_{k=1}^K (m_k + 1)}_{\text{number of exposures}} \right) - \left(\underbrace{\left[\sum_{k=1}^K m_k \right] + 1}_{\text{number of constraints}} \right) \\
& = \prod_{k=1}^K (m_k + 1) - \sum_{k=1}^K m_k - 1. \tag{A.3}
\end{aligned}$$

So $\hat{\Theta}$ is an affine independent set with dimension equal to the dimension of the set of LUEs, i.e. $\hat{\Theta}$ spans the set of LUEs. Hence, the set $\hat{\Theta}$ is an affine basis for the set of LUEs. \square

A.3 MIV LUEs

A.3.1 Optimization Problem

We solve the following optimization problem to find weights for exposures for a LUE that has minimum integrated variance. First, let $\alpha \in \Theta$ be the parameter corresponding to the baseline (i.e. when all exposure component values are 0) and $\theta_{k,j_k} \in \Theta$ be parameters corresponding to the k th exposure component, where $k \in \{1, \dots, K\}$,

when it is equal to j_k versus zero, where $j_k \in \{1, \dots, m_k\}$. Suppose the parameter of interest is $\theta_{1,m_1} \in \Theta$. To find a MIV LUE for a given prior, we find weights such that the integrated variance is minimized with respect to the linear unbiased constraints. Therefore, the optimization problem becomes:

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2} \int_{\Theta} \sum_{\vec{e}} p(\vec{e}) (w(\vec{e})Y(\vec{e}) - \theta_{1,m_1})^2 \pi(\theta') d\theta' + \lambda_1 \left(1 - \sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m_1\} \right) \\
&\quad - \sum_{m=1}^{m_1-1} \lambda_{2,m} \left(\sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m\} \right) - \lambda_3 \sum_{\vec{e}} p(\vec{e})w(\vec{e}) \\
&\quad - \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} \sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_k = j_k\} \\
&= \frac{1}{2} \left[\sum_{\vec{e}} p(\vec{e})w(\vec{e})^2 \int_{\Theta} Y(\vec{e})^2 \pi(\theta') d\theta' + \sum_{\vec{e}} p(\vec{e}) \int_{\Theta} \theta_{1,m_1}^2 \pi(\theta') d\theta' \right. \\
&\quad \left. - 2 \int_{\Theta} \theta_{1,m_1} \sum_{\vec{e}} p(\vec{e})w(\vec{e})Y(\vec{e})\pi(\theta') d\theta' \right] + \lambda_1 \left(1 - \sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m_1\} \right) \\
&\quad - \sum_{m=1}^{m_1-1} \lambda_{2,m} \left(\sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m\} \right) - \lambda_3 \sum_{\vec{e}} p(\vec{e})w(\vec{e}) \\
&\quad - \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} \sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_k = j_k\} \\
&= \frac{1}{2} \left[\sum_{\vec{e}} p(\vec{e})w(\vec{e})^2 \text{Var}(Y(\vec{e})) - \int_{\Theta} \theta_{1,m_1}^2 \pi(\theta') d\theta' \right] \\
&\quad + \lambda_1 \left(1 - \sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m_1\} \right) - \sum_{m=1}^{m_1-1} \lambda_{2,m} \left(\sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_1 = m\} \right) \\
&\quad - \lambda_3 \sum_{\vec{e}} p(\vec{e})w(\vec{e}) - \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} \sum_{\vec{e}} p(\vec{e})w(\vec{e})\mathbb{I}\{e_k = j_k\}
\end{aligned}$$

Taking the derivatives with respect to exposure $\vec{e} \in \mathcal{E}$ and setting the derivative to zero:

$$\begin{aligned}
0 &= \frac{\partial \mathcal{L}}{\partial w(\vec{e})} = p(\vec{e})w(\vec{e})\text{Var}(Y(\vec{e})) - \lambda_1 p(\vec{e})\mathbb{I}\{e_1 = m_1\} - \sum_{m=1}^{m_1-1} \lambda_{2,m} p(\vec{e})\mathbb{I}\{e_1 = m\} \\
&\quad - \lambda_3 p(\vec{e}) - \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} p(\vec{e})\mathbb{I}\{e_k = j_k\} \\
\Rightarrow p(\vec{e})w(\vec{e})\text{Var}(Y(\vec{e})) &= \lambda_1 p(\vec{e})\mathbb{I}\{e_1 = m_1\} + \sum_{m=1}^{m_1-1} \lambda_{2,m} p(\vec{e})\mathbb{I}\{e_1 = m\} \\
&\quad + \lambda_3 p(\vec{e}) + \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} p(\vec{e})\mathbb{I}\{e_k = j_k\} \\
w(\vec{e}) &= \frac{\lambda_1 \mathbb{I}\{e_1 = m_1\} + \sum_{m=1}^{m_1-1} \lambda_{2,m} \mathbb{I}\{e_1 = m\} + \lambda_3 + \sum_{k=2}^K \sum_{j_k=1}^{m_k} \lambda_{4,k,j_k} \mathbb{I}\{e_k = j_k\}}{\text{Var}(Y(\vec{e}))}
\end{aligned} \tag{A.4}$$

The MIV LUE problem is equivalent to the following matrix problem:

$$\begin{bmatrix}
p(\vec{e}_1)\text{Var}(Y(\vec{e}_1)) & \cdots 0 \cdots & 0 & p(\vec{e}_1)\mathbb{I}\{\theta_{1,m_1} \in \vec{e}_1\} & \cdots & p(\vec{e}_1)\mathbb{I}\{\theta_{K,m_K} \in \vec{e}_1\} & p(\vec{e}_1) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots 0 \cdots & p(\vec{e}_{|\mathcal{E}|})\text{Var}(Y(\vec{e}_{|\mathcal{E}|})) & p(\vec{e}_{|\mathcal{E}|})\mathbb{I}\{\theta_{1,m_1} \in \vec{e}_{|\mathcal{E}|}\} & \cdots & p(\vec{e}_{|\mathcal{E}|})\mathbb{I}\{\theta_{K,m_K} \in \vec{e}_{|\mathcal{E}|}\} & p(\vec{e}_{|\mathcal{E}|}) \\
p(\vec{e}_1)\mathbb{I}\{\theta_{1,m_1} \in \vec{e}_1\} & \cdots & p(\vec{e}_{|\mathcal{E}|})\mathbb{I}\{\theta_{1,m_1} \in \vec{e}_{|\mathcal{E}|}\} & 0 & \cdots & 0 & 0 \\
p(\vec{e}_1)\mathbb{I}\{\theta_{1,1} \in \vec{e}_1\} & \cdots & p(\vec{e}_{|\mathcal{E}|})\mathbb{I}\{\theta_{1,1} \in \vec{e}_{|\mathcal{E}|}\} & 0 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
p(\vec{e}_1)\mathbb{I}\{\theta_{K,m_K} \in \vec{e}_1\} & \cdots & p(\vec{e}_{|\mathcal{E}|})\mathbb{I}\{\theta_{K,m_K} \in \vec{e}_{|\mathcal{E}|}\} & 0 & \cdots & 0 & 0 \\
p(\vec{e}_1) & \cdots & p(\vec{e}_{|\mathcal{E}|}) & 0 & \cdots & 0 & 0
\end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \cdots \\ 0 \\ 1 \\ 0 \\ \cdots \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} w(\vec{e}_1) \\ \cdots \\ w(\vec{e}_{|\mathcal{E}|}) \\ \lambda_1 \\ \cdots \\ \lambda_{4,K,m_K} \\ \lambda_3 \end{bmatrix}, \tag{A.5}$$

where $\theta_{k,j_k} \in \vec{e}$ means that the parameter θ_{k,j_k} is a summand in the summation of parameters that equals to the potential outcome given exposure \vec{e} . Let matrix $\mathbf{P} = \begin{pmatrix} \mathbf{W} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix}$ be a block matrix of dimension $(|\mathcal{E}| + \sum_{k=1}^K m_k + 1) \times (|\mathcal{E}| + \sum_{k=1}^K m_k + 1)$, where \mathbf{W} is a $|\mathcal{E}| \times |\mathcal{E}|$ diagonal matrix and the j th diagonal

entry corresponding to \vec{e}_j for $j \in \{1, \dots, |\mathcal{E}|\}$ is

$$\mathbf{W}_{j,j} = p(\vec{e}_j)\text{Var}(Y(\vec{e}_j))$$

and \mathbf{C} is a $|\Theta| \times |\mathcal{E}|$ matrix of constraints given by Proposition 1 where the rows correspond to the parameters in Θ and the columns correspond to the exposures. We write Equation (A.5) as $\mathbf{P}^{-1}\mathbf{b} = \mathbf{w}$.

A.3.2 Characterizing MIV LUEs

Proof of Lemma 2.2. Let $\eta \in \mathbb{R}$ be such that $\eta \rightarrow \infty$ and $B \in \mathbb{R}^{|\Theta| \times |\Theta|}$ be a positive semi-definite matrix where elements $0 < b_{k,j} < \infty$ are small. Now consider the variance-covariance matrix $\tilde{\Sigma}_\eta = \eta\Sigma + B$. Then:

$$\begin{aligned} \lim_{\eta \rightarrow \infty} v_1^T \tilde{\Sigma}_\eta v_1 &= \lim_{\eta \rightarrow \infty} v_1^T (\eta\Sigma + B)v_1 \\ &= \lim_{\eta \rightarrow \infty} \eta \underbrace{v_1^T \Sigma v_1}_{=0} + v_1^T B v_1 \\ &= v_1^T B v_1 < \infty \end{aligned} \tag{A.6}$$

$$\begin{aligned} \lim_{\eta \rightarrow \infty} v_2^T \tilde{\Sigma}_\eta v_2 &= \lim_{\eta \rightarrow \infty} v_2^T (\eta\Sigma + B)v_2 \\ &= \lim_{\eta \rightarrow \infty} \eta \underbrace{v_2^T \Sigma v_2}_{=a} + v_2^T B v_2 \\ &= \lim_{\eta \rightarrow \infty} \underbrace{\eta}_{\rightarrow \infty} a + \underbrace{v_2^T B v_2}_{< \infty} \rightarrow \infty. \end{aligned} \tag{A.7}$$

□

Lemma A.1. Let the design p be such that $p(\vec{e}) > 0$ for all $\vec{e} \in \mathcal{E}$ and let $\text{Var}(Y(\vec{e})) > 0$ for all $\vec{e} \in \mathcal{E}$. The matrix $\mathbf{P} = \begin{pmatrix} \mathbf{W} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix}$ is full rank.

Proof. First \mathbf{W} is full rank since it is diagonal with positive diagonal entries. Furthermore, \mathbf{C} has full row rank because otherwise, the linear unbiased constraints given by Proposition 1 are redundant. If the constraints are redundant, we can remove a constraint, but the constraints for unbiasedness are minimal. That is, if we removed

a constraint, there are linear estimators that satisfy the remaining constraints but are not unbiased. Hence, \mathbf{C} must have full row rank in order to preserve unbiasedness. Suppose that $c = \begin{pmatrix} c_{\mathcal{E}} \\ c_{\Theta} \end{pmatrix} \in \mathbb{R}^{|\mathcal{E}|+|\Theta|}$ satisfies $c^T \mathbf{P} = 0$ where $c_{\mathcal{E}}$ is a vector of length $|\mathcal{E}|$ and c_{Θ} is a vector of length $|\Theta|$. Then:

$$c^T \mathbf{P} = c^T \begin{pmatrix} \mathbf{W} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} c_{\mathcal{E}}^T \mathbf{W} + c_{\Theta}^T \mathbf{C} \\ c_{\mathcal{E}}^T \mathbf{C}^T \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{|\mathcal{E}| \times 1} \\ \mathbf{0}_{|\Theta| \times 1} \end{pmatrix}.$$

We then solve for $c_{\mathcal{E}}$ and c_{Θ} :

$$\begin{aligned} \mathbf{0}_{|\mathcal{E}| \times 1} &= c_{\mathcal{E}}^T \mathbf{W} + c_{\Theta}^T \mathbf{C} \implies c_{\mathcal{E}}^T = -c_{\Theta}^T \mathbf{C} \mathbf{W}^{-1}, \\ \mathbf{0}_{|\Theta| \times 1} &= c_{\mathcal{E}}^T \mathbf{C}^T = -c_{\Theta}^T \mathbf{C} \mathbf{W}^{-1} \mathbf{C}^T \implies c_{\Theta}^T = \mathbf{0}_{|\Theta| \times 1}, \end{aligned}$$

where in the first line, we can take the inverse of \mathbf{W} since it is full rank and has positive diagonal entries, and in the second line, we multiply both sides by $(\mathbf{C} \mathbf{W}^{-1} \mathbf{C}^T)^{-1}$ where $\mathbf{C} \mathbf{W}^{-1} \mathbf{C}^T$ is full rank since \mathbf{C} has full row rank and \mathbf{W} is full rank and has positive diagonal entries. Since $c_{\Theta}^T = \mathbf{0}_{|\Theta| \times 1}$, then $c_{\mathcal{E}}^T = \mathbf{0}_{|\mathcal{E}| \times 1}$. Hence, all the rows in \mathbf{P} are linearly independent, and since \mathbf{P} is a square matrix, \mathbf{P} is full rank. \square

Proof of Theorem 2. The proof will proceed as follows. We will first partition the set of parameters and exposures into five sets. We will then show that if a specific one of these sets is empty, then $\text{supp}(\hat{\theta}) \subseteq \mathcal{E}'$. We will also show that $\text{supp}(\hat{\theta}) \subseteq \mathcal{E}'$ holds if that set is non-empty. Finally, we will argue that $\text{supp}(\hat{\theta}) = \mathcal{E}'$ under the conditions specified.

First, the MIV LUE problem is equivalent to the matrix equation:

$$\begin{bmatrix} p(\vec{e}_1) \text{Var}(Y(\vec{e}_1)) & \cdots & 0 & \cdots & p(\vec{e}_1) \mathbb{I}\{\theta_{1,m_1} \in \vec{e}_1\} & \cdots & p(\vec{e}_1) \mathbb{I}\{\theta_{K,m_K} \in \vec{e}_1\} & p(\vec{e}_1) \\ 0 & \cdots & 0 & \cdots & p(\vec{e}_{|\mathcal{E}|}) \mathbb{I}\{\theta_{1,m_1} \in \vec{e}_{|\mathcal{E}|}\} & \cdots & p(\vec{e}_{|\mathcal{E}|}) \mathbb{I}\{\theta_{K,m_K} \in \vec{e}_{|\mathcal{E}|}\} & p(\vec{e}_{|\mathcal{E}|}) \\ p(\vec{e}_1) \mathbb{I}\{\theta_{1,m_1} \in \vec{e}_1\} & \cdots & p(\vec{e}_{|\mathcal{E}|}) \mathbb{I}\{\theta_{1,m_1} \in \vec{e}_{|\mathcal{E}|}\} & \cdots & 0 & \cdots & 0 & 0 \\ p(\vec{e}_1) \mathbb{I}\{\theta_{1,1} \in \vec{e}_1\} & \cdots & p(\vec{e}_{|\mathcal{E}|}) \mathbb{I}\{\theta_{1,1} \in \vec{e}_{|\mathcal{E}|}\} & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p(\vec{e}_1) \mathbb{I}\{\theta_{K,m_K} \in \vec{e}_1\} & \cdots & p(\vec{e}_{|\mathcal{E}|}) \mathbb{I}\{\theta_{K,m_K} \in \vec{e}_{|\mathcal{E}|}\} & \cdots & 0 & \cdots & 0 & 0 \\ p(\vec{e}_1) & \cdots & p(\vec{e}_{|\mathcal{E}|}) & \cdots & 0 & \cdots & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} w(\vec{e}_1) \\ \vdots \\ w(\vec{e}_{|\mathcal{E}|}) \\ \lambda_1 \\ \vdots \\ \lambda_{4,K,m_K} \\ \lambda_3 \end{bmatrix},$$

where we write $\theta_{k,j_k} \in \vec{e}$ to mean that $\theta_{k,j_k} \in \Theta$ is in the sum of parameters in $Y(\vec{e})$. We write this as $\mathbf{P}^{-1} \mathbf{b} = \mathbf{w}$, where $\mathbf{P} = \begin{pmatrix} \mathbf{W} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix}$ is a block matrix of dimension $(|\mathcal{E}| + \sum_{k=1}^K m_k + 1) \times (|\mathcal{E}| + \sum_{k=1}^K m_k + 1)$. \mathbf{W} is a $|\mathcal{E}| \times |\mathcal{E}|$ diagonal matrix and the

j th diagonal entry corresponding to \vec{e}_j for $j \in \{1, \dots, |\mathcal{E}|\}$ is

$$\mathbf{W}_{j,j} = p(\vec{e}_j)\text{Var}(Y(\vec{e}_j))$$

and \mathbf{C} is a $|\Theta| \times |\mathcal{E}|$ matrix of constraints given by Proposition 1 where the rows correspond to the parameters in Θ and the columns correspond to the exposures.

Given $\mathcal{E}' \subseteq \mathcal{E}$, we arrange the exposures in \mathbf{P} , and the corresponding rows in \mathbf{b} and \mathbf{w} , such that exposures $\vec{e}_1, \dots, \vec{e}_{|\mathcal{E}'|} \in \mathcal{E}' \subseteq \mathcal{E}$ and $\vec{e}_{|\mathcal{E}'|+1}, \dots, \vec{e}_{|\mathcal{E}|} \in \mathcal{E} \setminus \mathcal{E}'$. For each $\vec{e} \in \mathcal{E}$, let $\vec{v}_{\vec{e}} \in \{0, 1\}^{|\Theta|}$ be such that $\vec{v}_{\vec{e}}^T \vec{v} = Y(\vec{e})$, where \vec{v} is the vector of parameters. Suppose that $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}) \cap \{\vec{v}_{\vec{e}}\}_{\vec{e} \in \mathcal{E}} = \{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}$. Note that $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'})$ is a linear subspace of $\mathbb{R}^{|\Theta|}$. Then there exists a positive semi-definite matrix Σ such that $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'}) = \text{Null}(\Sigma)$. In particular, $\Sigma = I - XX^T$, where the columns of X are vectors that form an orthonormal basis for $\text{span}(\{\vec{v}_{\vec{e}'}\}_{\vec{e}' \in \mathcal{E}'})$. By Lemma 2.2, there then exists a sequence of variance-covariance matrices $\tilde{\Sigma}_\eta$ such that $\lim_{\eta \rightarrow \infty} \vec{v}_{\vec{e}_j}^T \tilde{\Sigma}_\eta \vec{v}_{\vec{e}_j} < \infty$ for $j \in \{1, \dots, |\mathcal{E}'|\}$ and $\lim_{\eta \rightarrow \infty} \vec{v}_{\vec{e}_{j'}}^T \tilde{\Sigma}_\eta \vec{v}_{\vec{e}_{j'}} = \infty$ for $j' \in \{|\mathcal{E}'|+1, \dots, |\mathcal{E}|\}$.

Let the matrix \mathbf{P}_η be the same as \mathbf{P} , but with rows rearranged as described below and where the variances are given by variance-covariance matrix $\tilde{\Sigma}_\eta$. The MIV LUE problem then becomes $\mathbf{P}_\eta^{-1} \mathbf{b} = \mathbf{w}_\eta$. Let $\lim_{\eta \rightarrow \infty} \mathbf{P}_\eta^{-1} \mathbf{b} = \mathbf{w}^*$. We want to show that $\text{supp}(\mathbf{w}^*) = \mathcal{E}'$, and we first argue that $\text{supp}(\mathbf{w}^*) \subseteq \mathcal{E}'$. Note that $\text{supp}(\mathbf{w}^*) = \text{supp}(\hat{\theta})$ since $\hat{\theta}$ is given by \mathbf{w}^* .

