Introduction to Longitudinal Modified Treatment Policies

A solution for studying complex, continuous, and/or time-varying exposures

Kat Hoffman

2024-02-15

Overview

- Discussing a tutorial paper on Longitudinal Modified Treatment Policies
- Target audience: epidemiologists and applied statisticians
- Based on methodology proposed in *Díaz et al. (2021)*

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 118, NO. 542, 846–857: Theory and Methods
https://doi.org/10.1080/01621459.2021.1955691

Nonparametric Causal Effects Based on Longitudinal Modified Treatment Policies

Iván Díaz[a], Nicholas Williams[a], Katherine L. Hoffman[a], and Edward J. Schenck[b]

[a]Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine, New York; [b]Division of Pulmonary & Critical Care Medicine, Department of Medicine, Weill Cornell Medicine, New York

ABSTRACT

Most causal inference methods consider counterfactual variables under interventions that set the exposure to a fixed value. With continuous or multi-valued treatments or exposures, such counterfactuals may be of little practical interest because no feasible intervention can be implemented that would bring them about. Longitudinal modified treatment policies (LMTPs) are a recently developed nonparametric alternative that yield effects of immediate practical relevance with an interpretation in terms of meaningful interventions such as reducing or increasing the exposure by a given amount. LMTPs also have the advantage that they can be designed to satisfy the positivity assumption required for causal inference. We present a novel sequential regression formula that identifies the LMTP causal effect, study properties of the LMTP statistical estimand such as the efficient influence function and the efficiency bound, and propose four different estimators. Two of our estimators are efficient, and one is sequentially doubly robust in the sense that it is consistent if, for each time point, either an outcome regression or a treatment mechanism is consistently estimated. We perform numerical studies of the estimators, and present the results of our motivating study on hypoxemia and mortality in intubated Intensive Care Unit (ICU) patients. Software implementing our methods is provided in the form of the open source R package `lmtp` freely available on GitHub (*https://github.com/nt-williams/lmtp*) and CRAN.

ARTICLE HISTORY
Received June 2020
Accepted July 2021

KEYWORDS
Continuous exposures; Longitudinal data; Modified treatment policies; Sequential double robustness; Targeted minimum loss-based estimation

Methodology motivation

- Many causal inference methods and tutorials focus on binary exposures at a single time point
  - Continuous/multi-level categorical exposures common in applied research, but methods, software, teaching materials are more limited
  - Many studies have time-varying exposures, but methods are even less common

    If there are time-dependent confounders, we require special methods to appropriately estimate treatment effects

Methodology motivation

- Positivity assumption is essential to causal inference
  - Violations are common in cases of categorical and continuous exposures
  - Violations are exacerbated when there are multiple time points
- An active area of statistical research is defining, identifying, and estimating alternative causal estimands which may increase the likelihood of meeting the positivity assumption

One solution: Longitudinal Modified Treatment Policies (LMTPs)

- Diaz et al. (2021) proposed longitudinal interventions which depend on an individual's *natural value of treatment*
  - Natural value of treatment: the value treatment would take at time t if an intervention was discontinued right before time t
  - Provided identification result and doubly/sequentially robust estimation algorithms
- Methodology generalizes static, dynamic, and some stochastic interventions, so can accommodate:
  - binary, categorical, continuous, and multiple exposures
  - binary, continuous, time-to-event outcomes, competing risks, informative right-censoring, clustering
  - point-in-time and time-varying settings
- LMTPs help address violations of the positivity assumption, because we define an alternative interventions for which positivity holds by design

Tutorial organization

1. Review static and dynamic interventions, and introduce (longitudinal) modified treatment policies
2. High-level theory:
   - Identification in point-in-time and time-varying settings
   - Estimation procedures
3. Application:
   - Provide examples of research questions that could be (or already have!) been addressed using LMTPs
   - Illustrate application of an LMTP to estimate the effect of intubation timing on mortality in COVID-19 patients, using a real-world longitudinal observational data set

