Robust Causal Inference for EHR-based Studies of Point Exposures with Missingness in Eligibility Criteria

Luke Benz¹, Rajarshi Mukherjee¹, Rui Wang^{1,2,3}, David Arterburn⁴, Heidi Fischer⁵, Catherine Lee⁶, Susan M. Shortreed^{7,8}, Sebastien Haneuse^{*1}, and Alexander W. Levis^{*9}

¹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

²Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA, USA

³Department of Population Medicine, Harvard Medical School, Boston, MA, USA

⁴Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA

⁵Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, USA

⁶Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA

⁷Biostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA ⁸Department of Biostatistics, University of Washington School of Public Health, Seattle, WA, USA ⁹Department of Statistics & Data Science, Carnegie Mellon, Pittsburgh, PA, USA

April 24, 2025

Abstract

Missingness in variables that define study eligibility criteria is a seldom addressed challenge in electronic health record (EHR)-based settings. It is typically the case that patients with incomplete eligibility information are excluded from analysis without consideration of (implicit) assumptions that are being made, leaving study conclusions subject to potential selection bias. In an effort to ascertain eligibility for more patients, researchers may look back further in time prior to study baseline, and in using outdated values of eligibility-defining covariates may inappropriately be including individuals who, unbeknownst to the researcher, fail to meet eligibility at baseline. To the best of our knowledge, however, very little work has been done to mitigate these concerns. We propose a robust and efficient estimator of the causal average treatment effect on the treated, defined in the study eligible population, in cohort studies where eligibility-defining covariates are missing at random. The approach facilitates the use of flexible machine-learning strategies for component nuisance functions while maintaining appropriate convergence rates for valid asymptotic inference. EHR data from Kaiser Permanente are used as motivation as well as a basis for extensive simulations that verify robustness properties under various degrees of model misspecification. The data are also used to demonstrate the use of the method to analyze differences between two common bariatric surgical interventions for long-term weight and glycemic outcomes among a cohort of severely obese patients with type II diabetes mellitus.

 $^{^{\}ast}$ denotes co-last author (AWL and SH). This work was supported by NIH Grants R01 DK128150-01 (LB, RM, RW, DA, CL, HF, SMS, SH) and F31 DK141237-01 (LB)

Keywords: multiply robust, influence functions, missing data, bariatric surgery, diabetes

1 Introduction

Observational studies leveraging electronic health record (EHR) databases are seen as important alternatives to randomized clinical trials, particularly when trials may be infeasible due to financial, ethical, or logistical constraints (1-3). In contrast to data collected in randomized trials, however, EHR databases exist to record clinical activity and assist with patient billing, and thus are not collected with any specific research purpose in mind. As such, information that might be routinely collected in a trial or prospective observational study may be unavailable or outdated for some patients in EHR-based studies (4).

Well designed trials and observational studies require that subjects meet a set of study eligibility criteria (inclusion/exclusion criteria) to ensure analysis occurs on the intended population of interest. While measurements necessary to determine inclusion in a research study with primary data collection can be definitively ascertained through appropriate screening prior to enrollment (5), missingness in variables that define study eligibility criteria poses a serious challenge in EHR-based studies. In practice, subjects with incomplete eligibility data are almost always excluded from analysis (6, 7), although this can result in selection bias if subjects with observed eligibility data differ systematically from those whose eligibility status can not be determined (8, 9).

Superficially, missing eligibility criteria may seem like a standard missing data problem. Unlike other missing data problems, however, the concessions analysts make when faced with missing eligibility, such as looking back further in time to ascertain eligibility, may inadvertently include individuals who, unbeknownst to the analyst, are ineligible at study baseline. Our work is motivated by the use of EHR data to understand long-term outcomes following bariatric surgery (10–15) among patients with type II diabetes mellitus (T2DM). As such, it would be inappropriate to include patients with diabetes in the study.

To date, very few papers have considered the problem of selection bias due to missing eligibility data. In early work, Pan and Schaubel (16) introduced an approach similar to inverse probability weighting (IPW) for proportional hazards regression, and Heng et al. (17) proposed a sensitivity analysis where subjects with missing eligibility criteria were treated as eligible and later ineligible in two separate analyses, though neither paper had an explicit focus on causal contrasts. More recently, Tompsett et al. (8) and Austin et al. (18) proposed the use of multiple imputation for a single eligibility defining covariate, with the latter focused on settings where that eligibility covariate was also the primary exposure of interest. Benz et al. (9) introduced inverse probability of eligibility ascertainment weights within a larger IPW framework for common challenges in EHR-based studies.

A notable limitation of the existing literature on missing eligibility criteria is the need to specify all relevant models correctly (e.g., imputation, outcome, treatment, missingness probability) to ensure consistent estimation of causal contrasts (9, 18). On the other hand, estimating component models, or "nuisance functions", in a flexible/nonparametric manner via machine learning methods often yields slow rates of convergence, and thus invalid statistical inference (19, 20). Additionally, standard machine learning models often optimize loss functions that are not tied to specific causal contrasts, such as mean squared prediction error, and thus may result in bias in causal quantities utilizing those predictions (19, 21, 22). Tools from semiparametric theory can overcome some of these challenges, leading to estimators that are robust to various degrees of model misspecification, converge to desirable asymptotic distributions which facilitate valid statistical inference, and attain nonparametric efficiency bounds (19, 20, 22–27).

In this paper, we consider the setting where interest lies in using EHR data to estimate the causal average treatment effect of a point exposure on a given outcome among patients receiving a particular treatment in a targeted population of interest, articulated through a study's eligibility criteria. This is simply the causal average treatment effect on the treated (ATT) among study eligible subjects—as will be made clear, explicit conditioning on eligibility status makes apparent how restricting analysis to subjects with complete information on eligibility defining covariates can introduce selection bias. Using a novel factorization of the observed data likelihood, we identify this causal estimand and develop various estimators based on semiparametric theory that are robust and asymptotically normal, including one which attains a semiparametric efficiency bound.

The remainder of the paper is organized as follows. Section 2 provides background on bariatric surgery and elaborates on comparisons between two procedures among a population with obesity and T2DM. A detailed exploratory data analysis that motivates careful consideration of missing eligibility criteria is presented. Section 3 formalizes notation, assumptions, and the causal estimand of interest. Section 4 presents a new efficient influence function-based estimator and details its properties, while Section 5 explores finite sample properties of this estimator and others over numerous flexible estimation strategies in a simulation study tied closely to the motivating application. The use of this method in the motivating bariatric surgery study is presented in Section 6, and finally, Section 7 concludes with discussion. Detailed proofs are provided in the Supplementary Material.

2 Motivating Application

2.1 Bariatric Surgery for Patients with Obesity and T2DM

Bariatric surgery is an intervention to mitigate obesity and related comorbidities, with typical candidates having a body mass index (BMI) exceeding 35 kg/m² (28). The primary bariatric surgery procedures are Roux-en-Y gastric bypass (RYGB) and sleeve gastrectomy (SG), with SG having surpassed RYGB in frequency in the past 15 years (10, 29, 30). SG is a simpler, less invasive procedure that is offered by more surgeons, and is associated with fewer short-term post-surgical complications (31). On the other hand, patients undergoing SG may experience greater long-term re-operation risk (32) and less substantial long-term weight loss (11). Though comparisons between RYGB and SG have been published, evidence regarding their comparative effectiveness among patients with T2DM, particularly for long-term outcomes is not considered definitive (7, 33–36). For example, the American Diabetes Association recommends consideration of bariatric surgery for T2DM patients with sufficiently large BMI as a form of weight and glycemic control, but doesn't make recommendations about whether SG or RYGB should be preferred (37).

Two important clinical outcomes of interest for T2DM patients following bariatric surgery are relative weight change and remission of diabetes. These outcomes were previously examined in an EHR-based study by McTigue et al. (7), who found advantages to RYGB over SG on both outcomes among T2DM patients between 20-79 years of age with $BMI \geq .$ However, their analysis excluded at least 10% of surgical patients due to missing eligibility criteria, the implications of which were not explored.

With this backdrop, we investigate relative weight change and remission of diabetes at 3 years post surgery for a population of 14,809 patients with obesity and T2DM undergoing

RYGB or SG at one of three Kaiser Permanente (KP) sites (KP Washington, KP Northern California, KP Southern California) between 2008-2011. Given that the rise in popularity of SG has outpaced the rate at which long-term evidence has been generated for populations with diabetes, we are particularly interested in possible long-term benefits that T2DM patients who underwent RYGB would have forgone had they instead undergone SG.

Our study utilizes EHR data from DURABLE, an NIH-funded study examining longterm outcomes of bariatric surgery across KP sites (11–13, 15, 38). Following prior work (12, 15, 38), study eligibility criteria require that patients have T2DM, a BMI \geq 35 kg/m², and age between 19-79 years. When studying T2DM remission, we additionally require that patients have a DiaRem score of at least 3 points. DiaRem is a metric computed from pre-operative patient characteristics including age, hemoglobin A1c % (A1c), and medication usage, and relates to the likelihood of experiencing T2DM remission following bariatric surgery, with lower scores indicating greater chance of remission (39). This additional eligibility restriction places particular emphasis on patients less likely to experience remission (e.g., with greater disease severity), a population where differences between RYGB and SG on T2DM remission rates may be more substantial (7).

2.2 Ascertainment of Eligibility

Following previous DURABLE studies (12), we define T2DM as a measurement of A1c $\geq 6.5\%$ (fasting blood glucose $\geq 126 \text{ mg/dl}$), or via prescription for insulin or oral hypoglycemic medication. Metformin as the sole indicator of T2DM (e.g., subject 5 in Figure 1) is insufficient to establish diabetic status, and additionally requires an ICD-9 code of 250.x. Remission of T2DM is defined following Coleman et al. (12), as a measurement of A1c < 6.5\% following a period of at least 90 days without T2DM medication, or fasting



Figure 1: Hemoglobin A1c % (dark red dots) and diabetic medication usage for six patients undergoing bariatric surgery, information that establishes T2DM status, and thus study eligibility in the 24 months prior to surgery. A1c measurements are shown in relation to a cutoff of 6.5%, the typical clinical cutoff for T2DM. For medications, points indicate the start of a prescription while shaded bars indicate duration.

blood glucose < 126 mg/dl following a period of at least 7 days without T2DM medication.

In practice however, determining which patients have T2DM at the time of surgery using EHR data is a non-trivial task. To illustrate this complexity, Figure 1 depicts relevant diabetes information for six surgical patients. Some patients have frequent observations of A1c (e.g., subjects 1 and 2) making it straightforward to classify their diabetes status, with T2DM confirmed by active insulin prescription at the time of surgery in the case of subject 1. In contrast, other subjects have little to no information on A1c prior to surgery (e.g., 3-6). Despite having no pre-surgical A1c measure, subject 3 had an active prescription at the time of surgery of an oral hypogylcemic medication, allowing positive T2DM status to be established. Subjects 4 and 5 have no measures of A1c available prior to surgery, but unlike subject 3, had no active prescriptions at the time of surgery. Finally, subject 6 has multiple A1c measures exceeding 6.5% but all measures are more than 12 months outdated

by the time of surgery. Altogether, there is substantial heterogeneity in the patterns and frequency of observed information in the EHR making ascertainment of T2DM, and thus study eligibility, easy for some patients and difficult or impossible for others.

In determining how far to look back in time to ascertain study eligibility, different papers have used lookback windows of varying lengths (7, 9, 12, 15, 38), without consideration as to how such decisions might influence study conclusions. To better explore possible tradeoffs associated with this decision, we conduct analysis over a grid of 40 combinations of lookback windows for BMI (1, 3, 6, 12 months) and T2DM ascertainment via lab values (1, 3, 6, 12, 24 months) and medications (active prescription at the date of surgery, or active within 12 months prior to surgery).

Observable information about the joint distribution of indicators for eligibility ascertainment (R) and status (E) is shown in Figure 2. Panel A shows the number of surgical patients for each possible observable pair (R, E) across combinations of lookback windows. In the shortest of lookback window (1 month for BMI and diabetes labs; active T2DM prescription at time of surgery), eligibility status is ascertainable for just 4,912 of 14,809 patients (33%) and missing for 9,897 patients (67%). By contrast, eligibility status is deemed missing for just 361 patients (3.2%) when using the longest lookback windows, which is likely the impetus for certain works adopting longer lookback windows.

For each patient, we can compute in how many of the 40 combinations they are ascertained to be eligible (n_{11}) and ineligible (n_{10}) . Panel B of Figure 2 shows the number of patients with each unique combination of (n_{11}, n_{10}) . When $n_{10} = 0$ or $n_{11} = 0$, subjects are only ever ascertained to be eligible or ineligible, respectively, under combinations where their eligibility status is not missing. Of note, for many subjects both $n_{11} > 0$ and $n_{10} > 0$. That is, several subjects are ascertained to be ineligible under certain strategies



Figure 2: A) Joint distribution of eligibility ascertainment (R) and status (E) across 40 different possible ways to operationalize the study eligibility criteria. B) Distribution of (n_{11}, n_{10}) where n_{re} denotes the number of ways of operationalizing the study eligibility criteria that a subject has R = r, E = e. A similar figure for the remission outcome is available in the Supplementary Materials.

for applying inclusion criteria but eligible in others (rather than only eligible/missing or ineligible/missing). As such, simply applying the most relaxed operationalization of the eligibility criteria may inadequately reflect certain patients' eligibility status at surgery.

2.3 Unique Challenges of Missing Eligibility Data

Figures 1 and 2 serve to highlight that missing eligibility criteria is a distinct phenomenon from other forms of missing data for a few reasons. To begin with, complete case analyses that discard patients with missing outcomes or confounders do not risk inappropriately including patients in analysis. As is apparent from Figure 2, minimizing exclusions due to missing eligibility by looking back further in time may inadvertently include patients in the study population who are not study eligible at the time of surgery (i.e., do not have T2DM). On the other hand, looking back only a short time before surgery to collect the most accurate information yields rates of missingness as substantial as 70%. While confounders missing with this frequency could plausibly be dropped from analysis and treated as unmeasured, changing the eligibility criteria inherently changes the underlying analysis population and correspondingly, the scientific question of interest. Furthermore, when eligibility is missing with this frequency and longer lookbacks are employed, it can be tricky to determine the intended population to which study conclusions are applicable.

3 Problem Set-up

3.1 Notation

Let $A \in \{0, 1\}$ denote a binary treatment (e.g., RYGB vs. SG), $Y \in \mathbb{R}$ denote an outcome of interest (e.g., relative weight change, or remission of diabetes) and $\mathbf{L} \in \mathbb{R}^k$ denote a set of baseline covariates that are (assumed) fully sufficient for three purposes: (1) to define study eligibility, (2) control for confounding, and (3) predict missingness in eligibility status. We denote eligibility defining covariates by \mathbf{L}^e and let $E = g(\mathbf{L}^e, A)$ denote a binary indicator of eligibility status, where $g(\cdot)$ is some fixed and known eligibility rule. In the context of our motivating application, \mathbf{L}^e consists of BMI, age, A1c, DiaRem score (remission outcome only) and usage of insulin or oral hypogylcemic medications, and the eligibility defining rule is BMI $\geq 35 \text{ kg/m}^2$, $19 \leq \text{age} \leq 79$, and T2DM.

It is frequently the case that components of L^e are missing which precludes ascertainment of eligibility status E (Figure 1). We denote the subset of eligibility defining covariates with any missingness by $L_m^e \in \mathbb{R}^q$, and let $L^* = L \setminus L_m^e$. For example, age is an eligibility defining covariate but is never missing in our application, and thus is in L^* , while remaining eligibility defining covariates are missing to various degrees and therefore are part of L_m^e . Remaining covariates L^* include KP site, race, sex, age, estimated glomerular filtration rate (eGFR), self-reported smoking status, hypertension, dyslipidemia, and calender year of surgery. Finally, we let R be a binary indicator for each patient for whether or not the set of covariates L_m^e is completely observed, and consider a coarsened version of the observed data, with data units given by $O = (L^*, A, Y, R, RL_m^e)$. Some commentary on this possible discarding of information is offered in the discussion.

We observe a sample (O_1, \ldots, O_n) of n independent and identically distributed observations from some underlying and unknown data distribution P, with corresponding empirical distribution \mathbb{P}_n . For any P-integrable function f we define $Pf = \mathbb{E}_P[f] = \int f \, dP$ to denote the expectation of f under P, and analogously define $\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(O_i)$ to express the sample mean of f on data units (O_1, \ldots, O_n) . Lastly, we let Y(a) denote counterfactual outcomes, that is the outcome that would have been observed under treatment A = a (40). To simplify the exposition, we assume that there is no missing data outside of L_m^e (particularly in A, Y or L^*). In the discussion, we offer some considerations for situations where missing data also poses a challenge in outcomes and/or non-eligibility defining covariates.

3.2 Average Treatment Effect on the Treated and Study Eligible

The causal estimand of interest is the average treatment effect on treated individuals meeting inclusion criteria for the study eligible population (ATTE), which we define as

$$\theta_{\text{ATT}}^{\text{elig}} = \mathbb{E}[Y(1) - Y(0) \mid A = 1, E = 1]$$
(1)

Note, this is the estimand of interest for researchers targeting the ATT after applying carefully constructed inclusion/exclusion criteria. Explicitly conditioning on eligibility status E = 1 in Equation (1) makes clear how the target population (inadvertently) can change when those missing eligibility are discarded. For example, when researchers discard subjects without ascertainable eligibility status (R = 0), they are implicitly targeting the causal parameter $\theta_{ATT}^{CC} = \mathbb{E}[Y(1) - Y(0) \mid A = 1, E = 1, R = 1]$. In cases where $\theta_{ATT}^{CC} \neq \theta_{ATT}^{elig}$, that is, where the treatment effect among ascertainably eligible treated subjects differs from the treatment effect among all treated subjects meeting the study eligibility criteria, we say there is selection bias with respect to the ATTE.

3.3 Assumptions

We articulate two sets of assumptions to simultaneously address the potential for both selection bias and confounding. The first set assumptions are standard causal inference assumptions and need only hold among the study eligible population. This is particularly salient for other applications comparing treatment to no-treatment, where study eligibility is often designed to exclude patients who are not realistic candidates for treatment.

A1 (Consistency) Y(A) = Y when E = 1, almost surely

A2 (Positivity)
$$\exists \epsilon > 0$$
 such that $\epsilon \leq P(A = 1 \mid L^*, L_m^e, E = 1) \leq 1 - \epsilon$, almost surely

A3 (No Unmeasured Confounding)
$$Y(a) \perp A \mid L^*, L_m^e, E = 1 \text{ for } a \in \{0, 1\}$$

In the absence of missing eligibility, Assumptions 1-3 are sufficient to identify θ_{ATT}^{elig} , for example via the *g*-formula, $\mathbb{E}_P[Y \mid A = 1, E = 1] - \mathbb{E}_P[\mathbb{E}_P[Y \mid A = 0, \mathbf{L}^*, \mathbf{L}_m^e, E = 1] \mid A = 1]$ (40, 41). Given that \mathbf{L}_m^e , and thus *E*, may be missing, additional assumptions are required.

A4 (Eligibility Missing At Random) $R \perp\!\!\!\perp (Y, \boldsymbol{L}_m^e) \mid \boldsymbol{L}^*, A$

A5 (Complete Case Positivity) $\exists \epsilon > 0$ such that $\epsilon \leq P(R = 1 \mid L^*, A)$, almost surely

Assumption 4 is the core missing data assumption which states that whether or not a patient's eligibility status is observed is jointly independent of outcome and eligibility defining covariates, conditional on all other completely observed information. While the plausibility of various assumptions is application specific, Assumption 4 is most likely to be violated when there are either unobserved drivers of missingness or when the actual values of L_m^e influence how closely a patient is followed by their doctor.

Bariatric surgery is commonly preceded by a 3-6 month period in which a patient has greater levels of interaction with the health care system in an effort to monitor and encourage possible preoperative lifestyle changes, including weight loss and smoking cessation (42). Thus, it seems plausible that whether or not measures of BMI and A1c or medication usage are collected during this preoperative period would depend only on information recorded in the EHR. Furthermore postoperative outcomes 3 years after surgery seem unlikely to be related to the availability of BMI/T2DM information at the time of surgery except through patient demographics, or factors widely available in pre-surgical workup, such as diagnoses of obesity-related comorbidities like hypertension and dyslipidemia, smoking status, and measures of kidney function like serum creatinine and eGFR.

Alternative MAR assumptions and analysis strategies they might motivate, including the "removal" of Y from Assumption 4, are examined in the discussion. Though Assumptions 1-5 may be reasonably satisfied in our motivating study of bariatric surgery, the broad class of observational studies satisfying these assumptions generally only requires that there are no unmeasured confounders of treatment selection or drivers of missingness.

In theory, any given covariate could be relevant for only a single of the three stated purposes of L. As articulated in Assumptions 2-3, all of L is taken to be necessary insofar as the control of confounding is concerned. We feel this is appropriate as there exists ample evidence to suggest each covariate in L_m^e and L^* is a confounder for T2DM related outcomes following bariatric surgery (7, 12, 15, 38). Furthermore, in EHR-based studies, the exact set of covariates which predict missingness is unlikely to be fully understood by researchers, and thus it would be reasonable for analysts to utilize all available information in L^* when trying to account for missingness in L^e .

It is plausible that in other applications, a more refined partitioning of \boldsymbol{L} which further distinguishes confounders from covariates related only to eligibility status or ascertainment could yield slightly more parsimonious versions of Assumptions 2-5. In the Supplementary Materials, we offer some exploration of an alternative strategy to further partition \boldsymbol{L} , and comment on additionally complexities introduced.

One example of a covariate serving only a single of L's stated purposes occurs when study eligibility restricts to a single stratum of a categorical eligibility defining covariate (e.g., positive T2DM). In such cases, that covariate would no longer be a confounder among the study eligible population. Though we choose not to distinguish this further partitioning of L_m^e for ease of notation in subsequent sections, analysts might consider dropping covariates of this form when modeling component nuisance functions outlined in Table 1.

3.4 Factorization of the Observed Data Likelihood

In the presence of missing data, MAR assumptions allow for decompositions of the observed data likelihood that facilitate identification and estimation strategies for causal contrasts of interest (26). In the present context, Assumptions 4-5 motivate the following factorization:

$$p(O) = p(\mathbf{L}^*)p(A \mid \mathbf{L}^*)p(R \mid \mathbf{L}^*, A)p(\mathbf{L}_m^e \mid \mathbf{L}^*, A, R = 1)^R p(Y \mid \mathbf{L}^*, \mathbf{L}_m^e, A, R = 1)$$
(2)

Each component of Equation 2 is tied closely to nuisance functions which will play a critical role through the remainder of the work. Namely, $\pi(\mathbf{L}^*) = P(A = 1 | \mathbf{L}^*)$ is a regression for treatment probability conditional on non-eligibility defining covariates, and can be thought of as partial propensity score; $\eta(\mathbf{L}^*, A) = P(R = 1 | \mathbf{L}^*, A)$ is regression for being a complete case, which depends only on fully observed quantities; $\lambda_a(\mathbf{L}^e_m; \mathbf{L}^*) = p(\mathbf{L}^e_m | \mathbf{L}^*, A = a, R = 1)$ denotes the conditional density of partially observed eligibility defining covariates given treatment and fully observed covariates, among patients with complete information; and $\mu_a(\mathbf{L}^*, \mathbf{L}^e_m) = \mathbb{E}_P[Y | A = a, \mathbf{L}^*, \mathbf{L}^e_m, R = 1] =$ $\int y \, dP(y | \mathbf{L}^*, \mathbf{L}^e_m, A = a, R = 1)$ is the mean outcome conditional on all other information, among complete cases. Table 1 provides a summary of how these nuisance functions, as well as others introduced below, are used in various estimation strategies for the ATTE.

While alternative MAR assumptions could motivate alternative ways of factorizing the observed data likelihood, the factorization in Equation 2 has several appealing properties. Most important is that no component of the likelihood factorization conditions on a covariate stratum with incomplete information. As such, each component corresponds to a nuisance function which is readily estimable from the observed data. Moreover, by construction, components of the factorization are variationally independent, meaning any valid choice of conditional models for each of π , η , λ and μ can be used to construct a valid joint density (26, 43). Critically, this allows analysts the flexibility to choose any modeling strategy for each component nuisance functions separately, including both standard parametric models as well as modern machine learning techniques.

3.5 Identification of the ATTE

Under Assumptions 1-5 identification θ_{ATT}^{elig} is possible, as summarized in Theorem 1.