We define submatrices of \mathbf{P}_η as follows. First, the diagonal matrix \mathbf{W} can be decomposed as follows. Denote \mathbf{F}_η as the diagonal matrix of probabilities and variances for the potential outcomes that have a finite limiting variancee:

$$\mathbf{F}_\eta = \begin{bmatrix} p(\vec{e}_1)\text{Var}(Y(\vec{e}_1))_\eta & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & p(\vec{e}_{|\mathcal{E}'|})\text{Var}(Y(\vec{e}_{|\mathcal{E}'|}))_\eta \end{bmatrix}_{|\mathcal{E}'| \times |\mathcal{E}'|}. \quad (\text{A.8})$$

Similarly, denote \mathbf{N}_η as as the analogous matrix for potential outcomes that have a non-finite limiting variance:

$$\mathbf{N}_\eta = \begin{bmatrix} p(\vec{e}_{|\mathcal{E}'|+1})\text{Var}(Y(\vec{e}_{|\mathcal{E}'|+1}))_\eta & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & p(\vec{e}_{|\mathcal{E}|})\text{Var}(Y(\vec{e}_{|\mathcal{E}|}))_\eta \end{bmatrix}_{(|\mathcal{E}|-|\mathcal{E}'|) \times (|\mathcal{E}|-|\mathcal{E}'|)}. \quad (\text{A.9})$$

Next, denote the following subsets of parameters. Let $\Theta^N \subseteq \Theta$ denote the set of parameters where $\theta^N \in \Theta^N$ are such that $\theta^N \notin \vec{e}'$ for all $\vec{e}' \in \mathcal{E}'$. Here, we write

$\theta \notin \vec{e}$ to mean that θ is not in the sum of parameters in $Y(\vec{e})$. Note that for $\theta^N \in \Theta^N$, we have $\lim_{\eta \rightarrow \infty} \text{Var}(\theta^N) = \infty$ by Lemma 2.2. In addition, we will further divide the parameters in $\Theta^F = \Theta \setminus \Theta^N$ as $\Theta^F = \Theta^R \cup \Theta^{NR}$. Specifically, Θ^{NR} will be a maximal subset of Θ^F such that the submatrix of \mathbf{C} with rows given by Θ^{NR} and columns given by \mathcal{E}' is linearly independent.

The matrix $\mathbf{C} \in \mathbb{R}^{|\Theta| \times |\mathcal{E}|}$ contains submatrices corresponding to the linear unbiased constraints. We denote the constraint matrices as \mathbf{C}_e^p , where the subscript corresponds to the set of exposures e and superscript corresponds to the set of parameters p . For each $e \in \{N, F\}$ and $p \in \{N, NR, R\}$, we define \mathbf{C}_e^p to be the submatrix of \mathbf{C} containing linear unbiased constraints in which rows correspond to parameters in Θ^p and columns correspond to the exposures in \mathcal{E}^e . Here, $\mathcal{E}^F = \mathcal{E}'$ and $\mathcal{E}^N = \mathcal{E} \setminus \mathcal{E}'$.

For example \mathbf{C}_N^N is defined as follows:

$$\mathbf{C}_N^N = \begin{bmatrix} p(\vec{e}_{|\mathcal{E}'|+1})\mathbb{I}\{\theta_1^N \in \vec{e}_{|\mathcal{E}'|+1}\} & \cdots & p(\vec{e}_{|\mathcal{E}'|})\mathbb{I}\{\theta_1^N \in \vec{e}_{|\mathcal{E}'|}\} \\ \vdots & \ddots & \vdots \\ p(\vec{e}_{|\mathcal{E}'|+1})\mathbb{I}\{\theta_{|\Theta^N|}^N \in \vec{e}_{|\mathcal{E}'|+1}\} & \cdots & p(\vec{e}_{|\mathcal{E}'|})\mathbb{I}\{\theta_{|\Theta^N|}^N \in \vec{e}_{|\mathcal{E}'|}\} \end{bmatrix}_{|\Theta^N| \times (|\mathcal{E}| - |\mathcal{E}'|)}. \quad (\text{A.10})$$

Altogether, this results in the equation

$$\mathbf{P}_\eta^{-1} \mathbf{b} = \mathbf{w}_\eta \implies \begin{pmatrix} \mathbf{N}_\eta & \mathbf{C}_N^{NT} & \mathbf{0} & \mathbf{C}_N^{NR^T} & \mathbf{C}_N^{RT} \\ \mathbf{C}_N^N & \mathbf{0} & \mathbf{C}_F^N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_F^{NT} & \mathbf{F}_\eta & \mathbf{C}_F^{NR^T} & \mathbf{C}_F^{RT} \\ \mathbf{C}_N^{NR} & \mathbf{0} & \mathbf{C}_F^{NR} & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_N^R & \mathbf{0} & \mathbf{C}_F^R & \mathbf{0} & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0}_{|\mathcal{E}'| - |\mathcal{E}'| \times 1} \\ \mathbf{0}_{|\Theta^N| \times 1} \\ \mathbf{0}_{|\mathcal{E}'| \times 1} \\ 1 \\ \mathbf{0}_{|\Theta^{NR}| - 1 \times 1} \\ \mathbf{0}_{|\Theta^R| \times 1} \end{pmatrix} = \begin{pmatrix} \mathbf{w}(\vec{e}^N)_\eta \\ \lambda_\eta^N \\ \mathbf{w}(\vec{e}^F)_\eta \\ \lambda_{1\eta} \\ \lambda_\eta^{NR} \setminus \lambda_{1\eta} \\ \lambda_\eta^R \end{pmatrix}, \quad (\text{A.11})$$

where the 1 in \mathbf{b} and $\lambda_{1\eta}$ in \mathbf{w}_η corresponds to the constraint for $\theta_{1,m_1} \in \Theta^{NR}$.

Case 1 (Assume $\Theta^R = \emptyset$): We first consider the case when $\Theta^R = \emptyset$, so we can consider the linear equation

$$\begin{pmatrix} \mathbf{N}_\eta & \mathbf{C}_N^{NT} & \mathbf{0} & \mathbf{C}_N^{NR^T} \\ \mathbf{C}_N^N & \mathbf{0} & \mathbf{C}_F^N & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_F^{NT} & \mathbf{F}_\eta & \mathbf{C}_F^{NR^T} \\ \mathbf{C}_N^{NR} & \mathbf{0} & \mathbf{C}_F^{NR} & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0}_{|\mathcal{E}'| - |\mathcal{E}'| \times 1} \\ \mathbf{0}_{|\Theta^N| \times 1} \\ \mathbf{0}_{|\mathcal{E}'| \times 1} \\ 1 \\ \mathbf{0}_{|\Theta^{NR}| - 1 \times 1} \end{pmatrix} = \begin{pmatrix} \mathbf{w}(\vec{e}^N)_\eta \\ \lambda_\eta^N \\ \mathbf{w}(\vec{e}^F)_\eta \\ \lambda_{1\eta} \\ \lambda_\eta^{NR} \setminus \lambda_{1\eta} \end{pmatrix}. \quad (\text{A.12})$$

Note that since we can rearrange rows of \mathbf{P}_η such that \mathbf{P}_η has the form $\begin{pmatrix} \mathbf{W}_\eta & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix}$,

where $\mathbf{W}_\eta = \begin{pmatrix} \mathbf{N}_\eta & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_\eta \end{pmatrix}$ and $\mathbf{C} = \begin{pmatrix} \mathbf{C}_N^N & \mathbf{C}_F^N \\ \mathbf{C}_N^{NR} & \mathbf{C}_F^{NR} \end{pmatrix}$, and since variances are given by $\tilde{\Sigma}_\eta$, we have $\text{Var}(Y(\vec{e})) > 0$ for all $\vec{e} \in \mathcal{E}$. Then, by Lemma A.1, \mathbf{P}_η is full rank, and so \mathbf{P}_η is invertible with solution $\mathbf{w}_\eta = \mathbf{P}_\eta^{-1}\mathbf{b}$. Since $\mathbf{P}_\eta = \begin{bmatrix} \mathbf{A}_\eta & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D}_\eta \end{bmatrix}$ is a block matrix, where

$$\mathbf{A}_\eta = \begin{pmatrix} \mathbf{N}_\eta & \mathbf{C}_N^{NT} \\ \mathbf{C}_N^N & \mathbf{0} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{C}_N^{NR^T} \\ \mathbf{C}_F^N & \mathbf{0} \end{pmatrix}, \text{ and } \mathbf{D}_\eta = \begin{pmatrix} \mathbf{F}_\eta & \mathbf{C}_F^{NR^T} \\ \mathbf{C}_F^{NR} & \mathbf{0} \end{pmatrix},$$

we have the following formula for \mathbf{P}_η^{-1} :

$$\mathbf{P}_\eta^{-1} = \begin{bmatrix} \mathbf{A}_\eta^{-1} + \mathbf{A}_\eta^{-1}\mathbf{B}(\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{A}_\eta^{-1} & -\mathbf{A}_\eta^{-1}\mathbf{B}(\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{A}_\eta^{-1} & (\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1} \end{bmatrix}. \quad (\text{A.13})$$

First, we want to determine the vector of weights $\mathbf{w}^*(\vec{e}^N)$, where $\vec{e}^N \in \mathcal{E}^N$. We focus on the first $|\mathcal{E}|-|\mathcal{E}'|$ rows in \mathbf{P}_η^{-1} , i.e.

$$\begin{aligned} \mathbf{w}^*(\vec{e}^N) &= \lim_{\eta \rightarrow \infty} \left[\mathbf{A}_\eta^{-1} + \mathbf{A}_\eta^{-1}\mathbf{B}(\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{A}_\eta^{-1} \right]_{\text{first } |\mathcal{E}|-|\mathcal{E}'| \text{ rows}} \begin{bmatrix} \mathbf{0}_{|\mathcal{E}|-|\mathcal{E}'|\times 1} \\ \mathbf{0}_{|\Theta^N|\times 1} \end{bmatrix} \\ &+ \lim_{\eta \rightarrow \infty} \left[-\mathbf{A}_\eta^{-1}\mathbf{B}(\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1} \right]_{\text{first } |\mathcal{E}|-|\mathcal{E}'| \text{ rows}} \begin{bmatrix} \mathbf{0}_{|\mathcal{E}'|\times 1} \\ 1 \\ \mathbf{0}_{|\Theta^{NR}|-1\times 1} \end{bmatrix} \end{aligned} \quad (\text{A.14})$$

Since the first summand is multiplied by $\mathbf{0}_{|\mathcal{E}|-|\mathcal{E}'|+|\Theta^N|\times 1}$, we focus on the first $|\mathcal{E}|-|\mathcal{E}'|$ rows of the limit of $-\mathbf{A}_\eta^{-1}\mathbf{B}(\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1}$. Note that \mathbf{A}_η is also a block matrix, so \mathbf{A}_η^{-1} is equal to:

$$\mathbf{A}_\eta^{-1} = \begin{bmatrix} \mathbf{N}_\eta^{-1} + \mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT}(-\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT})^{-1}\mathbf{C}_N^N\mathbf{N}_\eta^{-1} & -\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT}(-\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT})^{-1} \\ -(-\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT})^{-1}\mathbf{C}_N^N\mathbf{N}_\eta^{-1} & (-\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT})^{-1} \end{bmatrix}. \quad (\text{A.15})$$

For now, we write $\mathbf{A}_\eta^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then, we have

$$\begin{aligned} -\mathbf{A}_\eta^{-1}\mathbf{B}(\mathbf{D}_\eta - \mathbf{B}^T\mathbf{A}_\eta^{-1}\mathbf{B})^{-1} &= -\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{C}_N^{NR^T} \\ \mathbf{C}_F^N & \mathbf{0} \end{bmatrix} \\ &\left(\left[\begin{bmatrix} \mathbf{F}_\eta & \mathbf{C}_F^{NR^T} \\ \mathbf{C}_F^{NR} & \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{C}_F^{NT} \\ \mathbf{C}_N^{NR} & \mathbf{0} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{C}_N^{NR^T} \\ \mathbf{C}_F^N & \mathbf{0} \end{bmatrix} \right)^{-1} \end{aligned}$$

$$\begin{aligned}
&= - \begin{bmatrix} b\mathbf{C}_F^N & a\mathbf{C}_N^{NR^T} \\ d\mathbf{C}_F^N & c\mathbf{C}_N^{NR^T} \end{bmatrix} \\
&\left(\begin{bmatrix} \mathbf{F}_\eta & \mathbf{C}_F^{NR^T} \\ \mathbf{C}_F^{NR} & \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{C}_F^{NT}c & \mathbf{C}_F^{NT}d \\ \mathbf{C}_N^{NR}a & \mathbf{C}_N^{NR}b \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{C}_N^{NR^T} \\ \mathbf{C}_F^N & \mathbf{0} \end{bmatrix} \right)^{-1} \\
&= - \begin{bmatrix} b\mathbf{C}_F^N & a\mathbf{C}_N^{NR^T} \\ d\mathbf{C}_F^N & c\mathbf{C}_N^{NR^T} \end{bmatrix} \\
&\times \begin{bmatrix} \mathbf{F}_\eta - \mathbf{C}_F^{NT}d\mathbf{C}_F^N & \mathbf{C}_F^{NR^T} - \mathbf{C}_F^{NT}c\mathbf{C}_N^{NR^T} \\ \mathbf{C}_F^{NR} - \mathbf{C}_N^{NR}b\mathbf{C}_F^N & -\mathbf{C}_N^{NR}a\mathbf{C}_N^{NR^T} \end{bmatrix}^{-1} \\
&= - \begin{bmatrix} \mathbf{0} & a\mathbf{C}_N^{NR^T} \\ \mathbf{0} & c\mathbf{C}_N^{NR^T} \end{bmatrix} \underbrace{\begin{bmatrix} \mathbf{F}_\eta & \mathbf{C}_F^{NR^T} \\ \mathbf{C}_F^{NR} & -\mathbf{C}_N^{NR}a\mathbf{C}_N^{NR^T} \end{bmatrix}^{-1}}_{= \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}} \\
&= - \begin{bmatrix} a\mathbf{C}_N^{NR^T}c' & a\mathbf{C}_N^{NR^T}d' \\ c\mathbf{C}_N^{NR^T}c' & c\mathbf{C}_N^{NR^T}d' \end{bmatrix}, \tag{A.16}
\end{aligned}$$

where $\mathbf{C}_F^N = \mathbf{0}$ since by definition of Θ^N , for all $\theta^N \in \Theta^N$, we have $\theta^N \notin \vec{e}$ for $\vec{e} \in \mathcal{E}'$. We are interested in the first $|\mathcal{E}| - |\mathcal{E}'|$ rows, but since the first $|\mathcal{E}'|$ columns are multiplied by 0, we focus on the last $|\Theta^{NR}|$ columns:

$$\begin{aligned}
-a\mathbf{C}_F^{NR^T}d' &= \mathbf{N}_\eta^{-1}\mathbf{C}_N^{NR^T} \left(\mathbf{C}_N^{NR}\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NR^T} \right. \\
&\quad - \mathbf{C}_N^{NR}\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT}(\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT})^{-1}\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NR^T} + \mathbf{C}_F^{NR}\mathbf{F}_\eta^{-1}\mathbf{C}_F^{NR^T} \Big)^{-1} \\
&\quad - \mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT}(\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT})^{-1}\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NR^T} \left(\mathbf{C}_N^{NR}\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NR^T} \right. \\
&\quad \left. - \mathbf{C}_N^{NR}\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT}(\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NT})^{-1}\mathbf{C}_N^N\mathbf{N}_\eta^{-1}\mathbf{C}_N^{NR^T} + \mathbf{C}_F^{NR}\mathbf{F}_\eta^{-1}\mathbf{C}_F^{NR^T} \right)^{-1}
\end{aligned} \tag{A.17}$$

Note that \mathbf{F}_η is full rank and by definition, \mathbf{C}_F^{NR} is also full row rank. The rank of the product $\mathbf{C}_F^{NR}\mathbf{F}_\eta^{-1}\mathbf{C}_F^{NR^T}$ is equal to $\min(\text{rank}(\mathbf{C}_F^{NR}), \text{rank}(\mathbf{F}_\eta))$. The rank of \mathbf{F}_η is $|\mathcal{E}'|$ and the rank of \mathbf{C}_F^{NR} is $|\Theta^{NR}|$. If $|\mathcal{E}'| < |\Theta^{NR}|$, then there are parameters such that they only appear in the same exposures, leading to linearly dependent constraints in \mathbf{C}_F^{NR} . This contradicts the definition of \mathbf{C}_F^{NR} , so $|\mathcal{E}'| \geq |\Theta^{NR}|$. Hence, $\text{rank}(\mathbf{C}_F^{NR}\mathbf{F}_\eta^{-1}\mathbf{C}_F^{NR^T}) = |\Theta^{NR}|$. So $\mathbf{C}_F^{NR}\mathbf{F}_\eta^{-1}\mathbf{C}_F^{NR^T}$ is full rank.

By the continuity of matrix inverse at full-rank matrices, we can exchange the limit and the inverse in Equation (A.17). Note that we can write the j th diagonal entry of \mathbf{N}_η as $p(\vec{e}_j) \left(\eta a_{\vec{e}_j} + \vec{v}_{\vec{e}_j}^T B \vec{v}_{\vec{e}_j} \right)$. Let $\tilde{\mathbf{N}}_\eta$ be the matrix with diagonal entries

$p(\vec{e}_j) \left(a_{\vec{e}_j} + \frac{1}{\eta} \vec{v}_{\vec{e}_j}^T B \vec{v}_{\vec{e}_j} \right)$ so that $\mathbf{N}_\eta = \eta \tilde{\mathbf{N}}_\eta$, i.e. $\mathbf{N}_\eta^{-1} = \frac{1}{\eta} \tilde{\mathbf{N}}_\eta^{-1}$. As $\lim_{\eta \rightarrow \infty} \left(\tilde{\mathbf{N}}_\eta^{-1} \right)_{j,j} = p(\vec{e}_j)^{-1} a_{\vec{e}_j}^{-1} < \infty$,

$$\begin{aligned}
& \lim_{\eta \rightarrow \infty} \left(\mathbf{C}_N^{NR} \mathbf{N}_\eta^{-1} \mathbf{C}_N^{NR^T} - \mathbf{C}_N^{NR} \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{NR^T} (\mathbf{C}_N^N \mathbf{N}_\eta^{-1} \mathbf{C}_N^{N^T})^{-1} \mathbf{C}_N^N \mathbf{N}_\eta^{-1} \mathbf{C}_N^{NR^T} \right) \\
&= \lim_{\eta \rightarrow \infty} \mathbf{C}_N^{NR} \frac{1}{\eta} \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{NR^T} \\
&- \lim_{\eta \rightarrow \infty} \mathbf{C}_N^{NR} \frac{1}{\eta} \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{NR^T} (\mathbf{C}_N^N \frac{1}{\eta} \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{N^T})^{-1} \mathbf{C}_N^N \frac{1}{\eta} \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{NR^T} \\
&= \mathbf{0}_{|\Theta^{NR}| \times |\Theta^{NR}|}.
\end{aligned} \tag{A.18}$$

Hence,

$$\begin{aligned}
\lim_{\eta \rightarrow \infty} a \mathbf{C}_N^{NR^T} d' &= \lim_{\eta \rightarrow \infty} -\frac{1}{\eta} \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{NR^T} \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T} \right)^{-1} \\
&+ \lim_{\eta \rightarrow \infty} \frac{1}{\eta} \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{NR^T} (\mathbf{C}_N^N \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{N^T})^{-1} \mathbf{C}_N^N \tilde{\mathbf{N}}_\eta^{-1} \mathbf{C}_N^{NR^T} \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T} \right)^{-1}.
\end{aligned} \tag{A.19}$$

Since each matrix above is bounded in η , $\lim_{\eta \rightarrow \infty} a \mathbf{C}_N^{NR^T} d' = \mathbf{0}_{|\mathcal{E}| - |\mathcal{E}'| \times |\Theta^{NR}|}$. Hence, $\mathbf{w}^*(\vec{e}^N) = \mathbf{0}_{|\mathcal{E}| - |\mathcal{E}'| \times 1}$. This establishes that $\text{supp}(\mathbf{w}^*) \subset \mathcal{E}'$ if $\Theta^R = \emptyset$.