Section 1

Notation and setup

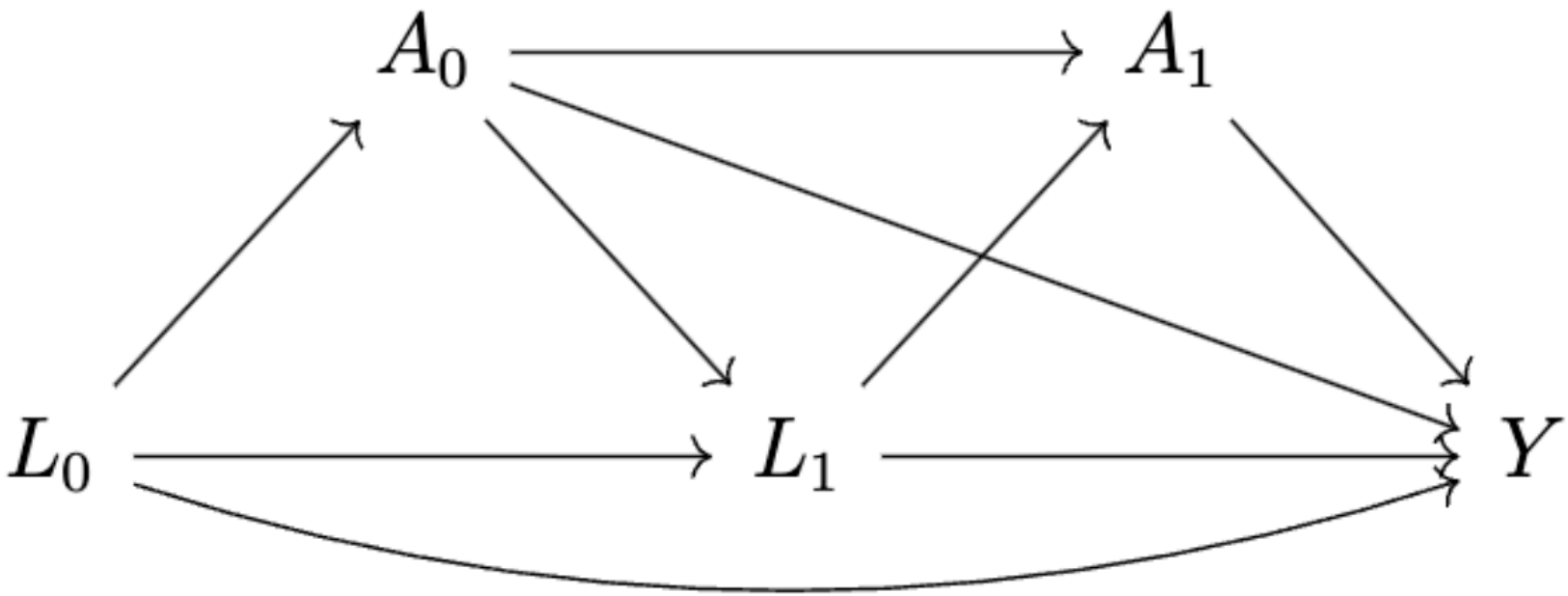
Notation

- $Z_1, ..., Z_n \overset{\text{iid}}{\sim} \mathsf{P}$
  - P represents a longitudinal process and may contain any number of time points, but for simplicity we will describe a distribution with only two time points, $t \in \{0, 1\}$
  - For each unit in the study, we observe a set of random variables $Z = (L_0, A_0, L_1, A_1, Y)$

Notation	Description	Structural Causal Equation
L_0	Baseline covariates	$L_0 \leftarrow f_{L_0}(U_{L_0})$
A_0	Treatment at first time point	$A_0 \leftarrow f_{A_0}(L_0, U_{A_0})$
L_1	Time-varying covariates	$L_1 \leftarrow f_{L_1}(L_0, A_0, U_{L_1})$
A_1	Treatment at second time point	$A_1 \leftarrow f_{A_1}(L_0, A_0, L_1, U_{A_1})$
Y	Outcome at defined study period end	$Y \leftarrow f_Y(L_0, A_0, L_1, A_1, U_Y)$

Directed Acyclic Graph (DAG)

Simple DAG omitting unmeasured confounders:



Could have many more time points, high dimensional variables, competing events, censoring nodes etc.