Function	Definition	Section Intro	oduced Mechanism(s)	$\widehat{\theta}_{\rm CC}$	$\widehat{\theta}_{\mathrm{IWOF}}$	${}_{ m R} \widehat{ heta}_{ m IF}$	$\widehat{\theta}_{\mathrm{EIF}}$
$\pi(L^*)$	$P(A=1 \mid \boldsymbol{L}^*)$	3.4	Treatment				
$\eta(\boldsymbol{L}^*,A)$	$P(R=1 \mid \boldsymbol{L}^*, A)$	3.4	Ascertainment		\checkmark	\checkmark	\checkmark
$\lambda_a(L^e_m;L^*)$	$P(\boldsymbol{L}_m^e ~ ~ \boldsymbol{L}^*, A=a, R=1)$	3.4	Imputation				
$\mu_a({oldsymbol L}^*,{oldsymbol L}^e_m)$	$\mathbb{E}[Y \mid \boldsymbol{L^*}, \boldsymbol{L^e_m}, A = a, R = 1]$	3.4	Outcome	\checkmark	\checkmark	\checkmark	\checkmark
$u(\pmb{L}^*, \pmb{L}^e_m)$	$P(A=1 \mid \boldsymbol{L^*}, \boldsymbol{L^e_m}, R=1)$	4.1	Treatment			\checkmark	\checkmark
$\varepsilon_a(L^*, Y)$	$P(E=1 \mid \boldsymbol{L}^*, Y, A=a, R=1)$	4.1	Imputation				\checkmark
$\xi(\pmb{L}^*,Y)$	$\mathbb{E}\big[E\mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e)\mid\boldsymbol{L}^*,Y,A=1,R=1\big]$] 4.1	Imputation/Outcome				\checkmark
$\gamma(\boldsymbol{L}^*,Y)$	$\mathbb{E}\Big[E\frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \mid \boldsymbol{L}^*, \boldsymbol{Y}, \boldsymbol{A}=\boldsymbol{0}, \boldsymbol{R}=1\Big]$] 4.1	Imputation/Treatment				\checkmark
$\chi(\pmb{L}^*,Y)$	$\mathbb{E}\left[E\frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}\mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) \mid \boldsymbol{L}^*, Y, A=0,\right]$	$R = 1] \qquad 4.1$	Imputation/Treatment/Outcom	е			\checkmark
$ u({m L}^*)$	$\mathbb{E}\left[E\left(Y-\mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e)\right) \mid \boldsymbol{L}^*, A=1, R=\right]$	1] 4.4	Imputation/Outcome			\checkmark	
$\omega_a({oldsymbol L}^*)$	$P(E = 1 \mid L^*, A = a, R = 1)$	4.4	Imputation			\checkmark	

Table 1: Definition of nuisance functions and their uses in various estimators of the ATTE

Theorem 1 Under Assumptions 1-5, θ_{ATT}^{elig} is identified by the functional $\theta(P) = \frac{\beta(P)}{\alpha(P)}$ where $\beta(P) = \mathbb{E}_P \Big[\frac{ARE}{\eta(\mathbf{L}^*, 1)} \Big(Y - \mu_0(\mathbf{L}^*, \mathbf{L}_m^e) \Big) \Big]$ and $\alpha(P) = \mathbb{E}_P \Big[\frac{ARE}{\eta(\mathbf{L}^*, 1)} \Big].$

Briefly, the result is obtained by showing that under Assumptions 4-5, $\theta(P)$ recovers the standard g-formula for the ATT under causal Assumptions 1-3 (44, 45).

4 Robust and Efficient Estimation of the ATTE

4.1 An Efficient Influence Function-Based Estimator of the ATTE

Given the form of $\theta(P)$, a natural choice of estimator for $\hat{\theta}$ is a plug-in estimator of the form $\hat{\beta}/\hat{\alpha}$, which in turn motivates the development of robust and efficient one-step estimators for $\alpha(P)$ and $\beta(P)$ (19, 46, 47). Towards that, we introduce the efficient influence functions of $\alpha(P)$ and $\beta(P)$ in Theorem 2.

Theorem 2 The efficient influence functions of $\alpha(P)$ and $\beta(P)$ at P in the semiparametric model induced by Assumption 4 (on the distribution of the coarsened observed data O) are given by

$$\dot{\alpha}_P^*(O) = A\left(1 - \frac{R}{\eta(\mathbf{L}^*, 1)}\right)\varepsilon_1(\mathbf{L}^*, Y) + \frac{ARE}{\eta(\mathbf{L}^*, 1)} - \alpha(P)$$
(3)

$$\dot{\beta}_{P}^{*}(O) = \frac{AR}{\eta(\boldsymbol{L}^{*},1)} \left[\left(\boldsymbol{E} - \varepsilon_{1}(\boldsymbol{L}^{*},Y) \right) \boldsymbol{Y} - \left(\boldsymbol{E}\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \xi(\boldsymbol{L}^{*},Y) \right) \right] + A \left(\varepsilon_{1}(\boldsymbol{L}^{*},Y)\boldsymbol{Y} - \xi(\boldsymbol{L}^{*},Y) \right) \\ - \frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)} \left[\boldsymbol{E} \frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \left(\boldsymbol{Y} - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \right) - \left(\gamma(\boldsymbol{L}^{*},Y)\boldsymbol{Y} - \chi(\boldsymbol{L}^{*},Y) \right) \right] \\ - (1-A)\frac{\eta(\boldsymbol{L}^{*},0)}{\eta(\boldsymbol{L}^{*},1)} \left(\gamma(\boldsymbol{L}^{*},Y)\boldsymbol{Y} - \chi(\boldsymbol{L}^{*},Y) \right) - \beta(P)$$

$$\tag{4}$$

Both $\dot{\alpha}_{P}^{*}$ and $\dot{\beta}_{P}^{*}$ introduce additional nuisance functions. Most directly interpretable are $u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e}) = P(A = 1 \mid \mathbf{L}^{*}, \mathbf{L}_{m}^{e}, R = 1)$, the propensity score of treatment among complete cases, and $\varepsilon_{a}(\mathbf{L}^{*}, Y) = P(E = 1 \mid \mathbf{L}^{*}, Y, A = a, R = 1)$, the probability of being eligible for the study conditional on all fully observed covariates, also among complete cases. Three additional nuisance functions appear in $\dot{\beta}_{P}^{*}$, which we refer to as nested nuisance functions given the appearance of simpler, previously defined nuisance functions, pre-multiplied by eligibility status E, within the outer most expectation of these more complex nuisance functions: $\xi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\mu_{0}(\mathbf{L}^{*}, \mathbf{L}_{m}^{e}) \mid \mathbf{L}^{*}, Y, A = 1, R = 1], \gamma(\mathbf{L}^{*}, Y) =$ $\mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1], \text{ and } \chi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1], \text{ and } \chi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1], \text{ and } \chi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1], \text{ ond } \chi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1], \text{ ond } \chi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1], \text{ ond } \chi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1], \text{ ond } \chi(\mathbf{L}^{*}, Y) = \mathbb{E}_{P}[E\frac{u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})}{1-u(\mathbf{L}^{*}, \mathbf{L}_{m}^{e})} \mid \mathbf{L}^{*}, Y, A = 0, R = 1].$ One way to interpret these nested nuisance functions is that they attempt to model treatment (e.g., via u) and/or outcome (e.g., via $\mu_{0})$ mechanisms by using marginalization to account for the fact that eligibility status remains unkn

Influence functions are by definition mean zero, so it is helpful to work with uncentered influence functions, $\dot{\alpha}_P(O) = \dot{\alpha}_P^*(O) + \alpha(P)$ and $\dot{\beta}_P(O) = \dot{\beta}_P^*(O) + \beta(P)$. With these, we propose the estimator $\hat{\theta}_{\text{EIF}}$ by taking the ratio of one-step estimators for $\beta(P)$ and $\alpha(P)$, which each correspond to the sample mean of an estimated uncentered influence function:

$$\widehat{\theta}_{\rm EIF} = \frac{\mathbb{P}_n[\dot{\beta}_{\widehat{P}}(O)]}{\mathbb{P}_n[\dot{\alpha}_{\widehat{P}}(O)]} \tag{5}$$

In Equation (5), the notation \hat{P} is introduced to indicate that component nuisance functions in Equations (3) and (4) are replaced by their corresponding estimated quantities. In simpler studies where there is no missing eligibility data and all subjects included in the study population are known to be eligible, $\hat{\theta}_{\text{EIF}}$ simplifies to the standard one-step (doubly-robust) estimator for the ATT (45). We demonstrate this correspondence in the Supplementary Materials.

4.2 Theoretical Properties of $\hat{\theta}_{\text{EIF}}$

That $\hat{\theta}_{\text{EIF}}$ uses one-step estimators based on efficient influence functions for $\beta(P)$ and $\alpha(P)$ rather than the efficient influence function for $\theta(P)$ directly, $\theta_P^*(O)$, is matter of practical convenience in implementation; asymptotically the two approaches are equivalent (46, 47). In any case, $\theta_P^*(O)$ is useful in characterizing the asymptotic behavior of $\hat{\theta}_{\text{EIF}}$.

Corollary 2.1 The efficient influence function of $\theta(P)$ at P in a semiparametric model induced by Assumption 4 is given by $\dot{\theta}_P^*(O) = \frac{1}{\alpha(P)} \left(\dot{\beta}_P(O) - \theta(P) \dot{\alpha}_P(O) \right)$

Corollary 2.1 follows from versions of the product rule and the chain rule for influence functions (19, 46). Another pair of important quantities necessary to characterize the asymptotic behavior of $\hat{\theta}_{\text{EIF}}$ are remainder terms $R_{\alpha}(\overline{P}, P)$ and $R_{\beta}(\overline{P}, P)$ from von Mises expansions (19, 24) for $\alpha(P)$ and $\beta(P)$, defined in the following lemma.

Lemma 1 $\alpha(P)$ and $\beta(P)$ satisfy the von Mises expansions

$$\alpha(\overline{P}) - \alpha(P) = -\int \dot{\alpha}_{\overline{P}}^*(o)dP(o) + R_{\alpha}(\overline{P}, P)$$
$$\beta(\overline{P}) - \beta(P) = -\int \dot{\beta}_{\overline{P}}^*(o)dP(o) + R_{\beta}(\overline{P}, P)$$

where the remainder terms (omitting inputs for brevity) are as follows:

$$R_{\alpha}(\overline{P}, P) = \mathbb{E}_{P} \left[A(\bar{\varepsilon}_{1} - \varepsilon_{1}) \left(1 - \frac{\eta_{1}}{\bar{\eta}_{1}} \right) \right]$$
$$R_{\beta}(\overline{P}, P) = \mathbb{E}_{P} \left[A \left(1 - \frac{\eta_{1}}{\bar{\eta}_{1}} \right) \left(Y(\bar{\varepsilon}_{1} - \varepsilon_{1}) - (\bar{\xi} - \xi) \right) - \frac{RE}{\bar{\eta}_{1}} (\mu_{0} - \bar{\mu}_{0}) \left(\frac{u(1 - \bar{u}) - \bar{u}(1 - u)}{1 - \bar{u}} \right) - (1 - A) \frac{(\eta_{0} - \bar{\eta}_{0})}{\bar{\eta}_{1}} \left(Y(\bar{\gamma} - \gamma) - (\bar{\chi} - \chi) \right) \right]$$

The rate of convergence for $\hat{\theta}_{\text{EIF}}$ is closely tied to the second-order remainder terms from the von Mises expansions in Lemma 1, and thus depends on products of errors in nuisance function estimation. Leveraging Lemma 1, we summarize the asymptotic behavior of $\hat{\theta}_{\text{EIF}}$ in the following theorem, where $||f||^2 = \mathbb{E}_P[f(O)^2]$ denotes the squared $L_2(P)$ norm.

Theorem 3 If $\|\dot{\alpha}_{\widehat{P}} - \dot{\alpha}_{P}\| = o_{P}(1), \|\dot{\beta}_{\widehat{P}} - \dot{\beta}_{P}\| = o_{P}(1), \alpha(P) > 0, and P\Big[\big|\mathbb{P}_{n}\big(\dot{\alpha}_{\widehat{P}}(O)\big)\big| \ge \epsilon\Big] = 1$ for some $\epsilon > 0$, then

$$\widehat{\theta}_{\mathrm{EIF}} - \theta(P) = \mathbb{P}_n[\dot{\theta}_P^*(O)] + O_P\Big(R_\alpha(\widehat{P}, P) + R_\beta(\widehat{P}, P)\Big) + o_P(n^{-1/2})$$

Moreover, if $R_{\alpha}(\hat{P}, P) + R_{\beta}(\hat{P}, P) = o_P(n^{-1/2})$ then $\sqrt{n} (\hat{\theta}_{\text{EIF}} - \theta(P)) \xrightarrow{d} \mathcal{N} (0, \text{Var}_P[\dot{\theta}_P^*(O)])$, whereby $\hat{\theta}_{\text{EIF}}$ attains the semiparametric efficiency bound induced by Assumption 4.

Corollary 3.1 Under the conditions of Theorem 3 and assuming that $P(\delta \le 1 - \hat{u} \le 1 - \delta)$ for some $\delta > 0$, $P(\hat{\eta}_1 > \epsilon) = 1$ for some $\epsilon > 0$, and $\mathbb{E}_P[Y^2] \le M < \infty$,

$$R_{\alpha}(\hat{P}, P) + R_{\beta}(\hat{P}, P) = O_{P} \bigg(\|\hat{\mu}_{0} - \mu_{0}\| \|\hat{u} - u\| + \|\hat{\eta}_{1} - \eta_{1}\| \Big\{ \|\hat{\varepsilon}_{1} - \varepsilon_{1}\| + \|\hat{\xi} - \xi\| \Big\} + \|\hat{\eta}_{0} - \eta_{0}\| \Big\{ \|\hat{\gamma} - \gamma\| + \|\hat{\chi} - \chi\| \Big\} \bigg)$$

Theorem 3 demonstrates that $\hat{\theta}_{\text{EIF}}$ converges faster than any component nuisance function, even when flexible modeling choices are used. As illustrated by Corollary 3.1, the rate of convergence for $\hat{\theta}_{\text{EIF}}$ is directly tied to error rates in estimation of component nuisance functions. When these product errors are $o_P(n^{-1/2})$ —for example, when all nuisance function error rates are $o_P(n^{-1/4})$ —then $\hat{\theta}_{\text{EIF}}$ is not only \sqrt{n} —consistent and asymptotically normal, but also attains the semiparametric efficiency bound induced by Assumption 4. In particular, careful examination of Corollary 3.1 illustrates that the remainder term $R_{\alpha} + R_{\beta}$ will be $o_P(n^{-1/2})$ if (i) the product of errors for the standard ATT nuisance functions (μ_0, u) , and (ii) the sum of product errors for missing data related nuisance functions $(\eta_0, \eta_1, \varepsilon_1, \xi, \gamma, \chi)$ are each $o_P(n^{-1/2})$. The required rate conditions (e.g., $o_P(n^{-1/4})$ for each nuisance function) are achievable under various structural assumptions, such as sparsity, smoothness, or additivity. Under such conditions, asymptotically valid Wald-style (1-q)-level confidence intervals are readily available, $\hat{\theta}_{\text{EIF}} \pm z_{1-q/2} \sqrt{\frac{1}{n} \mathbb{P}_n[\dot{\theta}_{\widehat{P}}^*(O)^2]}$, where z_q denotes the q^{th} quantile of the standard normal distribution.

4.3 Computation of $\hat{\theta}_{\text{EIF}}$

Routines for estimation of influence function-based estimators like $\hat{\theta}_{\text{EIF}}$ typically involve sample splitting or cross-fitting, particularly when using nonparametric and/or machine learning techniques to estimate component nuisance functions (20). In sample splitting, $(O_1, ..., O_n)$ is randomly split into two disjoint samples D_0, D_1 , with D_0 used to estimate component nuisance functions as well as any hyperparameters for relevant machine learning models. Models for nuisance functions trained on D_0 are subsequently applied to observations from the held out dataset D_1 . Finally, the roles of D_0 and D_1 are reversed, and estimation of $\hat{\theta}_{\text{EIF}}$ can proceed as in Equation 5 by averaging influence function contributions for each subject. Given that u and μ_0 appear in the respective targets for ξ, γ , and χ , estimation of these nested nuisance functions requires predictions of $\hat{\mu}_0$ and \hat{u} on training set D_0 to even train models for $\hat{\xi}, \hat{\gamma}$, and $\hat{\chi}$. Despite this complication, estimation of ξ, γ , and χ can proceed within traditional sample splitting routines. To illustrate this, we outline the entire procedure for computing $\hat{\theta}_{\text{EIF}}$ in a flexible manner in Algorithm 1.

We use k = 2 splits for illustrative purposes but analogous procedures work for k > 2.

Algorithm 1 (Computation of $\hat{\theta}_{\text{EIF}}$) Let D_0 and D_1 be two disjoint, independent splits of $(O_1, ..., O_n)$ of sizes n_0 and n_1 , respectively, with $n = n_0 + n_1$. The analytical procedure to compute the estimator $\hat{\theta}_{EIF}$ of $\theta(P)$ is given as follows:

- 1. Construct estimators $\hat{\eta}$, \hat{u} , $\hat{\varepsilon}_1$, and $\hat{\mu}_0$ on n_0 samples from D_0
- 2. For $O_1, O_2, ..., O_{n_0} \in D_0$, compute $\left\{ E \widehat{\mu}_0(\boldsymbol{L}_i^*, \boldsymbol{L}_{m_i}^e), E \frac{\widehat{u}(\boldsymbol{L}_i^*, \boldsymbol{L}_{m_i}^e)}{1 \widehat{u}(\boldsymbol{L}_i^*, \boldsymbol{L}_{m_i}^e)}, E \widehat{\mu}_0(\boldsymbol{L}_i^*, \boldsymbol{L}_{m_i}^e) \frac{\widehat{u}(\boldsymbol{L}_i^*, \boldsymbol{L}_{m_i}^e)}{1 \widehat{u}(\boldsymbol{L}_i^*, \boldsymbol{L}_{m_i}^e)} \right\}_{i=1}^{n_0}$, and use these quantities to construct estimators $\widehat{\xi}, \widehat{\gamma}$, and $\widehat{\chi}$, on n_0 samples from D_0 .
- 3. For $O_{n_0+1}, O_{n_0+2}, ..., O_n \in D_1$, construct plug-in estimates of uncentered influence functions $\{\dot{\alpha}_{\hat{P}_0}(O_i)\}_{i=n_0+1}^n$ and $\{\dot{\beta}_{\hat{P}_0}(O_i)\}_{i=n_0+1}^n$ using the nuisance estimators constructed in steps 1 and 2.
- 4. Repeat steps 1-3, swapping the roles of D_0 and D_1 , to create sets $\{\dot{\alpha}_{\widehat{P}}(O_i)\}_{i=1}^n = \{\dot{\alpha}_{\widehat{P}_0}(O_i)\}_{i=n_0+1}^n \cup \{\dot{\alpha}_{\widehat{P}_1}(O_i)\}_{i=1}^{n_0}$ and $\{\dot{\beta}_{\widehat{P}}(O_i)\}_{i=1}^n = \{\dot{\beta}_{\widehat{P}_0}(O_i)\}_{i=n_0+1}^n \cup \{\dot{\beta}_{\widehat{P}_1}(O_i)\}_{i=1}^{n_0}$ (e.g., retaining influence function contributions from corresponding out-of-sample nuisance function predictions).

5. Compute
$$\widehat{\alpha} = n^{-1} \sum_{i=1}^{n} \dot{\alpha}_{\widehat{P}}(O_i)$$
 and $\widehat{\beta} = n^{-1} \sum_{i=1}^{n} \dot{\beta}_{\widehat{P}}(O_i)$. Report $\widehat{\theta}_{EIF}$ as in Eq. 5.

6. Compute
$$\left\{\dot{\theta}_{\widehat{P}}^*(O_i) = \frac{1}{\widehat{\alpha}} \left(\dot{\beta}_{\widehat{P}}(O_i) - \frac{\widehat{\beta}}{\widehat{\alpha}} \dot{\alpha}_{\widehat{P}}(O_i)\right)\right\}_{i=1}^n$$
, and estimate $\widehat{Var}[\widehat{\theta}_{EIF}] = n^{-1} \mathbb{P}_n[\dot{\theta}_{\widehat{P}}^*(O)^2]$

We note that while neither outcome model, μ , nor propensity model, u, conditions on E = 1, all contributions of these nuisance functions in $\dot{\beta}_P(O)$ are pre-multiplied by eligibility indicator E, and thus only eligible subjects' values for these nuisances contribute to $\hat{\theta}_{\text{EIF}}$. Because both models already condition on R = 1, analysts might additionally restrict the conditioning set of these models to include E = 1 and better model the contributions of these nuisance functions for eligible patients if treatment and outcome mechanisms differ drastically for eligibility status.

4.4 Alternatives to $\hat{\theta}_{\text{EIF}}$

In completely nonparametric models, there is only a single influence function for a pathwise differentiable statistical functional, and thus the influence function is the efficient influence function. That is not the case in our problem, however, since Assumption 4 restricts the tangent space of possible nonparametric models. We derived the efficient influence functions $\dot{\alpha}_P^*$ and $\dot{\beta}_P^*$ by first finding nonparametric influence functions $\dot{\alpha}_P'^*$ and $\dot{\beta}_P'^*$ for $\alpha(P)$ and $\beta(P)$, and subsequently projecting them onto the tangent space (Supplementary Materials Section 3.3). The uncentered versions of these influence functions, $\dot{\alpha}_P'$ and $\dot{\beta}_P'$, are given by

$$\dot{\alpha}_P'(O) = A\left(1 - \frac{R}{\eta(\boldsymbol{L}^*, 1)}\right)\omega_1(\boldsymbol{L}^*) + \frac{ARE}{\eta(\boldsymbol{L}^*, 1)}$$

$$\dot{\beta}'_{P}(O) = A \left(1 - \frac{R}{\eta(\boldsymbol{L}^{*}, 1)} \right) \nu(\boldsymbol{L}^{*}) + \frac{RE}{\eta(\boldsymbol{L}^{*}, 1)} \left(Y - \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \right) \left[A - (1 - A) \frac{u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{1 - u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})} \right]$$

 $\dot{\alpha}'_P$ and $\dot{\beta}'_P$ contain fewer nested nuisance functions and thus motivate an estimator which is simpler to compute, using an analogous procedure to Algorithm 1 and Equation 5, and replacing $\dot{\alpha}_P$ and $\dot{\beta}_P$ with $\dot{\alpha}'_P$ and $\dot{\beta}'_P$, respectively.

$$\widehat{\theta}_{\mathrm{IF}} = \frac{\mathbb{P}_n[\dot{\beta}'_{\widehat{P}}(O)]}{\mathbb{P}_n[\dot{\alpha}'_{\widehat{P}}(O)]}$$

The estimator $\hat{\theta}_{IF}$ introduces two new nuisance functions, $\omega_a(\mathbf{L}^*) = P(E = 1 | \mathbf{L}^*, A = a, R = 1)$, and $\nu(\mathbf{L}^*) = \mathbb{E}_P[E(Y - \mu_0(\mathbf{L}^*, \mathbf{L}_m^e)) | \mathbf{L}^*, A = 1, R = 1]$. See that ω is very similar to ε in that it models eligibility among complete cases given completely observed information, only it doesn't use outcomes Y when modeling this probability. Though ν is a nested nuisance function, other nested nuisance functions ξ, γ and χ do not need to be estimated for $\hat{\theta}_{IF}$. While one would expect $\hat{\theta}_{IF}$ to be less efficient than $\hat{\theta}_{EIF}$, $\hat{\theta}_{IF}$ is relatively easier to implement, and could plausibly result in roughly the same (or less) small sample bias without much of a loss (if any) in terms of efficiency. Altogether, this trade-off is one we hope to explore and offer guidance on through this simulation study in Section 5.

Additionally, we consider two alternative estimators of $\theta(P)$ which are easier computationally than $\hat{\theta}_{\text{EIF}}$ or $\hat{\theta}_{\text{IF}}$. As a baseline, we consider $\hat{\theta}_{\text{CC}}$, the complete case g-formula estimator for the ATT, where C denotes the set of eligible complete cases undergoing RYGB:

$$\widehat{\theta}_{CC} = \frac{\mathbb{P}_n \left[ARE \left(Y - \widehat{\mu}_0(\boldsymbol{L}^*, \boldsymbol{L}_m^e) \right) \right]}{\mathbb{P}_n [ARE]} = \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \left(Y_i - \widehat{\mu}_0(\boldsymbol{L}_i^*, \boldsymbol{L}_{m_i}^e) \right)$$

This strategy is representative of the complete-case analyses that are common in practice, and would be valid if $\theta_{ATT}^{CC} = \theta_{ATT}^{elig}$.

Examination of the identification result in Theorem 1 additionally motivates the following inverse weighted outcome regression (IWOR) estimator, $\hat{\theta}_{\text{IWOR}}$.

$$\widehat{\theta}_{\text{IWOR}} = \frac{\mathbb{P}_n \left[\frac{ARE}{\widehat{\eta}(\boldsymbol{L}^*, 1)} (Y - \widehat{\mu}_0(\boldsymbol{L}^*, \boldsymbol{L}_m^e)) \right]}{\mathbb{P}_n \left[\frac{ARE}{\widehat{\eta}(\boldsymbol{L}^*, 1)} \right]}$$

Unlike $\hat{\theta}_{CC}$, which ignores the possibility of selection bias, $\hat{\theta}_{IWOR}$ accounts for the possibility of selection bias and requires only a single extra nuisance function, η .

Finally, it is worth noting that there is equivalent form of $\hat{\theta}_{\text{EIF}}$, which we denote $\tilde{\theta}_{\text{EIF}}$, that uses the nuisance functions π and λ instead of u, and thus is more directly tied to the likelihood factorization in Equation 2. Given that λ is a conditional density for \boldsymbol{L}_m^e , which itself may be multi-dimensional, ratios of the form $\frac{\lambda_1(\boldsymbol{L}_m^e;\boldsymbol{L}^*)}{\lambda_0(\boldsymbol{L}_m^e;\boldsymbol{L}^*)}$ (as is required by $\tilde{\theta}_{\text{EIF}}$) would be difficult to estimate via nonparametric methods. For completeness, we provide additional details on $\tilde{\theta}_{\text{EIF}}$ in the Supplementary Materials, and illustrate the connection between the conditional density ratio, $\frac{\lambda_1(\boldsymbol{L}_m^e;\boldsymbol{L}^*)}{\lambda_0(\boldsymbol{L}_m^e;\boldsymbol{L}^*)}$, and the propensity score ratio, $\frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}$.