Case 2 (Assume $\Theta^R \neq \emptyset$): Now we consider the case when $\Theta^R \neq \emptyset$, i.e. $\mathbf{C}_F^R \neq \mathbf{0}$. The matrix equation becomes

$$\begin{pmatrix} & & & & \begin{pmatrix} \mathbf{C}_N^{R^T} \\ \mathbf{0}_{|\Theta^N| \times |\Theta^R|} \\ \mathbf{C}_F^{R^T} \\ \mathbf{0}_{|\Theta^{NR}| \times |\Theta^R|} \\ \mathbf{0}_{|\Theta^R| \times |\Theta^R|} \end{pmatrix} \\ & \mathbf{P}_\eta & & & \\ \begin{pmatrix} \mathbf{C}_N^R & \mathbf{0}_{|\Theta^R| \times |\Theta^N|} & \mathbf{C}_F^R & \mathbf{0}_{|\Theta^R| \times |\Theta^{NR}|} \end{pmatrix} & & & & \end{pmatrix} \begin{pmatrix} \mathbf{w}_\eta \\ \lambda^R_\eta \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0}_{|\Theta^R| \times 1} \end{pmatrix}, \tag{A.20}$$

where \mathbf{P}_η , \mathbf{w}_η , and \mathbf{b} are matrices and vectors from Equation (A.12). Denote Equation (A.20) as $\tilde{\mathbf{P}}_\eta \tilde{\mathbf{w}}_\eta = \tilde{\mathbf{b}}$. Since we already showed that \mathbf{w}^* is the solution to the

matrix equation in the limit when $\Theta^R = \emptyset$, we have the following:

$$\tilde{\mathbf{P}}^* \begin{pmatrix} \mathbf{w}^* \\ \mathbf{0}_{|\Theta^R| \times 1} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{|\mathcal{E} \setminus \mathcal{E}'| \times 1} \\ \mathbf{0}_{|\Theta^N| \times 1} \\ \mathbf{0}_{|\mathcal{E}'| \times 1} \\ 1 \\ \mathbf{0}_{|\Theta^{NR}| - 1 \times 1} \\ \mathbf{C}_N^R \mathbf{w}^*(\vec{e}^N) + \mathbf{C}_F^R \mathbf{w}^*(\vec{e}^F) \end{pmatrix}, \quad (\text{A.21})$$

where $\mathbf{w}^*(\vec{e}^N)$ and $\mathbf{w}^*(\vec{e}^F)$ are the weights of the exposures in $\mathcal{E} \setminus \mathcal{E}'$ and \mathcal{E}' , respectively as given by \mathbf{w}^* and $\tilde{\mathbf{P}}^* = \lim_{\eta \rightarrow \infty} \tilde{\mathbf{P}}_\eta$. Recall that in the previous case, we showed that $\mathbf{w}^*(\vec{e}^N) = \mathbf{0}_{|\mathcal{E} \setminus \mathcal{E}'| \times 1}$. Furthermore, recall that by construction of \mathbf{C}_F^{NR} , $\mathbf{C}_F^R = \mathbf{T} \mathbf{C}_F^{NR}$, where the first column of \mathbf{T} only contains zeros since θ_{1,m_1} cannot be linearly dependent with another parameter. Otherwise, unbiasedness does not hold. Since $\mathbf{w}^*(\vec{e}^F)$ solves the matrix equation given by Equation (A.12), then $\mathbf{C}_F^R \mathbf{w}^*(\vec{e}^F) = \mathbf{0}_{|\Theta^R| \times 1}$. Then, $\tilde{\mathbf{P}}^* \tilde{\mathbf{w}}^* = \tilde{\mathbf{b}}$, where $\tilde{\mathbf{w}}^* = \begin{pmatrix} \mathbf{w}^* \\ \mathbf{0}_{|\Theta^R| \times 1} \end{pmatrix}$.

Let \mathbf{w} be the true solution in the limit to the problem $\tilde{\mathbf{P}}_\eta \tilde{\mathbf{w}}_\eta = \tilde{\mathbf{b}}$ as given by Equation (A.20). Then:

$$\tilde{\mathbf{P}}^* \mathbf{w} - \tilde{\mathbf{P}}^* \tilde{\mathbf{w}}^* = \tilde{\mathbf{P}}^* (\mathbf{w} - \tilde{\mathbf{w}}^*) = \mathbf{0}_{|\mathcal{E}| + |\Theta|}. \quad (\text{A.22})$$

By Lemma A.1, $\tilde{\mathbf{P}}^*$ is full rank, where now $\mathbf{C} = \begin{pmatrix} \mathbf{C}_N^N & \mathbf{C}_F^N \\ \mathbf{C}_N^R & \mathbf{C}_F^R \\ \mathbf{C}_N^{NR} & \mathbf{C}_F^{NR} \end{pmatrix}$. We can then multiply both sides by $\tilde{\mathbf{P}}^{*-1}$, and since all elements in $\tilde{\mathbf{P}}^{*-1}$ are finite, we have:

$$\mathbf{w} - \tilde{\mathbf{w}}^* = \mathbf{0}_{|\mathcal{E}| + |\Theta|}. \quad (\text{A.23})$$

Thus, in the limit, the solutions $\tilde{\mathbf{w}}^*$ and \mathbf{w} are the same, and we see that $\tilde{\mathbf{w}}^*(\vec{e}^N) = \mathbf{0}_{|\mathcal{E} \setminus \mathcal{E}'|}$ and $\tilde{\mathbf{w}}^*(\vec{e}^F)$ depends on $\mathbf{w}^*(\vec{e}^F)$.

Hence, if there exists $\mathcal{E}' \subseteq \mathcal{E}$ such that $\text{span}(\{\vec{v}_{\mathcal{E}'}\}_{\mathcal{E}' \in \mathcal{E}'} \cap \{\vec{v}_{\mathcal{E}}\}_{\mathcal{E} \in \mathcal{E}} = \{\vec{v}_{\mathcal{E}'}\}_{\mathcal{E}' \in \mathcal{E}'}$, then there exists a $\hat{\theta}$ with $\text{supp}(\hat{\theta}) \subseteq \mathcal{E}'$ and $\hat{\theta}$ is a limit of MIV LUEs.

Showing $\text{supp}(\mathbf{w}^*) = \mathcal{E}'$:

Finally, we want to determine the vector of weights $\mathbf{w}^*(\vec{e}^F)$, where $\vec{e}^F \in \mathcal{E}'$, which is given by:

$$\mathbf{w}^*(\vec{e}^F) = \lim_{\eta \rightarrow \infty} \left[(\mathbf{D}_\eta - \mathbf{B}^T \mathbf{A}_\eta^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{A}_\eta^{-1} \right]_{\text{first } |\mathcal{E}'| \text{ rows}} \begin{bmatrix} \mathbf{0}_{|\mathcal{E} \setminus \mathcal{E}'| \times 1} \\ \mathbf{0}_{|\Theta^N| \times 1} \end{bmatrix}$$

$$+ \lim_{\eta \rightarrow \infty} (\mathbf{D}_\eta - \mathbf{B}^T \mathbf{A}_\eta^{-1} \mathbf{B})_{\text{first } |\mathcal{E}'| \text{ rows}}^{-1} \begin{bmatrix} \mathbf{0}_{|\mathcal{E}'| \times 1} \\ 1 \\ \mathbf{0}_{|\Theta^{NR}|-1 \times 1} \end{bmatrix}. \quad (\text{A.24})$$

We focus on when $\Theta^R = \emptyset$ since we have shown that the weights for \vec{e}^F when $\Theta^R \neq \emptyset$ are the same in the limit as the weights when $\Theta^R = \emptyset$. Here, we focus on the first $|\mathcal{E}'|$ rows and last $|\Theta^{NR}|$ columns of $\lim_{\eta \rightarrow \infty} (\mathbf{D}_\eta - \mathbf{B}^T \mathbf{A}_\eta^{-1} \mathbf{B})^{-1}$. Note that $\mathbf{D}_\eta - \mathbf{B}^T \mathbf{A}_\eta^{-1} \mathbf{B}$ is a block matrix, so we are interested in the upper right block of the inverse. Again, we denote $\mathbf{A}_\eta^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then, using the right hand side of Equation (A.13), the upper right block of $(\mathbf{D}_\eta - \mathbf{B}^T \mathbf{A}_\eta^{-1} \mathbf{B})^{-1}$ is:

$$\begin{aligned} (\mathbf{D}_\eta - \mathbf{B}^T \mathbf{A}_\eta^{-1} \mathbf{B})_{\text{upper right block}}^{-1} &= \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \left(\mathbf{C}_N^{NR} a \mathbf{C}_N^{NR T} + \mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \right)^{-1} \\ &= \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \left(\mathbf{C}_N^{NR} \mathbf{N}_\eta^{-1} \mathbf{C}_N^{NR T} \right. \\ &\quad \left. - \mathbf{C}_N^{NR} \mathbf{N}_\eta^{-1} \mathbf{C}_N^{N T} (\mathbf{C}_N^N \mathbf{N}_\eta^{-1} \mathbf{C}_N^{N T})^{-1} \mathbf{C}_N^N \mathbf{N}_\eta^{-1} \mathbf{C}_N^{NR T} \right. \\ &\quad \left. + \mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \right)^{-1}. \end{aligned} \quad (\text{A.25})$$

Since $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T}$ is full rank as shown previously, we can take the limit inside the inverse. From Equation (A.18), we have

$$\lim_{\eta \rightarrow \infty} \left(\mathbf{C}_N^{NR} \mathbf{N}_\eta^{-1} \mathbf{C}_N^{NR T} - \mathbf{C}_N^{NR} \mathbf{N}_\eta^{-1} \mathbf{C}_N^{N T} (\mathbf{C}_N^N \mathbf{N}_\eta^{-1} \mathbf{C}_N^{N T})^{-1} \mathbf{C}_N^N \mathbf{N}_\eta^{-1} \mathbf{C}_N^{NR T} \right) = \mathbf{0}_{|\Theta^{NR}| \times |\Theta^{NR}|}.$$

Hence,

$$\mathbf{w}^*(\vec{e}^F) = \lim_{\eta \rightarrow \infty} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \right)^{-1} \begin{bmatrix} 1 \\ \mathbf{0}_{|\Theta^{NR}|-1 \times 1} \end{bmatrix}. \quad (\text{A.26})$$

Note that the (k, l) th entry of $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T}$ is given by:

$$\left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \right)_{k,l} = \sum_{j=1}^{|\mathcal{E}'|} \mathbb{I}\{\theta_k, \theta_l \in \vec{e}_j\} \frac{p(\vec{e}_j)}{\text{Var}(Y(\vec{e}_j))}, \quad (\text{A.27})$$

where the $k, l \in \{1, \dots, |\Theta^{NR}|\}$ indexes the different parameters in Θ^{NR} . If for every exposure $\vec{e}_j \in \mathcal{E}'$, we have

$$\lim_{\eta \rightarrow \infty} \sum_{k=1}^{|\Theta^{NR}|} \text{Adj} \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \right)_{k,1} \mathbb{I}\{\theta_k \in \vec{e}_j\} \neq 0, \quad (\text{A.28})$$

where $Adj \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \right)_{k,1}$ is the $(k,1)$ th entry of the adjugate matrix of $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T}$ corresponding of θ_{1,m_1} , then $\mathbf{w}^*(\vec{e}) \neq 0$. Hence, $\text{supp}(\mathbf{w}^*) = \mathcal{E}'$. \square

A.3.3 Example: Derivation of Weights for Six-Term Exposure

We show that in general $\mathbf{w}^*(\vec{e}^F) \neq 0$ through an example. Consider

$$\mathcal{E}^{\text{six term},m} = \{(0,0), (0,j), (m,0), (m,j), (m_1,0), (m_1,j)\},$$

where $m \in \{1, \dots, m_1 - 1\}, j \in \{1, \dots, m_2\}$ and consider a prior covariance-matrix Σ , where all prior variances of parameters are finite. Denote entries of the inverse of $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T}$ as $\tilde{a}_{k,j} = \frac{1}{\det(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T})} Adj \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR T} \right)_{k,j}$ where Adj is the adjugate. Then,

$$\begin{aligned} \mathbf{w}^*(\vec{e}^F) &= \begin{bmatrix} \frac{1}{p(0,0)\text{Var}(Y(0,0))} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{p(0,j)\text{Var}(Y(0,j))} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{p(m,0)\text{Var}(Y(m,0))} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{p(m,j)\text{Var}(Y(m,j))} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{p(m_1,0)\text{Var}(Y(m_1,0))} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{p(m_1,j)\text{Var}(Y(m_1,j))} \end{bmatrix} \\ &\times \begin{bmatrix} 0 & p(0,0) & 0 & 0 \\ 0 & p(0,j) & 0 & p(0,j) \\ 0 & p(m,0) & p(m,0) & 0 \\ 0 & p(m,j) & p(m,j) & p(m,j) \\ p(m_1,0) & p(m_1,0) & 0 & 0 \\ p(m_1,j) & p(m_1,j) & 0 & p(m_1,j) \end{bmatrix} \begin{bmatrix} \tilde{a}_{1,1} & \tilde{a}_{1,2} & \tilde{a}_{1,3} & \tilde{a}_{1,4} \\ \tilde{a}_{2,1} & \tilde{a}_{2,2} & \tilde{a}_{2,3} & \tilde{a}_{2,4} \\ \tilde{a}_{3,1} & \tilde{a}_{3,2} & \tilde{a}_{3,3} & \tilde{a}_{3,4} \\ \tilde{a}_{4,1} & \tilde{a}_{4,2} & \tilde{a}_{4,3} & \tilde{a}_{4,4} \end{bmatrix} \\ &= \begin{bmatrix} 0 & \frac{1}{\text{Var}(Y(0,0))} & 0 & 0 \\ 0 & \frac{1}{\text{Var}(Y(0,j))} & 0 & \frac{1}{\text{Var}(Y(0,j))} \\ 0 & \frac{1}{\text{Var}(Y(m,0))} & \frac{1}{\text{Var}(Y(m,0))} & 0 \\ 0 & \frac{1}{\text{Var}(Y(m,j))} & \frac{1}{\text{Var}(Y(m,j))} & \frac{1}{\text{Var}(Y(m,j))} \\ \frac{1}{\text{Var}(Y(m_1,0))} & \frac{1}{\text{Var}(Y(m_1,0))} & 0 & 0 \\ \frac{1}{\text{Var}(Y(m_1,j))} & \frac{1}{\text{Var}(Y(m_1,j))} & 0 & \frac{1}{\text{Var}(Y(m_1,j))} \end{bmatrix} \begin{bmatrix} \tilde{a}_{1,1} & \tilde{a}_{1,2} & \tilde{a}_{1,3} & \tilde{a}_{1,4} \\ \tilde{a}_{2,1} & \tilde{a}_{2,2} & \tilde{a}_{2,3} & \tilde{a}_{2,4} \\ \tilde{a}_{3,1} & \tilde{a}_{3,2} & \tilde{a}_{3,3} & \tilde{a}_{3,4} \\ \tilde{a}_{4,1} & \tilde{a}_{4,2} & \tilde{a}_{4,3} & \tilde{a}_{4,4} \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \frac{\tilde{a}_{2,1}}{\text{Var}(Y(0,0))} & \frac{\tilde{a}_{2,2}}{\text{Var}(Y(0,0))} & \frac{\tilde{a}_{2,3}}{\text{Var}(Y(0,0))} & \frac{\tilde{a}_{2,4}}{\text{Var}(Y(0,0))} \\ \frac{\tilde{a}_{2,1} + \tilde{a}_{4,1}}{\text{Var}(Y(0,j))} & \frac{\tilde{a}_{2,2} + \tilde{a}_{4,2}}{\text{Var}(Y(0,j))} & \frac{\tilde{a}_{2,3} + \tilde{a}_{4,3}}{\text{Var}(Y(0,j))} & \frac{\tilde{a}_{2,4} + \tilde{a}_{4,4}}{\text{Var}(Y(0,j))} \\ \frac{\tilde{a}_{2,1} + \tilde{a}_{3,1}}{\text{Var}(Y(m,0))} & \frac{\tilde{a}_{2,2} + \tilde{a}_{3,2}}{\text{Var}(Y(m,0))} & \frac{\tilde{a}_{2,3} + \tilde{a}_{3,3}}{\text{Var}(Y(m,0))} & \frac{\tilde{a}_{2,4} + \tilde{a}_{3,4}}{\text{Var}(Y(m,0))} \\ \frac{\tilde{a}_{2,1} + \tilde{a}_{3,1} + \tilde{a}_{4,1}}{\text{Var}(Y(m,j))} & \frac{\tilde{a}_{2,2} + \tilde{a}_{3,2} + \tilde{a}_{4,2}}{\text{Var}(Y(m,j))} & \frac{\tilde{a}_{2,3} + \tilde{a}_{3,3} + \tilde{a}_{4,3}}{\text{Var}(Y(m,j))} & \frac{\tilde{a}_{2,4} + \tilde{a}_{3,4} + \tilde{a}_{4,4}}{\text{Var}(Y(m,j))} \\ \frac{\tilde{a}_{1,1} + \tilde{a}_{2,1}}{\text{Var}(Y(m_1,0))} & \frac{\tilde{a}_{1,2} + \tilde{a}_{2,2}}{\text{Var}(Y(m_1,0))} & \frac{\tilde{a}_{1,3} + \tilde{a}_{2,3}}{\text{Var}(Y(m_1,0))} & \frac{\tilde{a}_{1,4} + \tilde{a}_{2,4}}{\text{Var}(Y(m_1,0))} \\ \frac{\tilde{a}_{1,1} + \tilde{a}_{2,1} + \tilde{a}_{4,1}}{\text{Var}(Y(m_1,j))} & \frac{\tilde{a}_{1,2} + \tilde{a}_{2,2} + \tilde{a}_{4,2}}{\text{Var}(Y(m_1,j))} & \frac{\tilde{a}_{1,3} + \tilde{a}_{2,3} + \tilde{a}_{4,3}}{\text{Var}(Y(m_1,j))} & \frac{\tilde{a}_{1,4} + \tilde{a}_{2,4} + \tilde{a}_{4,4}}{\text{Var}(Y(m_1,j))} \end{bmatrix}. \quad (\text{A.29})$$