Intervention notation

- H_t: history of data measured up to right before A_t
  - $H_0 = L_0$
  - $H_1 = (L_0, A_0, L_1)$
- Conceptualize treatment policies in terms of hypothetical interventions on nodes of the DAG
- Interventions: consider a user-given function $\mathrm{d}_0(a_0, h_0, \epsilon_0)$ which maps a treatment value a_0, a history h_0, and a possible randomizer ϵ_0 into a potential treatment value

Intervention notation

- Intervention $\mathsf{d}_0(a_0, h_0, \epsilon_0)$ defined by removing node A_0 from the DAG and replacing it with $A_0^{\mathsf{d}} = \mathsf{d}_0(A_0, H_0, \epsilon_0)$
  - This assignment generates counterfactual data:
    $H_1(A_0^{\mathsf{d}})$: counterfactual history
    $A_1(A_0^{\mathsf{d}})$: natural value of treatment, i.e. the value that treatment would have taken if the intervention is performed at time $t = 0$ but discontinued thereafter
- At time $t = 1$, the intervention is defined by a function $\mathsf{d}_1(a_1, h_1, \epsilon_1)$
  - However, at $t = 1$ (and all subsequent times if there are more than two time points), the function must be applied to both the natural value of treatment *and* the counterfactual history
  - Remove node A_1 from the DAG and replacing it with $A_1^{\mathsf{d}} = \mathsf{d}_1(A_1(A_0^{\mathsf{d}}), H_1(A_0^{\mathsf{d}}), \epsilon_1)$
- We refer to these longitudinal interventions, and the subsequent methods to identify and estimate effects under such interventions, as LMTPs

Section 2

Review of static and dynamic interventions

Static interventions

- All units receive the same treatment
  - For two time points, conceptualize a hypothetical world in which all units are treated at both time points ($\mathsf{d}_t = 1$ for $t \in \{0, 1\}$)
  - Contrast to a hypothetical world in which no units are treated at either time point ($\mathsf{d}_t = 0$ for $t \in \{0, 1\}$)
  - Gives rise to the well-known Average Treatment Effect (ATE)

Static intervention examples

- Hypothetical intervening on a population to:
  - treat everyone with a drug versus treat no one with a drug
  - enforce 30 minutes of moderate exercise for all individuals, every day
  - give all individuals an exact level of antibodies for a certain disease
  - setting a certain level of air quality each day, for all geographical areas of interest

Dynamic interventions

- Intervention depends only on a study unit's past covariates
  - Can include past treatment
- Often used in observational studies when study units need to meet an indication of interest for a treatment or policy to reasonably begin, e.g.
  - severity of illness indicator
  - socioeconomic threshold to begin a policy

Dynamic interventions

One of the first uses of dynamic interventions was in the context of HIV, where investigators were interested in the effect of initiating antiretroviral therapy for a person with HIV if their CD4 count falls below a threshold, e.g. 200 cells/μl (Hernán et al. 2006)

$$\mathsf{d}_t(h_t) = \begin{cases} 1 & \text{if } l_t^* < 200 \text{ for all } s \geq t \\ 0 & \text{otherwise,} \end{cases}$$

where L_t^* is a variable in H_t that denotes CD4 T-cell count

Dynamic intervention examples

Another example is studying the effect of initiating a corticosteroids regimen for COVID-19 patients (Hoffman et al. 2022)