5 Simulation Study

5.1 Data Generation

To investigate finite sample performance of $\hat{\theta}_{\text{EIF}}$ and $\hat{\theta}_{\text{IF}}$ we conducted a simulation study tied closely to our motivating bariatric surgery study. In this simulation study, interest lies in the effect of RYGB vs. SG (A) on percent weight change 3 years post surgery (Y) among patients with pre-diabetes or diabetes, defined as an A1c $\geq 5.7\%$ (\boldsymbol{L}_m^e). Covariates \boldsymbol{L}^* included health care site, sex, race, baseline BMI, age, smoking status, and eGFR.

To generate simulated datasets, we began by sampling covariate vectors L^* from the observed distribution in the DURABLE data to preserve the complex correlation structure of L^* . Simulated treatment, outcome, and eligibility status (both L_m^e and R) were generated using sampled covariates L^* and models for the nuisance functions tied directly to the likelihood factorization in Equation 2 $(\pi, \eta, \lambda, \mu)$. In particular, parametric models were fit on a sample of 16,461 surgical patients from the DURABLE database to inform this simulated data generation. Logistic regression was used for π and η , a gamma generalized linear model (GLM) was used for λ , and linear regression was used for μ . The outcome model μ was specified with several interactions between $A \times L_m^e$, $A \times L^*$, and $L_m^e \times L^*$ in order to introduce additional complexity and make μ difficult to estimate well, perhaps even with more flexible machine learning methods. Additional details, including complete tables of nuisance function coefficients used to generate simulated data are available in the Supplementary Materials.

1,000 simulated datasets were generated for each of n = 10,000 and n = 25,000 patients. These sample sizes are reasonable in the context many EHR-based settings, including our study of bariatric surgery for T2DM outcomes (n = 14,809), the work of McTigue et al. (n = 9,710) as well as previous DURABLE studies (n > 30,000) (11, 31). Across simulated datasets, 62% of patients were treated with RYGB (A = 1) and 70% of patients had ascertainable eligibility (R = 1), of whom 57% were eligible (i.e., P(E = 1 | R = 1)).

5.2 Estimators

When implementing $\hat{\theta}_{CC}$, we only estimated μ with the correct parametric specification, so that any resulting bias was due to selection bias and not model misspecification. When estimating $\hat{\theta}_{IWOR}$, we considered versions where both μ and η were correctly specified, as well as versions where only one of the two was correctly specified. For $\hat{\theta}_{IWOR}$ and the two influence function-based estimators, we applied six nonparametric strategies to each for flexible estimation of component nuisance functions using ensemble regression learners from the SuperLearner package in R (48, 49).

The six strategies arise from all combinations of two distinct sets of libraries included in the SuperLearner (SL) ensemble and three strategies for estimating the outcome regression among SG subjects, μ_0 . The first set of SL libraries (SL1) included random forest (RF) learners with several hyperparameter combinations, along with learners for linear model (LM)/GLM, generalized additive model (GAM) (for binary variables), and multivariate adaptive polynomial regression spline (Polymars). The second set (SL2) dropped LM/GLM learners to test estimator behavior when the correct parametric model was not directly among the family of candidate learners.

Given the complexity of interactions in the outcome model, we considered stratification (e.g., estimating μ_0 only on subjects with A = 0) and augmenting the design matrix with all $A \times \mathbf{L}$ interactions as alternatives to fitting a single outcome model μ on all complete cases. The rationale is that fitting a single outcome model might inadequately capture the full scope of interactions in μ , and thus there could be some degree of model misspecification, which may affect various estimators in different ways (Supplemental Figure S1).

5.3 Results

Table 2 presents relative bias and standard errors for all estimators, as well as coverage of 95% confidence intervals for the two influence function-based estimators, $\hat{\theta}_{\rm EIF}$ and $\hat{\theta}_{\rm IF}$. Unsurprisingly, $\hat{\theta}_{\rm CC}$ was biased by over 15%, even with correctly specified outcome model μ , thereby demonstrating that selection bias can be a prominent concern when discarding subjects with incomplete eligibility information. Performance of $\hat{\theta}_{\rm IWOR}$ was very dependent on model specification. When μ and η were both correctly specified, $\hat{\theta}_{\rm IWOR}$ was unbiased, but when either nuisance function was misspecified the resulting biases of 17% and 24%, respectively, were worse than that of $\hat{\theta}_{\rm CC}$. Using nonparametric SuperLearner ensembles to estimate μ, η did not guarantee unbiasedness, with plug-in bias ranging from 2-19%, with bias worse in cases with greater estimation error for $\hat{\mu}_0$ (Supplemental Figure S1).

Both influence function-based estimators were unbiased in all six nonparametric estimation strategies and attained nominal coverage for 95% confidence intervals, despite a degree of error in estimating $\hat{\mu}_0$, and possibly other nuisance functions. Under each strategy, $\hat{\theta}_{IF}$ had a standard error between 4-14% larger than the corresponding standard error of $\hat{\theta}_{EIF}$, demonstrating the additional modeling complexities did indeed increase efficiency. Of note, the standard error of $\hat{\theta}_{IWOR}$ with correctly specified parametric models for μ, η only matched the standard error for the $\hat{\theta}_{EIF}$ estimation strategy producing the smallest standard error for n = 10,000, and was 5% larger for n = 25,000.

Altogether, these simulations demonstrate that both $\hat{\theta}_{IF}$ and $\hat{\theta}_{EIF}$ are implementable with flexible, nonparametric estimation strategies for component nuisance functions, even

ts $n = 25,000$ Patients	verage %-Bias SD Coverage	— 15.7 2.39e-03 —	0.1 2.72e-03	16.8 2.43e-03	24.0 2.43e-03	— 17.9 3.18e-03 —	1.8 2.82e-03	2.2 2.86e-03	— 11.0 4.53e-03 —	— 11.7 3.34e-03 —	— 13.3 4.78e-03 —	95.0 0.1 3.21e-03 95.6		93.8 0.1 2.92e-03 94.7	93.8 0.1 2.92e-03 94.7 93.3 0.1 2.91e-03 94.3	93.8 0.1 2.92e-03 94.7 93.3 0.1 2.91e-03 94.3 95.1 0.2 3.21e-03 94.8	93.8 0.1 2.92e-03 94.7 93.3 0.1 2.91e-03 94.3 95.1 0.2 3.21e-03 94.8 95.1 0.2 3.21e-03 94.8 93.8 0.3 3.16e-03 94.8	93.8 0.1 2.92e-03 94.7 93.3 0.1 2.91e-03 94.3 95.1 0.2 3.21e-03 94.8 93.8 0.3 3.16e-03 94.8 94.5 0.5 3.20e-03 94.6	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0.1 $2.92e-03$ 94.7 93.8 0.1 $2.91e-03$ 94.7 95.1 0.2 $3.21e-03$ 94.8 95.1 0.2 $3.21e-03$ 94.8 93.8 0.3 $3.16e-03$ 93.8 94.5 0.5 $3.20e-03$ 94.6 94.2 -0.2 $2.83e-03$ 93.8 94.7 -0.1 $2.59e-03$ 95.4	0.1 $2.92e-03$ 94.7 93.8 0.1 $2.91e-03$ 94.7 95.1 0.2 $3.21e-03$ 94.8 95.1 0.2 $3.21e-03$ 94.8 93.8 0.3 $3.16e-03$ 93.8 94.5 0.5 $3.20e-03$ 94.6 94.2 -0.2 $2.83e-03$ 93.8 94.7 -0.1 $2.59e-03$ 95.4 94.7 -0.1 $2.58e-03$ 95.0	0.1 $2.92e-03$ 94.7 93.8 0.1 $2.91e-03$ 94.7 95.1 0.1 $2.91e-03$ 94.3 95.1 0.2 $3.21e-03$ 94.8 93.8 0.3 $3.16e-03$ 93.8 94.5 0.3 $3.16e-03$ 94.6 94.2 0.5 $3.20e-03$ 94.6 94.2 -0.2 $2.83e-03$ 93.8 94.7 -0.1 $2.59e-03$ 95.4 94.7 -0.1 $2.58e-03$ 95.0 93.6 -0.1 $2.92e-03$ 94.8	0.1 $2.92e-03$ 94.7 93.8 0.1 $2.91e-03$ 94.7 95.1 0.1 $2.91e-03$ 94.3 95.1 0.2 $3.21e-03$ 94.8 93.8 0.3 $3.16e-03$ 94.6 94.5 0.5 $3.20e-03$ 94.6 94.7 -0.2 $2.83e-03$ 93.8 94.7 -0.1 $2.59e-03$ 95.4 94.7 -0.1 $2.58e-03$ 95.4 94.7 -0.1 $2.92e-03$ 94.8 94.7 -0.1 $2.92e-03$ 94.8 94.1 0.1 $2.98e-03$ 94.8 94.1 0.1 $2.98e-03$ 94.8
000 Patient	SD Cov	72e-03	35e-03	77e-03	76e-03	30e-03	56e-03 -	50e-03)9e-03	78e-03 -	26e-03		<u>t</u> 0e-03 у	10e-03 y 88e-03 9	We-U3 y 38e-03 9 32e-03 9	l0e-03 y 38e-03 9 32e-03 9 34e-03 9	10e-03 y 38e-03 9 32e-03 9 34e-03 9 40e-03 9	10e-03 y 88e-03 9 82e-03 9 84e-03 9 10e-03 9 86e-03 9	Uce-0.3 y See-0.3 9 See-0.3 9 J4e-0.3 9 Mee-0.3 9 See-0.3 9 See-0.3 9 See-0.3 9	Uc-03 y 38e-03 9 38e-03 9 34e-03 9 66-03 9 36e-03 9 36e-03 9	Weel Weel <th< th=""><th>$\begin{array}{c ccccccccccccccccccccccccccccccccccc$</th><th>$\begin{array}{c ccccc} & y \\ 10e-0.3 & y \\ 38e-0.3 & 9 \\ 32e-0.3 & 9 \\ 10e-0.3 & 9 \\ 36e-0.3 & 9 \\ 36e-0.3 & 9 \\ 36e-0.3 & 9 \\ 34e-0.3 & 9 \\ 34e-0.3 & 9 \\ 19e-0.3 & 9 \\ 19e-0.3 & 9 \\ 19e-0.3 & 9 \\ 10e-0.3 & 9$</th></th<>	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccc} & y \\ 10e-0.3 & y \\ 38e-0.3 & 9 \\ 32e-0.3 & 9 \\ 10e-0.3 & 9 \\ 36e-0.3 & 9 \\ 36e-0.3 & 9 \\ 36e-0.3 & 9 \\ 34e-0.3 & 9 \\ 34e-0.3 & 9 \\ 19e-0.3 & 9 \\ 19e-0.3 & 9 \\ 19e-0.3 & 9 \\ 10e-0.3 & 9$
n = 10, 0	%-Bias	15.4 3.7	0.0 4.3	16.5 3.7	23.8 3.7	19.1 4.3	3.7 4.5	3.8 4.5	7.9 5.9	17.1 4.7	17.9 5.2		-0.2 5.4	-0.2 5.4 -0.2 4.8	-0.2 5.4 -0.2 4.8 -0.1 4.9	-0.2 5.4 -0.2 4.8 -0.1 4.9 -0.3 5.3 0.3 5.3	$\begin{array}{cccc} -0.2 & 5.4 \\ -0.2 & 4.8 \\ -0.1 & 4.9 \\ 0.3 & 5.3 \\ 0.4 & 5.4 \\ \end{array}$	$\begin{array}{cccc} -0.2 & 5.4 \\ -0.2 & 4.8 \\ -0.1 & 4.9 \\ 0.3 & 5.3 \\ 0.4 & 5.4 \\ 0.5 & 5.3 \\ 0.5 & 5.3 \end{array}$	$\begin{array}{cccc} -0.2 & 5.4 \\ -0.2 & 4.8 \\ -0.1 & 4.9 \\ 0.3 & 5.3 \\ 0.4 & 5.4 \\ 0.5 & 5.3 \\ 0.5 & 5.3 \\ -0.4 & 4.8 \end{array}$	$\begin{array}{cccc} -0.2 & 5.4 \\ -0.2 & 4.8 \\ -0.1 & 4.9 \\ 0.3 & 5.3 \\ 0.4 & 5.4 \\ 0.5 & 5.3 \\ 0.5 & 5.3 \\ -0.4 & 4.8 \\ -0.1 & 4.3 \end{array}$	$\begin{array}{cccc} -0.2 & 5.4 \\ -0.2 & 4.8 \\ -0.1 & 4.9 \\ 0.3 & 5.3 \\ 0.4 & 5.4 \\ 0.4 & 5.4 \\ 0.5 & 5.3 \\ -0.4 & 4.8 \\ -0.1 & 4.3 \\ -0.1 & 4.3 \\ \end{array}$	$\begin{array}{cccc} -0.2 & 5.4 \\ -0.2 & 4.8 \\ -0.1 & 4.9 \\ 0.3 & 5.3 \\ 0.5 & 5.3 \\ 0.5 & 5.3 \\ 0.5 & 5.3 \\ -0.4 & 4.8 \\ -0.1 & 4.3 \\ -0.1 & 4.3 \\ 0.0 & 5.0 \end{array}$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
	True $\mu/\eta^{\mathbf{c}}$	π	η/μ	π	h																		
	$\mu_0 {f Strategy}^{f b}$	1		1		1	5	3	1	2	3		1	2	3 2 1	1 3 2 1	1 2 3 2 2	1 2 2 3 3 3	1 2 1 3 2 1 1 3 2 1 3 1 1	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	1 2 1 3 2 1 3 2 1 3 2 1 1	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$
	$SL \ Libs^{a}$		SL1							SL1	SL1	SL1	SL1	SL1 SL2	SL1 SL2	SL1 SL2 SL2 SL1	SL1 - SL1 - SL1 - SL2 - SL2 - SL1 -	SL1 SL2 SL1 SL1 SL2 SL1	SL1				
	Strategy	Parametric		$\operatorname{Parametric}$			Vonparametric —									Nonparametric -	Nonparametric -	Nonparametric -	Nonparametric	Nonparametric -	Nonparametric -	Nonparametric - Nonparametric -	Nonparametric - Nonparametric -
			$\theta_{\rm CC}$ $\hat{\theta}_{\rm IWOR}$ N																				

^a SL Libs = SuperLearner libraries: S1 = Random Forest, LM/GLM, GAM, Polymars; SL2 = Random Forest, GAM, Polymars ^b μ_0 Strategy: (1) Fit single $\hat{\mu}$ on A = 0, 1 together (2) Fit $\hat{\mu}_0$ on A = 0 only (stratification) (3) Specify all $A \times \mathbf{L}$ interactions in design matrix for $\hat{\mu}$ ^c Correctly specified parametric models for μ and/or η

Table 2: Comparison of estimators of $\theta(P)$ in simulation study

in the presence of complex nested nuisance functions. The estimators retain good finite sample properties for sample sizes typical of EHR-studies for bariatric surgery. Finally, these simulations demonstrate that $\hat{\theta}_{\rm EIF}$ attains noticeable efficiency gains if one is willing to estimate slightly more complex nuisance functions (cf. $\hat{\theta}_{\rm IF}$).

6 Comparison of Bariatric Surgical Procedures

6.1 Methodological Details

Finally, we return to the bariatric surgery study introduced and discussed in Section 2.1. For both the relative weight change and T2DM remission outcomes, we present results for each estimator proposed in Section 4 across all 40 operationalizations of the study eligibility criteria (Figure 2). Both influence-function based estimators, $\hat{\theta}_{EIF}$ and $\hat{\theta}_{IF}$, were computed using the SL1 set of SuperLearner libraries described in Section 5.2 to estimate relevant nuisance functions. The complete-case estimator, $\hat{\theta}_{CC}$, was computed using a linear model for μ with pre-specified interactions between surgery type and select L^* covariates. Finally, the inverse-weighted outcome regression estimator, $\hat{\theta}_{IWOR}$, used the same μ specification as $\hat{\theta}_{CC}$, and used a main-effects logistic regression for η . 95% confidence intervals for both $\hat{\theta}_{CC}$ and $\hat{\theta}_{IWOR}$ were computed via bootstrapping using a normal approximation (50). Additional methodological details are available in the Supplementary Materials, including model specifications for each estimation procedure.

6.2 Study Results

To get a sense of how eligibility-related nuisance function estimates varied by lookback window (i.e., period of time one looks back to ascertain study eligibility), Figure 3 presents the distribution of estimates of complete case probability $\hat{\eta}$ and eligibility probability $\hat{\varepsilon}$ used in computing $\hat{\theta}_{\text{EIF}}$. The distribution of complete case probability displayed greater sensitivity to the lookback length for diabetes labs than for BMI. This is not entirely surprising as A1c measures were less commonly available (68.5% missing within 1 month prior to surgery, 19.2% missing within 6 months) than BMI measures (16.2% missing within 1 month prior to surgery, 1.5% within 6 months), and thus T2DM was the main criterion driving missingness in eligibility. Similarity between the rows of Figure 3 suggests that ascertainment of T2DM was more affected by availability of lab measures than medications. The covariate-adjusted average difference in observing complete eligibility information between RYGB and SG, $\mathbb{P}_n[\hat{\eta}_1(\mathbf{L}^*) - \hat{\eta}_0(\mathbf{L}^*)]$, ranged from -6.1% to 7.6%. Differential missingness by procedure was most prominent when the length of time for BMI and T2DM lookback windows differed.

Average values of ε_1 ranged from 0.613 in the most stringent settings with to 0.300 in the most relaxed, indicating that a higher proportion of subjects with complete information were judged to be eligible in lookback windows with narrower scope. A related consideration is whether the plausibility of MAR Assumption 4 also varies by lookback length. Given the increased interaction a patient has with the health care system prior to surgery (42), it is certainly plausible that Assumption 4 is most likely to hold within 6 months of surgery and may be less likely to hold when looking 1-2 years prior to surgery to ascertain eligibility.

Point estimates and 95% confidence intervals for $\hat{\theta}_{CC}$, $\hat{\theta}_{IWOR}$, $\hat{\theta}_{IF}$, and $\hat{\theta}_{EIF}$ across application of eligibility for both outcomes are presented in Figure 4. Point estimates for $\hat{\theta}_{CC}$ ranged from -7.1% to -6.2% for 3-year post surgical weight change and 3.7% to 8.2% for T2DM remission, indicating that eligible subjects undergoing RYGB (with complete information) experienced greater weight loss and rates of remission than had they undergone SG instead. $\hat{\theta}_{IWOR}$ had the same range for weight change estimates as $\hat{\theta}_{CC}$, but indicated



Figure 3: Distributions of select nuisance function estimates from $\hat{\theta}_{\text{EIF}}$ related to ascertainment $(\hat{\eta})$ and eligibility $(\hat{\varepsilon})$ for relative weight change outcome. An analogous figure for T2DM remission is available in the Supplementary Materials. Results for BMI lookback of 3 months are similar to those of 1 month BMI lookback, and omitted for space considerations.

greater differences in T2DM remission (range 5.2% to 10.3%). Point estimates for $\hat{\theta}_{\text{EIF}}$ were attenuated towards the null by an average of 19.5% for relative weight change compared to $\hat{\theta}_{\text{CC}}$ (range -6.6% to -4.5%), and 34.1% for T2DM remission (range 0.1% to 7.4%). Estimated standard errors were on average 5.3% larger for $\hat{\theta}_{\text{IF}}$ than corresponding standard error estimates for $\hat{\theta}_{\text{EIF}}$. Point estimates from influence-function based estimators were most similar to those of $\hat{\theta}_{\text{CC}}$ and $\hat{\theta}_{\text{IWOR}}$ in longer lab lookbacks, with differences exceeding 6.5% in a 1-month T2DM labs lookback window. While $\hat{\theta}_{\text{EIF}}$ still shows significant weight loss benefits to RYGB relative to SG, there is less evidence to suggest an RYGB advantage for remission of T2DM, at least among patients with a DiaRem score of ≥ 3 .

In addition to motivations related to the plausibility of MAR Assumption 4 and reducing the risk of inappropriately including ineligible patients, one might also be willing to adopt a more stringent lookback window for diabetes labs given the short time frame used in the definition of the remission outcome. Altogether, if shorter lookback time frames are preferred, then we might conclude that patients undergoing RYGB would not be giving up a substantial decline in T2DM remission rate had they received SG instead. As such, remission of diabetes may be less significant of a factor in determining which surgical procedure to recommend to prospective patients, as least relative to factors like durability of weight loss and risk of complications associated with surgery.



Figure 4: Point estimates and 95% confidence intervals for four estimators of the average treatment effect bariatric surgery type on eligible RYGB patients. Estimates are presented for difference in % weight change and diabetes remission rate 3 years post surgery. Results for BMI lookback of 3 months are similar to those of 1 month BMI lookback, and omitted for space considerations.

Finally, we note that in Figure 4, 95% confidence intervals for $\hat{\theta}_{\text{EIF}}$ and $\hat{\theta}_{\text{IF}}$ do not monotonically get tighter with less stringent operationalization of study eligibility. There seems to be a complex interplay between increasing the number of subjects with complete information and the complexity of modeling certain nuisance functions. While increasing the "R = 1" fraction corresponded to a greater number of patients deemed eligible (Figure 2), more stringent lookbacks yielded in a smaller fraction of patients with complete eligibility information (Figure 3). In particular, nested nuisances (γ, ξ, χ, ν) model compound outcomes which pre-multiply by E, and thus may simultaneously have to contend with greater degrees of zero-inflation and extreme values (Supplemental Figure S4).

7 Discussion

When designing and analyzing observational studies from EHR databases, the potential for confounding bias is often considered paramount, with missing data frequently viewed as a secondary consideration. While analysts might consider methods for dealing with missing data in certain covariates or outcomes after finalizing their analysis population, it is rare that they account for the possibility that discarding subjects without ascertainable eligibility when building an analysis population can introduce selection bias. Towards addressing this form of selection bias, we introduced a pair of estimators, $\hat{\theta}_{\rm EIF}$ and $\hat{\theta}_{\rm IF}$ which are robust to various degrees of model misspecification, attain \sqrt{n} -convergence to normal distributions even under flexible machine learning modeling strategies, and in the case of $\hat{\theta}_{\rm EIF}$, can achieve semiparametric efficiency bounds.

In prior work, the multiple imputation strategy of Tompsett et al. (8) and Austin et el. (18) requires modeling $\lambda(\boldsymbol{L}_m^e; \boldsymbol{L}^*)$ directly. Such a task does not readily facilitate nonparametric modeling strategies, particularly when \boldsymbol{L}_m^e is multidimensional. Of greater practical concern however is how this strategy would be implemented in EHR-studies where 70% of subjects are missing eligibility (Figure 1). Benz et al. (9) used IPW based on $\eta(\boldsymbol{L}^*, A)$, which is a similar idea to $\hat{\theta}_{\text{IWOR}}$. Their work focused on sequential target trial emulations (an observational study design for time-to-event endpoints) and additional challenges inherent to that study design, including time-varying eligibility. Extensions of our work to designs where study eligibility can vary over time is a valuable direction for future research.

Critical to addressing the possibility of selection bias due to missing eligibility data is the MAR assumption for L_m^e (Assumption 4). Within our MAR statement we also assumed that conditionally on fully observed covariates and treatment, eligibility ascertainment was independent of the outcome. The set of studies that work with Assumption 4 is quite broad, including Benz et al. (9), Tompsett et al. (8), and Austin et al. (18). Relaxation of this assumption to remove Y from the joint independence is possible but would ultimately entail integration of several nuisance functions over the entire conditional distribution of L_m^e , similar to the strategy used by Levis et al. (26) in the case of missing confounders. Given the presence of multiple nested nuisance functions, this solution would likely pose significant computational challenges, to the point where it might inhibit the practical use of the resulting method.

For the purposes of developing $\hat{\theta}_{\text{EIF}}$ and $\hat{\theta}_{\text{IF}}$, we worked with a coarsened version of the data, treating \boldsymbol{L}_m^e as fully observed or completely missing. One benefit of such a decision is that it avoids the complexities associated with non-monotone missingness. Assessing the plausibility of a version of Assumption 4 considering $2^q - 1$ possible missingness patterns may often be very challenging. More practically, the best existing methods for non-monotone missingness under MAR (51) rely on parametric models which condition on unobservable covariate strata (26). In practice, analysts may consider treating R as an indicator for whether E is determined given the available information in \boldsymbol{L}_m^e to reduce information loss, particularly if the portions of \boldsymbol{L}_m^e which are missing are not confounders.

The focus of this work was on missingness in eligibility-defining covariates L^e , and thus implicit to this work was the assumption that there was no missingness in other variables. In practice, that is unlikely to be the case in complex EHR-based studies, and missing data can affect ascertainment of confounders, treatment, and outcomes, perhaps simultaneously. In cases where outcomes Y or non-eligibility related covariates L^* are missing with less frequency than L^e (Figure 2), imputation may be a reasonable approach. We adopted this approach in our study and comment on specifics in the Supplementary Materials. If the primary source of missing data came from L^* , the path of least resistance might be to include the subset of L^* which was missing within L^e_m , but such an approach is likely to introduce non-monotone missingness when multiple components of L^* are missing, not to mention reducing the likelihood that Assumption 4 holds.