The weights $\mathbf{w}^*(\vec{e}^F)$ are given by the entries in the first column, and so $\mathbf{w}^*(\vec{e}^F)$ is non-zero if the corresponding entries of the inverse of $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T}$ are non-zero. We focus on $\mathbf{w}^*(0,0) = \frac{\tilde{a}_{2,1}}{\text{Var}(Y(0,0))}$. Here $\tilde{a}_{2,1} = \frac{1}{\det(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T})} \text{Adj} \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T} \right)_{2,1}$. Since $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T}$ is full rank, the determinant is non-zero, and so we focus on the adjugate term in terms of the minor, denoted by $\mathbf{M}_{i,j}$:

$$\begin{aligned} \text{Adj} \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T} \right)_{2,1} &= -\mathbf{M}_{2,1} \\ &= - \begin{vmatrix} \frac{p(m_1,0)}{\text{Var}(Y(m_1,0))} + \frac{p(m_1,j)}{\text{Var}(Y(m_1,j))} & 0 & \frac{p(m_1,j)}{\text{Var}(Y(m_1,j))} \\ \frac{p(m,0)}{\text{Var}(Y(m,0))} + \frac{p(m,j)}{\text{Var}(Y(m,j))} & \frac{p(m,0)}{\text{Var}(Y(m,0))} + \frac{p(m,j)}{\text{Var}(Y(m,j))} & \frac{p(m,j)}{\text{Var}(Y(m,j))} \\ \frac{p(0,j)}{\text{Var}(Y(0,j))} + \frac{p(m_1,j)}{\text{Var}(Y(m_1,j))} + \frac{p(m,j)}{\text{Var}(Y(m,j))} & \frac{p(m,j)}{\text{Var}(Y(m,j))} & \frac{p(0,j)}{\text{Var}(Y(0,j))} + \frac{p(m_1,j)}{\text{Var}(Y(m_1,j))} + \frac{p(m,j)}{\text{Var}(Y(m,j))} \end{vmatrix} \\ &= - \left(\frac{p(m_1,0)p(m,0)p(0,j)}{\text{Var}(Y(m_1,0))\text{Var}(Y(m,0))\text{Var}(Y(0,j))} + \frac{p(m_1,0)p(m,0)p(m,j)}{\text{Var}(Y(m_1,0))\text{Var}(Y(m,0))\text{Var}(Y(m,j))} \right. \\ &\quad + \frac{p(m_1,0)p(m,0)p(m_1,j)}{\text{Var}(Y(m_1,0))\text{Var}(Y(m,0))\text{Var}(Y(m_1,j))} + \frac{p(m_1,0)p(m,j)p(0,j)}{\text{Var}(Y(m_1,0))\text{Var}(Y(m,j))\text{Var}(Y(0,j))} \\ &\quad \left. + \frac{p(m_1,0)p(m,j)p(m_1,j)}{\text{Var}(Y(m_1,0))\text{Var}(Y(m,j))\text{Var}(Y(m_1,j))} + \frac{p(m_1,j)p(m,0)p(m,j)}{\text{Var}(Y(m_1,j))\text{Var}(Y(m,0))\text{Var}(Y(m,j))} \right). \end{aligned} \quad (\text{A.30})$$

Thus, we would need to set at least two probabilities of exposures to be zero in order for $\mathbf{w}^*(0,0) = 0$. This holds similarly for other parameters. Hence, for typical choices of the design probabilities and for priors where all variances are finite, $w(\vec{e}^F) \neq 0$, i.e. $\text{supp}(\hat{\theta}) = \mathcal{E}^{\text{six term},m}$.

A.4 Example: Six-Term Exposure Set

Proof of Corollary 3. Consider the exposure set

$$\mathcal{E}^{\text{six term},m} = \{(0,0), (0,j), (m_1,0), (m_1,j), (m,0), (m,j)\},$$

where $j \in \{1, \dots, m_2\}$ and $m \in \{1, \dots, m-1\}$. Note that $\{\vec{v}_{\bar{e}'}\}_{\bar{e}' \in \mathcal{E}^{\text{six term},m}} = \text{span}(\{\vec{v}_{\bar{e}'}\}_{\bar{e}' \in \mathcal{E}^{\text{six term},m}}) \cap \{\vec{v}_{\bar{e}}\}_{\bar{e} \in \mathcal{E}}$. By Theorem 2, there exists an estimator $\hat{\theta}$ such that it is a MIV LUE, for a given prior variance-covariance matrix, and $\text{supp}(\hat{\theta}) = \mathcal{E}^{\text{six term},m}$. Since $\hat{\theta}$ is a LUE, there are weights $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ such that

$$\begin{aligned} \hat{\theta} &= \alpha_1 (HT_{(m_1,0)} - HT_{(0,0)}) + \alpha_2 (HT_{(m_1,j)} - HT_{(0,j)}) \\ &\quad + \alpha_3 (HT_{(m_1,j)} - HT_{(m,j)} + HT_{(m,0)} + HT_{(0,0)}), \end{aligned} \quad (\text{A.31})$$

where the three ALUEs form a basis for six-term estimators. We know that α_1 and α_2 can equal to 1 since the two two-term estimators are also MIV LUEs, but $\alpha_3 \neq 0$ because the four-term estimator is not a MIV LUE. However, exposures in the support for the four-term estimator can still contribute to MIV LUEs. We investigate this contribution by finding the maximum of the weight α_3 .

First, we want to solve for the weights of the exposures in $\mathcal{E}^{\text{six term},m}$. From the proof of Theorem 2, we know that the weights are given by

$$w(0,0) = \frac{\tilde{a}_{2,1}}{\text{Var}(Y(0,0))} \quad (\text{A.32})$$

$$w(0,j) = \frac{\tilde{a}_{2,1} + \tilde{a}_{4,1}}{\text{Var}(Y(0,j))} \quad (\text{A.33})$$

$$w(m,0) = \frac{\tilde{a}_{2,1} + \tilde{a}_{3,1}}{\text{Var}(Y(m,0))} \quad (\text{A.34})$$

$$w(m,j) = \frac{\tilde{a}_{2,1} + \tilde{a}_{3,1} + \tilde{a}_{4,1}}{\text{Var}(Y(m,j))} \quad (\text{A.35})$$

$$w(m_1,0) = \frac{\tilde{a}_{1,1} + \tilde{a}_{2,1}}{\text{Var}(Y(m_1,0))} \quad (\text{A.36})$$

$$w(m_1,j) = \frac{\tilde{a}_{1,1} + \tilde{a}_{2,1} + \tilde{a}_{4,1}}{\text{Var}(Y(m_1,j))}, \quad (\text{A.37})$$

where the terms $\tilde{a}_{i,j}$ are the limit of terms in the adjugate matrix divided by the determinant of $\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NR^T}$ and the potential outcome variances are given by the prior variance-covariance matrix. Suppose the variances of the potential outcomes

are all finite. We first compute the determinant. We write $r(\vec{e}) = \frac{p(\vec{e})}{\text{Var}(Y(\vec{e}))}$:

$$\begin{aligned}
\det \left(\mathbf{C}_F^{NR} \mathbf{F}_\eta^{-1} \mathbf{C}_F^{NRT} \right) &= [r(m_1, 0) + r(m_1, j)] \\
&\times \left\{ \left[r(0, 0) + r(0, j) + r(m_1, 0) + r(m_1, j) + r(m, 0) + r(m, j) \right] \right. \\
&\times \left[r(m, 0) + r(m, j) \right] \left[r(0, j) + r(m, j) + r(m_1, j) \right] \\
&- \left[r(0, j) + r(m, j) + r(m_1, j) \right]^2 \left[r(m, 0) + r(m, j) \right] \\
&+ \left[r(m, 0) + r(m, j) \right] r(m, j) \left[r(0, j) + r(m, j) + r(m_1, j) \right] \\
&- \left[r(m, 0) + r(m, j) \right]^2 \left[r(0, j) + r(m, j) + r(m_1, j) \right] \\
&+ \left[r(0, j) + r(m, j) + r(m_1, j) \right] \left[r(m, 0) + r(m, j) \right] r(m, j) \\
&- \left[r(0, 0) + r(0, j) + r(m_1, 0) \right. \\
&\left. + r(m_1, j) + r(m, 0) + r(m, j) \right] r(m, j)^2 \left. \right\} \\
&- \left\{ \left[r(m_1, 0) + r(m_1, j) \right] \left[r(m, 0) + r(m, j) \right] \right. \\
&\times \left[r(0, j) + r(m, j) + r(m_1, j) \right] \\
&- r(m_1, j) \left[r(m, 0) + r(m, j) \right] \left[r(0, j) + r(m_1, j) + r(m, j) \right] \\
&+ r(m_1, j) r(m, j) \left[r(m, 0) + r(m, j) \right] \\
&\left. - r(m, j)^2 \left[r(m_1, 0) + r(m_1, j) \right] \right\} \\
&- r(m_1, j) \left\{ \left[r(m_1, 0) + r(m_1, j) \right] \left[r(m, 0) + r(m, j) \right] r(m, j) \right. \\
&- r(m_1, j) \left[r(m, 0) + r(m, j) \right]^2 \\
&\left. + r(m_1, j) \left[r(0, 0) + r(0, j) + r(m_1, 0) \right] \right\}
\end{aligned}$$

$$\begin{aligned}
& + r(m_1, j) + r(m, 0) + r(m, j) \Big] \\
& \times \left[r(m, 0) + r(m, j) \right] \\
& - \left[r(m_1, 0) + r(m_1, j) \right] \left[r(m, 0) + r(m, j) \right] \\
& \times \left[r(0, j) + r(m, j) + r(m_1, j) \right] \Big\} \\
& = r(m_1, 0) \left[r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, j)r(m_1, j) \right] \\
& + r(m_1, j) \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \\
& \times \left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0)r(m, 0)r(0, j) \right. \\
& + r(0, 0)r(m, 0)r(m, j) + r(0, 0)r(m, j)r(0, j) \\
& \left. + r(0, j)r(m, 0)r(m, j) \right].
\end{aligned}$$

The different entries of the adjugate matrix that are needed to compute the exposure weights are as follows:

$$\begin{aligned}
\tilde{A}_{1,1} &= r(0, 0)r(m, 0)r(0, j) + r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, 0)r(m, j) \\
&+ r(0, 0)r(m, j)r(0, j) + r(0, 0)r(m, j)r(m_1, j) + r(m_1, 0)r(m, 0)r(0, j) \\
&+ r(m_1, 0)r(m, 0)r(m_1, j) + r(m_1, 0)r(m, 0)r(m, j) + r(m_1, 0)r(m, j)r(0, j) \\
&+ r(m_1, 0)r(m, j)r(m_1, j) + r(0, j)r(m, 0)r(m, j) + r(m_1, j)r(m, 0)r(m, j) \\
\tilde{A}_{2,1} &= - \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, 0)r(m, j) + r(m_1, 0)r(m, 0)r(m_1, j) \right. \\
&\left. + r(m_1, 0)r(m, j)r(0, j) + r(m_1, 0)r(m, j)r(m_1, j) + r(m_1, j)r(m, 0)r(m, j) \right] \\
\tilde{A}_{3,1} &= r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, 0)r(m, j) + r(m_1, 0)r(m, 0)r(m_1, j) \\
&+ r(m_1, j)r(m, j)r(0, 0) + r(m_1, j)r(m, j)r(m_1, 0) + r(m_1, j)r(m, 0)r(m, j) \\
\tilde{A}_{4,1} &= \left[r(m, 0) + r(m, j) \right] \left[r(m_1, 0)r(0, j) - r(m_1, j)r(0, 0) \right].
\end{aligned}$$

Using the adjugate entries and the determinant, the weights $\alpha_1, \alpha_2, \alpha_3$ are then as

follows:

$$\begin{aligned}
\alpha_1 = & r(m_1, 0) \left\{ r(0, 0)r(m, 0)r(0, j) + r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& \left. + r(0, 0)r(m, j)r(0, j) + r(0, 0)r(m, j)r(m_1, j) + r(0, j)r(m, 0)r(m, j) \right\} \\
& \times \left\{ r(m_1, 0) \left[r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, j)r(m_1, j) \right] \right. \\
& \left. + r(m_1, j) \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \right. \\
& \left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0)r(m, 0)r(0, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& \left. \left. + r(0, 0)r(m, j)r(0, j) + r(0, j)r(m, 0)r(m, j) \right] \right\}^{-1} \tag{A.38}
\end{aligned}$$

$$\begin{aligned}
\alpha_2 = & r(0, j) \left\{ r(m, 0)r(m_1, j)r(0, 0) + r(m, j)r(m_1, j)r(0, 0) + r(m_1, 0)r(m, 0)r(m, j) \right. \\
& \left. + r(m_1, 0)r(m, 0)r(m_1, j) + r(m_1, 0)r(m, j)r(m_1, j) + r(m_1, j)r(m, 0)r(m, j) \right\} \\
& \times \left\{ r(m_1, 0) \left[r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, j)r(m_1, j) \right] \right. \\
& \left. + r(m_1, j) \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \right. \\
& \left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0)r(m, 0)r(0, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& \left. \left. + r(0, 0)r(m, j)r(0, j) + r(0, j)r(m, 0)r(m, j) \right] \right\}^{-1} \tag{A.39}
\end{aligned}$$

$$\begin{aligned}
\alpha_3 = & 1 - \alpha_1 - \alpha_2 \\
= & r(m, 0)r(m, j) \left\{ r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) \right\} \\
& \times \left\{ r(m_1, 0) \left[r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, j)r(m_1, j) \right] \right. \\
& \left. + r(m_1, j) \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \right. \\
& \left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0)r(m, 0)r(0, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& \left. \left. + r(0, 0)r(m, j)r(0, j) + r(0, j)r(m, 0)r(m, j) \right] \right\}^{-1}. \tag{A.40}
\end{aligned}$$

We focus on the α_3 weight, and we want to find the maximum of this weight. To have an idea on how α_3 depends on the different parameters, we take the partial derivatives of the weight by the different parameter variances.

Case 1 (Change of α_3 as $\text{Var}(\theta_{1,m}) \rightarrow \infty$). We can rewrite α_3 as follows:

$$\begin{aligned} \alpha_3 = & \left\{ r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) \right\} \\ & \times \left\{ r(m_1, 0) \left[r(0, 0) \frac{\text{Var}(Y(m, j))}{p(m, j)} r(m_1, j) + r(0, 0) \frac{\text{Var}(Y(m, 0))}{p(m, 0)} r(m_1, j) \right] \right. \\ & + r(m_1, j) \left[r(m_1, 0) \frac{\text{Var}(Y(m, j))}{p(m, j)} r(0, j) + r(m_1, 0) \frac{\text{Var}(Y(m, 0))}{p(m, 0)} r(0, j) \right] \\ & \left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0) \frac{\text{Var}(Y(m, j))}{p(m, j)} r(0, j) + r(0, 0) \right. \\ & \left. \left. + r(0, 0) \frac{\text{Var}(Y(m, 0))}{p(m, 0)} r(0, j) + r(0, j) \right] \right\}^{-1} \end{aligned}$$

Since the $\text{Var}(\theta_{1,m})$ only appears in the denominator, then α_3 decreases and approaches 0 as $\text{Var}(\theta_{1,m}) \rightarrow \infty$.

For the following cases, for simplicity we use *denominator* and *numerator* to denote the denominator and numerator of α_3 , respectively.

Case 2 (Change of α_3 as $\text{Var}(\theta_{1,m_1}) \rightarrow \infty$). The partial derivative of α_3 with respect to $\text{Var}(\theta_{1,m_1})$ is as follows:

$$\begin{aligned} \frac{\partial \alpha_3}{\partial \text{Var}(\theta_{1,m_1})} = & - \frac{r(m, 0)r(m, j) \left[\frac{r(0,0)r(m_1,j)}{\text{Var}(Y(m_1,j))} - \frac{r(0,j)r(m_1,0)}{\text{Var}(Y(m_1,0))} \right]}{\text{denominator}} \\ & + \frac{\text{numerator}}{\text{denominator}^2} \\ & \times \left\{ \frac{r(m_1, 0)}{\text{Var}(Y(m_1, 0))} \left[r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, j)r(m_1, j) \right] \right. \\ & + r(m_1, 0) \left[\frac{r(0, 0)r(m, 0)r(m_1, j)}{\text{Var}(Y(m_1, j))} + \frac{r(0, 0)r(m, j)r(m_1, j)}{\text{Var}(Y(m_1, j))} \right] \\ & + \frac{r(m_1, j)}{\text{Var}(Y(m_1, j))} \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \\ & \left. + r(m_1, j) \left[\frac{r(m_1, 0)r(m, 0)r(0, j)}{\text{Var}(Y(m_1, 0))} + \frac{r(m_1, 0)r(m, j)r(0, j)}{\text{Var}(Y(m_1, 0))} \right] \right\} \end{aligned}$$

$$\begin{aligned}
& + \left[\frac{r(m_1, 0)}{\text{Var}(Y(m_1, 0))} + \frac{r(m_1, j)}{\text{Var}(Y(m_1, j))} \right] \\
& \times \left[r(0, 0)r(m, 0)r(0, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& \left. + r(0, 0)r(m, j)r(0, j) + r(m, 0)r(0, j)r(m, j) \right] \Big\}
\end{aligned}$$

There are two terms in the derivative with different signs. Thus, the sign of the partial derivative depends on which term is larger. Here, the first term is roughly equivalent to the $\frac{\text{numerator}}{\text{denominator}}$, while the second term is slightly larger than $\frac{\text{numerator}}{\text{denominator}}$ since there are additional terms in the numerator. Hence, the second term should roughly be larger in magnitude than the first term, and so the sign of the partial derivative is determined by the second term. The sign of the second term is determined by the sign of the numerator: $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j)$. Hence, if $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) > 0$, then α_3 increases as $\text{Var}(\theta_{1, m_1}) \rightarrow \infty$. If $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) < 0$, then α_3 decreases as $\text{Var}(\theta_{1, m_1}) \rightarrow \infty$.