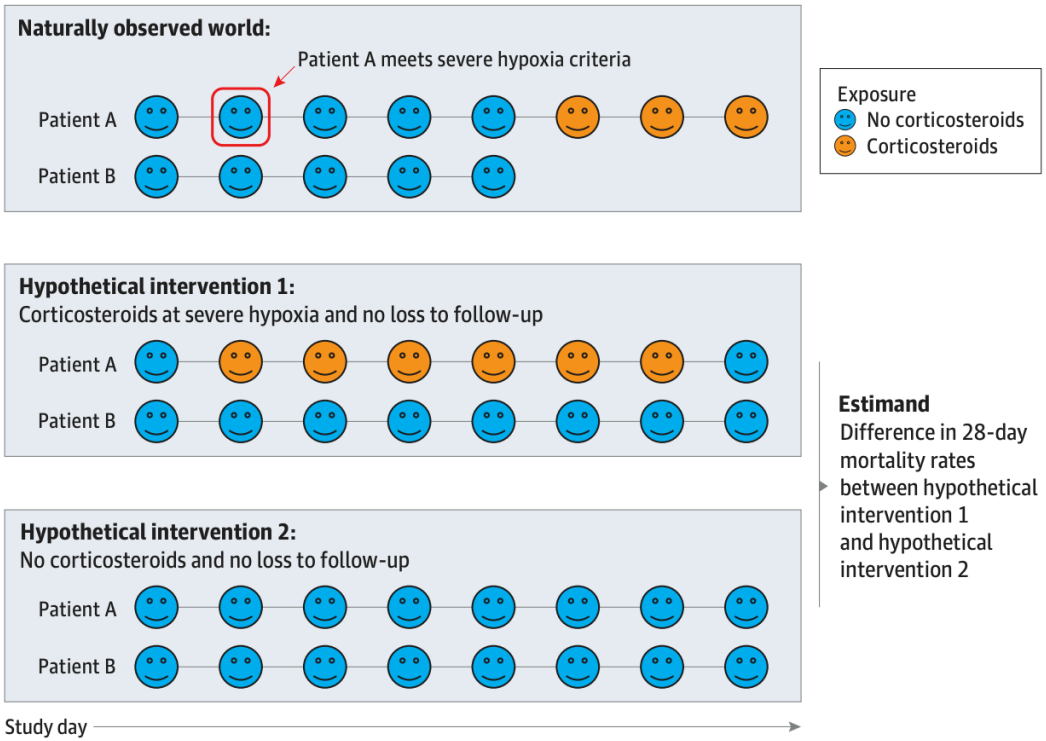
We estimated mortality under a hypothetical policy where corticosteroids are administered for six days if and when a COVID-19 patient first meets a severity of illness criteria (i.e. low levels of blood oxygen)

$$\mathsf{d}_t(h_t) = \begin{cases} 1 & \text{if } l_s^* = 1 \text{ for any } s \in \{t-5, \dots, t\} \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where L_t^* is a variable in H_s that denotes the first instance of low levels of blood oxygen.

Dynamic intervention examples

Figure 1. Illustrated Example of 2 Patients Under the 2 Hypothetical Treatment Regimens of the Target Trial Emulation



Patient A reached severe hypoxia criteria at study day 2 and was followed the entire study duration. Patient B never reached severe hypoxia criteria and was lost to follow-up after 5 study days. Under the dynamic corticosteroids regimen (intervention 1), patient A received 6 days of corticosteroids, and under intervention 2, they received no corticosteroids. Patient B did not receive corticosteroids under either treatment regimen; however, in both hypothetical worlds, they were observed for the entire study duration.

Section 3

Modified Treatment Policies

Modified Treatment Policy

- While static and dynamic interventions are considered **deterministic**, MTPs are part of more general class of **stochastic** interventions
- Intervention function $\mathrm{d}_t(a_t, h_t, \epsilon_t)$ non-trivially depends on the natural value of treatment a_t, and perhaps on h_t and/or ϵ_t
- Various types of MTPs have been proposed over the years (sometimes called "interventions which depend on the natural value of treatment")

MTP examples: threshold functions

Threshold function: all natural exposure values which fall outside of a certain boundary are intervened upon to meet a constant value