When missing outcomes are of concern, analysts may consider $\hat{\theta}_{IF}$ instead of $\hat{\theta}_{EIF}$, as the former does not use Y in the conditioning set of any component nuisance function. One area of future work is the case where Y and L^e have a form of monotone missingness, for example when L^e is needed for the definition of Y as in the case for weight change post surgery. Under this scenario, it seems possible to develop similar influence function-based estimation strategies, albeit with additional assumptions.

Code Availability

All code for analysis and simulations is made available on GitHub at https://github. com/lbenz730/semiparametric_missing_elig.

References

- [1] Kathy L Hudson and Francis S Collins. The 21st Century Cures Act? A view from the NIH. New England Journal of Medicine, 376(2):111–113, 2017.
- [2] National Center for Research Resources. Widening the use of electronic health record data for research. Videocast, 2009. http://videocast.nih.gov/summary.asp? (accessed 27 Nov 2013).
- [3] Miguel Hernán and James Robins. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, 183(8):758–764, 2016.
- [4] Sebastien Haneuse and Michael Daniels. A general framework for considering selection bias in EHR-based studies: what data are observed and why? *eGEMs*, 4(1), 2016.
- [5] U.S. Food and Drug Administration. Screening tests prior to study enrollment: Guidance for institutional review boards and clinical investigators. Technical report, January 1998.
- [6] Goodarz Danaei, Luis A Rodríguez, Ofelia F Cantero, et al. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Statistical methods in medical research*, 22(1):70–96, 2013.
- [7] Kathleen M. McTigue, Rebecca Wellman, Eric Nauman, et al. Comparing the 5-year diabetes outcomes of sleeve gastrectomy and gastric bypass: The pcornet bariatric study. JAMA Surgery, 155(10):1–9, 2020.
- [8] D. Tompsett, A. Zylbersztejn, P. Hardelid, and B. De Stavola. Target trial emulation and bias through missing eligibility data: An application to a study of palivizumab for the prevention of hospitalization due to infant respiratory illness. *American Journal* of Epidemiology, 192(4):600–611, 2023.
- [9] Luke Benz, Rajarshi Mukjerkee, Rui Wang, et al. Adjusting for Selection Bias due to Missing Eligibility Criteria in Emulated Target Trials. American Journal of Epidemiology, 2024.
- [10] D. Arterburn and A. Courcoulas. Bariatric surgery for obesity and metabolic conditions in adults. BMJ, 349:g3961, 2014.
- [11] David E Arterburn, Eric Johnson, Karen J Coleman, et al. Weight outcomes of sleeve gastrectomy and gastric bypass compared to nonsurgical treatment. *Annals of Surgery*, 274(6):e1269–e1276, 2020.
- [12] Karen J. Coleman, Sebastien Haneuse, Eric Johnson, et al. Long-term microvascular disease outcomes in patients with type 2 diabetes after bariatric surgery: Evidence for the legacy effect of surgery. *Diabetes care*, 39(8):1400–1407, 2016.
- [13] K. J. Coleman, Y. H. Shu, H. Fischer, et al. Bariatric surgery and risk of death in persons with chronic kidney disease. Annals of Surgery, 276(6):e784–e791, 2022.
- [14] Anita P Courcoulas, Nadine J Christian, Steven H Belle, Paul D Berk, David R Flum, Lorena Garcia, Mary Horlick, Melissa A Kalarchian, Wendy C King, James E Mitchell, Emma J Patterson, John R Pender, Alfons Pomp, Walter J Pories, Richard C Thirlby, Susan Z Yanovski, Bruce M Wolfe, and Longitudinal Assessment of Bariatric Surgery (LABS) Consortium. Weight change and health outcomes at 3 years after bariatric surgery among individuals with severe obesity. JAMA, 310(22):2416–2425, 2013.

- [15] R. O'Brien, E. Johnson, S. Haneuse, et al. Microvascular outcomes in patients with diabetes after bariatric surgery versus usual care: A matched cohort study. Annals of Internal Medicine, 169(5):300–310, 2018.
- [16] Qianying Pan and Douglas E. Schaubel. Proportional hazards regression in the presence of missing study eligibility information. *Lifetime Data Analysis*, 20(3):424–443, 2014.
- [17] D.Y.C. Heng, T.K. Choueiri, B.I. Rini, et al. Outcomes of patients with metastatic renal cell carcinoma that do not meet eligibility criteria for clinical trials. Annals of Oncology, 25(1):149–154, 2014.
- [18] Peter C Austin, Daniele Giardiello, and Stef van Buuren. Impute-then-exclude versus exclude-then-impute: Lessons when imputing a variable used both in cohort creation and as an independent variable in the analysis model. *Statistics in Medicine*, 42(10):1525–1541, 2023.
- [19] Edward H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review, 2023.
- [20] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, et al. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018.
- [21] Susan M. Shortreed and Ashkan Ertefaie. Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*, 73(4):1111–1122, 2017.
- [22] Edward H Kennedy. Semiparametric theory and empirical processes in causal inference. *Statistical causal inferences and their applications in public health research*, pages 141–167, 2016.
- [23] Anastasios Tsiatis. Semiparametric Theory and Missing Data. Springer Series in Statistics. Springer, New York, 2006.
- [24] Aad W Van der Vaart. Asymptotic statistics, volume 3. Cambridge university press, 2000.
- [25] Mark J Laan and James M Robins. Unified methods for censored longitudinal data and causality. Springer, 2003.
- [26] Alexander W Levis, Rajarshi Mukherjee, Rui Wang, and Sebastien Haneuse. Robust causal inference for point exposures with missing confounders. *Canadian Journal of Statistics*, page e11832, 2024.
- [27] Iván Díaz. Machine learning in the estimation of causal effects: targeted minimum loss-based estimation and double/debiased machine learning. *Biostatistics (Oxford, England)*, 21(2):353–358, 2020.
- [28] Kaiser Permanente. Weight-loss (bariatric) surgery. https://healthy.kaiserpermanente.org/health-wellness/healthencyclopedia/he.weight-loss-bariatric-surgery.abq5755, 2023.
- [29] S. Toh, L. J. Rasmussen-Torvik, E. E. Harmata, et al. The national patient-centered clinical research network (pcornet) bariatric study cohort: rationale, methods, and baseline characteristics. *JMIR Research Protocols*, 6(12):e222, 2017.
- [30] B. N. Reames, J. F. Finks, D. Bacal, A. M. Carlin, and J. B. Dimick. Changes in bariatric surgery procedure use in michigan, 2006-2013. *JAMA*, 312(9):959–961, 2014.

- [31] Ron A. Li, Lisa Liu, David Arterburn, et al. Five-year longitudinal cohort study of reinterventions after sleeve gastrectomy and roux-en-y gastric bypass. *Annals of* surgery, 273(4):758–765, 2021.
- [32] Marko Kraljević, Julian Süsstrunk, Bettina Karin Wölnerhanssen, et al. Long-term outcomes of laparoscopic roux-en-y gastric bypass vs laparoscopic sleeve gastrectomy for obesity: The sm-boss randomized clinical trial. *JAMA Surgery*, 2025.
- [33] Amanda Jiménez, Roser Casamitjana, Lílliam Flores, et al. Long-term effects of sleeve gastrectomy and roux-en-y gastric bypass surgery on type 2 diabetes mellitus in morbidly obese subjects. *Annals of surgery*, 256(6):1023–1029, 2012.
- [34] Francesca Abbatini, Mario Rizzello, Giovanni Casella, et al. Long-term effects of laparoscopic sleeve gastrectomy, gastric bypass, and adjustable gastric banding on type 2 diabetes. *Surgical endoscopy*, 24:1005–1010, 2010.
- [35] Wei-Jei Lee, Keong Chong, Kong-Han Ser, et al. Gastric bypass vs sleeve gastrectomy for type 2 diabetes mellitus: a randomized controlled trial. Archives of surgery, 146(2):143–148, 2011.
- [36] Paulina Salminen, Mika Helmiö, Jari Ovaska, et al. Effect of laparoscopic sleeve gastrectomy vs laparoscopic roux-en-y gastric bypass on weight loss at 5 years among patients with morbid obesity: the sleevepass randomized clinical trial. Jama, 319(3):241– 254, 2018.
- [37] American Diabetes Association Professional Practice Committee. 8. obesity and weight management for the prevention and treatment of type 2 diabetes: Standards of care in diabetes-2024. *Diabetes Care*, 47(Supplement_1), 2024.
- [38] D. P. Fisher, E. Johnson, S. Haneuse, et al. Association between bariatric surgery and macrovascular disease outcomes in patients with type 2 diabetes and severe obesity. *JAMA*, 320(15):1570–1582, 2018.
- [39] Christopher D. Still, Gerald C. Wood, Philip Benotti, et al. Preoperative prediction of type 2 diabetes remission after roux-en-y gastric bypass surgery: a retrospective cohort study. *The Lancet Diabetes & Endocrinology*, 2(1):38–45, Jan 2014.
- [40] MA Hernán and J Robins. *Causal Inference: What if.* Boca Raton: Chapman & Hall/CRC, 2024.
- [41] James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9–12):1393–1512, 1986.
- [42] A. L. Madenci, K. E. Kurgansky, B. A. Dickerman, et al. Estimating the effect of bariatric surgery on cardiovascular events using observational data? *Epidemiology*, 35(5):721–729, 2024.
- [43] K Evans, I Fulcher, and E Tchetgen Tchetgen. A coherent likelihood parametrization for doubly robust estimation of a causal effect with missing confounders. *arXiv preprint arXiv:2007.10393*, 2020.
- [44] A. Wang, R.A. Nianogo, and O.A. Arah. G-computation of average treatment effects on the treated and the untreated. *BMC Medical Research Methodology*, 17(3):1–8, 2017.
- [45] E. H. Kennedy, A. Sjölander, and D. S. Small. Semiparametric causal inference in matched cohort studies. *Biometrika*, 102(3):739–746, September 2015.

- [46] Kenta Takatsu, Alexander W Levis, Edward Kennedy, et al. Doubly robust machine learning-based estimation methods for instrumental variables with an application to surgical care for cholecystitis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, page qnae089, 2024.
- [47] Alexander W. Levis, Edward H. Kennedy, and Luke Keele. Nonparametric identification and efficient estimation of causal effects with instrumental variables, 2024.
- [48] Mark J van der Laan, Eric C Polley, and Alan E Hubbard. Super learner. *Statistical applications in genetics and molecular biology*, 6:Article25, 2007.
- [49] Eric Polley, Erin LeDell, Chris Kennedy, and Mark van der Laan. Superlearner: Super learner prediction. https://CRAN.R-project.org/package=SuperLearner, 2023. R package version 2.0-28.1.
- [50] Larry Wasserman. All of Statistics: A Concise Course in Statistical Inference. Springer Science & Business Media, 2004.
- [51] Bingkai Sun and Eric J. Tchetgen Tchetgen. On inverse probability weighting for nonmonotone missing at random data. *Journal of the American Statistical Association*, 113(521):369–379, 2018.

Supplementary Marterial for "Robust Causal Inference for EHR-based Studies of Point Exposures with Missingness in Eligibility Criteria"

Luke Benz¹, Rajarshi Mukherjee¹, Rui Wang^{1,2,3}, David Arterburn⁴, Heidi Fischer⁵, Catherine Lee⁶, Susan M. Shortreed^{7,8}, Sebastien Haneuse^{*1}, and Alexander W. Levis^{*9}

¹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA ²Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA, USA ³Department of Population Medicine, Harvard Medical School, Boston, MA, USA

⁴Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA

⁵Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, USA

⁶Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA

⁷Biostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA ⁸Department of Biostatistics, University of Washington School of Public Health, Seattle, WA, USA ⁹Department of Statistics & Data Science, Carnegie Mellon, Pittsburgh, PA, USA

April 24, 2025

Contents

S1 Identification of θ_{ATT}^{elig} (Proof of Theorem 1)	2
S2 Alternative Representations of $\hat{\theta}_{\text{EIF}}$ and $\hat{\theta}_{\text{IF}}$ S2.1 Reparameterization of $\dot{\beta}_P(O)$ and $\dot{\beta}'_P(O)$ S2.2 Correspondence Between $\hat{\theta}_{\text{EIF}}$ and One-Step Estimation of the ATT	4 4 5
S3 Proofs of Theoretical Properties S3.1 von Mises Expansion of $\alpha(P)$ and $\beta(P)$ with $\dot{\alpha}_{P}^{*}(O)$ and $\dot{\beta}_{P}^{*}(O)$ (Proof of Lemma 1)	6
S4 Additional Simulation Information S4.1 Simulation Data Generating Process S4.2 Simulation Parameters	20 20 22
S5 Additional Data Application Information S5.1 Methodological Details	23 23 25
S6 Alternative Covariate Partitioning S6.1 Notation and Assumptions S6.2 Identification of θ_{ATT}^{elig} Under Alternative Covariate Partition	27 27 28

* denotes co-last author (AWL and SH).

S1 Identification of θ_{ATT}^{elig} (Proof of Theorem 1)

In this section, we will prove Theorem 1, that under Assumptions 1-5, $\theta_{ATT}^{elig} = \mathbb{E}_P[Y(1) - Y(0) | A = 1, E = 1]$ is identified by the functional

$$\theta(P) = \frac{\beta(P)}{\alpha(P)} = \frac{\mathbb{E}_P\left[\frac{ARE}{\eta(\boldsymbol{L}^*,1)} \left(Y - \mu_0(\boldsymbol{L}^*, \boldsymbol{L}_m^e)\right)\right]}{\mathbb{E}_P\left[\frac{ARE}{\eta(\boldsymbol{L}^*,1)}\right]}$$

Lemma S1 Let A, B, C be random variables, with $C \in \{0, 1\}$. Then

$$\mathbb{E}[A \mid B, C = 1] = \frac{\mathbb{E}[AC \mid B]}{P(C = 1 \mid B)}$$

This result follows immediately from the law of total expectation by observing that

$$\mathbb{E}[AC \mid B] = \mathbb{E}[A \mid B, C = 1]P(C = 1 \mid B) + 0 \times P(C = 0 \mid B)$$

Using this lemma, and Assumptions 1-5, we have that

$$P(E = 1 \mid A = 1) = \mathbb{E}_{P}[\mathbb{E}_{P}(E \mid \boldsymbol{L}^{*}, A = 1) \mid A = 1)]$$

$$= \mathbb{E}_{P}[\mathbb{E}_{P}(E \mid \boldsymbol{L}^{*}, A = 1, R = 1) \mid A = 1)] \quad (A4, A5)$$

$$= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\frac{RE}{\eta(\boldsymbol{L}^{*}, 1)} \mid \boldsymbol{L}^{*}, A = 1\right) \mid A = 1\right] \quad (\text{Lemma S1, Defn. of } \eta)$$

$$= \mathbb{E}_{P}\left[\frac{RE}{\eta(\boldsymbol{L}^{*}, 1)} \mid A = 1\right]$$

$$= \frac{\mathbb{E}_{P}\left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)}\right]}{P(A = 1)} \quad (\text{Lemma S1})$$

$$= \frac{\alpha(P)}{P(A = 1)}$$

Next, we have

$$\begin{split} \mathbb{E}_{P}[EY \mid A = 1] &= \mathbb{E}_{P}[\mathbb{E}_{P}(EY \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1) \mid A = 1)] \\ &= \mathbb{E}_{P}[E\mathbb{E}_{P}(Y \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1) \mid A = 1)] \quad (E = g(\boldsymbol{L}^{e}, A), \text{ fixed function of } \boldsymbol{L}^{e}, A) \\ &= \mathbb{E}_{P}[E\mathbb{E}_{P}(Y \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1, R = 1) \mid A = 1)] \quad (A4, A5) \\ &= \mathbb{E}_{P}\left[E\mathbb{E}_{P}\left(\frac{RY}{P(R = 1 \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1) \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1\right) \mid A = 1\right] \quad (Lemma S1) \\ &= \mathbb{E}_{P}\left[\frac{E}{\eta(\boldsymbol{L}^{*}, 1)}\mathbb{E}_{P}[RY \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1] \mid A = 1\right] \quad (A4, A5) \\ &= \mathbb{E}_{P}\left[\frac{RE}{\eta(\boldsymbol{L}^{*}, 1)}Y \mid A = 1\right] \\ &= \frac{\mathbb{E}_{P}\left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)}Y \mid A = 1\right] \\ &= \frac{\mathbb{E}_{P}\left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)}Y \mid A = 1\right] \\ &= \mathbb{E}_{P}\left[\mathbb{E}_{P}(A = 1) \quad (Lemma S1) \\ \\ \mathbb{E}_{P}[E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1] = \mathbb{E}_{P}\left[\frac{\mathbb{E}_{P}(R \mid \boldsymbol{L}^{*}, A = 1, \boldsymbol{L}_{m}^{e})}{\eta(\boldsymbol{L}^{*}, 1)} E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1\right] \quad (A4, A5) \\ &= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\frac{R}{\eta(\boldsymbol{L}^{*}, 1)}E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1\right) \mid A = 1\right] \\ &= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\frac{RE}{\eta(\boldsymbol{L}^{*}, 1)}E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1\right] \end{aligned}$$

$$\mathbb{E}_{P}[E\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \mid A=1] = \mathbb{E}_{P}\left[\frac{\mathbb{E}_{P}[\boldsymbol{L}^{*},\boldsymbol{1}] - (\boldsymbol{L}^{*},\boldsymbol{1})}{\eta(\boldsymbol{L}^{*},\boldsymbol{1})} E\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \mid A=1\right] \quad (A4, A5)$$

$$= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\frac{R}{\eta(\boldsymbol{L}^{*},\boldsymbol{1})} E\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \mid \boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}, A=1\right) \mid A=1\right]$$

$$= \mathbb{E}_{P}\left[\frac{RE}{\eta(\boldsymbol{L}^{*},\boldsymbol{1})} \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \mid A=1\right]$$

$$= \frac{\mathbb{E}_{P}\left[\frac{ARE}{\eta(\boldsymbol{L}^{*},\boldsymbol{1})} \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right]}{P(A=1)}$$

Applying Assumptions 1-3, the causal inference assumptions among study eligible subjects, we have

$$\mathbb{E}_P[Y(1) \mid A = 1, E = 1] = \mathbb{E}_P[Y \mid A = 1, E = 1] = \frac{\mathbb{E}_P[EY \mid A = 1]}{P(E = 1 \mid A = 1)}$$

$$\begin{split} \mathbb{E}_{P}[Y(0) \mid A = 1, E = 1] &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y(0) \mid A = 1, E = 1, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1, E = 1] \\ &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y(0) \mid A = 0, E = 1, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1, E = 1] \quad (A2, A3) \\ &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y \mid A = 0, E = 1, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1, E = 1] \quad (A1) \\ &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y \mid A = 0, E, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1, E = 1] \\ &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y \mid A = 0, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1, E = 1] \quad (E = g(\boldsymbol{L}^{e}, A) \text{ fixed}) \\ &= \mathbb{E}_{P}[\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1, E = 1] \quad (A4) \\ &= \frac{\mathbb{E}_{P}[E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 1]}{P(E = 1 \mid A = 1)} \quad (\text{Lemma S1}) \end{split}$$

Combining pieces, we obtain the desired result, namely

$$\begin{aligned} \theta_{\text{ATT}}^{\text{eng}} &= \mathbb{E}_{P}[Y(1) - Y(0) \mid A = 1, E = 1] \\ &= \frac{\mathbb{E}_{P}\left[E\left(Y - \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})\right) \mid A = 1\right]}{P(E = 1 \mid A = 1)} \\ &= \frac{\mathbb{E}_{P}\left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)}\left(Y - \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})\right)\right] \middle/ P(A = 1)}{\alpha(P)/P(A = 1)} \\ &= \frac{\beta(P)}{\alpha(P)} = \theta(P) \end{aligned}$$

S2 Alternative Representations of $\hat{\theta}_{\text{EIF}}$ and $\hat{\theta}_{\text{IF}}$

S2.1 Reparameterization of $\dot{\beta}_P(O)$ and $\dot{\beta}'_P(O)$

The likelihood factorization in Equation (2) of the main paper was closely tied to four nuisance functions, $(\pi, \eta, \lambda, \mu)$. Perhaps somewhat surprisingly then, π and λ did not appear in any estimator for $\theta(P)$. In this section, we show how both nuisance functions actually are relevant to estimation of $\theta(P)$, and why we chose an alternative nuisance function parametrization which did not require estimation of either π or λ .

Using Bayes' theorem, we see the following relationship between λ , η , π , and the complete case propensity score u.

$$\begin{aligned} \frac{\lambda_1(\boldsymbol{L}_m^e; \boldsymbol{L}^*)}{\lambda_0(\boldsymbol{L}_m^e; \boldsymbol{L}^*)} &= \frac{p(\boldsymbol{L}_m^e \mid \boldsymbol{L}^*, A = 1, R = 1)}{p(\boldsymbol{L}_m^e \mid \boldsymbol{L}^*, A = 0, R = 1)} \\ &= \frac{p(\boldsymbol{L}_m^e, \boldsymbol{L}^*, A = 1, R = 1)/p(\boldsymbol{L}^*, A = 1, R = 1)}{p(\boldsymbol{L}_m^e, \boldsymbol{L}^*, A = 0, R = 1)/p(\boldsymbol{L}^*, A = 0, R = 1)} \\ &= \frac{p(A = 1 \mid \boldsymbol{L}_m^e, \boldsymbol{L}^*, R = 1)/p(\boldsymbol{L}^*, A = 0, R = 1)}{p(A = 0 \mid \boldsymbol{L}_m^e, \boldsymbol{L}^*, R = 1)/p(\boldsymbol{L}^*, A = 0, R = 1)} \\ &= \frac{u(\boldsymbol{L}^*, \boldsymbol{L}_m^e)}{1 - u(\boldsymbol{L}^*, \boldsymbol{L}_m^e)} \frac{[1 - \pi(\boldsymbol{L}^*)]\eta(\boldsymbol{L}^*, 0)}{\pi(\boldsymbol{L}^*)\eta(\boldsymbol{L}^*, 1)} \end{aligned}$$

This identity yields alternative parametrizations of $\dot{\beta}_P(O)$ and $\dot{\beta}'_P(O)$, respectively, as follows:

$$\begin{split} \dot{\beta}_{P}(O) &= \frac{AR}{\eta(\boldsymbol{L}^{*},1)} \bigg[\bigg(E - \varepsilon_{1}(\boldsymbol{L}^{*},Y) \bigg) Y - \bigg(E\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \xi(\boldsymbol{L}^{*},Y) \bigg) \bigg] + A \bigg(\varepsilon_{1}(\boldsymbol{L}^{*},Y)Y - \xi(\boldsymbol{L}^{*},Y) \bigg) \\ &- \frac{(1-A)R}{\eta(\boldsymbol{L}^{*},0)} \frac{\pi(\boldsymbol{L}^{*})}{1-\pi(\boldsymbol{L}^{*})} \bigg[E \frac{\lambda_{1}(\boldsymbol{L}_{m}^{e};\boldsymbol{L}^{*})}{\lambda_{0}(\boldsymbol{L}_{m}^{e};\boldsymbol{L}^{*})} \Big(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Big) - \Big(\gamma'(\boldsymbol{L}^{*},Y)Y - \chi'(\boldsymbol{L}^{*},Y) \Big) \bigg] \\ &- (1-A) \frac{\pi(\boldsymbol{L}^{*})}{1-\pi(\boldsymbol{L}^{*})} \Big(\gamma'(\boldsymbol{L}^{*},Y)Y - \chi'(\boldsymbol{L}^{*},Y) \Big) \\ \dot{\beta}_{P}'(O) &= A \bigg(1 - \frac{R}{\eta(\boldsymbol{L}^{*},1)} \bigg) \nu(\boldsymbol{L}^{*}) + RE \bigg(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \bigg) \bigg[\frac{A}{\eta(\boldsymbol{L}^{*},1)} - \frac{(1-A)}{\eta(\boldsymbol{L}^{*},0)} \frac{\pi(\boldsymbol{L}^{*})}{1-\pi(\boldsymbol{L}^{*})} \frac{\lambda_{1}(\boldsymbol{L}_{m}^{e};\boldsymbol{L}^{*})}{\lambda_{0}(\boldsymbol{L}_{m}^{e};\boldsymbol{L}^{*})} \bigg] \end{split}$$

where

$$\gamma'(\boldsymbol{L}^*, Y) = \mathbb{E}\Big[E\frac{\lambda_1(\boldsymbol{L}_m^e; \boldsymbol{L}^*)}{\lambda_0(\boldsymbol{L}_m^e; \boldsymbol{L}^*)} \mid \boldsymbol{L}^*, A = 0, R = 1, Y\Big]$$

and

$$\chi'(\boldsymbol{L}^*, Y) = \mathbb{E}\Big[E\frac{\lambda_1(\boldsymbol{L}^e_m; \boldsymbol{L}^*)}{\lambda_0(\boldsymbol{L}^e_m; \boldsymbol{L}^*)}\mu_0(\boldsymbol{L}^*, \boldsymbol{L}^e_m) \mid \boldsymbol{L}^*, A = 0, R = 1, Y\Big]$$

are versions of γ and χ induced by the switch from propensity score odds $\frac{u(\boldsymbol{L}^*, \boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*, \boldsymbol{L}_m^e)}$ to density ratio $\frac{\lambda_1(\boldsymbol{L}_m^e; \boldsymbol{L}^*)}{\lambda_0(\boldsymbol{L}_m^e; \boldsymbol{L}^*)}$.