Case 3 (Change of α_3 As $\text{Var}(\theta_{2, j}) \rightarrow \infty$). The partial derivative of α_3 with respect to $\text{Var}(\theta_{2, j})$ is as follows:

$$\begin{aligned}
\frac{\partial \alpha_3}{\partial \text{Var}(\theta_{2, j})} = & - \left(\frac{\frac{r(m, 0)r(m, j) \left[r(0, 0)r(m_1, j) - r(0, j)r(m_1, 0) \right]}{\text{Var}(Y(m, j))}}{\text{denominator}} \right. \\
& \left. + \frac{r(m, 0)r(m, j) \left[\frac{r(0, 0)r(m_1, j)}{\text{Var}(Y(m_1, j))} - \frac{r(0, j)r(m_1, 0)}{\text{Var}(Y(0, j))} \right]}{\text{denominator}} \right) \\
& + \frac{\text{numerator}}{\text{denominator}^2} \\
& \times \left\{ r(m_1, 0) \left[\frac{r(0, 0)r(m, 0)r(m_1, j)}{\text{Var}(Y(m_1, j))} \right. \right. \\
& \left. \left. + \frac{r(0, 0)r(m, j)r(m_1, j)}{\text{Var}(Y(m, j))\text{Var}(Y(m_1, j))} [\text{Var}(Y(m, j)) + \text{Var}(Y(m_1, j))] \right] \right. \\
& \left. + \frac{r(m_1, j)}{\text{Var}(Y(m_1, j))} \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \right\}
\end{aligned}$$

$$\begin{aligned}
& + r(m_1, j) \left[\frac{r(m_1, 0)r(m, 0)r(0, j)}{\text{Var}(Y(0, j))} \right. \\
& + \left. \frac{r(m_1, 0)r(m, j)r(0, j)}{\text{Var}(Y(m, j))\text{Var}(Y(0, j))} [\text{Var}(Y(m, j)) + \text{Var}(Y(0, j))] \right] \\
& + \left[\frac{r(m_1, j)}{\text{Var}(Y(m_1, j))} \right] \left[r(0, 0)r(m_1, 0)r(0, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& + \left. r(0, 0)r(m, j)r(0, j) + r(m, 0)r(0, j)r(m, j) \right] \\
& + \left[r(m_1, 0) + r(m_1, j) \right] \\
& \times \left[\frac{r(0, 0)r(m, 0)r(0, j)}{\text{Var}(Y(0, j))} + \frac{r(0, 0)r(m, 0)r(m, j)}{\text{Var}(Y(m, j))} \right. \\
& + \frac{r(0, 0)r(m, j)r(0, j)}{\text{Var}(Y(0, j))\text{Var}(Y(m, j))} [\text{Var}(Y(0, j)) + \text{Var}(Y(m, j))] \\
& + \left. \frac{r(m, 0)r(0, j)r(m, j)}{\text{Var}(Y(0, j))\text{Var}(Y(m, j))} [\text{Var}(Y(0, j)) + \text{Var}(Y(m, j))] \right] \Bigg\}.
\end{aligned}$$

Here, the first term is roughly equivalent to the $2 \times \frac{\text{numerator}}{\text{denominator}}$, while the second term is slightly larger than $\frac{\text{numerator}}{\text{denominator}}$ since there are additional terms in the denominator. The actual sign of the partial derivative will depend on the prior variances and probabilities of exposures. However, generally, the first term should be larger in magnitude than the second term, and so the sign of the partial derivative is determined by the first term. The sign of the first term is equal to the negative sign of the numerator, which is determined by: $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j)$. Hence, if $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) > 0$, then α_3 decreases as $\text{Var}(\theta_{2,j}) \rightarrow \infty$. If $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) < 0$, then α_3 increases as $\text{Var}(\theta_{2,j}) \rightarrow \infty$.

Case 4 (Change of α_3 As $\text{Var}(\alpha) \rightarrow \infty$). The partial derivative of α_3 with respect to $\text{Var}(\alpha)$ is as follows:

$$\begin{aligned}
\frac{\partial \alpha_3}{\partial \text{Var}(\alpha)} &= - \frac{1}{\text{denominator}} \left\{ \frac{r(m, 0)r(m, j) [\text{Var}(Y(m, 0)) + \text{Var}(Y(m, j))]}{\text{Var}(Y(m, 0))\text{Var}(Y(m, j))} \right. \\
& \times [r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j)] \\
& + r(m, 0)r(m, j) \left[\frac{r(0, 0)r(m_1, j)}{\text{Var}(Y(0, 0))\text{Var}(Y(m_1, j))} [\text{Var}(Y(0, 0)) + \text{Var}(Y(m_1, j))] \right. \\
& \left. \left. - \frac{r(0, j)r(m_1, 0)}{\text{Var}(Y(m_1, 0))\text{Var}(Y(0, j))} [\text{Var}(Y(m_1, 0)) + \text{Var}(Y(0, j))] \right] \right\}
\end{aligned}$$

$$\begin{aligned}
& + \frac{\text{numerator}}{\text{denominator}^2} \\
& \times \left\{ \frac{r(m_1, 0)}{\text{Var}(Y(m_1, 0))} \left[r(0, 0)r(m, 0)r(m_1, j) + r(0, 0)r(m, j)r(m_1, j) \right] \right. \\
& + r(m_1, 0) \left[\frac{r(0, 0)r(m, 0)r(m_1, j) [\text{Var}(Y(0, 0)) + \text{Var}(Y(m, 0)) + \text{Var}(Y(m_1, j))]}{\text{Var}(Y(0, 0))\text{Var}(Y(m, 0))\text{Var}(Y(m_1, j))} \right. \\
& + \left. \frac{r(0, 0)r(m, j)r(m_1, j) [\text{Var}(Y(0, 0)) + \text{Var}(Y(m, j)) + \text{Var}(Y(m_1, j))]}{\text{Var}(Y(0, 0))\text{Var}(Y(m, j))\text{Var}(Y(m_1, j))} \right] \\
& + \frac{r(m_1, j)}{\text{Var}(Y(m_1, j))} \left[r(m_1, 0)r(m, 0)r(0, j) + r(m_1, 0)r(m, j)r(0, j) \right] \\
& + r(m_1, j) \left[\frac{r(m_1, 0)r(m, 0)r(0, j) [\text{Var}(Y(m_1, 0)) + \text{Var}(Y(m, 0)) + \text{Var}(Y(0, j))]}{\text{Var}(Y(m_1, 0))\text{Var}(Y(m, 0))\text{Var}(Y(0, j))} \right. \\
& + \left. \frac{r(m_1, 0)r(m, j)r(0, j) [\text{Var}(Y(m_1, 0)) + \text{Var}(Y(m, j)) + \text{Var}(Y(0, j))]}{\text{Var}(Y(m_1, 0))\text{Var}(Y(m, j))\text{Var}(Y(0, j))} \right] \\
& + \left[\frac{r(m_1, 0)}{\text{Var}(Y(m_1, 0))} + \frac{r(m_1, j)}{\text{Var}(Y(m_1, j))} \right] \\
& \times \left[r(0, 0)r(m_1, 0)r(0, j) + r(0, 0)r(m, 0)r(m, j) \right. \\
& + \left. r(0, 0)r(m, j)r(0, j) + r(m, 0)r(0, j)r(m, j) \right] \\
& + \left[r(m_1, 0) + r(m_1, j) \right] \\
& \times \left[\frac{r(0, 0)r(m, 0)r(0, j) [\text{Var}(Y(0, 0)) + \text{Var}(Y(m, 0)) + \text{Var}(Y(0, j))]}{\text{Var}(Y(0, 0))\text{Var}(Y(m, 0))\text{Var}(Y(0, j))} \right. \\
& + \frac{r(0, 0)r(m, 0)r(m, j) [\text{Var}(Y(0, 0))\text{Var}(Y(m, 0))\text{Var}(Y(m, j))]}{\text{Var}(Y(0, 0))\text{Var}(Y(m, 0))\text{Var}(Y(m, j))} \\
& + \frac{r(0, 0)r(m, j)r(0, j) [\text{Var}(Y(0, 0)) + \text{Var}(Y(0, j)) + \text{Var}(Y(m, j))]}{\text{Var}(Y(0, 0))\text{Var}(Y(0, j))\text{Var}(Y(m, j))} \\
& + \left. \frac{r(m, 0)r(0, j)r(m, j) [\text{Var}(Y(m, 0)) + \text{Var}(Y(0, j)) + \text{Var}(Y(m, j))]}{\text{Var}(Y(m, 0))\text{Var}(Y(0, j))\text{Var}(Y(m, j))} \right] \left. \right\}.
\end{aligned}$$

Here, the first term is roughly equivalent to the $2 \times \frac{\text{numerator}}{\text{denominator}}$, while the second term is slightly larger than $2 \times \frac{\text{numerator}}{\text{denominator}}$. The second term might generally be larger in magnitude than the first term since there are more terms. The sign of the partial derivative is determined by the second term. The sign of the second term is equal to the sign of the numerator, which is determined by: $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j)$. Hence, if $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) > 0$, then α_3 increases as $\text{Var}(\theta_{2,j}) \rightarrow \infty$. If $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) < 0$, then α_3 decreases as $\text{Var}(\theta_{2,j}) \rightarrow \infty$.

Putting everything together: We can now work to maximize α_3 . We consider the case when $r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) > 0$. We first take $\lim \text{Var}(\theta_{1,m}) \rightarrow 0$:

$$\begin{aligned}
\lim_{\text{Var}(\theta_{1,m}) \rightarrow 0} \alpha_3 &= \lim_{\text{Var}(\theta_{1,m}) \rightarrow 0} \left\{ r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) \right\} \\
&\times \left\{ r(m_1, 0) \left[r(0, 0) \frac{\text{Var}(Y(m, j))}{p(m, j)} r(m_1, j) + r(0, 0) \frac{\text{Var}(Y(m, 0))}{p(m, 0)} r(m_1, j) \right] \right. \\
&+ r(m_1, j) \left[r(m_1, 0) \frac{\text{Var}(Y(m, j))}{p(m, j)} r(0, j) + r(m_1, 0) \frac{\text{Var}(Y(m, 0))}{p(m, 0)} r(0, j) \right] \\
&\left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0) \frac{\text{Var}(Y(m, j))}{p(m, j)} r(0, j) + r(0, 0) \right. \\
&\left. \left. + r(0, 0) \frac{\text{Var}(Y(m, 0))}{p(m, 0)} r(0, j) + r(0, j) \right] \right\}^{-1} \\
&= \lim_{\text{Var}(\theta_{1,m}) \rightarrow 0} \left\{ r(m_1, j)r(0, 0) - r(m_1, 0)r(0, j) \right\} \\
&\times \left\{ r(m_1, 0) \left[r(0, 0) \frac{\text{Var}(\alpha) + \text{Var}(\theta_{2,j})}{p(m, j)} r(m_1, j) + r(0, 0) \frac{\text{Var}(\alpha)}{p(m, 0)} r(m_1, j) \right] \right. \\
&+ r(m_1, j) \left[r(m_1, 0) \frac{\text{Var}(\alpha) \text{Var}(\theta_{2,j})}{p(m, j)} r(0, j) + r(m_1, 0) \frac{\text{Var}(\alpha)}{p(m, 0)} r(0, j) \right] \\
&\left[r(m_1, 0) + r(m_1, j) \right] \left[r(0, 0) \frac{\text{Var}(\alpha) + \text{Var}(\theta_{2,j})}{p(m, j)} r(0, j) + r(0, 0) \right. \\
&\left. \left. + r(0, 0) \frac{\text{Var}(\alpha)}{p(m, 0)} r(0, j) + r(0, j) \right] \right\}^{-1} \\
&= \left\{ p(m, j)p(m, 0) \left[p(m_1, j)p(0, 0) \text{Var}(Y(m_1, 0)) \text{Var}(Y(0, j)) \right. \right. \\
&\left. \left. - p(m_1, 0)p(0, j) \text{Var}(Y(m_1, j)) \text{Var}(Y(0, 0)) \right] \right\} \\
&\times \left\{ \text{Var}(Y(0, j)) \left[p(m_1, 0)p(0, 0) \text{Var}(Y(0, j))p(m_1, j)p(m, 0) \right. \right. \\
&+ p(0, 0)p(m_1, j) \text{Var}(Y(0, 0))p(m, j)p(m_1, 0) \left. \right] \\
&+ \text{Var}(Y(0, 0)) \left[p(m_1, j)p(m_1, 0)p(0, j)p(m_1, 0) \text{Var}(Y(0, j)) \right. \\
&+ p(m_1, 0) \text{Var}(Y(0, 0))p(0, j)p(m_1, j)p(m, j) \left. \right] \\
&+ \text{Var}(Y(m_1, j)) \left[p(m_1, 0)p(0, 0)p(0, j)p(m, 0) \text{Var}(Y(0, j)) \right.
\end{aligned}$$

$$\begin{aligned}
& + p(m_1, 0)p(0, 0)p(m, j)p(m, 0)\text{Var}(Y(0, j)) \\
& + p(0, 0)p(0, j)p(m_1, 0)\text{Var}(Y(0, 0))p(m, j) \\
& + p(0, j)p(m_1, 0)\text{Var}(Y(0, 0))p(m, j)p(m, 0) \Big] \\
& + \text{Var}(Y(m_1, 0)) \Big[p(m_1, j)p(0, 0)p(0, j)p(m, 0)\text{Var}(Y(0, j)) \\
& + p(m_1, j)p(0, 0)p(m, j)p(m, 0)\text{Var}(Y(0, j)) \\
& + p(0, 0)p(0, j)p(m_1, j)\text{Var}(Y(0, 0))p(m, j) \\
& + p(0, j)p(m_1, j)\text{Var}(Y(0, 0))p(m, j)p(m, 0) \Big] \Big\}^{-1}.
\end{aligned}$$

Then, we take the limit of the term as $\text{Var}(\theta_{1,m_1}) \rightarrow \infty$. However, since $\text{Var}(\theta_{1,m_1})$ appears in both the numerator and denominator, the limit will lead to $\frac{\infty}{\infty}$. Thus, we use L'Hopital's rule and take the limit of the partial derivative of the numerator and denominator with respect to $\text{Var}(\theta_{1,m_1})$ as $\text{Var}(\theta_{1,m_1}) \rightarrow \infty$:

$$\begin{aligned}
\lim_{\substack{\text{Var}(\theta_{1,m}) \rightarrow 0 \\ \text{Var}(\theta_{1,m_1}) \rightarrow \infty}} a_3 = & \left\{ p(m, j)p(m, 0) \Big[p(m_1, j)p(0, 0)\text{Var}(Y(0, j)) \right. \\
& \left. - p(m_1, 0)p(0, j)\text{Var}(Y(0, 0)) \Big] \right\} \\
& \times \left\{ \text{Var}(Y(0, j)) \Big[p(m_1, 0)p(0, 0)p(0, j)p(m, 0) \right. \\
& + p(m_1, 0)p(0, 0)p(m, j)p(m, 0) \\
& + p(m_1, j)p(0, 0)p(0, j)p(m, 0) + p(m_1, j)p(0, 0)p(m, j)p(m, 0) \Big] \\
& + \text{Var}(Y(0, 0)) \Big[p(0, 0)p(0, j)p(m_1, 0)p(m, j) \\
& + p(0, j)p(m_1, 0)p(m, j)p(m, 0) \\
& \left. + p(0, 0)p(0, j)p(m_1, j)p(m, j) + p(0, j)p(m_1, j)p(m, j)p(m, 0) \right] \Big\}^{-1}.
\end{aligned}$$

To maximize the limit of the term, we can set $\text{Var}(\alpha) \rightarrow 0$ so that we are not subtracting any terms. Note that in addition, we would need $\text{Var}(\theta_{2,j}) < \infty$. Then,

taking the limit as $\text{Var}(\alpha) \rightarrow 0$, we get:

$$\lim_{\substack{\text{Var}(\theta_{1,m}) \rightarrow 0 \\ \text{Var}(\theta_{1,m_1}) \rightarrow \infty \\ \text{Var}(\alpha) \rightarrow 0}} a_3 = \left\{ p(m, j)p(m, 0)p(m_1, j)p(0, 0)\text{Var}(\theta_{2,j}) \right\} \times \\ \left\{ \text{Var}(\theta_{2,j}) \left[p(m_1, 0)p(0, 0)p(0, j)p(m, 0) + p(m_1, 0)p(0, 0)p(m, j)p(m, 0) \right. \right. \\ \left. \left. + p(m_1, j)p(0, 0)p(0, j)p(m, 0) + p(m_1, j)p(0, 0)p(m, j)p(m, 0) \right] \right\}^{-1}.$$

Since there is a $\text{Var}(\theta_{2,j})$ in both the denominator and numerator, we get the following:

$$\lim_{\substack{\text{Var}(\theta_{1,m}) \rightarrow 0 \\ \text{Var}(\theta_{1,m_1}) \rightarrow \infty \\ \text{Var}(\alpha) \rightarrow 0}} a_3 = \frac{p(m, j)p(m_1, j)}{p(m_1, 0)p(0, j) + p(m_1, 0)p(m, j) + p(m_1, j)p(0, j) + p(m_1, j)p(m, j)}.$$

□

A.5 Simulations from an Erdős-rényi Network

We also sampled networks from an Erdős-rényi distribution where the probability of an edge is 0.25 (denoted as ER(0.25)). In particular, we sampled an ER(0.25) directed network of sizes $n = 10, 20, \dots, 50$. Figure A.1 shows a directed network with 40 nodes. Note that in an ER(0.25) graph, units may have different degrees, with an expected degree being $(n - 1) \times 0.25$. Hence, an ER(0.25) graph is generally more dense than a k -regular graph. Since units have different degrees, each unit is affected differently by other units, and so unlike in a k -regular graph, each unit may contribute to the estimate of the average interference effect differently in a ER(0.25) network.

Figure A.2 shows the IMSEs for the different estimators as the number of units increase when the true mean interference effect is zero and additivity holds. Note that as the number of units increase, the number of edges also increase in an Erdős-rényi network. Hence, the IMSEs increase with the number of units, unlike in the k -regular

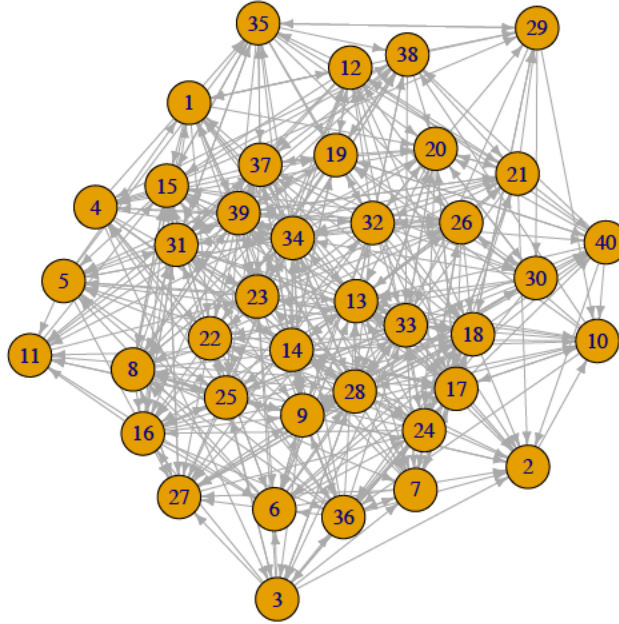


Figure A-1: Directed Erdős-rényi network with 40 nodes and probability of an edge is 0.25.

graph. Instead, the increases in IMSEs are similar to the case of the k -regular graphs when the graph becomes more dense. Furthermore, the IMSEs of the estimators in the ER(0.25) network are higher than the IMSEs in the k -regular graphs. However, in general, M_{Ind} , M_{Dil} , and HT_{Avg} still outperform the two-term estimators, with the IMSE of M_{Ind} , M_{Dil} , and HT_{Avg} being very close as in the case of the k -regular network.