- Could be used to assess the effect of lifestyle interventions, for example, intervening on individuals' average number of drinks per week and estimating the risk of coronary heart disease (Taubman et al. 2009)
- If we categorize drinks per week as 1 = "none," 2 = "1-5," 3 = "6-10," 4 = "11-15," and 5 = ">25", and we intervene to lower all individuals in the highest two drinks-per-week categories to "6-10," we can consider that intervention in notation as,

$$\mathsf{d}_t(a_t) = \begin{cases} a_t & \text{if } a_t < 4 \\ 3 & \text{otherwise.} \end{cases}$$

MTP examples: stochastic policy

- Another example of an MTP: a hypothetical policy in which half of all current smokers quit smoking forever (Robins, Hernán, and Siebert 2004)
  - Motivated by the infeasibility of studying a world in which all current smokers quit smoking forever, since genetics, environment, and many other factors will always create some portion of current smokers who will never quit
  - Letting A_t denote a random variable denoting smoking and ϵ_t a random draw from a uniform distribution in $(0, 1)$,

$$\mathsf{d}_t(a_t, \epsilon_t) = \begin{cases} 0 & \text{if } \epsilon_t < 0.5 \text{ and } a_t = 1 \\ a_t & \text{otherwise,} \end{cases}$$

MTP examples: shift functions

Shift functions assign treatment by modifying the natural value of the exposure by some constant δ

- This intervention can be additive onto the exposure value, such as estimating the effect of a hypothetical intervention to reduce lung cancer resection surgeries lasting longer than 60 minutes by 15 minutes (Haneuse and Rotnitzky 2013)

$$\mathsf{d}_t(a_t) = \begin{cases} a_t & \text{if } a_t \leq 60 \\ a_t - 15 & \text{otherwise.} \end{cases}$$

MTP examples: shift functions

- A shift function can also change the exposure on a multiplicative scale
  - For example, studying the effect of an intervention which doubles the number of street lights for roads with less than 10 lights per mile on nighttime automobile accidents

$$\mathsf{d}_t(a_t) = \begin{cases} a_t & \text{if } a_t \geq 10 \\ 2a_t & \text{otherwise.} \end{cases}$$

Section 4

Causal estimand

Causal estimand

- Once an intervention is specified, the counterfactual outcomes of observations under a specific d are denoted as $Y(\bar{A}_\tau^{\mathsf{d}})$, where $\bar{A}$ indicates the history of measurements of A for all time points, i.e. $\bar{A} = (A_1, \ldots, A_\tau)$
- Causal effects are defined as a distribution of contrasts of $Y(\bar{A}_\tau^{\mathsf{d}})$ under different interventions, d' and $\mathsf{d}^\star$
- In this tutorial, we focus on $\mathsf{E}[Y(\bar{A}_\tau^{\mathsf{d}'}) - Y(\bar{A}_\tau^{\mathsf{d}^\star})]$ as our causal estimand of interest
  - The functions d' and $\mathsf{d}^\star$ may be any type of intervention, including "no intervention"

Section 5

Identification

Identification

- Now, we need to write our counterfactual expectation $\mathrm{E}[Y(\bar{A}_{\tau}^{\mathrm{d}'})]$ as a formula that depends only on the observed data distribution—i.e., an identifying formula
- This requires assumptions, some of which are untestable with the data available
  - The mathematically rigorous form of the assumptions is given elsewhere (Richardson and Robins 2013) (Díaz et al. 2021), but we state them in the tutorial in simple terms

Identification assumptions

Positivity

- If it is possible to find an observation with history h_t with an exposure of a_t, then it is also possible to find an observation with history h_t with an exposure (a_t, h_t, ϵ_t)

Strong sequential randomization

- All common causes of the intervention variable A_t and $(U_{L,t+1}, U_{A,t+1})$ are measured and recorded in H_t.
  - Generally satisfied if H_t contains all common causes of A_t and $(L_{t+1}, A_{t+1}, \ldots, L_\tau, A_\tau, Y)$, where τ is the last time point in the study

Weak sequential randomization

- All common causes of the intervention variable A_t and $(U_{L,t+1})$ are measured and recorded in H_t

Identification assumptions

- While static and dynamic interventions require the weak version of sequential randomization, LMTPs require the strong version
- Violations to the positivity assumption can be **structural** or **practical**
  - **Structural**: certain characteristics of an individual or unit which will never yield receipt of the treatment assignment under the intervention. This violation will not improve even with an infinite sample size.
  - **Practical**: Due to random chance or small datasets, there are certain covariate combinations with zero or near-zero predicted probabilities of treatment.