In fact, the above versions of $\dot{\beta}_P(O)$ and $\dot{\beta}'_P(O)$ were the original expressions we derived, and follow much more directly from the likelihood factorization in Equation (2). The presence of density ratio $\frac{\lambda_1(L_m^e; L^*)}{\lambda_0(L_m^e; L^*)}$ makes nonparametric estimation more challenging given that both \boldsymbol{L}_m^e and \boldsymbol{L}^* can be multidimensional and in general, conditional density estimation is a very challenging statistical problem. While other techniques for density ratio estimation exist (1), our reparameterization follows the same technique as Diaz et al. (2).

When estimating $\theta(P)$ using a one-step estimator for $\beta(P)$ based on this alternative representation of $\dot{\beta}_P(O)$ or $\dot{\beta}'_P(O)$, we denote such estimators by $\dot{\beta}_{\tilde{P}}(O)$ and $\dot{\beta}'_{\tilde{P}}(O)$, respectively, and corresponding estimators for $\theta(P)$ by $\tilde{\theta}_{\text{EIF}}$ and $\tilde{\theta}_{\text{IF}}$. A choice to use $\tilde{\theta}_{\text{EIF}}$ and $\tilde{\theta}_{\text{IF}}$ rather than $\hat{\theta}_{\text{EIF}}$ or $\hat{\theta}_{\text{IF}}$ might be more appropriate if analysts had understanding of the exact process giving rise to eligibility defining covariates L^e . For completeness, we provide summary of the asymptotics of these alternative estimators in Section S3.5.

S2.2 Correspondence Between $\hat{\theta}_{EIF}$ and One-Step Estimation of the ATT

In the absence of missing eligibility data, R = 1 for all subjects. Suppose further that all n remaining subjects are eligible for the study (E = 1). Such might be the case if analysis is restricted to n eligible subjects in an EHR database with N > n subjects. This is analogous to the typical way cohorts are built in observational studies for point exposures if missing eligibility criteria is not an issue. We will show informally that when (R = 1, E = 1) for all n subjects on which we observe data units $(O_1, ..., O_n)$, $\hat{\theta}_{\text{EIF}}$ simplifies to the usual one-step estimator for the average treatment effect on the treated (ATT). Under this scenario

- R = 1, E = 1 for all subjects
- $\eta(L^*, A) = 1$ for $A \in \{0, 1\}$ (All *n* patients are complete cases)
- $\varepsilon_a(\mathbf{L}^*, Y) = 1$ for $a \in \{0, 1\}$ (All *n* patients are study eligible)

$$\begin{split} \mathbb{P}_{n}[\dot{\beta}_{P}(O)] &= \mathbb{P}_{n} \Biggl\{ \frac{AR}{\hat{\eta}(\boldsymbol{L}^{*},1)} \Biggl[\Biggl(\boldsymbol{E} - \hat{\varepsilon}_{1}(\boldsymbol{L}^{*},Y) \Biggr) Y - \Bigl(\boldsymbol{E}\hat{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \hat{\xi}(\boldsymbol{L}^{*},Y) \Biggr) \Biggr] \\ &+ A \Biggl(\hat{\varepsilon}_{1}(\boldsymbol{L}^{*},Y) Y - \hat{\xi}(\boldsymbol{L}^{*},Y) \Biggr) \\ &- \frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)} \Biggl[\boldsymbol{E} \frac{\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \Biggl(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Biggr) - \Bigl(\gamma(\boldsymbol{L}^{*},Y) Y - \chi(\boldsymbol{L}^{*},Y) \Biggr) \Biggr] \\ &- (1-A) \frac{\eta(\boldsymbol{L}^{*},0)}{\eta(\boldsymbol{L}^{*},1)} \Biggl(\gamma(\boldsymbol{L}^{*},Y) Y - \chi(\boldsymbol{L}^{*},Y) \Biggr) \Biggr\} \\ &= \mathbb{P}_{n} \Biggl\{ -A \Bigl(\hat{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \hat{\xi}(\boldsymbol{L}^{*},Y) \Bigr) + A \Bigl(Y - \hat{\xi}(\boldsymbol{L}^{*},Y) \Bigr) \\ &- (1-A) \Biggl[\frac{\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \Bigl(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Biggr) - \Bigl(\gamma(\boldsymbol{L}^{*},Y) Y - \chi(\boldsymbol{L}^{*},Y) \Bigr) \Biggr] \\ &- (1-A) \Biggl[\frac{\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \Biggl(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Biggr) - \Bigl(\gamma(\boldsymbol{L}^{*},Y) Y - \chi(\boldsymbol{L}^{*},Y) \Biggr) \Biggr] \\ &= \mathbb{P}_{n} \Biggl\{ A \Bigl(Y - \hat{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Bigr) - (1-A) \Biggl[\frac{\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \Bigl(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Biggr) \Biggr] \Biggr\} \\ &= \mathbb{P}_{n} \Biggl\{ \Biggl\{ \Biggl\{ Y - \hat{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Biggr\} - (1-A) \Biggl[\frac{\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-\hat{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \Biggl\} \Biggr\}$$

Similarly,

$$\mathbb{P}_{n}[\dot{\alpha}_{P}(O)] = \mathbb{P}_{n}\left\{ A\left(1 - \frac{R}{\widehat{\eta}(\boldsymbol{L}^{*}, 1)}\right)\widehat{\varepsilon}_{1}(\boldsymbol{L}^{*}, Y) + \frac{ARE}{\widehat{\eta}(\boldsymbol{L}^{*}, 1)} \right\} = \mathbb{P}_{n}[A]$$

Thus

$$\widehat{\theta}_{\mathrm{EIF}} = \frac{\mathbb{P}_n[\dot{\beta}_P(O)]}{\mathbb{P}_n[\dot{\alpha}_P(O)]} = \mathbb{P}_n\left\{\frac{1}{\mathbb{P}_n[A]}\left(Y - \widehat{\mu}_0(\boldsymbol{L}^*, \boldsymbol{L}_m^e)\right)\left[A - (1 - A)\left(\frac{\widehat{u}(\boldsymbol{L}^*, \boldsymbol{L}_m^e)}{1 - \widehat{u}(\boldsymbol{L}^*, \boldsymbol{L}_m^e)}\right)\right]\right\}$$

which is nothing more than the one-step estimator (eg., doubly-robust) for the ATT (3–5). A very similar argument yields the same result for $\hat{\theta}_{IF}$.

S3 Proofs of Theoretical Properties

S3.1 von Mises Expansion of $\alpha(P)$ and $\beta(P)$ with $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ (Proof of Lemma 1)

Before proving Lemma 1, we will note that

$$\alpha(P) = \mathbb{E}_{P} \left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)} \right]$$
$$= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*}, 1)} \mathbb{E}_{P}[E \mid \boldsymbol{L}^{*}, A = 1, R = 1, Y] \right]$$
$$= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*}, 1)} \varepsilon_{1}(\boldsymbol{L}^{*}, Y) \right]$$
$$= \mathbb{E}_{P}[A\varepsilon_{1}(\boldsymbol{L}^{*}, Y)]$$

where the last line follows by applying iterated expectation conditioning on L^* , A, Y and than applying Assumption 4 (MAR) to note that $\mathbb{E}_P[R \mid L^*, A = 1, Y] = \mathbb{E}_P[R \mid L^*, A = 1] = \eta(L^*, 1)$. Similarly, we note that

$$\beta(P) = \mathbb{E}_{P} \left[\frac{ARE}{\eta(\boldsymbol{L}^{*},1)} \left(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \right) \right]$$
$$= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*},1)} \mathbb{E}_{P} \left[E \left(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \right) \middle| \boldsymbol{L}^{*}, A = 1, R = 1, Y \right] \right]$$
$$= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*},1)} \left(Y \varepsilon_{1}(\boldsymbol{L}^{*},Y) - \xi(\boldsymbol{L}^{*},Y) \right) \right] \quad (\text{Defn. of } \varepsilon_{a},\xi)$$
$$= \mathbb{E}_{P} \left[A \left(Y \varepsilon_{1}(\boldsymbol{L}^{*},Y) - \xi(\boldsymbol{L}^{*},Y) \right) \right]$$

where again the last line follows by applying iterated expectation conditioning on L^* , A, Yand than applying Assumption 4 (MAR) to note that $\mathbb{E}_P[R \mid L^*, A = 1, Y] = \mathbb{E}_P[R \mid L^*, A = 1] = \eta(L^*, 1)$. Now to prove Lemma 1:

$$\begin{aligned} R_{\alpha}(\overline{P},P) &= \alpha(\overline{P}) - \alpha(P) + \int \dot{\alpha}_{\overline{P}}^{*}(o)dP(o) \\ &= \alpha(\overline{P}) - \alpha(P) + \mathbb{E}_{P} \left[A \left(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y) + \frac{ARE}{\overline{\eta}(\boldsymbol{L}^{*},1)} - \alpha(\overline{P}) \right] \\ &= \mathbb{E}_{P} \left[A \left(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y) + \frac{ARE}{\overline{\eta}(\boldsymbol{L}^{*},1)} - A\varepsilon_{1}(\boldsymbol{L}^{*},Y) \right] \\ &= \mathbb{E}_{P} \left[A \left(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y) + \frac{AR\varepsilon_{1}(\boldsymbol{L}^{*},Y)}{\overline{\eta}(\boldsymbol{L}^{*},1)} - A\varepsilon_{1}(\boldsymbol{L}^{*},Y) \right] \quad (\text{It. Exp. on } \boldsymbol{L}^{*}, A, R, Y) \\ &= \mathbb{E}_{P} \left[A \left(1 - \frac{\eta(\boldsymbol{L}^{*},1)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \bar{\varepsilon}_{1}(\boldsymbol{L}^{*}) + A \frac{\eta(\boldsymbol{L}^{*},1)\varepsilon_{1}(\boldsymbol{L}^{*})}{\overline{\eta}(\boldsymbol{L}^{*},1)} - A\varepsilon_{1}(\boldsymbol{L}^{*}) \right] \quad (\text{It. Exp. on } \boldsymbol{L}^{*}, A, Y + \text{A4}) \\ &= \mathbb{E}_{P} \left[A (\bar{\varepsilon}_{1} - \varepsilon_{1}) \left(1 - \frac{\eta_{1}}{\overline{\eta}_{1}} \right) \right] \end{aligned}$$

$$\begin{split} R_{\beta}(\overline{P},P) &= \beta(\overline{P}) - \beta(P) + \int \dot{\beta}_{\overline{P}}^{*}(o)dP(o) \\ &= \beta(\overline{P}) - \beta(P) + \\ \mathbb{E}_{P}\bigg[\frac{AR}{\overline{\eta}(\boldsymbol{L}^{*},1)}\bigg\{\Big(E - \bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y)\Big)Y - \Big(E\overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \bar{\xi}(\boldsymbol{L}^{*},Y)\Big)\bigg\} + A\Big(\bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y)Y - \bar{\xi}(\boldsymbol{L}^{*},Y)\Big) \\ &- \frac{(1-A)R}{\overline{\eta}(\boldsymbol{L}^{*},1)}\bigg\{E\frac{\overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-\overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\Big(Y - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\Big) - \Big(\overline{\gamma}(\boldsymbol{L}^{*},Y)Y - \overline{\chi}(\boldsymbol{L}^{*},Y)\Big)\bigg\} \\ &- (1-A)\frac{\overline{\eta}(\boldsymbol{L}^{*},0)}{\overline{\eta}(\boldsymbol{L}^{*},1)}\Big(\overline{\gamma}(\boldsymbol{L}^{*},Y)Y - \overline{\chi}(\boldsymbol{L}^{*},Y)\Big) - \beta(\overline{P})\bigg] \\ &= \mathbb{E}_{P}\bigg[\frac{AR}{\overline{\eta}(\boldsymbol{L}^{*},1)}\bigg\{\Big(E - \bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y)\Big)Y - \Big(E\overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \bar{\xi}(\boldsymbol{L}^{*},Y)\Big)\bigg\} + A\Big(\bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y)Y - \bar{\xi}(\boldsymbol{L}^{*},Y)\Big) \\ &- \frac{(1-A)R}{\overline{\eta}(\boldsymbol{L}^{*},1)}\bigg\{E\frac{\overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-\overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\Big(Y - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\Big) - \Big(\overline{\gamma}(\boldsymbol{L}^{*},Y)Y - \overline{\chi}(\boldsymbol{L}^{*},Y)\Big)\bigg\} \\ &- (1-A)\frac{\overline{\eta}(\boldsymbol{L}^{*},0)}{\overline{\eta}(\boldsymbol{L}^{*},1)}\Big(\overline{\gamma}(\boldsymbol{L}^{*},Y)Y - \overline{\chi}(\boldsymbol{L}^{*},Y)\Big) - A\Big(\varepsilon_{1}(\boldsymbol{L}^{*},Y)Y - \xi(\boldsymbol{L}^{*},Y)\Big)\bigg]\bigg\} \end{split}$$

Adding and subtracting the term $E\mu_0(\boldsymbol{L}^*, \boldsymbol{L}_m^e)$ on line 1 and applying iterated expectation (conditioning on $\boldsymbol{L}^*, A, Y, \boldsymbol{L}_m^e$ in line 2) yields

$$\begin{aligned} R_{\beta}(\overline{P},P) &= \mathbb{E}_{P} \left[\frac{AR}{\overline{\eta}(\boldsymbol{L}^{*},1)} \left\{ \left(E - \bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y) \right) Y - \left(E \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) + E \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - E \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \bar{\xi}(\boldsymbol{L}^{*},Y) \right) \right\} \\ &- \frac{(1-A)R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \left\{ E \frac{\overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1 - \overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \left(\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \right) - \left(\overline{\gamma}(\boldsymbol{L}^{*},Y)Y - \overline{\chi}(\boldsymbol{L}^{*},Y) \right) \right\} \\ &- (1-A)\frac{\overline{\eta}(\boldsymbol{L}^{*},0)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \left(\overline{\gamma}(\boldsymbol{L}^{*},Y)Y - \overline{\chi}(\boldsymbol{L}^{*},Y) \right) \\ &- A \Big(\varepsilon_{1}(\boldsymbol{L}^{*},Y)Y - \xi(\boldsymbol{L}^{*},Y) \Big) + A \Big(\overline{\varepsilon}_{1}(\boldsymbol{L}^{*},Y)Y - \overline{\xi}(\boldsymbol{L}^{*},Y) \Big) \end{aligned}$$

Next, we rearrange and combine terms and begin simplifying. In particular, we apply additional iterated expectations in lines 1 (on L^* , A, R, Y) and 3 (on L^* , A, Y, followed by Assumption 4) of the below expression, and finally add and subtract $\gamma(L^*, Y)Y$ in the final term below.

$$R_{\beta}(\overline{P}, P) = \mathbb{E}_{P} \left[A \left(1 - \frac{\eta(\boldsymbol{L}^{*}, 1)}{\overline{\eta}(\boldsymbol{L}^{*}, 1)} \right) \left\{ Y \left(\bar{\varepsilon}_{1}(\boldsymbol{L}^{*}, Y) - \varepsilon_{1}(\boldsymbol{L}^{*}, Y) \right) - \left(\bar{\xi}(\boldsymbol{L}^{*}, Y) - \xi(\boldsymbol{L}^{*}, Y) \right) \right\} - \frac{RE}{\overline{\eta}(\boldsymbol{L}^{*}, 1)} \left(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) - \overline{\mu}_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \right) \left\{ A - (1 - A) \frac{\overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{1 - \overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})} \right\} - (1 - A) \frac{\left(\eta(\boldsymbol{L}^{*}, 0) - \overline{\eta}(\boldsymbol{L}^{*}, 0) \right)}{\overline{\eta}(\boldsymbol{L}^{*}, 1)} \left(\overline{\gamma}(\boldsymbol{L}^{*}, Y)Y + \gamma(\boldsymbol{L}^{*}, Y)Y - \gamma(\boldsymbol{L}^{*}, Y)Y - \overline{\chi}(\boldsymbol{L}^{*}, Y) \right) \right]$$

Applying iterated expectation once more (to line two, on $\boldsymbol{L}^*, A, R, \boldsymbol{L}_m^e$), we arrive at the

desired result.

$$\begin{split} R_{\beta}(\overline{P},P) &= \mathbb{E}_{P} \bigg[A \bigg(1 - \frac{\eta(\boldsymbol{L}^{*},1)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg) \bigg\{ Y \Big(\bar{\varepsilon}_{1}(\boldsymbol{L}^{*},Y) - \varepsilon_{1}(\boldsymbol{L}^{*},Y) \Big) - \Big(\bar{\xi}(\boldsymbol{L}^{*},Y) - \xi(\boldsymbol{L}^{*},Y) \Big) \bigg\} \\ &- \frac{RE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \Big(\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Big) \bigg\{ \frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \big(1 - \overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \big) - \overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \big(1 - u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \big)}{1 - \overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \bigg) - (1 - A) \frac{\Big(\eta(\boldsymbol{L}^{*},0) - \overline{\eta}(\boldsymbol{L}^{*},0) \Big)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg\{ Y \Big(\overline{\gamma}(\boldsymbol{L}^{*},Y) - \gamma(\boldsymbol{L}^{*},Y) \Big) - \Big(\overline{\chi}(\boldsymbol{L}^{*},Y) - \chi(\boldsymbol{L}^{*},Y) \Big) \bigg\} \bigg] \\ &= \mathbb{E}_{P} \bigg[A \bigg(1 - \frac{\eta_{1}}{\overline{\eta_{1}}} \bigg) \Big(Y(\overline{\varepsilon}_{1} - \varepsilon_{1}) - (\overline{\xi} - \xi) \Big) - \frac{RE}{\overline{\eta}} \big(\mu_{0} - \overline{\mu}_{0} \big) \bigg(\frac{u(1 - \overline{u}) - \overline{u}(1 - u)}{1 - \overline{u}} \bigg) \\ &- (1 - A) \frac{\big(\eta_{0} - \overline{\eta}_{0} \big)}{\overline{\eta_{1}}} \Big(Y(\overline{\gamma} - \gamma) - (\overline{\chi} - \chi) \Big) \bigg] \end{split}$$

This finishes the proof of Lemma 1, by showing that the claimed von Mises expansions are satisfied. An immediate consequence of Lemma 1—invoking Lemma 2 in (6)—is that $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ are influence functions at P in a semiparametric model induced by Assumption 4. As mentioned in the main paper, Assumption 4 restricts the tangent space of the model, and as such demonstrating that the claimed von Mises expansions are satisfied is not sufficient to prove that $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ the **efficient** influence functions. In order to prove Theorem 2, that $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ are indeed the efficient influence functions in the induced semiparmetric model, we will first show that $\dot{\alpha}_P'(O)$ and $\dot{\beta}_P'(O)$ satisfy a different set of von Mises expansions for $\alpha(P)$ and $\beta(P)$, respectively, and thus are themselves influence functions (Section S3.2). Then, we will go through the process of projecting $\dot{\alpha}_P'(O)$ and $\dot{\beta}_P'(O)$, thereby proving $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ are indeed the efficient influence function yields $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P'(O)$, thereby proving $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ are indeed the efficient influence functions (Section S3.3)

S3.2 von Mises Expansion of $\alpha(P)$ and $\beta(P)$ with $\dot{\alpha}_P^{\prime*}(O)$ and $\dot{\beta}_P^{\prime*}(O)$

Lemma S2 $\alpha(P)$ and $\beta(P)$ satisfy the von Mises expansions

$$\alpha(\overline{P}) - \alpha(P) = -\int \dot{\alpha}_{\overline{P}}^{\prime*}(o)dP(o) + R_{\alpha}^{\prime}(\overline{P}, P)$$
$$\beta(\overline{P}) - \beta(P) = -\int \dot{\beta}_{\overline{P}}^{\prime*}(o)dP(o) + R_{\beta}^{\prime}(\overline{P}, P)$$

where the remainder terms (omitting inputs for brevity) are as follows:

$$R'_{\alpha}(\overline{P},P) = \mathbb{E}_{P}\left[A(\overline{\omega}_{1}-\omega_{1})\left(1-\frac{\eta_{1}}{\overline{\eta}_{1}}\right)\right]$$
$$R'_{\beta}(\overline{P},P) = \mathbb{E}_{P}\left[A\left(1-\frac{\eta_{1}}{\overline{\eta}_{1}}\right)(\overline{\nu}-\nu) - \frac{RE}{\overline{\eta}}(\mu_{0}-\overline{\mu}_{0})\left(\frac{u(1-\overline{u})-\overline{u}(1-u)}{1-\overline{u}}\right)\right]$$

The proof of Lemma S2 is very similar to that of Lemma 1, from the previous section. Before proving Lemma S2, we note that

$$\alpha(P) = \mathbb{E}_{P} \left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)} \right]$$
$$= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*}, 1)} \mathbb{E}_{P}[E \mid \boldsymbol{L}^{*}, A = 1, R = 1] \right]$$
$$= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*}, 1)} \omega_{1}(\boldsymbol{L}^{*}) \right]$$
$$= \mathbb{E}_{P}[A\omega_{1}(\boldsymbol{L}^{*})]$$

where the last line follows by applying iterated expectation conditioning on L^*, A . Additionally, we note that

$$\begin{split} \beta(P) &= \mathbb{E}_{P} \left[\frac{ARE}{\eta(\boldsymbol{L}^{*},1)} \Big(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Big) \right] \\ &= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*},1)} \mathbb{E}_{P} \left[E \Big(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \Big) \Big| \boldsymbol{L}^{*}, A = 1, R = 1 \right] \right] \\ &= \mathbb{E}_{P} \left[\frac{AR}{\eta(\boldsymbol{L}^{*},1)} \nu(\boldsymbol{L}^{*}) \right] \quad (\text{Defn. of } \nu) \\ &= \mathbb{E}_{P} [A\nu(\boldsymbol{L}^{*})] \quad (\text{It. Exp. on } \boldsymbol{L}^{*}, A) \end{split}$$

Now to prove Lemma S2

$$\begin{aligned} R'_{\alpha}(\overline{P},P) &= \alpha(\overline{P}) - \alpha(P) + \int \dot{\alpha}_{\overline{P}}^{\prime*}(o)dP(o) \\ &= \alpha(\overline{P}) - \alpha(P) + \mathbb{E}_{P} \left[A \left(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \overline{\omega}_{1}(\boldsymbol{L}^{*}) + \frac{ARE}{\overline{\eta}(\boldsymbol{L}^{*},1)} - \alpha(\overline{P}) \right] \\ &= \mathbb{E}_{P} \left[A \left(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \overline{\omega}_{1}(\boldsymbol{L}^{*}) + \frac{ARE}{\overline{\eta}(\boldsymbol{L}^{*},1)} - A\omega_{1}(\boldsymbol{L}^{*}) \right] \\ &= \mathbb{E}_{P} \left[A \left(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \overline{\omega}_{1}(\boldsymbol{L}^{*}) + \frac{AR\omega_{1}(\boldsymbol{L}^{*})}{\overline{\eta}(\boldsymbol{L}^{*},1)} - A\omega_{1}(\boldsymbol{L}^{*}) \right] \quad (\text{It. Exp. on } \boldsymbol{L}^{*}, A, R) \\ &= \mathbb{E}_{P} \left[A \left(1 - \frac{\eta(\boldsymbol{L}^{*},1)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \right) \overline{\omega}_{1}(\boldsymbol{L}^{*}) + A \frac{\eta(\boldsymbol{L}^{*},1)\omega_{1}(\boldsymbol{L}^{*})}{\overline{\eta}(\boldsymbol{L}^{*},1)} - A\omega_{1}(\boldsymbol{L}^{*}) \right] \quad (\text{It. Exp. on } \boldsymbol{L}^{*}, A) \\ &= \mathbb{E}_{P} \left[A (\overline{\omega}_{1} - \omega_{1}) \left(1 - \frac{\eta_{1}}{\overline{\eta}_{1}} \right) \right] \end{aligned}$$

Next we have that

$$\begin{split} R'_{\beta}(\overline{P},P) &= \beta(\overline{P}) - \beta(P) + \int \dot{\beta}_{\overline{P}}^{\prime*}(o)dP(o) \\ &= \beta(\overline{P}) - \beta(P) + \mathbb{E}_{P} \bigg[A \bigg(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg) \overline{\nu}(\boldsymbol{L}^{*}) + \frac{ARE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg(Y - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \bigg) \\ &- \frac{(1 - A)RE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \frac{\overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1 - \overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \bigg(Y - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \bigg) - A\nu(\boldsymbol{L}^{*}) \bigg] \\ &= \mathbb{E}_{P} \bigg[A \bigg(1 - \frac{R}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg) \overline{\nu}(\boldsymbol{L}^{*}) - A\nu(\boldsymbol{L}^{*}) + \frac{ARE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg(Y - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) + \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \bigg) \\ &- \frac{(1 - A)RE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \frac{\overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1 - \overline{u}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \bigg(Y - \overline{\mu}_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) + \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \bigg) \end{split}$$

Now rearranging terms, and applying iterated expectations, we have

$$\begin{split} R'_{\beta}(\overline{P},P) &= \mathbb{E}_{P} \bigg[A \bigg(1 - \frac{\eta(\boldsymbol{L}^{*},1)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg) \Big(\overline{\nu}(\boldsymbol{L}^{*}) - \nu(\boldsymbol{L}^{*}) \Big) & (\text{It. Exp. on } \boldsymbol{L}^{*}, A, R, \text{ then again on } \boldsymbol{L}^{*}, A) \\ &+ \frac{ARE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) - \overline{\mu}_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})) \bigg) \\ &- \frac{(1-A)RE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \frac{\overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{1 - \overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})} \bigg(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) - \overline{\mu}_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \bigg) \bigg] & (\text{It. Exp. on } \boldsymbol{L}^{*}, A, R, R, \boldsymbol{L}_{m}^{e}) \\ &= \mathbb{E}_{P} \bigg[A \bigg(1 - \frac{\eta(\boldsymbol{L}^{*},1)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg) \bigg(\overline{\nu}(\boldsymbol{L}^{*}) - \nu(\boldsymbol{L}^{*}) \bigg) + \frac{RE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) - \overline{\mu}_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \bigg) \bigg\{ A - (1-A) \frac{\overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{1 - \overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})} \bigg\} \bigg] \\ &= \mathbb{E}_{P} \bigg[A \bigg(1 - \frac{\eta(\boldsymbol{L}^{*},1)}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg) \bigg(\nu(\boldsymbol{L}^{*}) - \nu(\boldsymbol{L}^{*}) \bigg) \\ &+ \frac{RE}{\overline{\eta}(\boldsymbol{L}^{*},1)} \bigg(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) - \overline{\mu}_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) - \bigg(1 - u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \bigg) \bigg] \bigg(\frac{\overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{1 - \overline{u}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})} \bigg\} \bigg] \end{split}$$

where the final line follows from iterated expectations on L^* , L_m^e , R. Omitting conditioning terms we see that this error term is of second order

$$\mathbb{E}_{P}\left[A\left(1-\frac{\eta_{1}}{\bar{\eta}_{1}}\right)(\nu-\bar{\nu})+\frac{RE(\mu_{0}-\bar{\mu}_{0})}{\bar{\eta}_{1}}\left\{\frac{u(1-\bar{u})-\bar{u}(1-u)}{1-\bar{u}}\right\}\right]$$

This proves Lemma S2 and shows that both $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^{\prime*}(O)$ are valid influence functions for $\alpha(P)$ and $\beta(P)$ at P, respectively, as were $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$. In Section S3.3, we will go through the process of projecting $\dot{\alpha}_P^{\prime*}(O)$ and $\dot{\beta}_P^{\prime*}(O)$ onto the tangent space of the model, and show that projection yields $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$, thereby proving $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ are indeed the efficient influence functions of $\alpha(P)$ and $\beta(P)$.