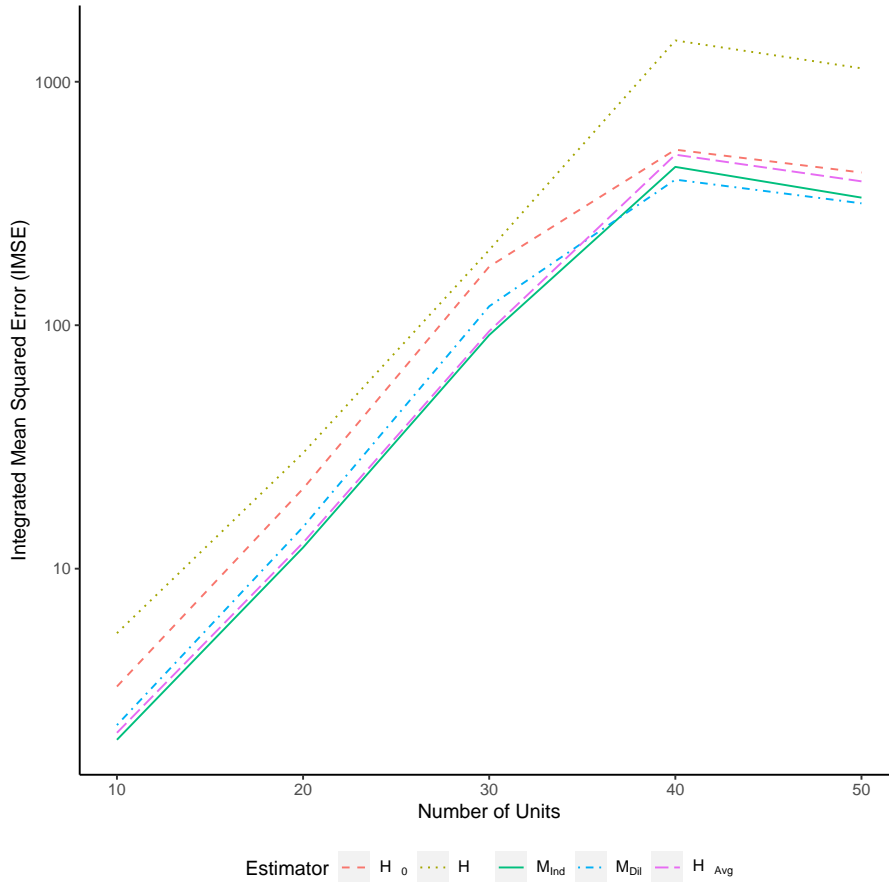


Figure A-2: IMSE for estimators (indicated by color and line type) as the number of units vary (indicated by x-axis) when additivity holds and mean interference effect is zero for ER(0.25) network.

Figure A-3 shows the IMSEs for the estimators for different interference and interaction effect sizes for a 40-node ER(0.25) network. Again, the IMSEs are generally higher than the IMSEs in the k -regular graphs. As in the k -regular network, the IMSEs of all estimators increase as the mean interference increases since we assumed a zero-mean prior for the parameters. There are some instances when the multi-term MIV LUEs outperform HT_0 , such as when the interference and interaction effect is low. However, unlike in the k -regular network, as the mean interference effect increases, the multi-term MIV LUEs have higher IMSEs than HT_0 besides M_{Dil} . This suggests that in the presence of heterogeneity in the degree distributions of the nodes,

the multi-term MIV LUEs are not as robust to additivity as in the case when the degree distributions are more homogenous. Despite this, the multi-term MIV LUEs still outperform HT_{Avg} and HT_1 , suggesting that there might still be some benefit in using the multi-term MIV LUEs.

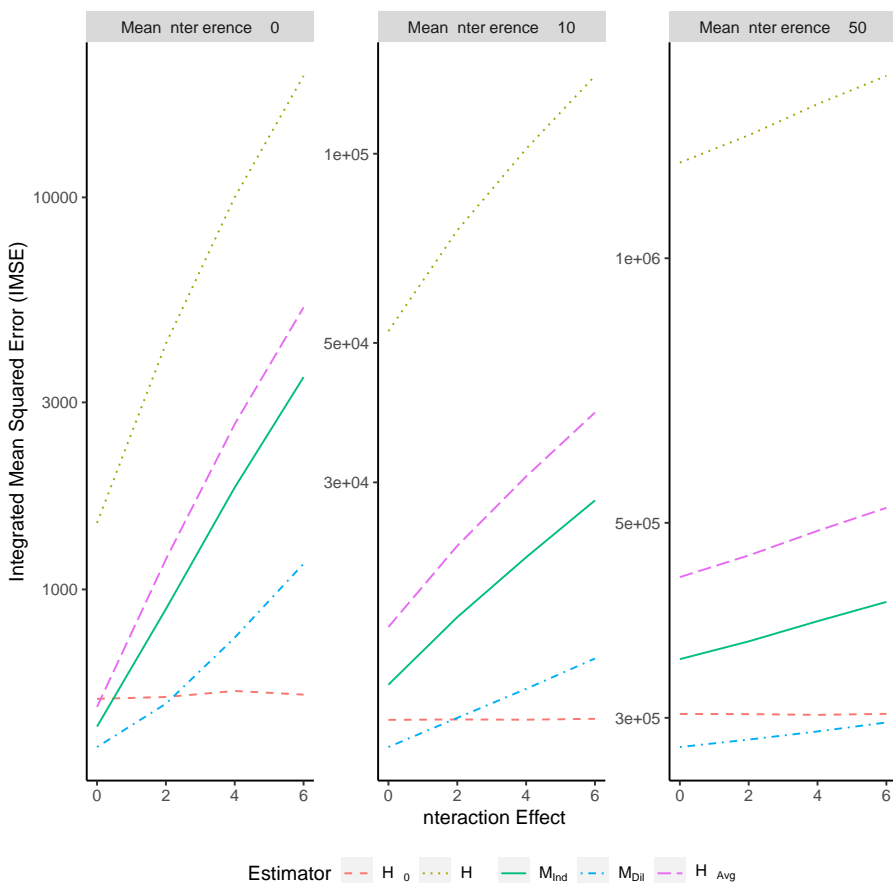


Figure A.3: IMSE for estimators (indicated by color and line type) as the interaction effect varies (indicated by x-axis) for different mean interference effects (indicated by the panels) for ER(0.25) network with 40 nodes.

Appendix B

Treatment Effect of Drug-Induced Homicide Prosecutions Reported by the Media on Drug Overdose Deaths

B.1 Intervention Dates

Table B.1: The intervention date (year-month) of DIH prosecutions reported by the media in each state and the number of DIH prosecutions reported by the media in each state.

State	Intervention Date	Number of DIH prosecutions
Alabama	2008-04	11
Alaska	2016-06	6
Arizona	2012-10	5
Arkansas	2011-11	4
California	2001-10	51
Colorado	2011-08	27
Connecticut	2008-01	39
Delaware	2016-10	3
Florida	2001-02	102
Georgia	2000-04	24
Hawaii	NA	NA
Idaho	2015-09	7
Illinois	2002-07	298
Indiana	2010-04	50
Iowa	2004-11	29
Kansas	2007-12	6
Kentucky	2011-09	34
Louisiana	2004-03	58
Maine	2009-01	15
Maryland	2002-08	54
Massachusetts	2008-01	25
Michigan	2005-08	99
Minnesota	2007-11	116
Mississippi	2014-08	1
Missouri	2005-08	35
Montana	2004-01	9
Nebraska	2013-01	1
Nevada	2005-10	10
New Hampshire	2010-06	37
New Jersey	2002-08	114
New Mexico	2010-02	4
New York	2007-10	83
North Carolina	2005-04	82
North Dakota	2013-01	37
Ohio	2000-03	356
Oklahoma	2011-05	31
Oregon	2009-10	16
Pennsylvania	2001-02	652
Rhode Island	2014-02	2
South Carolina	2016-03	9
South Dakota	2017-02	11
Tennessee	2008-03	82
Texas	2001-08	33
Utah	2005-07	16
Vermont	2009-07	12
Virginia	2003-09	59
Washington	2008-10	57
West Virginia	2013-03	31
Wisconsin	2004-06	338
Wyoming	2009-08	19

Table B.2: The intervention dates (year-month) for the other relevant policy measures in each state: naloxone access law where pharmacists a) can dispense naloxone without prescription (NAL: can dispense) and b) cannot dispense naloxone without prescription (NAL: cannot dispense) c) medical marijuana law (MML) d) recreational marijuana law (RML) e) Prescription Drug Monitoring Program (PDMP) f) 911 Good Samaritan Law (911 GSL) f) Medicaid expansion (Medicaid).

State	NAL: can dispense	NAL: cannot dispense	MML	RML	PDMP	911 GSL	Medicaid
Alabama	2015-06	NA	NA	NA	2005-05	2015-06	NA
Alaska	2016-03	NA	1999-03	2015-02	2008-09	2008-09	2015-09
Arizona	2016-08	NA	2010-12	NA	2007-09	2018-04	2014-01
Arkansas	2015-07	NA	2016-11	NA	2011-03	2015-07	2014-01
California	2014-01	2008-01	1996-11	2017-01	1939-01	2013-01	2010-11
Colorado	2015-04	2013-05	2000-12	2012-12	2005-06	2012-05	2014-01
Connecticut	2015-06	2003-10	2012-05	NA	2006-06	2011-10	2010-04
Delaware	2014-08	NA	2011-07	NA	2010-07	2013-08	2014-01
Florida	2016-07	2015-06	2017-01	NA	2009-06	2012-10	NA
Georgia	2014-04	NA	NA	NA	2011-05	2014-04	NA
Hawaii	2016-06	NA	2000-06	NA	1943-01	2015-07	2014-01
Idaho	2015-07	NA	NA	NA	1967-01	2018-07	NA
Illinois	2010-01	NA	2014-01	NA	1961-01	2012-06	2014-01
Indiana	2015-04	NA	NA	NA	1997-01	2014-03	2015-02
Iowa	2016-05	NA	NA	NA	2006-05	2018-07	2014-01
Kansas	2017-07	NA	NA	NA	2008-07	NA	NA
Kentucky	2013-06	NA	NA	NA	1998-07	2015-03	2014-01
Louisiana	2015-08	NA	NA	NA	2006-07	2014-08	2016-07
Maine	2015-10	2014-04	1999-12	2017-01	2003-06	NA	2019-01
Maryland	2015-10	2013-10	2013-10	NA	2011-05	2009-10	2014-01
Massachusetts	2014-07	2012-08	2013-01	2016-12	1992-01	2012-08	2014-01
Michigan	2017-03	2014-10	2008-12	NA	1988-01	2017-01	2014-04
Minnesota	2014-05	NA	2004-05	NA	2007-07	2014-07	2010-03
Mississippi	2015-07	NA	NA	NA	2005-01	2015-07	NA
Missouri	2016-08	NA	NA	NA	NA	2017-08	NA
Montana	2017-05	NA	2004-11	NA	2011-07	2017-05	2016-01
Nebraska	NA	2015-05	NA	NA	2011-04	2017-08	NA
Nevada	2015-10	NA	2001-10	2017-01	1995-06	2015-10	2014-01
New Hampshire	2015-06	NA	2013-07	NA	2012-06	2015-09	2014-08
New Jersey	2013-07	NA	2010-10	NA	2008-01	2013-05	2011-04
New Mexico	2014-03	2001-04	2007-07	NA	2004-07	2007-06	2014-01
New York	2014-06	2006-04	2014-07	NA	1972-01	2011-09	2014-01
North Carolina	2013-04	NA	NA	NA	2005-08	2013-04	NA
North Dakota	2015-08	NA	2016-12	NA	2005-12	2015-08	2014-01
Ohio	2015-07	2014-03	2016-09	NA	2005-05	2016-09	2014-01
Oklahoma	2014-11	2013-11	NA	NA	1990-05	NA	NA
Oregon	2013-06	NA	1998-12	2015-07	2009-07	2016-01	2014-01
Pennsylvania	2014-12	NA	2016-05	NA	1972-01	2014-12	2015-01
Rhode Island	2014-03	2012-06	2006-01	NA	1978-01	2012-06	2014-01
South Carolina	2016-06	2015-06	NA	NA	2006-06	2017-06	NA
South Dakota	2016-07	NA	NA	NA	2010-03	2017-07	NA
Tennessee	2014-07	NA	NA	NA	2003-01	2015-07	NA
Texas	2015-09	NA	NA	NA	1981-09	NA	NA
Utah	2016-05	2014-05	NA	NA	1995-01	2014-03	NA
Vermont	2013-07	NA	2004-07	NA	2006-05	2013-06	2014-01
Virginia	2015-04	2013-07	NA	NA	2002-04	2015-07	2019-01
Washington	2015-07	2010-06	1998-12	2012-12	2007-07	2010-06	2011-01
West Virginia	2016-06	2015-05	NA	NA	1995-07	2015-06	2014-01
Wisconsin	2014-04	NA	NA	NA	2010-05	2014-04	NA
Wyoming	2017-07	NA	NA	NA	2003-03	NA	NA

B.2 States and their U.S. Regions

Table B.3: The states considered in the analysis with their U.S. regions.

Northeast	Midwest	South	West
Connecticut	Illinois	Alabama	Alaska
Maine	Iowa	Arkansas	Arizona
Massachusetts	Indiana	Delaware	California
New Hampshire	Kansas	Florida	Colorado
New Jersey	Michigan	Georgia	Hawaii
New York	Minnesota	Kentucky	Idaho
Pennsylvania	Missouri	Louisiana	Montana
Rhode Island	Nebraska	Maryland	New Mexico
Vermont	North Dakota	Mississippi	Nevada
	Ohio	North Carolina	Oregon
	South Dakota	Oklahoma	Utah
	Wisconsin	South Carolina	Washington
		Tennessee	Wyoming
		Texas	
		Virginia	
		West Virginia	

B.3 Sandwich Estimator

B.3.1 Constant Treatment Effect Model

First, the Mean Value Theorem states that for a function f that is continuous on $[a, b]$ and differentiable on (a, b) for $a, b \in \mathbb{R}$ and $a < b$:

$$f'(c)(b - a) = f(b) - f(a)$$

for some value $c \in (a, b)$. From the unbiased estimating equations, we let $f(\Theta) = \sum_{s,t} \vec{Z}_{s,t} (Y_{s,t} - p_{s,t,\Theta})$, where $Y_{s,t}$ is the risk of unintentional drug overdose death in state s at time interval t and $p_{s,t,\Theta^*} = \text{expit}(\vec{Z}_{s,t}^T \Theta^*)$ is the underlying probability of unintentional drug overdose deaths given by true parameter Θ^* . Then,

$$f'(\tilde{\Theta}) (\hat{\Theta} - \Theta^*) = \sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit}(\vec{Z}_{s,t}^T \hat{\Theta}) \right) - \sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit}(\vec{Z}_{s,t}^T \Theta^*) \right), \quad (\text{B.1})$$

where $\tilde{\Theta}$ is between $\hat{\Theta}$ and Θ^* . Since $\hat{\Theta} \xrightarrow{p} \Theta^*$ and $\tilde{\Theta}$ is between $\hat{\Theta}$ and Θ^* , then $\tilde{\Theta} \xrightarrow{p} \Theta^*$. Note that the derivative f' is given by:

$$\begin{aligned}
f'(\Theta) &= \frac{\partial}{\partial \Theta} \sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta \right) \right) \\
&= - \sum_{s,t} \frac{\partial}{\partial \Theta} \vec{Z}_{s,t} \text{expit} \left(\vec{Z}_{s,t}^T \Theta \right) \\
&= - \sum_{s,t} \vec{Z}_{s,t} \frac{\partial}{\partial \Theta} \frac{\exp \left(\vec{Z}_{s,t}^T \Theta \right)}{1 + \exp \left(\vec{Z}_{s,t}^T \Theta \right)} \\
&= - \sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T \left(\frac{\exp \left(\vec{Z}_{s,t}^T \Theta \right)}{1 + \exp \left(\vec{Z}_{s,t}^T \Theta \right)} - \frac{\exp \left(\vec{Z}_{s,t}^T \Theta \right)^2}{\left(1 + \exp \left(\vec{Z}_{s,t}^T \Theta \right) \right)^2} \right) \\
&= - \sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T \left(p_{s,t,\Theta} (1 - p_{s,t,\Theta}) \right).
\end{aligned}$$

Substituting in $-\sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T (p_{s,t,\Theta^*} (1 - p_{s,t,\Theta^*}))$ for the derivative, we can rearrange Equation (B.1):

$$\begin{aligned}
\sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \hat{\Theta} \right) \right) &= \sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \\
&\quad - \sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T (p_{s,t,\Theta^*} (1 - p_{s,t,\Theta^*})) \left(\hat{\Theta} - \Theta^* \right). \quad (\text{B.2})
\end{aligned}$$

Note that the left hand side of Equation (B.2) is 0 since $\hat{\Theta}$ solves the unbiased estimating equation. Then,

$$\left(\hat{\Theta} - \Theta^* \right) = \underbrace{\left(\sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T (p_{s,t,\Theta^*} (1 - p_{s,t,\Theta^*})) \right)^{-1}}_{=C^{-1}} \sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right). \quad (\text{B.3})$$

Note that C is a positive semi-definite matrix since $\vec{Z}_{s,t} \vec{Z}_{s,t}^T$ is positive semi-definite and $p_{s,t,\Theta^*} (1 - p_{s,t,\Theta^*}) \geq 0$. Since C^{-1} is easy to estimate, we focus on the limiting

distribution of $\sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right)$.

First, we find the expected value. By linearity of expectations, we focus on the individual s, t terms:

$$\mathbb{E} \left(\vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \right) = \mathbb{E} \left(\vec{Z}_{s,t} \mathbb{E} \left(Y_{s,t} - p_{s,t,\Theta^*} \mid \vec{Z}_{s,t} = \vec{Z}_{s,t} \right) \right) = 0,$$

where the first equality holds by the Law of Total Expectations and the last line holds by Assumption 3.2. Since the expected value is zero, we focus on the second moment to find the variance.

$$\begin{aligned} & \mathbb{E} \left(\left(\sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \right) \left(\sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \right)^T \right) \\ &= \sum_{s,t} \sum_{s',t'} \mathbb{E} \left(\vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \left(Y_{s',t'} - \text{expit} \left(\vec{Z}_{s',t'}^T \Theta^* \right) \right) \vec{Z}_{s',t'}^T \right). \end{aligned}$$

We now consider two cases: 1) when $s \neq s'$ and without loss of generality, $t' \leq t$ and 2) when $s = s'$ and without loss of generality $t' < t$.

Case 1 ($s \neq s', t' \leq t$).

$$\begin{aligned} & \mathbb{E} \left(\vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \left(Y_{s',t'} - \text{expit} \left(\vec{Z}_{s',t'}^T \Theta^* \right) \right) \vec{Z}_{s',t'}^T \right) \\ &= \mathbb{E} \left(\mathbb{E} \left[\vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \left(Y_{s',t'} - \text{expit} \left(\vec{Z}_{s',t'}^T \Theta^* \right) \right) \vec{Z}_{s',t'}^T \mid \vec{Z}_{s,t}, \vec{Z}_{s',t'}, Y_{s',t'} \right] \right) \\ &= \mathbb{E} \left(\vec{Z}_{s,t} \mathbb{E} \left[Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \mid \vec{Z}_{s,t}, \vec{Z}_{s',t'}, Y_{s',t'} \right] \left(Y_{s',t'} - \text{expit} \left(\vec{Z}_{s',t'}^T \Theta^* \right) \right) \vec{Z}_{s',t'}^T \right) \\ &= \mathbb{E} \left(\vec{Z}_{s,t} \mathbb{E} \left[Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \mid \vec{Z}_{s,t} \right] \left(Y_{s',t'} - \text{expit} \left(\vec{Z}_{s',t'}^T \Theta^* \right) \right) \vec{Z}_{s',t'}^T \right) \\ &= 0, \end{aligned}$$

where the first equality holds by Law of Total Expectations, the third equality holds by Assumption 3.1, and the last line holds by Assumption 3.2.