Positivity, cont.

- Positivity violations increase the finite bias and variance of estimates and severely threaten the validity of casual inference analyses when not addressed
- For time-varying treatments, positivity must be maintained at each time point
- By design, non-static interventions (e.g. dynamic treatment rules, MTPs) may help define estimands with plausible positivity, since the function d can be modified to affect the exposure of only observations which are not subject to positivity violations

Example of redefining an estimand

- Think of a world in which a continuous exposure is observed at some fixed value higher or lower than it was factually observed for every unit in the study
  - EX: surgery times were 15 minutes shorter for all lung resection biopsies
- This hypothetical intervention is destined for structural positivity violations, because at the lowest end of the observed exposure range, there will by definition be no support for the intervened exposure level $\mathrm{d}(a_t)$ (much less conditional on the observation's history h_t)
- This can be avoided by constraining the range of a_t affected by the hypothetical intervention, so that no $\mathrm{d}(a_t)$ values are produced outside the observed range of A
  - Could modify intervention to accommodate any other remaining structural or practical positivity violations

Section 6

Identification formula

Identification formula

- Under positivity and strong (or weak) sequential randomization assumptions, the estimand is identified by the generalized g-formula
- A re-expression of this generalized g-formula involves recursively defining the expected outcome under the intervention conditional on the observation's observed exposure and history
  - beginning at the final time point, and proceeding until the earliest time point

Identification formula

1. Start with the conditional expectation of the outcome Y given $A_1 = a_1$ and $H_1 = h_1$. Let this function be denoted $Q_1(a_1, h_1)$
2. Evaluate the above conditional expectation of Y if A_1 were changed to A_1^{d}, which results in a pseudo outcome $\tilde{Y}_1 = Q_1(A_1^{\mathsf{d}}, H_1)$
3. Let the true expectation of $\tilde{Y}_1$ conditional on $A_0 = a_0$ and $H_0 = h_0$ be denoted $Q_0(a_0, h_0)$
4. Evaluate the above expectation of $\tilde{Y}_1$ if A_0 were changed to A_0^{d}, which results in $\tilde{Y}_0 = Q_0(A_0^{\mathsf{d}}, H_0)$
5. Under the identifying assumptions, we have $\mathsf{E}[Y(\bar{A}_\tau^{\mathsf{d}})] = \mathsf{E}[\tilde{Y}_0]$

Section 7

Estimation

References

For more information:

Díaz, Iván, Nicholas Williams, Katherine L. Hoffman, and Edward J. Schenck. 2021. "Nonparametric Causal Effects Based on Longitudinal Modified Treatment Policies." *Journal of the American Statistical Association*, September, 1–16. https://doi.org/10.1080/01621459.2021.1955691.

Haneuse, Sebastian, and Andrea Rotnitzky. 2013. "Estimation of the Effect of Interventions That Modify the Received Treatment." *Statistics in Medicine* 32 (30): 5260–77.

Hernán, Miguel A, Emilie Lanoy, Dominique Costagliola, and James M Robins. 2006. "Comparison of Dynamic Treatment Regimes via Inverse Probability Weighting." *Basic & Clinical Pharmacology & Toxicology* 98 (3): 237–42.

Hoffman, Katherine L, Edward J Schenck, Michael J Satlin, William Whalen, Di Pan, Nicholas Williams, and Iván Díaz. 2022. "Comparison of a Target Trial Emulation Framework Vs Cox Regression to Estimate the Association of Corticosteroids with COVID-19 Mortality." *JAMA Network Open* 5 (10): e2234425–25.

Richardson, Thomas S, and James M Robins. 2013. "Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality." *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper* 128 (30): 65–78.