S3.3 Derivation of $\dot{\alpha}_P^*(O)$ and $\dot{\beta}_P^*(O)$ From $\dot{\alpha}_P(O)$ and $\dot{\beta}_P(O)$ (Proof of Theorem 2)

The observed data distribution P is restricted by the Assumption 4, namely that $R \perp$ $Y \mid L^*, A$. As such the tangent space of the model is also restricted as follows, assuming that L_m^e is completely observed or completely missing. We let Λ_P denote the tangent space of the model, which by Lemma 24 of (7) decomposes into orthogonal subspaces as follows:

$$\Lambda_P = \Lambda_{\boldsymbol{L}^*,A} \oplus \Lambda_{R|\boldsymbol{L}^*,A} \oplus \Lambda_{Y|\boldsymbol{L}^*,A} \oplus \Lambda_{R\boldsymbol{L}_m^e|\boldsymbol{L}^*,A,R,Y}$$

where in the above

$$\Lambda_{W|V} = \{ f(w, v) \in L_2^0(P) \mid \mathbb{E}_P[f \mid V] = 0 \}$$

Next, we introduce the projection operator Π . For any function $f \in L_2(P)$,

$$\Pi(f, \Lambda_{W|V}) = \mathbb{E}_P[f(O) \mid W, V] - \mathbb{E}_P[f(O) \mid V]$$

Projection of $\alpha_P^{\prime*}(O)$ **onto** Λ_P

First notice that because $A\left(1-\frac{R}{\eta(L^*,1)}\right)\omega_1(L^*)$ is mean zero given L^*, A , we have that

$$\alpha_P^{\prime*}(O) = \underbrace{A\left(1 - \frac{R}{\eta(\boldsymbol{L}^*, 1)}\right)\omega_1(\boldsymbol{L}^*)}_{\in \Lambda_{R|\boldsymbol{L}^*, A}} + \frac{ARE}{\eta(\boldsymbol{L}^*, 1)} - \alpha(P)$$

Thus, we only need to project the latter two terms on Λ_P . Furthermore, notice that term $\alpha(P)$ is a constant (given the distribution P), it will cancel out in projection onto every orthogonal subspace with conditioning statements, and thus we need only keep track of it when projecting onto orthogonal subspace without conditioning statements.

$$\Pi\left(\frac{ARE}{\eta(\boldsymbol{L}^*,1)} - \alpha(P) \mid \Lambda_{\boldsymbol{L}^*,A,R}\right) = \mathbb{E}\left[\frac{ARE}{\eta(\boldsymbol{L}^*,1)} - \alpha(P) \mid \boldsymbol{L}^*,A,R\right] = \frac{AR}{\eta(\boldsymbol{L}^*,1)}\omega_1(\boldsymbol{L}^*) - \alpha(P)$$

$$\begin{split} &\Pi\left(\frac{ARE}{\eta(\boldsymbol{L}^*,1)} - \alpha(P) \mid \Lambda_{Y|\boldsymbol{L}^*,A}\right) \\ &= \mathbb{E}\left[\frac{ARE}{\eta(\boldsymbol{L}^*,1)} \mid \boldsymbol{L}^*, A, Y\right] - \mathbb{E}\left[\frac{ARE}{\eta(\boldsymbol{L}^*,1)} \mid \boldsymbol{L}^*, A\right] \\ &= \mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^*,1)}\mathbb{E}[E \mid \boldsymbol{L}^*, A = 1, Y, R = 1] \mid \boldsymbol{L}^*, A, Y\right] - \mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^*,1)}\mathbb{E}[E \mid \boldsymbol{L}^*, A, R = 1] \mid \boldsymbol{L}^*, A\right] \\ &= \mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^*,1)}\varepsilon_1(\boldsymbol{L}^*, Y) \mid \boldsymbol{L}^*, A, Y\right] - \mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^*,1)}\omega_1(\boldsymbol{L}^*) \mid \boldsymbol{L}^*, A\right] \\ &= A\left(\varepsilon_1(\boldsymbol{L}^*, Y) - \omega_1(\boldsymbol{L}^*)\right) \quad (A4 \text{ and defn. of } \eta) \end{split}$$

$$\begin{split} \Pi \bigg(\frac{ARE}{\eta(\boldsymbol{L}^*, 1)} - \alpha(P) \mid \Lambda_{\boldsymbol{L}_m^e \mid \boldsymbol{L}^*, A, R, Y} \bigg) &= \mathbb{E} \bigg[\frac{ARE}{\eta(\boldsymbol{L}^*, 1)} \mid \boldsymbol{L}^*, A, R, Y, \boldsymbol{L}_m^e \bigg] - \mathbb{E} \bigg[\frac{ARE}{\eta(\boldsymbol{L}^*, 1)} \mid \boldsymbol{L}^*, A, R, Y \bigg] \\ &= \frac{ARE}{\eta(\boldsymbol{L}^*, 1)} - \frac{AR}{\eta(\boldsymbol{L}^*, 1)} \mathbb{E}[E \mid \boldsymbol{L}^*, A = 1, R = 1, Y] \\ &= \frac{ARE}{\eta(\boldsymbol{L}^*, 1)} - \frac{AR}{\eta(\boldsymbol{L}^*, 1)} \varepsilon_1(\boldsymbol{L}^*, Y) \end{split}$$

Summing terms, we have

$$\Pi\left(\alpha_P^{\prime*}(O),\Lambda_P\right) = A\left(1 - \frac{R}{\eta(\boldsymbol{L}^*,1)}\right)\varepsilon_1(\boldsymbol{L}^*,Y) + \frac{ARE}{\eta(\boldsymbol{L}^*,1)} - \alpha(P)$$

which is exactly $\alpha_P^*(O)$. Thus, $\alpha_P^*(O)$ is the efficient influence function of $\alpha(P)$.

Projection of $\beta_P^{\prime*}(O)$ onto Λ_P

We begin by presenting $\dot{\beta}_P^{*}(O)$ with terms slightly rearranged, which will facilitate the projection of terms onto the requisite orthogonal subspaces.

$$\begin{split} \dot{\beta}_{P}^{\prime*}(O) &= A \bigg(1 - \frac{R}{\eta(\boldsymbol{L}^{*}, 1)} \bigg) \nu(\boldsymbol{L}^{*}) + \frac{RE}{\eta(\boldsymbol{L}^{*}, 1)} \bigg(Y - \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \bigg) \bigg[A - (1 - A) \frac{u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{1 - u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})} \bigg] - \beta(P) \\ &= \frac{AR}{\eta(\boldsymbol{L}^{*}, 1)} \bigg\{ E \Big(Y - \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \Big) - \nu(\boldsymbol{L}^{*}) \bigg\} + \underbrace{A\nu(\boldsymbol{L}^{*}) - \beta(P)}_{\in \Lambda_{\boldsymbol{L}^{*}, A}} \\ &- \frac{(1 - A)R}{\eta(\boldsymbol{L}^{*}, 1)} \frac{u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{1 - u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})} \bigg\{ E \Big(Y - \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \Big) \bigg\} \end{split}$$

where $A\nu(\mathbf{L}^*) - \beta(P) \in \Lambda_{\mathbf{L}^*,A}$ as $\mathbb{E}_P[A\nu(\mathbf{L}^*)] = \beta(P)$. Thus we only need we will project the components of the first and third term on each portion of the tangent space.

$$\begin{split} \Pi\left(\frac{AR}{\eta(\boldsymbol{L}^*,1)}\left\{E\left(\boldsymbol{Y}-\boldsymbol{\mu}_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e)\right)-\boldsymbol{\nu}(\boldsymbol{L}^*)\right\} \mid \Lambda_{\boldsymbol{L}^*,A,R}\right) &= \mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^*,1)}\left\{E\left(\boldsymbol{Y}-\boldsymbol{\mu}_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e)\right)-\boldsymbol{\nu}(\boldsymbol{L}^*)\right\} \mid \boldsymbol{L}^*,A,R\right] \\ &= \frac{AR}{\eta(\boldsymbol{L}^*,1)}\left\{\mathbb{E}\left[E\left(\boldsymbol{Y}-\boldsymbol{\mu}_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e)\right) \mid \boldsymbol{L}^*,A,R\right]-\boldsymbol{\nu}(\boldsymbol{L}^*)\right\} \\ &= \frac{AR}{\eta(\boldsymbol{L}^*,1)}\left(\boldsymbol{\nu}(\boldsymbol{L}^*)-\boldsymbol{\nu}(\boldsymbol{L}^*)\right) \\ &= 0 \end{split}$$

$$\begin{split} &\Pi\left(\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}\mid\Lambda_{\boldsymbol{L}_{m}^{e}|\boldsymbol{L}^{*},A,R,Y}\right)\\ &=\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}-\mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}\mid\boldsymbol{L}^{*},A,R,Y\right]\\ &=\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}-\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{Y\mathbb{E}[E\mid\boldsymbol{L}^{*},A,R,Y]-\mathbb{E}[E\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\mid\boldsymbol{L}^{*},A,R,Y]-\nu(\boldsymbol{L}^{*})\right\}\\ &=\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}-\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{\varepsilon_{1}(\boldsymbol{L}^{*},Y)Y-\xi(\boldsymbol{L}^{*},Y)-\nu(\boldsymbol{L}^{*})\right\}\\ &=\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{\left(E-\varepsilon_{1}(\boldsymbol{L}^{*},Y)\right)Y-\left(E\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})-\xi(\boldsymbol{L}^{*},Y)\right)\right\}\end{split}$$

$$\begin{split} \Pi\left(\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}\mid\Lambda_{Y|\boldsymbol{L}^{*},A}\right)\\ &=\mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}\mid\boldsymbol{L}^{*},A,Y\right]-\mathbb{E}\left[\frac{AR}{\eta(\boldsymbol{L}^{*},1)}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}\mid\boldsymbol{L}^{*},A\right]\\ &=\mathbb{E}\left[A\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}\mid\boldsymbol{L}^{*},A,Y,R=1\right]-\mathbb{E}\left[A\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)-\nu(\boldsymbol{L}^{*})\right\}\mid\boldsymbol{L}^{*},A,R=1\right] \quad (S1)\\ &=A\left(Y\mathbb{E}[E\mid\boldsymbol{L}^{*},A=1,Y,R=1]-\mathbb{E}[E\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\mid\boldsymbol{L}^{*},A=1,Y,R=1]-\nu(\boldsymbol{L}^{*})\right)\\ &-A\left(\mathbb{E}\left[E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)\mid\boldsymbol{L}^{*},A=1,R=1\right]-\nu(\boldsymbol{L}^{*})\right)\\ &=A\left(\varepsilon_{1}(\boldsymbol{L}^{*},Y)Y-\xi(\boldsymbol{L}^{*},Y)-\nu(\boldsymbol{L}^{*})\right)-A\left(\nu(\boldsymbol{L}^{*})-\nu(\boldsymbol{L}^{*})\right)\end{aligned}$$

Now for the final term, $\frac{(1-A)R}{\eta(\boldsymbol{L}^*,1)} \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \left\{ E\left(Y - \mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e)\right) \right\}$

$$\begin{split} \Pi & \left(\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)} \frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \left\{ E \left(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \right) \right\} \middle| \Lambda_{\boldsymbol{L}^{*},A,R} \right) \\ &= \mathbb{E} \left[\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)} \frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \left\{ E \left(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \right) \right\} \middle| \boldsymbol{L}^{*},A,R \right] \\ &= \mathbb{E} \left[\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)} \frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} E \mathbb{E} [Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \mid \boldsymbol{L}^{*},A = 0, R = 1, \boldsymbol{L}_{m}^{e}] \mid \boldsymbol{L}^{*},A,R \right] \\ &= \mathbb{E} \left[\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)} \frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} E \left(\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \right) \right] \\ &= 0 \end{split}$$

$$\begin{split} &\Pi\left(\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)}\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)\right\}\mid\Lambda_{\boldsymbol{L}_{m}^{e}}\mid\boldsymbol{L}^{*},A,R,Y\right)\\ &=\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)}\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)\right\}-\mathbb{E}\left[\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)}\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)\right\}\mid\boldsymbol{L}^{*},A,R,Y\right]\\ &=\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)}\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\left\{E\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\right)\right\}\\ &-\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)}\left\{Y\mathbb{E}\left[E\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\mid\boldsymbol{L}^{*},A,R,Y\right]-\mathbb{E}\left[E\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\mid\boldsymbol{L}^{*},A,R,Y\right]\right\}\\ &=\frac{(1-A)R}{\eta(\boldsymbol{L}^{*},1)}\left\{E\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\left(Y-\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})-\left(\gamma(\boldsymbol{L}^{*},Y)Y-\chi(\boldsymbol{L}^{*},Y)\right)\right)\right\}\end{split}$$

$$\begin{split} \Pi \bigg(\frac{(1-A)R}{\eta(\boldsymbol{L}^*,1)} \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \bigg\{ E \Big(Y - \mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) \Big) \bigg\} \mid \Lambda_{\boldsymbol{Y} \mid \boldsymbol{L}^*,\boldsymbol{A}} \bigg) \\ &= \mathbb{E} \bigg[\frac{(1-A)R}{\eta(\boldsymbol{L}^*,1)} \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \bigg\{ E \Big(Y - \mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) \Big) \bigg\} \mid \boldsymbol{L}^*,\boldsymbol{A},\boldsymbol{Y} \bigg] \\ &- \mathbb{E} \bigg[\frac{(1-A)R}{\eta(\boldsymbol{L}^*,1)} \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \bigg\{ E \Big(Y - \mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) \Big) \bigg\} \mid \boldsymbol{L}^*,\boldsymbol{A} \bigg] \\ &= (1-A) \frac{\eta(\boldsymbol{L}^*,0)}{\eta(\boldsymbol{L}^*,1)} \bigg\{ Y \mathbb{E} \bigg[E \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \mid \boldsymbol{L}^*,\boldsymbol{A} = 0, R = 1, Y \bigg] - \mathbb{E} \bigg[E \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) \mid \boldsymbol{L}^*,\boldsymbol{A} = 0, R = 1, Y \bigg] \bigg\} \\ &- (1-A) \frac{\eta(\boldsymbol{L}^*,0)}{\eta(\boldsymbol{L}^*,1)} \bigg\{ \mathbb{E} \bigg[E Y \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \mid \boldsymbol{L}^*,\boldsymbol{A} = 0, R = 1 \bigg] - \mathbb{E} \bigg[E \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) \mid \boldsymbol{L}^*,\boldsymbol{A} = 0, R = 1 \bigg] \bigg\} \\ &= (1-A) \frac{\eta(\boldsymbol{L}^*,0)}{\eta(\boldsymbol{L}^*,1)} \bigg\{ \mathbb{E} \bigg[E Y \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \mid \boldsymbol{L}^*,\boldsymbol{A} = 0, R = 1 \bigg] - \mathbb{E} \bigg[E \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \mid \boldsymbol{L}^*,\boldsymbol{A} = 0, R = 1 \bigg] \bigg\} \end{split}$$

The second/third to last lines follow from Lemma S1. That the 2nd term disappears in the final line is because

$$\begin{split} & \mathbb{E}\bigg[EY\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \mid \boldsymbol{L}^{*}, A = 0, R = 1\bigg] - \mathbb{E}\bigg[E\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) \mid \boldsymbol{L}^{*}, A = 0, R = 1\bigg] \\ & = \mathbb{E}\bigg[E\bigg(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\bigg)\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \mid \boldsymbol{L}^{*}, A = 0, R = 1\bigg] \\ & = \mathbb{E}\bigg[\mathbb{E}\bigg[E\bigg(Y - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\bigg)\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \mid \boldsymbol{L}^{*}, A = 0, R = 1, \mathbf{L}_{m}^{e}\bigg] \mid \boldsymbol{L}^{*}, A = 0, R = 1\bigg] \\ & = \mathbb{E}\bigg[\mathbb{E}\bigg[E\bigg(\mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e}) - \mu_{0}(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})\bigg)\frac{u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})}{1-u(\boldsymbol{L}^{*},\boldsymbol{L}_{m}^{e})} \mid \boldsymbol{L}^{*}, A = 0, R = 1\bigg] \\ & = 0 \end{split}$$

Putting the pieces together we have that

$$\Pi\left(\beta_P^{\prime*}(O),\Lambda_P\right) = \frac{AR}{\eta(\boldsymbol{L}^*,1)} \left[\left(E - \varepsilon_1(\boldsymbol{L}^*,Y) \right) Y - \left(E\mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) - \xi(\boldsymbol{L}^*,Y) \right) \right] \\ + A \left(\varepsilon_1(\boldsymbol{L}^*,Y)Y - \xi(\boldsymbol{L}^*,Y) \right) \\ - \frac{(1-A)R}{\eta(\boldsymbol{L}^*,1)} \left[E \frac{u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)}{1-u(\boldsymbol{L}^*,\boldsymbol{L}_m^e)} \left(Y - \mu_0(\boldsymbol{L}^*,\boldsymbol{L}_m^e) \right) - \left(\gamma(\boldsymbol{L}^*,Y)Y - \chi(\boldsymbol{L}^*,Y) \right) \right] \\ - (1-A) \frac{\eta(\boldsymbol{L}^*,0)}{\eta(\boldsymbol{L}^*,1)} \left(\gamma(\boldsymbol{L}^*,Y)Y - \chi(\boldsymbol{L}^*,Y) \right) - \beta(P)$$

which is exactly $\beta_P^*(O)$. Thus, $\beta_P^*(O)$ is the efficient influence function of $\beta(P)$, concluding the proof of Theorem 2.

S3.4 Asymptotic Behavior of $\hat{\theta}_{\text{EIF}}$ (Proof of Theorem 3)

Proof of Theorem 3

To prove Theorem 3, we follow arguments used for ratio parameters by (Kennedy et al., 2023) (Theorem 3) (6) and Levis et al. (Theorem 1) (8). For ease of notation, we use the shorthand $P[f] = \mathbb{E}_P[f(O)]$ for the mean of any function f under distribution P. We then have that

$$\begin{aligned} \widehat{\theta}_{\text{EIF}} &- \theta(P) = \frac{\mathbb{P}_n[\dot{\beta}_{\widehat{P}}(O)]}{\mathbb{P}_n[\dot{\alpha}_{\widehat{P}}(O)]} - \frac{P[\dot{\beta}_P(O)]}{P[\dot{\alpha}_P(O)]} \\ &= \frac{1}{\mathbb{P}_n[\dot{\alpha}_{\widehat{P}}(O)]} \left\{ \mathbb{P}_n[\dot{\beta}_{\widehat{P}}(O)] - P[\dot{\beta}_P(O)] - \theta(P) \Big(\mathbb{P}_n[\dot{\alpha}_{\widehat{P}}(O)] - P[\dot{\alpha}_P(O)] \Big) \right\} \end{aligned}$$

Now applying the results of the von Mises expansions in Lemma 2, we have the following decompositions:

$$\mathbb{P}_{n}[\dot{\beta}_{\widehat{P}}(O)] - P[\dot{\beta}_{P}(O)] = (\mathbb{P}_{n} - P)[\dot{\beta}_{P}(O)] + (\mathbb{P}_{n} - P)[\dot{\beta}_{\widehat{P}}(O) - \dot{\beta}_{P}(O)] + R_{\beta}(\widehat{P}, P)$$
$$\mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)] - P[\dot{\alpha}_{P}(O)] = (\mathbb{P}_{n} - P)[\dot{\alpha}_{P}(O)] + (\mathbb{P}_{n} - P)[\dot{\alpha}_{\widehat{P}}(O) - \dot{\alpha}_{P}(O)] + R_{\alpha}(\widehat{P}, P)$$

Because our estimation procedure for $\hat{\theta}_{\text{EIF}}$ leverages sample splitting (Algorithm 1), Lemma 2 of (Kennedy, 2020) guarantees that $(\mathbb{P}_n - P)[\dot{\beta}_{\hat{P}}(O) - \dot{\beta}_P(O)]$ and $(\mathbb{P}_n - P)[\dot{\alpha}_{\hat{P}}(O) - \dot{\alpha}_P(O)]$ are both $o_P(n^{-1/2})$ (9). Plugging in these decompositions, and using the fact that $\mathbb{P}_n[\dot{\alpha}_{\hat{P}}(O)]$ is bounded away from zero, we have that

$$\widehat{\theta}_{\text{EIF}} - \theta(P) = \frac{1}{\mathbb{P}_n[\dot{\alpha}_{\widehat{P}}(O)]} (\mathbb{P}_n - P) \left(\dot{\beta}_P(O) - \theta(P) \dot{\alpha}_P(O) \right) + O_P(R_\alpha + R_\beta) + o_P(n^{-1/2})$$

Next, we observe that

$$\mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)] - \alpha(P) = \mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)] - P[\dot{\alpha}_{P}(O)] \\
= (\mathbb{P}_{n} - P)(\dot{\alpha}_{P}(O)) + (\mathbb{P}_{n} - P)(\dot{\alpha}_{\widehat{P}}(O) - \dot{\alpha}_{P}(O)) + P(\dot{\alpha}_{\widehat{P}}(O) - \dot{\alpha}_{P}(O)) \\
= O_{P}(n^{-1/2}) + O_{P}(n^{-1/2}) + o_{P}(1) \\
= o_{P}(1)$$
(S1)

where the third equality follows from the central limit theorem (term 1), Lemma 2 of (Kennedy, 2020) (9) (term 2), and that for term 3, $\left|P(\dot{\alpha}_{\hat{P}}(O) - \dot{\alpha}_P(O))\right| \leq \|\dot{\alpha}_{\hat{P}}(O) - \dot{\alpha}_P(O))\| \leq \|\dot{\alpha}_{\hat{P}}(O) - \dot{\alpha}_P(O)\| = o_P(1)$ by assumption. Recalling from Corollary 2.1 that $\dot{\theta}_P^*(O) = \frac{1}{\alpha(P)} (\dot{\beta}_P(O) - \theta(P)\dot{\alpha}_P(O))$, observe that

$$\frac{1}{\mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)]} (\mathbb{P}_{n} - P) (\dot{\beta}_{P}(O) - \theta(P)\dot{\alpha}_{P}(O)) - (\mathbb{P}_{n} - P) (\theta_{P}^{*}(O))$$

$$= \left(\frac{1}{\mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)]} - \frac{1}{\alpha(P)}\right) (\mathbb{P}_{n} - P) (\dot{\beta}_{P}(O) - \theta(P)\dot{\alpha}_{P}(O))$$

$$= \frac{\alpha(P) - \mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)]}{\mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)]\alpha(P)} \cdot O_{p}(n^{-1/2}) \quad \text{(Central Limit Theorem)}$$

$$= o_{p}(1)O_{p}(n^{-1/2}) \quad \text{(Equation (S1) and } \mathbb{P}_{n}[\dot{\alpha}_{\widehat{P}}(O)] \text{ bounded away from zero)}$$

$$= o_{P}(n^{-1/2})$$

Combining pieces, we have the desired result, that

$$\widehat{\theta}_{\text{EIF}} - \theta(P) = \mathbb{P}_n[\dot{\theta}_P^*(O)] + O_P(R_\alpha + R_\beta) + o_P(n^{-1/2})$$