Case 2 ($s = s', t' < t$).

$$\begin{aligned} & \mathbb{E} \left(\vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \left(Y_{s,t'} - \text{expit} \left(\vec{Z}_{s,t'}^T \Theta^* \right) \right) \vec{Z}_{s,t'}^T \right) \\ &= \mathbb{E} \left(\mathbb{E} \left[\vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \left(Y_{s,t'} - \text{expit} \left(\vec{Z}_{s,t'}^T \Theta^* \right) \right) \vec{Z}_{s,t'}^T \mid \vec{Z}_{s,t}, \vec{Z}_{s,t'}, Y_{s,t'} \right] \right) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left(\vec{Z}_{s,t} \mathbb{E} \left[\left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \mid \vec{Z}_{s,t}, \vec{Z}_{s,t'}, Y_{s,t'} \right] \left(Y_{s,t'} - \text{expit} \left(\vec{Z}_{s,t'}^T \Theta^* \right) \right) \vec{Z}_{s,t'}^T \right) \\
&= \mathbb{E} \left(\vec{Z}_{s,t} \mathbb{E} \left[Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \mid \vec{Z}_{s,t} \right] \left(Y_{s,t'} - \text{expit} \left(\vec{Z}_{s,t'}^T \Theta^* \right) \right) \vec{Z}_{s,t'}^T \right) \\
&= 0,
\end{aligned}$$

where the first equality holds by Law of Total Expectations, the third equality holds by Assumption 3.1, and the last line holds by Assumption 3.2.

In both cases, we see that when $s \neq s'$ and $t' \leq t$ or when $s = s'$ and $t' < t$, the expectation is equal to 0. Hence, the expectation is only non-zero for when $s = s'$ and for $t = t'$, i.e.

$$\begin{aligned}
&\mathbb{E} \left(\left(\sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \right) \left(\sum_{s,t} \vec{Z}_{s,t} \left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right) \right)^T \right) \\
&= \sum_{s,t} \mathbb{E} \left(\left(Y_{s,t} - \text{expit} \left(\vec{Z}_{s,t}^T \Theta^* \right) \right)^2 \vec{Z}_{s,t} \vec{Z}_{s,t}^T \right).
\end{aligned}$$

The Sandwich estimator is then given by

$$\begin{aligned}
\hat{\Sigma} &= \frac{N}{N-d} \left(\sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T p_{s,t,\hat{\Theta}} \left(1 - p_{s,t,\hat{\Theta}} \right) \right)^{-1} \left(\sum_{s,t} \left(Y_{s,t} - p_{s,t,\hat{\Theta}} \right)^2 \vec{Z}_{s,t} \vec{Z}_{s,t}^T \right) \\
&\quad \times \left(\left(\sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T p_{s,t,\hat{\Theta}} \left(1 - p_{s,t,\hat{\Theta}} \right) \right)^{-1} \right)^T \tag{B.4}
\end{aligned}$$

where we estimated Θ^* by $\hat{\Theta}$. In addition, we included a factor of $\frac{N}{N-d}$ for bias correction, where N is equal to the number of state-time combinations and d is equal to the number of parameters (Li and Redden, 2015).

B.3.2 Treatment Effect That Depends on Exposure Duration

The sandwich estimator under the GAM with a linear link is derived in a similar manner as in the constant treatment effect model. The main difference is we now,

the unbiased estimating equation gives:

$$f(\Theta) = \sum_{s,t} \vec{Z}_{s,t} \left(\log(Y_{s,t}) - \vec{Z}_{s,t}^T \Theta \right).$$

Hence, the Sandwich estimator for the event study model is given by:

$$\hat{\Sigma} = \frac{N}{N-d} (C)^{-1} \left(\sum_{s,t} \left(\log(Y_{s,t}) - \vec{Z}_{s,t}^T \hat{\Theta} \right)^2 \vec{Z}_{s,t} \vec{Z}_{s,t}^T \right) (C^{-1})^T, \quad (\text{B.5})$$

where $C = \sum_{s,t} \vec{Z}_{s,t} \vec{Z}_{s,t}^T$ now and $\vec{Z}_{s,t}$ is given by Equation (3.19).

B.4 Estimation of Attributable Deaths

B.4.1 Constant Treatment Effect Model

The estimated number of attributable deaths in a state s is given by Equations (3.11) and (3.12):

$$n_{s,t,\text{attributable deaths}} = n_{s,t,\text{observed overdose deaths}} - n_{s,t,\text{population}} \times \hat{p}_{s,t}^{(a_{s,t}=0)},$$

where $n_{s,t,\text{population}}$ is the population in state s at time t and $\hat{p}_{s,t}^{(a_{s,t}=0)}$ is the estimated probability of unintentional drug overdose deaths had the intervention not occurred.

For states s and time intervals t , our model (Equation (3.6)) implied that

$$\text{logit} \left(\mathbb{E} \left(Y_{s,t} \mid \vec{Z}_{s,t} \right) \right) = \alpha_s + \gamma_{r(s)}(t) + N_t \psi + X_{s,t} \delta + A_{s,t} \beta.$$

If no intervention occurred in state s at time interval t , we have $A_{s,t} = 0$, and so

$$\text{logit}(p_{s,t}^{(a_{s,t}=0)}) = \alpha_s + \gamma_{r(s)}(t) + N_t \psi + X_{s,t} \delta. \quad (\text{B.6})$$

Let $\phi_{s,t} = \alpha_s + \gamma_{r(s)}(t) + N_t\psi + X_{s,t}\delta$, and so the logit of probability of an overdose death had the intervention not occurred is:

$$\text{logit}(p_{s,t}^{(a_{s,t}=0)}) = \phi_{s,t}. \quad (\text{B.7})$$

We can estimate $\text{logit}(p_{s,t}^{(a_{s,t}=0)})$ using $\hat{\beta}$ and $\hat{\phi}_{s,t}$, which are estimates for β and $\phi_{s,t}$, respectively. However, note that we can also estimate $\text{logit}(p_{s,t})$ by finding the logit of proportion of people who died from a drug overdose, i.e. $\text{logit}\left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\right)$. Then, we have the following:

$$\begin{aligned} A_{s,t}\hat{\beta} + \hat{\phi}_{s,t} &= \text{logit}(\hat{p}_{s,t}) \approx \text{logit}\left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\right) \\ \Rightarrow \hat{\phi}_{s,t} &\approx \text{logit}\left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\right) - A_{s,t}\hat{\beta} \\ \Rightarrow \hat{p}_{s,t}^{(a_{s,t}=0)} &= \text{expit}(\hat{\phi}_{s,t}) \approx \text{expit}\left(\text{logit}\left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\right) - A_{s,t}\hat{\beta}\right), \quad (\text{B.8}) \end{aligned}$$

where in the last line, we used Equation (B.7). Equation (B.8) provides an estimate for the probability of a drug overdose death in state s , had the intervention not occurred. Combining Equations (3.11), (3.12), and (B.8):

$$\begin{aligned} \hat{n}_{s,t,\text{attributable deaths}} &= n_{s,t,\text{observed overdose deaths}} \\ &\quad - n_{s,t,\text{population}} \times \text{expit}\left(\text{logit}\left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\right) - A_{s,t}\hat{\beta}\right). \end{aligned} \quad (\text{B.9})$$

To find a 95% confidence interval for the number of attributable deaths, we substituted $\hat{\beta}$ by the 95% confidence interval upper and lower limits of β .

B.4.2 Treatment Effect That Depends on Exposure Duration

For states s and time intervals t , we assumed from Equation (3.18) that:

$$\mathbb{E}(\log(Y_{s,t})|\vec{Z}_{s,t}) = \alpha_s + \gamma_{r(s)}(t) + X_{l,s,t}\delta_{0,l} + K_{l,s,t}\delta_{1,l} + A_{s,t}\beta_0 + K_{s,t}\beta_1,$$

where $Y_{s,t}$ is the risk of unintentional drug overdose deaths in state s at time interval t . If no intervention occurred in state s at time interval t , we have $A_{s,t} = 0$ and $K_{s,t} = 0$, and so the log probability of unintentional drug overdose deaths had intervention not occurred is

$$\mathbb{E}(\log(p_{s,t}^{(a_{s,t}=0)})|\vec{Z}_{s,t}) = \alpha_s + \gamma_{r(s)}(t) + X_{l,s,t}\delta_{0,l} + K_{l,s,t}\delta_{1,l}, \quad (\text{B.10})$$

where $p_{s,t}^{(a_{s,t}=0)}$ denotes the probability of unintentional drug overdose deaths had intervention not occurred. We denote $\alpha_s + \gamma_{r(s)}(t) + X_{l,s,t}\delta_{0,l} + K_{l,s,t}\delta_{1,l}$ by $\phi_{s,t}$. We estimated the probability of unintentional drug overdose deaths had intervention not occurred in a similar manner as in the previous section using $\hat{\phi}_{s,t}$:

$$\begin{aligned} A_{s,t}\hat{\beta}_0 + K_{s,t}\hat{\beta}_1 &= \hat{\phi}_{s,t} \approx \log\left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\right) \\ &\Rightarrow \hat{\phi}_{s,t} \approx \log\left(\frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\right) - A_{s,t}\hat{\beta}_0 - K_{s,t}\hat{\beta}_1 \\ \Rightarrow \hat{p}_{s,t}^{(a_{s,t}=0)} &= \exp(\hat{\phi}_{s,t}) \approx \frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}\exp\left(-A_{s,t}\hat{\beta}_0 - K_{s,t}\hat{\beta}_1\right). \end{aligned} \quad (\text{B.11})$$

We then estimated the number of attributable deaths using $\hat{p}_{s,t}^{(a_{s,t}=0)}$:

$$\begin{aligned} \hat{n}_{s,t,\text{attributable deaths}} &= n_{s,t,\text{observed overdose deaths}} \\ &\quad - n_{s,t,\text{population}} \times \frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}} \\ &\quad \times \exp\left(-A_{s,t}\hat{\beta}_0 - K_{s,t}\hat{\beta}_1\right). \end{aligned} \quad (\text{B.12})$$

Since the number of attributable deaths is estimated using both $\hat{\beta}_0$ and $\hat{\beta}_1$, we

need to account for the covariance between the coefficients when computing the lower and upper bounds of the 95% confidence interval of the number of attributable deaths. We first denote $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ and denote $\hat{\beta}$ as the estimate of β . We also use Σ_β to denote the variance-covariance matrix of the coefficients β_0 and β_1 . From the theory of unbiased estimating equations, we have

$$\sqrt{n} \left(\hat{\beta} - \beta \right) \xrightarrow{D} \mathcal{N}(0, \Sigma_\beta).$$

Using the Delta Method, we have

$$\sqrt{n} \left(g(\hat{\beta}) - g(\beta) \right) \xrightarrow{D} \mathcal{N}(0, J_{\beta,t} \Sigma_\beta J_{\beta,t}^T),$$

where for a time interval t ,

$$\begin{aligned} g(\beta) &= \sum_s n_{\text{attributable deaths},s,t} \\ &= \sum_s n_{\text{population},s,t} (p_{s,t} - p_{s,t} \exp(-A_{s,t}\beta_0 - K_{s,t}\beta_1)), \end{aligned}$$

where we estimate $p_{s,t}$ using $\hat{p}_{s,t}^{\text{obs}} = \frac{n_{s,t,\text{observed overdose deaths}}}{n_{s,t,\text{population}}}$, the observed risk of unintentional drug overdose deaths, and $J_{\beta,t}$ is the Jacobian matrix. The Jacobian matrix is given by the following:

$$J_{\beta,t} = \begin{pmatrix} \frac{\partial g(\beta)}{\partial \beta_0} & \frac{\partial g(\beta)}{\partial \beta_1} \end{pmatrix},$$

where

$$\frac{\partial g(\beta)}{\partial \beta_0} = \sum_s n_{\text{population},s,t} p_{s,t} \exp(-A_{s,t}\beta_0 - K_{s,t}\beta_1) A_{s,t}$$

and

$$\frac{\partial g(\beta)}{\partial \beta_1} = \sum_s n_{\text{population},s,t} p_{s,t}^{\text{obs}} \exp(-A_{s,t}\beta_0 - K_{s,t}\beta_1) K_{s,t}.$$

Hence, the lower and upper limits of the 95% confidence interval of the attributable deaths in at time interval t are given by:

$$\sum_s \hat{n}_{\text{attributable deaths},s,t} \pm \frac{1.96}{\sqrt{n}} \sqrt{J_{\beta,t} \Sigma_{\beta} J_{\beta,t}^T}.$$

B.5 Results for Constant Treatment Effect

Table B.4: Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for logistic GAM when assuming constant treatment effects and assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years.

	Logistic GAM with Lasting Effect (95% Confidence Interval)	Logistic GAM with Two-Year Effect (95% Confidence Interval)
NAL: can dispense	0.976 (0.927, 1.028)	0.966 (0.917, 1.018)
NAL: cannot dispense	1.008 (0.963, 1.056)	1.003 (0.958, 1.049)
MML	1.065 (1.019, 1.113)	1.070 (1.024, 1.118)
RML	0.964 (0.913, 1.018)	0.962 (0.911, 1.015)
911 GSL	1.034 (0.994, 1.077)	1.037 (0.996, 1.079)
PDMP	0.981 (0.939, 1.024)	0.983 (0.942, 1.026)
Medicaid expansion	1.104 (1.061, 1.149)	1.108 (1.064, 1.154)
DIH prosecutions reported by media	1.064 (1.024, 1.106)	1.059 (1.026, 1.094)
Number of treated states	1.004 (0.990, 1.017)	0.998 (0.990, 1.005)

Table B.5: Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for GAM with linear link function when assuming constant treatment effects and assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years.

	GAM with Linear Link with Lasting Effect (95% Confidence Interval)	GAM with Linear Link with Two-Year Effect (95% Confidence Interval)
NAL: can dispense	0.923 (0.863, 0.987)	0.919 (0.860, 0.983)
NAL: cannot dispense	0.999 (0.936, 1.065)	0.994 (0.932, 1.061)
MML	1.213 (1.133, 1.298)	1.215 (1.134, 1.301)
RML	0.895 (0.832, 0.962)	0.891 (0.829, 0.958)
911 GSL	1.055 (0.999, 1.115)	1.059 (1.003, 1.119)
PDMP	0.858 (0.813, 0.906)	0.857 (0.812, 0.905)
Medicaid expansion	1.095 (1.040, 1.153)	1.094 (1.038, 1.152)
DIH prosecutions reported by media	0.975 (0.931, 1.021)	1.014 (0.974, 1.056)
Number of treated states	0.996 (0.979, 1.013)	0.995 (0.978, 1.012)

B.6 Results of Sensitivity Analyses for Model where Treatment Effect Depends on Exposure Duration

Table B.6: Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for logistic GAM when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years.

	Logistic GAM with Lasting Effect (95% Confidence Interval)	Logistic GAM with Two-Year Effect (95% Confidence Interval)
Exposure to NAL: can dispense	1.016 (0.873, 1.182)	0.987 (0.842, 1.158)
Linear effect of NAL: can dispense	0.979 (0.956, 1.002)	0.975 (0.947, 1.005)
Exposure to NAL: cannot dispense	1.064 (0.891, 1.270)	1.055 (0.862, 1.291)
Linear effect of NAL: cannot dispense	0.998 (0.981, 1.015)	0.997 (0.982, 1.013)
Exposure to MML	1.034 (0.881, 1.214)	1.051 (0.862, 1.281)
Linear effect of MML	0.996 (0.981, 1.012)	0.994 (0.975, 1.014)
Exposure to RML	1.002 (0.858, 1.170)	0.998 (0.805, 1.236)
Linear effect of RML	0.980 (0.949, 1.012)	0.988 (0.957, 1.020)
Exposure to 911 GSL	1.026 (0.911, 1.156)	1.033 (0.916, 1.165)
Linear effect of 911 GSL	1.014 (0.994, 1.034)	1.013 (0.992, 1.034)
Exposure to PDMP	0.968 (0.817, 1.146)	0.991 (0.772, 1.273)
Linear effect of PDMP	1.005 (0.983, 1.027)	1.006 (0.979, 1.034)
Exposure to Medicaid expansion	1.070 (0.953, 1.202)	1.070 (0.936, 1.223)
Linear effect of Medicaid expansion	1.014 (0.995, 1.034)	1.018 (0.994, 1.042)
Exposure to DIH prosecutions	0.981 (0.835, 1.154)	1.051 (0.931, 1.187)
Linear effect of DIH prosecutions	0.985 (0.967, 1.004)	0.991 (0.977, 1.005)

Table B.7: Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for GAM with linear link function when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years. We excluded the last six years of data.

	GAM with Linear Link with Lasting Effect (95% Confidence Interval)	GAM with Linear Link with Two-Year Effect (95% Confidence Interval)
Exposure to NAL: can dispense	0.836 (0.663, 1.055)	0.794 (0.623, 1.012)
Linear effect of NAL: can dispense	0.943 (0.885, 1.005)	0.956 (0.894, 1.022)
Exposure to NAL: cannot dispense	1.162 (0.847, 1.595)	1.159 (0.847, 1.587)
Linear effect of NAL: cannot dispense	0.957 (0.932, 0.982)	0.959 (0.934, 0.984)
Exposure to MML	1.264 (0.988, 1.619)	1.260 (0.984, 1.614)
Linear effect of MML	0.984 (0.968, 1.001)	0.985 (0.968, 1.002)
Exposure to RML	0.779 (0.518, 1.172)	0.825 (0.559, 1.219)
Linear effect of RML	1.015 (0.876, 1.175)	1.014 (0.873, 1.177)
Exposure to 911 GSL	1.144 (0.900, 1.455)	1.136 (0.894, 1.444)
Linear effect of 911 GSL	1.010 (0.971, 1.050)	1.012 (0.973, 1.053)
Exposure to PDMP	0.884 (0.755, 1.036)	0.899 (0.766, 1.053)
Linear effect of PDMP	1.011 (0.993, 1.030)	1.013 (0.995, 1.032)
Exposure to Medicaid expansion	1.034 (0.847, 1.264)	1.037 (0.845, 1.271)
Linear effect of Medicaid expansion	1.016 (0.957, 1.079)	1.022 (0.958, 1.089)
Exposure to DIH prosecutions	0.995 (0.843, 1.175)	1.041 (0.912, 1.188)
Linear effect of DIH prosecutions	0.983 (0.969, 0.997)	0.981 (0.967, 0.995)

Table B.8: Estimated risk ratios and 95% confidence intervals for relevant policy measures and DIH prosecutions reported by the media for logistic GAM when assuming treatment effect lasts until at least 2019 and when assuming treatment effect lasts only for two years. We excluded the last six years of data.