Proof of Corollary 3.1 Since $A \in \{0, 1\}$ is bounded, and $P(\hat{\eta}_1 > \epsilon) = 1$ for some $\varepsilon > 0$, we have that

$$|R_{\alpha}(\widehat{P}, P)| = \left| \mathbb{E}_{P} \left[A(\widehat{\varepsilon}_{1} - \varepsilon_{1}) \left(1 - \frac{\eta_{1}}{\widehat{\eta}_{1}} \right) \right] \right|$$

$$\leq \mathbb{E}_{P} \left[\left| \frac{A}{\widehat{\eta}_{1}} (\widehat{\varepsilon}_{1} - \varepsilon_{1}) (\widehat{\eta}_{1} - \widehat{\eta}_{1}) \right| \right]$$

$$\leq \frac{1}{\epsilon} \|\widehat{\varepsilon}_{1} - \varepsilon_{1}\| \|\widehat{\eta}_{1} - \widehat{\eta}_{1}\| \quad \text{(Cauchy-Schwarz)}$$

$$= O_{P}(\|\widehat{\varepsilon}_{1} - \varepsilon_{1}\| \|\widehat{\eta}_{1} - \widehat{\eta}_{1}\|)$$

Similarly, since $\mathbb{E}_P[Y^2] \leq M < \infty$ is bounded by some constant M, we have that

$$\begin{aligned} \left| \mathbb{E}_{P} \left[A \left(1 - \frac{\eta_{1}}{\widehat{\eta}_{1}} \right) \left(Y(\widehat{\varepsilon}_{1} - \varepsilon_{1}) \right) \right] \right| &\leq \mathbb{E}_{P} \left[\left| \frac{AY}{\widehat{\eta}_{1}} (\widehat{\eta}_{1} - \eta_{1}) (\widehat{\varepsilon}_{1} - \varepsilon_{1}) \right| \right] \\ &\leq \frac{\sqrt{M}}{\epsilon} \| \widehat{\varepsilon}_{1} - \varepsilon_{1} \| \| \widehat{\eta}_{1} - \widehat{\eta}_{1} \| \quad \text{(Cauchy-Schwarz)} \\ &= O_{P} (\| \widehat{\varepsilon}_{1} - \varepsilon_{1} \| \| \widehat{\eta}_{1} - \widehat{\eta}_{1} \|) \\ \\ \left| \mathbb{E}_{P} \left[-A \left(1 - \frac{\eta_{1}}{\widehat{\eta}_{1}} \right) (\widehat{\xi} - \xi) \right] \right| &\leq \mathbb{E}_{P} \left[\frac{1}{\widehat{\eta}_{1}} \left| (\widehat{\eta}_{1} - \eta_{1}) (\widehat{\xi} - \xi) \right| \right] \\ &\leq \frac{1}{\epsilon} \| \widehat{\varepsilon}_{1} - \varepsilon_{1} \| \| \widehat{\xi} - \xi \| \quad \text{(Cauchy-Schwarz)} \\ &= O_{P} (\| \widehat{\varepsilon}_{1} - \varepsilon_{1} \| \| \widehat{\xi} - \xi \|) \end{aligned}$$

Furthermore, since $P[\delta \le 1 - \hat{u} \le 1 - \delta]$ for some $\delta > 0$, and $R, E \in \{0, 1\}$, we have that

$$\left| \mathbb{E}_{P} \left[\frac{RE}{\widehat{\eta}_{1}} (\mu_{0} - \widehat{\mu}_{0}) \left(\frac{u(1 - \widehat{u}) + \widehat{u}(1 - u)}{1 - \widehat{u}} \right) \right] \right| \leq \mathbb{E}_{P} \left[\left| \frac{RE}{\widehat{\eta}_{1}(1 - \widehat{u})} (\mu_{0} - \widehat{\mu}_{0}) \left(u(1 - \widehat{u}) - \widehat{u}(1 - u) \right) \right| \right]$$
$$\leq \frac{1}{\epsilon \delta} \| \widehat{\mu}_{0} - \mu_{0} \| \| u(1 - \widehat{u}) - \widehat{u}(1 - u) \| \quad (\text{Cauchy-Schwarz})$$
$$= O_{P} (\| \widehat{\mu}_{0} - \mu_{0} \| \| \widehat{u} - u \|)$$

Finally

$$\begin{split} \left| \mathbb{E}_{P} \left[(1-A) \frac{(\eta_{0} - \widehat{\eta}_{0})}{\widehat{\eta}_{1}} \left(Y(\widehat{\gamma} - \gamma) \right) \right| &\leq \mathbb{E}_{P} \left[\left| \frac{(1-A)Y}{\widehat{\eta}_{1}} (\widehat{\eta}_{0} - \eta_{0}) (\widehat{\gamma} - \gamma) \right| \right] \\ &\leq \frac{\sqrt{M}}{\epsilon} \| \widehat{\eta}_{0} - \eta_{0} \| \| \widehat{\gamma} - \gamma \| \quad \text{(Cauchy-Schwarz)} \\ &= O_{P} (\| \widehat{\eta}_{0} - \eta_{0} \| \| \widehat{\gamma} - \gamma \|) \end{split}$$

$$\left| \mathbb{E}_{P} \left[-(1-A) \frac{(\eta_{0} - \widehat{\eta}_{0})}{\widehat{\eta}_{1}} (\widehat{\chi} - \chi) \right| \leq \mathbb{E}_{P} \left[\left| \frac{(1-A)}{\widehat{\eta}_{1}} (\widehat{\eta}_{0} - \eta_{0}) (\widehat{\chi} - \chi) \right| \right] \\ \leq \frac{1}{\epsilon} \|\widehat{\eta}_{0} - \eta_{0}\| \|\widehat{\chi} - \chi\| \quad \text{(Cauchy-Schwarz)} \\ = O_{P}(\|\widehat{\eta}_{0} - \eta_{0}\| \|\widehat{\chi} - \chi\|)$$

Combining pieces we have that

$$R_{\beta}(\widehat{P},P) = O_{P}\left(\|\widehat{\mu}_{0} - \mu_{0}\|\|\widehat{u} - u\| + \|\widehat{\eta}_{1} - \eta_{1}\|\left\{\|\widehat{\varepsilon}_{1} - \varepsilon_{1}\| + \|\widehat{\xi} - \xi\|\right\} + \|\widehat{\eta}_{0} - \eta_{0}\|\left\{\|\widehat{\gamma} - \gamma\| + \|\widehat{\chi} - \chi\|\right\}\right)$$

and thus

$$R_{\alpha}(\hat{P},P) + R_{\beta}(\hat{P},P) = O_{P}\left(\|\hat{\mu}_{0} - \mu_{0}\|\|\hat{u} - u\| + \|\hat{\eta}_{1} - \eta_{1}\|\left\{\|\hat{\varepsilon}_{1} - \varepsilon_{1}\| + \|\hat{\xi} - \xi\|\right\} + \|\hat{\eta}_{0} - \eta_{0}\|\left\{\|\hat{\gamma} - \gamma\| + \|\hat{\chi} - \chi\|\right\}\right)$$

S3.5 Summary of Asymptotic Behavior of $\hat{\theta}_{EIF}$, $\hat{\theta}_{IF}$, and $\hat{\theta}_{IF}$

In this section, we briefly summarize the asymptotic behavior of $\theta_{\rm EIF}$, as well as $\theta_{\rm IF}$ and $\theta_{\rm IF}$. We omit formal proofs, but note that justification of results in this section are nearly identical to the proofs of Theorem 3 and Corollary 3.1.

Asymptotic Behavior of $\tilde{\theta}_{\text{EIF}}$

Lemma S3 If $\|\dot{\alpha}_{\widehat{P}} - \dot{\alpha}_{P}\| = o_{P}(1), \|\dot{\beta}_{\widetilde{P}} - \dot{\beta}_{P}\| = o_{P}(1), \alpha(P) > 0, and P\Big[\big|\mathbb{P}_{n}\big(\dot{\alpha}_{\widehat{P}}(O)\big)\big| \geq \epsilon\Big] = 1 \text{ for some } \epsilon > 0, \text{ then}$

$$\widetilde{\theta}_{\text{EIF}} - \theta(P) = \mathbb{P}_n[\dot{\theta}_P^*(O)] + O_P\Big(R_\alpha(\widehat{P}, P) + \widetilde{R}_\beta(\widehat{P}, P)\Big) + o_P(n^{-1/2})$$

where remainder term, omitting inputs for brevity, is given by

$$\begin{split} \widetilde{R}_{\beta}(\overline{P},P) &= \mathbb{E}_{P} \left[A \left(1 - \frac{\eta_{1}}{\overline{\eta}_{1}} \right) \left(\left(\overline{\varepsilon}_{1} - \varepsilon_{1} \right) - \left(\overline{\xi} - \xi \right) \right) \\ &+ \frac{RE(\mu_{0} - \overline{\mu}_{0})}{\lambda_{1}\eta_{1}\pi + \lambda_{0}\eta_{0}(1 - \pi)} \left\{ \lambda_{1}\pi \left(\frac{\eta_{1}}{\overline{\eta}_{1}} - 1 \right) + \left(\lambda_{1}\pi - \overline{\lambda}_{1}\overline{\pi} \right) + \overline{\lambda}_{1}\overline{\pi} \left(1 - \frac{\eta_{0}\lambda_{0}(1 - \pi)}{\overline{\eta}_{0}\overline{\lambda}_{0}(1 - \overline{\pi})} \right) \right\} \\ &+ \left(1 - A \right) \frac{\overline{\pi}}{1 - \overline{\pi}} \left\{ Y(\overline{\gamma}' - \gamma') - \left(\overline{\chi}' - \chi' \right) \right\} \right] \end{split}$$

Moreover, if $R_{\alpha}(\widehat{P}, P) + \widetilde{R}_{\beta}(\widehat{P}, P) = o_P(n^{-1/2})$ then $\sqrt{n} (\widetilde{\theta}_{\text{EIF}} - \theta(P)) \xrightarrow{d} \mathcal{N}(0, \operatorname{Var}_P[\dot{\theta}_P^*(O)])$, attaining the semiparametric efficiency bound induced by Assumption 4.

Corollary S3.1 Under the conditions of Lemma S3 and assuming that $P(\delta \leq 1 - \hat{\pi} \leq 1 - \delta)$ for some $\delta > 0$, $P(\hat{\eta}_a > \epsilon) = 1$ for some $\epsilon > 0, a \in \{0, 1\}$, $P(\hat{\lambda}_0 > c) = 1$ for some c > 0and $\mathbb{E}_P[Y^2] \leq M < \infty$

$$R_{\alpha}(\widehat{P}, P) + \widetilde{R}_{\beta}(\widehat{P}, P) = O_{P} \left(\|\widehat{\eta}_{1} - \eta_{1}\| \left\{ \|\widehat{\varepsilon}_{1} - \varepsilon_{1}\| + \|\widehat{\xi} - \xi\| \right\} + \|\widehat{\eta}_{0} - \eta_{0}\| \left\{ \|\widehat{\gamma} - \gamma\| + \|\widehat{\chi} - \chi\| \right\} + \|\widehat{\mu}_{0} - \mu_{0}\| \left\{ \|\widehat{\eta}_{1} - \eta\| + \|\widehat{\lambda}_{1} - \lambda_{1}\| + \|\widehat{\pi} - \pi\| + \|\widehat{\eta}_{0} - \eta_{0}\| + \|\widehat{\lambda}_{0} - \lambda_{0}\| \right\} \right)$$

Asymptotic Behavior of $\hat{\theta}_{IF}$

Analogous to Corollary 2.1, we note that an influence function for $\theta(P)$ follows directly from $\dot{\alpha}'_P(O)$ and $\dot{\beta}'_P(O)$ by $\dot{\theta}'^*_P(O) = \frac{1}{\alpha(P)} (\dot{\beta}'_P(O) - \theta(P)\dot{\alpha}'_P(O)).$

Lemma S4 If $\|\dot{\alpha}'_{\widehat{P}} - \dot{\alpha}_{P}\| = o_{P}(1), \|\dot{\beta}'_{\widehat{P}} - \dot{\beta}_{P}\| = o_{P}(1), \alpha(P) > 0, and P\Big[\left| \mathbb{P}_{n} (\dot{\alpha}'_{\widehat{P}}(O)) \right| \geq \epsilon \Big] = 1 \text{ for some } \epsilon > 0, \text{ then}$ $\widehat{\theta}_{IF} - \theta(P) = \mathbb{P}_{n}[\dot{\theta}'^{*}_{P}(O)] + O_{P} \Big(R'_{\alpha}(\widehat{P}, P) + R'_{\beta}(\widehat{P}, P) \Big) + o_{P}(n^{-1/2})$

Moreover, if $R'_{\alpha}(\widehat{P}, P) + R'_{\beta}(\widehat{P}, P) = o_P(n^{-1/2})$ then $\sqrt{n}(\widehat{\theta}_{\mathrm{IF}} - \theta(P)) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}_P[\dot{\theta}'^*_P(O)])$

Corollary S4.1 Under the conditions of Lemma S4 and Corollary 3.1

$$R'_{\alpha}(\widehat{P},P) + R'_{\beta}(\widehat{P},P) = O_P\left(\|\widehat{\eta}_1 - \eta_1\|\left\{\|\widehat{\nu} - \nu\| + \|\widehat{\omega}_1 - \omega_1\|\right\} + \|\widehat{\mu}_0 - \mu_0\|\|\widehat{u} - u\|\right)$$

Asymptotic Behavior of $\tilde{\theta}_{IF}$

Lemma S5 If $\|\dot{\alpha}_{\widehat{P}} - \dot{\alpha}_{P}\| = o_{P}(1), \|\dot{\beta}_{\widetilde{P}} - \dot{\beta}_{P}\| = o_{P}(1), \alpha(P) > 0, and P\Big[\big|\mathbb{P}_{n}\big(\dot{\alpha}_{\widehat{P}}(O)\big)\big| \geq \epsilon\Big] = 1$ for some $\epsilon > 0$, then

$$\widetilde{\theta}_{\rm IF} - \theta(P) = \mathbb{P}_n[\dot{\theta}_P^*(O)] + O_P\Big(R'_\alpha(\widehat{P}, P) + \widetilde{R}'_\beta(\widehat{P}, P)\Big) + o_P(n^{-1/2})$$

where remainder term, omitting inputs for brevity, is given by

$$\widetilde{R}_{\beta}'(\overline{P},P) = \mathbb{E}_{P}\left[A\left(1-\frac{\eta_{1}}{\overline{\eta}_{1}}\right)(\nu-\overline{\nu}) + \frac{RE(\mu_{0}-\overline{\mu}_{0})}{\lambda_{1}\eta_{1}\pi + \lambda_{0}\eta_{0}(1-\pi)}\left\{\lambda_{1}\pi\left(\frac{\eta_{1}}{\overline{\eta}_{1}}-1\right) + (\lambda_{1}\pi-\overline{\lambda}_{1}\overline{\pi}) + \overline{\lambda}_{1}\overline{\pi}\left(1-\frac{\eta_{0}\lambda_{0}(1-\pi)}{\overline{\eta}_{0}\overline{\lambda}_{0}(1-\overline{\pi})}\right)\right\}\right]$$

Moreover, if $R'_{\alpha} + \widetilde{R}'_{\beta} = o_P(n^{-1/2})$ then $\sqrt{n} (\widetilde{\theta}_{\mathrm{IF}} - \theta(P)) \xrightarrow{d} \mathcal{N} (0, \mathrm{Var}_P[\dot{\theta}'^*_P(O)]).$

Corollary S5.1 Under the conditions of Lemma S3 and Corollary S3.1

$$R'_{\alpha}(\widehat{P}, P) + \widetilde{R}'_{\beta}(\widehat{P}, P) = O_{P} \left(\|\widehat{\eta}_{1} - \eta_{1}\| \left\{ \|\widehat{\omega}_{1} - \omega_{1}\| + \|\widehat{\nu} - \nu\| \right\} + \|\widehat{\eta}_{0} - \eta_{0}\| \left\{ \|\widehat{\gamma} - \gamma\| + \|\widehat{\chi} - \chi\| \right\} + \|\widehat{\mu}_{0} - \mu_{0}\| \left\{ \|\widehat{\eta}_{1} - \eta\| + \|\widehat{\lambda}_{1} - \lambda_{1}\| + \|\widehat{\pi} - \pi\| + \|\widehat{\eta}_{0} - \eta_{0}\| + \|\widehat{\lambda}_{0} - \lambda_{0}\| \right\} \right)$$

There are a few comments worth pointing out regarding these asymptotic results. To begin with $\hat{\theta}_{\text{EIF}}$ and $\tilde{\theta}_{\text{EIF}}$ have the same asymptotic distribution, and likewise with $\hat{\theta}_{\text{IF}}$ and $\tilde{\theta}_{\text{IF}}$. Of course, the asymptotic results rely on slightly different assumptions regarding respective von Mises remainder terms. Moreover, performance may not be the same in finite samples. Given the appearance of λ in remainder terms for both $\tilde{\theta}_{\text{EIF}}$ and $\tilde{\theta}_{\text{IF}}$, these estimators may be unrealistic in practice unless \boldsymbol{L}_m^e is low dimensional and one is willing to entertain a parametric model for \boldsymbol{L}_m^e .

It's also worth noting that $\hat{\theta}_{IF}$ and $\hat{\theta}_{IF}$ do not attain the semiparametric efficiency bound induced by Assumption 4. Nevertheless, studying the form of $R'_{\alpha}(\hat{P}, P) + R'_{\beta}(\hat{P}, P)$, there is some clear appeal to the estimator $\hat{\theta}_{IF}$. To begin with, consistency only requires that η_1 and one of the typical ATT nuisance functions are consistent, even if nested nuisance function ν is inconsistent. This is in contrast to $\hat{\theta}_{EIF}$ which requires η_1 and η_0 to both be consistent if any nested nuisance function is inconsistent, which could be more challenging if there is significant differential missingness by treatment status, and/or interactions between A and L^* . Altogether, the optimal choice of estimator may be problem specific based on which assumptions an analyst is most likely to entertain, as well as the complexity of component nuisance functions.

S4 Additional Simulation Information

S4.1 Simulation Data Generating Process

Simulated data were generated following the factorization of the observed data likelihood according to the following steps. In the below steps, we let β_f and X_f respectively denote the coefficient vector and design matrix relevant to nuisance function $f(\cdot)$.

- 1. Draw L^* from the observed set of non-eligibility defining covariates in the DURABLE database.
- 2. Sample surgery type $A \mid \boldsymbol{L}^* \sim \text{Bernoulli}(\pi(\boldsymbol{L}^*))$ where $\text{logit}[\pi(\boldsymbol{L}^*)] = \boldsymbol{\beta}_{\pi}^T \boldsymbol{X}_{\pi}$
- 3. Sample complete case indicator $R \mid \mathbf{L}^*, A \sim \text{Bernoulli}(\eta(L^*, A))$ where $\text{logit}[\eta(\mathbf{L}^*, A)] = \beta_{\eta}^T \mathbf{X}_{\eta}$
- 4. Sample A1c according to $(L_m^e 3) \mid \mathbf{L}^*, A, R = 1 \sim \text{Gamma}(\alpha_{\lambda}, b(\mathbf{L}^*, A))$ with rate parameter $b(\mathbf{L}^*, A) = \frac{\alpha_{\lambda}}{\mathbb{E}_P[L_m^e \mid \mathbf{L}^*, A, R=1]}$, and we parameterize $\log(\mathbb{E}_P[L_m^e \mid \mathbf{L}^*, A, R=1]) = \boldsymbol{\beta}_{\lambda}^T \boldsymbol{X}_{\lambda}$
- 5. Sample 3 year weight loss $Y \mid \boldsymbol{L}^*, A, R = 1, L_m^e \sim \mathcal{N}(\mu_A(\boldsymbol{L}^*, L_m^e), \sigma_y^2)$ where $\mu_A(\boldsymbol{L}^*, L_m^e) = \boldsymbol{\beta}_{\mu}^T \boldsymbol{X}_{\mu}$
- 6. For subjects with R = 0, set the value of L_m^e to be missing (NA).

Figure S1 illustrates estimates of $\hat{\mu}_0$ on a single simulated dataset across the six flexible strategies outlined in the main paper in comparison to estimates of $\hat{\mu}_0$ from the correctly



Figure S1: Estimates of $\hat{\mu}_0$ on a single simulated dataset across 6 nonparametric strategies compared with estimates of $\hat{\mu}_0$ from the correctly specified parametric model. Points are stratified by surgical site and self-reported smoking status, two covariates in L^* which have interactions with surgical procedure Ain μ .

specified parametric model. Clearly there is some variability in how well these strategies can estimate the complex form of μ , and thus perhaps the performance of different estimators for $\theta(P)$.

	Model Information							
Component Model	Coefficient	Covariate	Value					
-		(Intercept)	0.96					
		site[NC]	-0.64					
		site[SC]	-0.96					
		gender	2.7×10^{-2}					
		race	0.35					
True true and	0	baseline bmi	1.6×10^{-2}					
Ireatment	β_{π}	smoking status[former]	-0.29					
		smoking status [never]	-0.23					
		smoking status [no self report]	-0.32					
		haseline age	3×10^{-3}					
			5×10^{-3}					
		eGFR	-5 × 10 °					
		(Intercept)	0.30					
			-0.38					
		site[SC]	0.79					
		gender	-0.15					
		race	0.10					
Missingness	B	baseline_bmi	-2×10^{-2}					
Wissingliess	μη	smoking_status[former]	0.44					
		smoking_status[never]	0.32					
		smoking_status[no_self_report]	-2.58					
		baseline_age	1.1×10^{-2}					
		eGFR	-1×10^{-4}					
		bs_type	0.50					
		(Intercept)	1.06					
		site[NC]	0.23					
		site[SC]	-0.24					
		gender	-0.10					
		race	-6.9×10^{-2}					
		baseline_bmi	-7.5×10^{-3}					
Eligibility Defining Covariate		I(baseline_bmi^2)	1×10^{-4}					
	β_{λ}	smoking status[former]	-5.7×10^{-2}					
		smoking status [never]	-7.6×10^{-2}					
		gmoking gtatug [no go]f roport]	0.4×10^{-2}					
		Smoking_status[no_sell_report]	-9.4×10^{-3}					
		Daseline_age	9.2×10^{-4}					
		eGFR	7×10^{-4}					
		bs_type	0.10					
	α_{λ}		4.83					
		(Intercept)	-0.24					
		bs_type	3.3×10^{-2}					
		site[NC]	0.18					
		site[SC]	0.14					
		gender	-0.14					
		race	-1.5×10^{-2}					
		baseline_bmi	-3.8×10^{-3}					
		<pre>smoking_status[former]</pre>	3.8×10^{-2}					
		<pre>smoking_status[never]</pre>	4.9×10^{-2}					
		<pre>smoking_status[no_self_report]</pre>	-0.15					
		baseline_age	9.7×10^{-4}					
Outcomo	$oldsymbol{eta}_{\mu}$	eGFR	1.4×10^{-4}					
Outcome		haseline alc	2.2×10^{-4}					
		ha tupo hagolino ala	2.2×10^{-3}					
		mondon hossiling at	1 2 2 10-3					
		gender:paseline_aic	4.0×10^{-3}					
		gender:baseline_bmi	2 × 10 °					
		<pre>smoking_status[no_self_report]:bs_type</pre>	0.17					
		smoking_status[never]:bs_type	-2.4×10^{-2}					
		smoking_status[former]:bs_type	-2.4×10^{-2}					
		site[NC]:bs_type	-0.12					
		site[SC]:bs_type	-0.10					
	σ_y^2		1×10^{-2}					

S4.2 Simulation Parameters

 Table S1: Coefficients values used to generate simulated datasets

S5 Additional Data Application Information

S5.1 Methodological Details

In this section, we provide some additional details on the data application which were omitted from the main text due to space constraints.

Model Specifications

- L^* : surgery site, race, sex, age, estimated glomerular filtration rate (eGFR), self reported smoking status, hypertension, dyslipidemia, and calender year of surgery
- L_m^e : baseline BMI, baseline A1c, T2DM medication usage, DiaRem score (remission outcome only)
- A: RYGB (A = 1) vs. SG (A = 0)
- Y: % weight change at 3 years post surgery (continuous), remission of T2DM at any point within 3 years post surgery (binary)

As mentioned in the main text, T2DM status is a discrete covariate which has no variance among the study eligible population, and thus it is not included in modeling any component nuisance functions, even though it may reasonably be considered in L_m^e if desired. While several medication types were considered for establishment of T2DM status, when modeling nuisance functions, we only considered an indicator of insulin usage in the corresponding lookback window, as insulin is typically utilized when other common T2DM medications fail to achieve desired level of gylcemic control (10).

 $\hat{\theta}_{\text{EIF}}$ and $\hat{\theta}_{\text{EF}}$ were estimated with the SL1 set of SuperLearner libraries. For nuisance functions with continuous modeling targets $(\mu, \xi, \gamma, \chi, \nu)$, these libraries included random forest (SL.ranger), linear models (SL.1m), and multivariate adaptive polynomial regression spline (SL.polymars). For binary modeling targets $(\eta, u, \varepsilon, \omega)$, these libraries included random forest (SL.ranger), generalized linear models (SL.glm), and generlized additive models (SL.gam). Rather than just use a single candidate random forest learner, all nuisance models used 27 candidate random forest learners over a grid of three important hyperparameters.

- mtry (number of predictors that will be randomly sampled at each split): $\{0.5, 1, 2\} \times \lfloor \sqrt{p} \rfloor$, where p denotes the rank of the corresponding design matrix
- num.trees (number of trees used for prediction): {250, 500, 1000}
- min.node.size (minimal node size to split at): {5,30,50}

Finally, both outcome model μ and propensity score u were trained only using eligible subjects. Given that the contributions of these nuisance functions to $\beta(P)$ are only nonzero for eligible subjects, this is strictly decision of whether one believes they can model these contributions better by including ineligible subjects, and thus increasing sample size, or by getting a smaller but more narrow training set for the model which better reflects the patient who ultimately contribute. $\hat{\theta}_{CC}$ and $\hat{\theta}_{IWOR}$ used a linear model for the outcome regression, also fit among ascertainably eligible subjects by definition. The following $A \times L$ interactions were specified, motivated by existing bariatric surgery literature (11, 12).