	Logistic GAM with Lasting Effect (95% Confidence Interval)	Logistic GAM with Two-Year Effect (95% Confidence Interval)
Exposure to NAL: can dispense	0.956 (0.798, 1.146)	0.903 (0.7509, 1.0865)
Linear effect of NAL: can dispense	0.966 (0.919, 1.015)	0.974 (0.9259, 1.0239)
Exposure to NAL: cannot dispense	1.109 (0.867, 1.418)	1.092 (0.8568, 1.3917)
Linear effect of NAL: cannot dispense	0.986 (0.971, 1.001)	0.987 (0.9713, 1.0025)
Exposure to MML	1.021 (0.840, 1.241)	1.026 (0.8458, 1.2434)
Linear effect of MML	0.995 (0.978, 1.012)	0.993 (0.9763, 1.0094)
Exposure to RML	0.854 (0.628, 1.162)	0.927 (0.6944, 1.2372)
Linear effect of RML	0.983 (0.879, 1.099)	0.974 (0.8705, 1.0897)
Exposure to 911 GSL	0.974 (0.843, 1.125)	0.970 (0.8401, 1.1198)
Linear effect of 911 GSL	1.001 (0.976, 1.026)	1.001 (0.9763, 1.0263)
Exposure to PDMP	0.955 (0.812, 1.124)	0.992 (0.8423, 1.1676)
Linear effect of PDMP	1.007 (0.986, 1.029)	1.009 (0.9884, 1.0306)
Exposure to Medicaid expansion	1.048 (0.903, 1.216)	1.027 (0.8882, 1.1867)
Linear effect of Medicaid expansion	1.017 (0.974, 1.063)	1.021 (0.9760, 1.0681)
Exposure to DIH prosecutions	1.015 (0.880, 1.171)	1.040 (0.9324, 1.1610)
Linear effect of DIH prosecutions	0.982 (0.968, 0.997)	0.985 (0.9712, 0.9995)

Appendix C

Structural Nested Risk Ratio Models

C.1 Mimicking Potential Outcomes

Proof of Theorem 4. We prove that

$$\mathbb{E}(H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t) = \mathbb{E}(Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))} \mid \bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t)$$

for $k = 1, \dots, K$ and $t \leq k - 1$ using backwards induction.

Case: $t = k-1$

By the definition of $H_{s,k}(k)$ and Consistency Assumption 4.3:

$$H_{s,k}(k) = Y_{s,k} = Y_{s,k}^{(g(\bar{a}_{k-1}), g(\bar{a}_{k-1}))}.$$

Then,

$$\begin{aligned} & \mathbb{E} \left(H_{s,k}(k-1) \mid \bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1} \right) \\ &= \mathbb{E} \left(\frac{H_{s,k}(k)}{\gamma_k^{k-1}(\bar{X}_{s,k-1}, \bar{A}_{s,k-1})} \mid \bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1} \right) \\ &= \frac{\mathbb{E} \left(\mathbb{E} \left[H_{s,k}(k) \mid \bar{X}_{s,k}, \bar{A}_{s,k} \right] \mid \bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1} \right)}{\gamma_k^{k-1}(\bar{X}_{s,k-1}, \bar{A}_{s,k-1})} \\ &= \frac{\mathbb{E} \left(\mathbb{E} \left[Y_{s,k}^{(g(\bar{a}_{k-1}), g(\bar{a}_{k-1}))} \mid \bar{X}_{s,k}, \bar{A}_{s,k} \right] \mid \bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1} \right)}{\gamma_k^{k-1}(\bar{X}_{s,k-1}, \bar{A}_{s,k-1})} \\ &= \frac{\mathbb{E} \left(Y_{s,k}^{(g(\bar{a}_{k-1}), g(\bar{a}_{k-1}))} \mid \bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1} \right)}{\gamma_k^{k-1}(\bar{X}_{s,k-1}, \bar{A}_{s,k-1})} \\ &= \mathbb{E} \left(Y_{s,k}^{(g(\bar{a}_{k-2}), g(\bar{a}_{k-2}), g(\bar{a}_{k-2}))} \mid \bar{X}_{s,k-1}, \bar{A}_{s,k-1} = \bar{a}_{k-1} \right), \end{aligned}$$

where we use the Law of Iterated Expectations for the third and fifth lines, the fourth

line is from the definition of $H_{s,k}(k)$, and the last line is a result of the definition of γ_k^{k-1} , as given by Equation (4.1). So Equation (4.4) holds for $t = k - 1$.

Induction Hypothesis: We assume that for $t = k - 2, k - 3, \dots, m + 1$:

$$\mathbb{E} \left(H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t \right) = \mathbb{E} \left(Y_{s,k}^{(g(\bar{a}_{t-1}), g(\bar{a}_{t-1}), \dots, g(\bar{a}_{t-1}))} \mid \bar{X}_{s,t}, \bar{A}_{s,t} = \bar{a}_t \right).$$

Case: $t = m$

$$\begin{aligned} & \mathbb{E} \left(H_{s,k}(m) \mid \bar{X}_{s,m}, \bar{A}_{s,m} = \bar{a}_m \right) \\ &= \mathbb{E} \left(\frac{H_{s,k}(m+1)}{\gamma_k^m(\bar{X}_{s,m}, \bar{A}_{s,m})} \mid \bar{X}_{s,m}, \bar{A}_{s,m} = \bar{a}_m \right) \\ &= \frac{\mathbb{E} \left(\mathbb{E} \left[H_{s,k}(m+1) \mid \bar{X}_{s,m+1}, \bar{A}_{s,m+1} \right] \mid \bar{X}_{s,m}, \bar{A}_{s,m} = \bar{a}_m \right)}{\gamma_k^m(\bar{X}_{s,m}, \bar{A}_{s,m})} \\ &= \frac{\mathbb{E} \left(\mathbb{E} \left[Y_{s,k}^{(g(\bar{a}_{m-1}), g(\bar{a}_m), \dots, g(\bar{a}_m))} \mid \bar{X}_{s,m+1}, \bar{A}_{s,m+1} \right] \mid \bar{X}_{s,m}, \bar{A}_{s,m} = \bar{a}_m \right)}{\gamma_k^m(\bar{X}_{s,m}, \bar{A}_{s,m})} \\ &= \frac{\mathbb{E} \left(Y_{s,k}^{(g(\bar{a}_{m-1}), g(\bar{a}_m), \dots, g(\bar{a}_m))} \mid \bar{X}_{s,m}, \bar{A}_{s,m} = \bar{a}_m \right)}{\gamma_k^m(\bar{X}_{s,m}, \bar{A}_{s,m})} \\ &= \mathbb{E} \left(Y_{s,k}^{(g(\bar{a}_{m-1}), g(\bar{a}_{m-1}), \dots, g(\bar{a}_{m-1}))} \mid \bar{X}_{s,m}, \bar{A}_{s,m} = \bar{a}_m \right), \end{aligned}$$

where we use the Law of Iterated Expectations for the third and fifth lines, the fourth line is a result of the induction hypothesis, and the last line is a result of the definition of γ_k^m as given by Equation (4.1). So Equation (4.4) holds for $t = m$. Since k was not chosen specifically, Theorem 4 holds by backwards induction. \square

C.2 Unbiased Estimating Equations

Proof of Theorem 5.

$$\begin{aligned} & \mathbb{E} \left(\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) H_{s,k}(t) (g(\bar{A}_{s,t}) - \lambda_{s,t}) \right) \\ &= \mathbb{E} \left(\mathbb{E} \left[\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) H_{s,k}(t) (g(\bar{A}_{s,t}) - \lambda_{s,t}) \mid \bar{X}_{s,t}, \bar{A}_{s,t} \right] \right) \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left(\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) \mathbb{E} \left[H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t} \right] (g(\bar{A}_{s,t}) - \lambda_{s,t}) \right) \\
&= \mathbb{E} \left(\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) \underbrace{\mathbb{E} \left[H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right]}_{\text{Assumption 4.2 and Theorem 4}} (g(\bar{A}_{s,t}) - \lambda_{s,t}) \right) \\
&= \mathbb{E} \left(\mathbb{E} \left[\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) \mathbb{E} \left[H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right] \right. \right. \\
&\quad \left. \left. \times (g(\bar{A}_{s,t}) - \lambda_{s,t}) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right] \right) \\
&= \mathbb{E} \left(\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \bar{q}_k^t(\bar{X}_{s,t}, \bar{A}_{s,t-1}) \mathbb{E} \left[H_{s,k}(t) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right] \right) \\
&\quad \times \mathbb{E} \left[(g(\bar{A}_{s,t}) - \lambda_{s,t}) \mid \bar{X}_{s,t}, \bar{A}_{s,t-1} \right] \\
&= \vec{0},
\end{aligned}$$

□

where the first and the fourth equalities hold by the Law of Iterated Expectations, the third equality holds by Assumption 4.2 and Theorem 4, and the last line holds for a correctly specified model for $\lambda_{s,t}$.

C.3 One-parameter Model with Binary Treatment

Since the one-parameter model for the treatment effect γ_k^t is linear in ψ , we can solve the unbiased estimating equations for a closed form solution for ψ .

$$\begin{aligned}
0 &= \sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} \left(\frac{Y_{s,k}}{\exp(\hat{\psi})} \mathbb{I}\{k > T_s^*\} + Y_{s,k} \mathbb{I}\{k \leq T_s^*\} \right) \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t}) \\
&= \sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \exp(-\hat{\psi}) \mathbb{I}\{k > T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t}) \\
&\quad + \sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k \leq T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t})
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \exp(-\hat{\psi}) \sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k > T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t}) \\
&= - \sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k \leq T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t}) \\
\Rightarrow \exp(-\hat{\psi}) &= - \frac{\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k \leq T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t})}{\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k > T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t})} \\
\Rightarrow \hat{\psi} &= -\log \left(- \frac{\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k \leq T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t})}{\sum_s \sum_{k=1}^K \sum_{t=0}^{k-1} Y_{s,k} \mathbb{I}\{k > T_s^*\} \mathbb{I}\{\bar{A}_{s,t-1} = \bar{0}\} (A_{s,t} - \hat{p}_{s,t})} \right).
\end{aligned}$$

References

21 U.S. Code §841 (1986).

Abouk, R., Pacula, R. L., and Powell, D. (2019). Association between state laws facilitating pharmacy distribution of naloxone and risk of fatal overdose. *Journal of the American Medical Association Internal Medicine*, 179(6):805–811.

Aronow, P. M., Samii, C., et al. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947.

Athey, S., Eckles, D., and Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240.

Baca, C. T. and Grant, K. J. (2007). What heroin users tell us about overdose. *Journal of Addictive Diseases*, 26(4):63–68.

Baker, P., Beletsky, L., Avalos, L., Venegas, C., Rivera, C., Strathdee, S. A., and Cepeda, J. (2019). Policing practices and hiv risk among people who inject drugs—a systematic literature review. *Available at SSRN 3401985*.

Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M., and Emch, M. E. (2020). Causal inference from observational studies with clustered interference, with application to a cholera vaccine study. *The Annals of Applied Statistics*, 14(3):1432–1448.

- Beletsky, L., Grau, L. E., White, E., Bowman, S., and Heimer, R. (2011). The roles of law, client race and program visibility in shaping police interference with the operation of us syringe exchange programs. *Addiction*, 106(2):357–365.
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.
- Besag, J. and Kempton, R. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics*, pages 231–251.
- Bickel, P. J. and Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes I-II package*. Chapman and Hall/CRC.
- Borusyak, K. and Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.
- Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.
- Carroll, J. J., Ostrach, B., Wilson, L., Getty, R., Bennett, J., and Dunlap, J. L. (2021). Drug induced homicide laws may worsen opioid related harms: an example from rural north carolina. *International Journal of Drug Policy*, 97:103406.
- Cox, D. R. (1958). *Planning of experiments*. Wiley.
- David, O. and Kempton, R. A. (1996). Designs for interference. *Biometrics*, pages 597–606.
- De Chaisemartin, C. and d’Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

- De Chaisemartin, C. and d'Haultfoeuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2):999–1028.
- Eckles, D., Karrer, B., and Ugander, J. (2017). Design and analysis of experiments in networks: reducing bias from interference. *Journal of Causal Inference*, 5(1).
- Feng, P., Zhou, X.-H., Zou, Q.-M., Fan, M.-Y., and Li, X.-S. (2012). Generalized propensity score for estimating the average treatment effect of multiple treatments. *Statistics in Medicine*, 31(7):681–697.
- Ferrão, M. E. and Goldstein, H. (2014). Adjusting for differential misclassification in multilevel models: the relationship between child exposure to smoke and cognitive development. *Quality & Quantity*, 48(1):251–258.
- Forastiere, L., Airoidi, E. M., and Mealli, F. (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918.
- Friedman, J., Beletsky, L., and Schriger, D. L. (2020a). Overdose-related cardiac arrests observed by emergency medical services during the us covid-19 epidemic. *Journal of the American Medical Association Psychiatry*.
- Friedman, J., Syvertsen, J. L., Bourgois, P., Bui, A., Beletsky, L., and Pollini, R. (2020b). Intersectional structural vulnerability to abusive policing among people who inject drugs: a mixed methods assessment in california’s central valley. *International Journal of Drug Policy*, 87:102981.
- Frisch, R. and Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, pages 387–401.

- Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, pages 467–480.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC press.
- Health in Justice Action Lab (2021). Drug induced homicide. <https://www.healthinjustice.org/drug-induced-homicide>.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.
- Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.
- Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.
- Kershner, R. P. and Federer, W. T. (1981). Two-treatment crossover designs for estimating a variety of effects. *Journal of the American Statistical Association*, 76(375):612–619.
- Kesmodel, U. S. (2018). Information bias in epidemiological studies with a special focus on obstetrics and gynecology. *Acta Obstetricia et Gynecologica Scandinavica*, 97(4):417–423.

- LaSalle, L. (2017). An overdose death is not murder: why drug-induced homicide laws are counterproductive and inhumane. *Drug Policy Alliance*.
- Lechner, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric evaluation of labour market policies*, pages 43–58. Springer.
- Lechner, M. (2002). Program heterogeneity and propensity score matching: an application to the evaluation of active labor market policies. *Review of Economics and Statistics*, 84(2):205–220.
- Li, P. and Redden, D. T. (2015). Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes. *Statistics in Medicine*, 34(2):281–296.
- Li, W., Sussman, D. L., and Kolaczyk, E. D. (2021). Causal inference under network interference with noise. *arXiv preprint arXiv:2105.04518*.
- Liu, L., Hudgens, M. G., Saul, B., Clemens, J. D., Ali, M., and Emch, M. E. (2019). Doubly robust estimation in observational studies with partial interference. *Stat*, 8(1):e214.
- Lok, J., Gill, R., Van Der Vaart, A., and Robins, J. (2004). Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica*, 58(3):271–295.
- Lok, J. J. (2007). Structural nested models and standard software: A mathematical foundation through partial likelihood. *Scandinavian Journal of Statistics*, 34(1):186–206.

- Lok, J. J. and DeGruttola, V. (2012). Impact of time to start treatment following infection with application to initiating haart in hiv-positive patients. *Biometrics*, 68(3):745–754.
- Lok, J. J. et al. (2021). Optimal estimation of coarse structural nested mean models with application to initiating art in hiv infected patients. *arXiv preprint arXiv:2106.12677*.
- Lopez, M. J. and Gutman, R. (2017). Estimation of causal effects with multiple treatments: a review and new ideas. *Statistical Science*, pages 432–454.
- McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, 32(19):3388–3414.
- McClellan, C., Lambdin, B. H., Ali, M. M., Mutter, R., Davis, C. S., Wheeler, E., Pemberton, M., and Kral, A. H. (2018). Opioid-overdose laws association with opioid use and overdose mortality. *Addictive Behaviors*, 86:90–95.
- Miettinen, O. S. (1974). Proportion of disease caused or prevented by a given exposure, trait or intervention. *American Journal of Epidemiology*, 99(5):325–332.
- Moradzadeh, R., Mansournia, M. A., Baghfalaki, T., Nadrian, H., Gustafson, P., and McCandless, L. C. (2018). The impact of maternal smoking during pregnancy on childhood asthma: adjusted for exposure misclassification; results from the national health and nutrition examination survey, 2011–2012. *Annals of Epidemiology*, 28(10):697–703.
- New Jersey Revised Statutes §2C:35-1.1 (1987).

- Neyman, J. S. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. (translated and edited by dm dabrowska and tp speed, statistical science (1990), 5, 465-480). *Annals of Agricultural Sciences*, 10:1–51.
- Pacula, R. L., Powell, D., Heaton, P., and Sevigny, E. L. (2015). Assessing the effects of medical marijuana laws on marijuana use: the devil is in the details. *Journal of Policy Analysis and Management*, 34(1):7–31.
- People v. Boand, 838 N.E.2d (2005).
- Pollini, R. A., McCall, L., Mehta, S. H., Celentano, D. D., Vlahov, D., and Strathdee, S. A. (2006). Response to overdose among injection drug users. *American Journal of Preventive Medicine*, 31(3):261–264.
- Powell, D., Pacula, R. L., and Jacobson, M. (2018). Do medical marijuana laws reduce addictions and deaths related to pain killers? *Journal of Health Economics*, 58:29–42.
- Rassen, J. A., Shelat, A. A., Franklin, J. M., Glynn, R. J., Solomon, D. H., and Schneeweiss, S. (2013). Matching by propensity score in cohort studies with three treatment groups. *Epidemiology*, pages 401–409.
- Rassen, J. A., Solomon, D. H., Glynn, R. J., and Schneeweiss, S. (2011). Simultaneously assessing intended and unintended treatment effects of multiple treatment options: a pragmatic “matrix design”. *Pharmacoepidemiology and Drug Safety*, 20(7):675–683.
- Robins, J. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79(2):321–334.

- Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods*, 23(8):2379–2412.
- Robins, J. M. (1998). Structural nested failure time models. *Encyclopedia of Biostatistics*, 6:4372–4389.
- Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese*, pages 151–179.
- Robins, J. M., Blevins, D., Ritter, G., and Wulfsohn, M. (1992). G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of aids patients. *Epidemiology*, pages 319–336.
- Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.
- Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Roth, J., Sant’Anna, P. H., Bilinski, A., and Poe, J. (2022). What’s trending in difference-in-differences? a synthesis of the recent econometrics literature. *arXiv preprint arXiv:2201.01194*.

- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.
- Rubin, D. B. (1980). Randomization analysis of experimental data: the fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593.
- Schneider, K. E., Park, J. N., Allen, S. T., Weir, B. W., and Sherman, S. G. (2020). Knowledge of good samaritan laws and beliefs about arrests among persons who inject drugs a year after policy change in baltimore, maryland. *Public Health Reports*, 135(3):393–400.
- Snow, J. (1856). On the mode of communication of cholera. *Edinburgh Medical Journal*, 1(7):668.
- Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.
- Sussman, D. L. and Airoldi, E. M. (2017). Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*.
- Tchetgen, E. J. T. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75.
- Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the*

- 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337. ACM.
- U.S. Census (Accessed: August 2018). 2010 census regions and divisions of the united states. <https://www.census.gov/geographies/reference-maps/2010/geo/2010-census-regions-and-divisions-of-the-united-states.html>.
- Van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 1. Cambridge university press.
- Vansteelandt, S. and Joffe, M. (2014). Structural nested models and g-estimation: the partially realized promise. *Statistical Science*, 29(4):707–731.
- Volkow, N. (2020). Collision of the covid-19 and addiction epidemics. *Annals of Internal Medicine*.
- Wakeman, S. E., Green, T. C., and Rich, J. (2020). An overdose surge will compound the covid-19 pandemic if urgent action is not taken. *Nature Medicine*, pages 1–2.
- Walker, J. (2017). Prosecutors treat opioid overdoses as homicides, snagging friends, relatives. <https://www.wsj.com/articles/prosecutors-treat-opioid-overdoses-as-homicides-snagging-friends-relatives-1513538404>.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- Xu, J., Murphy, S., and Kochanek, K. (2021). Deaths: Final data for 2019. *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 70(8):1–87.
- Yang, S. and Lok, J. J. (2018). Sensitivity analysis for unmeasured confounding in coarse structural nested mean models. *Statistica Sinica*, 28(4):1703.

CURRICULUM VITAE