- % Weight Change: Sex, Age, Race, Baseline BMI
- Diabetes Remission: Sex, Age, Baseline A1c, DiaRem

Additional Missing Data Considerations

The focus of this work was on missingness in eligibility defining covariates L^e , and thus implicit to this work was the assumption that there was no missingness in other variables. In practice, in EHR-based observational studies, there is likely to be missingness elsewhere, which was the case in our data application. While the missingness was not nearly as pervasive as in eligibility defining covariates, we describe additional missing data challenges and our approach in dealing with such challenges in the data application.

In our application on bariatric surgery, there was no missingness in treatment A, which may not be the case in all applications. Only a single covariate in L^* , baseline eGFR, was missing for a small number of subjects. Estimated glucose filtration rate is a function of serum creatinine, age, and sex (13), so whenever eGFR was missing it because a patient's serum creatinine value was not available. Similar to previous work by Benz et al. (14), we imputed missing serum creatinine values (prior to operationalizing the eligibility criteria) using a gamma GLM and used the imputed value to compute eGFR.

Recall that covariates \mathbf{L}_m^e are used in two nuisance functions, $\mu_a(\mathbf{L}^*, \mathbf{L}_m^e)$ and $u(\mathbf{L}^*, \mathbf{L}_m^e)$, and that furthermore, in our application, we estimated these models using eligible complete cases (E = 1, R = 1) rather than all complete cases (R = 1) given that their respective influence function contributions are multiplied by E and thus do not contribute if E = 0. Nevertheless, because T2DM status could be established without a baseline A1c measure for a small subset of patients, baseline A1c was not available for use as an important confounder. For study eligible patients without available A1c, we imputed A1c values using a gamma GLM as in previous work (14). Imputation was not used to retroactively determine study eligibility (eg., to determine T2DM status or DiaRem score).

Finally, weight change at 3 years was computed following the method of Thaweethai et al. (15), and could be missing if patients were missing a baseline BMI and/or were missing a weight measure within \pm 6 months of 3 years post-surgery. Given that outcomes Y are in the conditioning set for several nuisance functions used by $\hat{\theta}_{\text{EIF}}$, we decided to impute weight outcomes when missing so that we could illustrate the use of $\hat{\theta}_{\text{EIF}}$. In our application, % weight change outcomes were missing far less frequently than study eligibility. In other applications, outcome missingness might be more pervasive, which might motivate analysts to consider using $\hat{\theta}_{\text{IF}}$ instead of $\hat{\theta}_{\text{EIF}}$, as the former does not use Y in the conditioning set of any component nuisance function. We used a two tiered approach towards imputing outcomes: if baseline BMI was available, we imputed % weight change at 3 years via linear regression using A, L^* , and baseline BMI as predictors; if baseline BMI was not available we imputed % weight change at 3 years via linear regression using A, L^* .

That our (eligibility) complete case analysis closely matches the results reported by McTigue et al. (12) suggests that the small amount of imputation did not substantively drive any observed results.

S5.2 Additional Figures



Figure S2: Distribution of (n_{11}, n_{10}) for diabetes remission outcome, where n_{re} denotes the number of ways of operationalizing the study eligibility criteria that a subject has R = r, E = e. This figure is analogous to Figure 2 in the main text, which shows the same distribution for the relative weight change outcome.



Figure S3: Distributions of select nuisance function estimates from $\hat{\theta}_{\text{EIF}}$ related to ascertainment $(\hat{\eta})$ and eligibility $(\hat{\varepsilon})$ for diabetes remission outcome.



Figure S4: Distributions of nested nuisance function estimates from $\hat{\theta}_{\text{EIF}}$ ($\hat{\xi}, \hat{\gamma}, \hat{\chi}$) and $\hat{\theta}_{\text{IF}}$ ($\hat{\nu}$) for relative weight change outcome.

S6 Alternative Covariate Partitioning

S6.1 Notation and Assumptions

In this section, we explore an alternative way of partitioning baseline covariates \boldsymbol{L} . In particular, we show that though this partitioning may facilitate assumptions which are slightly more parsimonious than those in the main paper, identification of θ_{ATT}^{elig} is much more complex. In this exploration, we partition \boldsymbol{L} as follows:

- $L_m^{e,\bar{c}}$: Eligibility defining covariates with some degree of missingness, which are not confounders
- $L_m^{e,c}$: Eligibility defining covariates with some degree of missingness, which are confounders
- $L^{*,c}$: Completely observed covariates which are confounders
- $L^{*,\bar{c}}$: Completely observed covariates which are not confounders (but may for example be necessary to predict missingness in L^e

The connection to this covariate partitioning and the one utilized in the majority of the work is given by $\boldsymbol{L}_m^e = (\boldsymbol{L}_m^{e,\bar{c}}, \boldsymbol{L}_m^{e,c})$ and $\boldsymbol{L}^* = (\boldsymbol{L}^{*,c}, \boldsymbol{L}^{*,\bar{c}})$ Assumptions are as follows:

Assumption 1: Y(A) = Y | E = 1Assumption 3': $Y(a) \perp A | \mathbf{L}^{*,c}, \mathbf{L}_m^{e,c}, E = 1$ for $a \in \{0, 1\}$ Assumption 2': $\exists \epsilon > 0$ such that $0 < \epsilon \leq P(A = 1 | \mathbf{L}^{*,c}, \mathbf{L}_m^{e,c}, E = 1) \leq 1 - \epsilon < 1$, almost surely Assumption 4: $R \perp (Y, \mathbf{L}_m^e) | \mathbf{L}^*, A$ Assumption 5: $\exists \epsilon > 0$ such that $0 < \epsilon \leq P(R = 1 | \mathbf{L}^*, A)$, almost surely

While Assumptions 1, 4 and 5 remain unchanged, Assumptions 2' and 3' only require confounders $L_m^{e,c}$ and $L^{*,c}$ rather than all of L.

We introduce some additional nuisance functions for identification of θ_{ATT}^{elig} in the following section. Not all of these nuisance functions appear in the identification result, but several are useful shorthand for quantities which appear at various points in the derivation process.

$$\begin{split} \Lambda_{a}(\boldsymbol{L}_{m}^{e,c};\boldsymbol{L}^{*,c}) &= p(E=1,\boldsymbol{L}_{m}^{e,c} \mid A=a,\boldsymbol{L}^{*,c},R=1) \\ \delta_{a}(\boldsymbol{L}_{m}^{e,c};\boldsymbol{L}^{*,c}) &= \int_{\mathcal{L}^{*,\bar{c}}} p(\boldsymbol{L}_{m}^{e,c} \mid A=a,\boldsymbol{L}^{*,c},\boldsymbol{\ell}^{*,\bar{c}},R=1) d\boldsymbol{\ell}^{*,\bar{c}} \\ \kappa(\boldsymbol{L}^{*,c}) &= P(A=1 \mid \boldsymbol{L}^{*,c}) \\ \rho(\boldsymbol{L}_{m}^{e,c},\boldsymbol{L}^{*,c}) &= P(A=1 \mid E=1,\boldsymbol{L}_{m}^{e,c},\boldsymbol{L}^{*,c},R=1) \\ \sigma(\boldsymbol{L}^{*,c}) &= P(A=1 \mid \boldsymbol{L}^{*,c},R=1) \end{split}$$

S6.2 Identification of θ_{ATT}^{elig} Under Alternative Covariate Partition

Theorem S1 Under Assumptions 1, 2', 3', 4, and 5, θ_{ATT}^{elig} is identified by $\frac{\zeta(P)}{\alpha(P)}$ where

$$\zeta(P) = \mathbb{E}_P \left[REY \left\{ \frac{A}{\eta(\boldsymbol{L}^*, 1)} - \frac{(1-A)}{\eta(\boldsymbol{L}^*, 0)} \cdot \frac{\kappa(\boldsymbol{L}^{*,c})\rho(\boldsymbol{L}^{e,c}_m, \boldsymbol{L}^{*,c})(1-\sigma(\boldsymbol{L}^{*,c}))}{(1-\kappa(\boldsymbol{L}^{*,c}))(1-\rho(\boldsymbol{L}^{e,c}_m, \boldsymbol{L}^{*,c}))\sigma(\boldsymbol{L}^{*,c})} \right\} \right]$$
$$\alpha(P) = \mathbb{E}_P \left[\frac{ARE}{\eta(\boldsymbol{L}^*, 1)} \right]$$

<u>**Proof**</u>: We first note that the following result holds unchanged, given that it only relied on consistency (A1) and Lemma S1.

$$\mathbb{E}_{P}[Y(1) \mid A = 1, E = 1] = \mathbb{E}_{P}[Y \mid A = 1, E = 1] = \frac{\mathbb{E}_{P}[EY \mid A = 1]}{P(E = 1 \mid A = 1)}$$

Moreover, the identification of $\frac{\mathbb{E}_{P}[EY \mid A=1]}{P(E=1 \mid A=1)}$ remains unchanged. In the notation of this updated covariate partitioning, we have

$$P(E = 1 \mid A = 1) = \mathbb{E}_{P}[\mathbb{E}_{P}(E \mid \boldsymbol{L}^{*}, A = 1) \mid A = 1)]$$

$$= \mathbb{E}_{P}[\mathbb{E}_{P}(E \mid \boldsymbol{L}^{*}, A = 1, R = 1) \mid A = 1] \quad (A4, A5)$$

$$= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\frac{RE}{\eta(\boldsymbol{L}^{*}, 1)} \mid \boldsymbol{L}^{*}, A = 1\right) \mid A = 1\right] \quad (Lemma \ S1, Defn. \text{ of } \eta)$$

$$= \mathbb{E}_{P}\left[\frac{RE}{\eta(\boldsymbol{L}^{*}, 1)} \mid A = 1\right]$$

$$= \frac{\mathbb{E}_{P}\left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)}\right]}{P(A = 1)} \quad (Lemma \ S1)$$

$$= \frac{\alpha(P)}{P(A = 1)}$$

$$\begin{split} \mathbb{E}_{P}[EY \mid A = 1] &= \mathbb{E}_{P}[\mathbb{E}_{P}(EY \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1) \mid A = 1)] \\ &= \mathbb{E}_{P}[E\mathbb{E}_{P}(Y \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1) \mid A = 1)] \quad (E = g(\boldsymbol{L}^{e}, A), \text{ fixed function of } \boldsymbol{L}^{e}, A) \\ &= \mathbb{E}_{P}[E\mathbb{E}_{P}(Y \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1, R = 1) \mid A = 1)] \quad (A4, A5) \\ &= \mathbb{E}_{P}\bigg[E\mathbb{E}_{P}\bigg(\frac{RY}{P(R = 1 \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1)} \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1\bigg) \mid A = 1\bigg] \quad (\text{Lemma S1}) \\ &= \mathbb{E}_{P}\bigg[\frac{E}{\eta(\boldsymbol{L}^{*}, 1)}\mathbb{E}_{P}[RY \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}, A = 1] \mid A = 1\bigg] \quad (A4, A5) \\ &= \mathbb{E}_{P}\bigg[\frac{RE}{\eta(\boldsymbol{L}^{*}, 1)}Y \mid A = 1\bigg] \\ &= \frac{\mathbb{E}_{P}\bigg[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)}Y \mid A = 1\bigg] \\ &= \frac{\mathbb{E}_{P}\bigg[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)}Y\bigg] \quad (\text{Lemma S1}) \end{split}$$

Thus we have that

$$\mathbb{E}_P[Y(1) \mid A = 1, E = 1] = \frac{\mathbb{E}_P\left[\frac{ARE}{\eta(\boldsymbol{L}^*, 1)}Y\right]}{\alpha(P)}$$

Things become more complicated for $\mathbb{E}_P[Y(0) \mid A = 1, E = 1]$ because the set of covariates to satisfy the assumption of no unmeasured confounding $(\boldsymbol{L}_m^{e,c}, \boldsymbol{L}^{*,c})$ is no longer sufficient to define eligibility or satisfy MAR (A4). In particular

$$\begin{split} & \mathbb{E}_{P}[Y(0) \mid A = 1, E = 1] \\ &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y(0) \mid A = 1, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}) \mid A = 1, E = 1] \\ &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y(0) \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}) \mid A = 1, E = 1] \quad (A2', A3') \\ &= \mathbb{E}_{P}[\mathbb{E}_{P}(Y \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}) \mid A = 1, E = 1] \quad (A1) \\ &= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\mathbb{E}_{P}[Y \mid A = 0, E = 1, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}] \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}\right) \mid A = 1, E = 1\right] \\ &= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\mathbb{E}_{P}[Y \mid A = 0, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}] \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}\right) \mid A = 1, E = 1\right] \quad (E = g(\boldsymbol{L}^{e}, A), \text{ fixed}) \\ &= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\mathbb{E}_{P}[Y \mid A = 0, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}] \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}\right) \mid A = 1, E = 1\right] \quad (A4, A5) \\ &= \mathbb{E}_{P}\left[\mathbb{E}_{P}\left(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}\right) \mid A = 1, E = 1\right] \end{split}$$

Repeated application of Lemma S1 yields

$$\begin{split} & \mathbb{E}_{P} \left[\mathbb{E}_{P} \left(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c} \right) \mid A = 1, E = 1 \right] \\ &= \mathbb{E}_{P} \left[\frac{E}{P(E = 1 \mid A = 1)} \mathbb{E}_{P} \left(\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \mid A = 0, E = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c} \right) \mid A = 1 \right] \quad \text{(Lemma S1)} \\ &= \mathbb{E}_{P} \left[\frac{EA}{\alpha(P)} \mathbb{E}_{P} \left(\frac{E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{P(E = 1 \mid A = 0, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})} \mid A = 0, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c} \right) \right] \quad \text{(Lemma S1)} \\ &= \mathbb{E}_{P} \left[\frac{EA}{\alpha(P)} \mathbb{E}_{P} \left(\frac{(1 - A)}{P(A = 0 \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})} \frac{E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e})}{P(E = 1 \mid A = 0, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})} \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c} \right) \right] \quad \text{(Lemma S1)} \\ &= \mathbb{E}_{P} \left[\frac{\mathbb{E}_{P} [EA \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}]}{\alpha(P)} \frac{(1 - A)}{P(A = 0 \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})} \frac{E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e,c})}{P(E = 1 \mid A = 0, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})} \right] \\ &= \mathbb{E}_{P} \left[\frac{1 - A}{\alpha(P)} \cdot \frac{P(A = 1 \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})P(E = 1 \mid A = 1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})}{P(A = 0 \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})P(E = 1 \mid A = 0, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})} E\mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \right] \end{aligned}$$

Next we see that

$$P(E = 1 \mid A = a, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}) = \frac{p(E = 1, \boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c})}{p(\boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c})} \quad \text{(Bayes Rule)}$$

$$= \frac{p(E = 1, \boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c})}{\int_{\mathcal{L}^{*,\bar{c}}} p(\boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c}, \boldsymbol{\ell}^{*,\bar{c}}) d\boldsymbol{\ell}^{*,\bar{c}}}$$

$$= \frac{p(E = 1, \boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c}, R = 1)}{\int_{\mathcal{L}^{*,\bar{c}}} p(\boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c}, \boldsymbol{\ell}^{*,\bar{c}}, R = 1) d\boldsymbol{\ell}^{*,\bar{c}}} \quad (A4, A5)$$

$$= \frac{\Lambda_{a}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}{\delta_{a}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}$$

Similarly,

$$P(A = a \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}) = \frac{p(\boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c}) P(A = a \mid \boldsymbol{L}^{*,c})}{\sum_{a'=0}^{1} p(\boldsymbol{L}_{m}^{e,c} \mid A = a', \boldsymbol{L}^{*,c}) P(A = a' \mid \boldsymbol{L}^{*,c})} \quad (\text{Bayes Rule})$$

$$= \frac{P(A = a \mid \boldsymbol{L}^{*,c}) \int_{\mathcal{L}^{*,\bar{c}}} p(\boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c}, \boldsymbol{\ell}^{*,\bar{c}}) d\boldsymbol{\ell}^{*,\bar{c}}}{\sum_{a'=0}^{1} P(A = a' \mid \boldsymbol{L}^{*,c}) \int_{\mathcal{L}^{*,\bar{c}}} p(\boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c}, \boldsymbol{\ell}^{*,\bar{c}}) d\boldsymbol{\ell}^{*,\bar{c}}}$$

$$= \frac{P(A = a \mid \boldsymbol{L}^{*,c}) \int_{\mathcal{L}^{*,\bar{c}}} p(\boldsymbol{L}_{m}^{e,c} \mid A = a, \boldsymbol{L}^{*,c}, \boldsymbol{\ell}^{*,\bar{c}}) d\boldsymbol{\ell}^{*,\bar{c}}}{\sum_{a'=0}^{1} P(A = a' \mid \boldsymbol{L}^{*,c}) \int_{\mathcal{L}^{*,\bar{c}}} p(\boldsymbol{L}_{m}^{e,c} \mid A = a', \boldsymbol{L}^{*,c}, \boldsymbol{\ell}^{*,\bar{c}}, R = 1) d\boldsymbol{\ell}^{*,\bar{c}}} \quad (A4, A5)$$

$$= \frac{P(A = a \mid \boldsymbol{L}^{*,c}) \delta_a(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}{\sum_{a'=0}^{1} P(A = a' \mid \boldsymbol{L}^{*,c}) \delta_{a'}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}$$

Thus our expression simplifies to

$$\begin{split} & \mathbb{E}_{P} \left[\frac{1-A}{\alpha(P)} \cdot \frac{P(A=1 \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}) P(E=1 \mid A=1, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})}{P(A=0 \mid \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c}) P(E=1 \mid A=0, \boldsymbol{L}^{*,c}, \boldsymbol{L}_{m}^{e,c})} E \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \right] \\ &= \mathbb{E}_{P} \left[\frac{1-A}{\alpha(P)} \cdot \frac{\kappa(\boldsymbol{L}^{*,c}) \Lambda_{1}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}{(1-\kappa(\boldsymbol{L}^{*,c})) \Lambda_{0}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})} E \mu_{0}(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}) \right] \\ &= \mathbb{E}_{P} \left[\frac{1-A}{\alpha(P)} \cdot \frac{\kappa(\boldsymbol{L}^{*,c}) \Lambda_{1}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}{(1-\kappa(\boldsymbol{L}^{*,c})) \Lambda_{0}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})} E \mathbb{E}_{P} \left(\frac{R}{\eta(\boldsymbol{L}^{*}, 0)} Y \mid A=0, \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e} \right) \right] \quad \text{(Lemma S1)} \\ &= \mathbb{E}_{P} \left[\frac{1-A}{\alpha(P)} \cdot \frac{\kappa(\boldsymbol{L}^{*,c}) \Lambda_{1}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}{(1-\kappa(\boldsymbol{L}^{*,c})) \Lambda_{0}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})} E \mathbb{E}_{P} \left(\frac{R(1-A)}{\eta(\boldsymbol{L}^{*}, 0)(1-u(\boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e}))} Y \mid \boldsymbol{L}^{*}, \boldsymbol{L}_{m}^{e} \right) \right] \quad \text{(Lemma S1)} \\ &= \mathbb{E}_{P} \left[\frac{(1-A)REY}{\eta(\boldsymbol{L}^{*}, 0)} \cdot \frac{\kappa(\boldsymbol{L}^{*,c}) \Lambda_{1}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})}{(1-\kappa(\boldsymbol{L}^{*,c})) \Lambda_{0}(\boldsymbol{L}_{m}^{e,c}; \boldsymbol{L}^{*,c})} \right] \Big/ \alpha(P) \end{split}$$

Finally, we use a trick similar to the one in Section S2.1 to re-express the density ratio $\frac{\Lambda_1(L_m^{e,c}; L^{*,c})}{\Lambda_0(L_m^{e,c}; L^{*,c})}$ as a ratio of treatment probabilities.

$$\begin{split} \frac{\Lambda_1(\boldsymbol{L}_m^{e,c};\boldsymbol{L}^{*,c})}{\Lambda_0(\boldsymbol{L}_m^{e,c};\boldsymbol{L}^{*,c})} &= \frac{P(E=1,\boldsymbol{L}_m^{e,c}\mid A=1,\boldsymbol{L}^{*,c},R=1)}{P(E=1,\boldsymbol{L}_m^{e,c},A=1,\boldsymbol{L}^{*,c},R=1)} \\ &= \frac{P(E=1,\boldsymbol{L}_m^{e,c},A=1,\boldsymbol{L}^{*,c},R=1)/P(A=1,\boldsymbol{L}^{*,c},R=1)}{P(E=1,\boldsymbol{L}_m^{e,c},A=0,\boldsymbol{L}^{*,c},R=1)/P(A=0,\boldsymbol{L}^{*,c},R=1)} \\ &= \frac{P(A=1\mid E=1,\boldsymbol{L}_m^{e,c},\boldsymbol{L}^{*,c},R=1)/P(A=1\mid \boldsymbol{L}^{*,c},R=1)}{P(A=0\mid E=1,\boldsymbol{L}_m^{e,c},\boldsymbol{L}^{*,c},R=1)/P(A=0\mid \boldsymbol{L}^{*,c},R=1)} \\ &= \frac{\rho(\boldsymbol{L}_m^{e,c},\boldsymbol{L}^{*,c})(1-\sigma(\boldsymbol{L}^{*,c}))}{\sigma(\boldsymbol{L}^{*,c})(1-\rho(\boldsymbol{L}_m^{e,c},\boldsymbol{L}^{*,c}))} \end{split}$$

Combining piece, we obtain the desired result, that

$$\begin{aligned} \theta_{\text{ATT}}^{\text{elig}} &= \mathbb{E}_{P}[Y(1) \mid A = 1, E = 1] - \mathbb{E}_{P}[Y(0) \mid A = 1, E = 1] \\ &= \frac{1}{\alpha(P)} \left\{ \mathbb{E}_{P} \left[\frac{ARE}{\eta(\boldsymbol{L}^{*}, 1)} Y \right] - \mathbb{E}_{P} \left[\frac{(1 - A)REY}{\eta(\boldsymbol{L}^{*}, 0)} \cdot \frac{\kappa(\boldsymbol{L}^{*,c})\rho(\boldsymbol{L}_{m}^{e,c}, \boldsymbol{L}^{*,c})(1 - \sigma(\boldsymbol{L}^{*,c}))}{(1 - \kappa(\boldsymbol{L}^{*,c}))(1 - \rho(\boldsymbol{L}_{m}^{e,c}, \boldsymbol{L}^{*,c}))\sigma(\boldsymbol{L}^{*,c})} \right] \right\} \\ &= \frac{\zeta(P)}{\alpha(P)} \end{aligned}$$

Though θ_{ATT}^{elig} can still be identified under this alternative covariate partitioning, the derivation of any influence function would likely entail something similar to Λ_a or δ_a , possibility in a nested form. As such, further pursuing (efficient) influence function-based estimation techniques may yield estimators which are computationally very difficult to work with in practice.

References

- [1] Masashi Sugiyama, Taiji Suzuki, and Takafumi Kanamori. *Density ratio estimation in machine learning*. Cambridge University Press, 2012.
- [2] Iván Díaz, Nicholas Williams, and Edward J. Hoffman Katherine L., Schenck. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal* of the American Statistical Association, 118(542):846–857, 2023.
- [3] E. H. Kennedy, A. Sjölander, and D. S. Small. Semiparametric causal inference in matched cohort studies. *Biometrika*, 102(3):739–746, September 2015.
- [4] Andrea Mercatanti and Fan Li. Do debit cards increase household spending? evidence from a semiparametric causal analysis of a survey. 2014.
- [5] Erica EM Moodie, Olli Saarela, and David A Stephens. A doubly robust weighting estimator of the average treatment effect on the treated. *Stat*, 7(1):e205, 2018.
- [6] EH Kennedy, S Balakrishnan, and LA Wasserman. Semiparametric counterfactual density estimation. *Biometrika*, page asad017, 2023.
- [7] Andrea Rotnitzky and Ezequiel Smucler. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *Journal of Machine Learning Research*, 21:1–86, 2020.
- [8] Alexander W. Levis, Edward H. Kennedy, and Luke Keele. Nonparametric identification and efficient estimation of causal effects with instrumental variables, 2024.
- [9] Edward H Kennedy. Efficient nonparametric causal inference with missing exposure information. *The International Journal of Biostatistics*, 16(1), 2020.
- [10] S Thota and A Akbar. Insulin, 2023.
- [11] David E Arterburn, Eric Johnson, Karen J Coleman, et al. Weight outcomes of sleeve gastrectomy and gastric bypass compared to nonsurgical treatment. *Annals of Surgery*, 274(6):e1269–e1276, 2020.
- [12] Kathleen M. McTigue, Rebecca Wellman, Eric Nauman, et al. Comparing the 5-year diabetes outcomes of sleeve gastrectomy and gastric bypass: The pornet bariatric study. JAMA Surgery, 155(10):1–9, 2020.
- [13] National Kidney Foundatation. CKD-EPI Creatinine Equation. https://www.kidney.org/content/ckd-epi-creatinine-equation-2021, 2021.
- [14] Luke Benz, Rajarshi Mukjerkee, Rui Wang, et al. Adjusting for Selection Bias due to Missing Eligibility Criteria in Emulated Target Trials. American Journal of Epidemiology, 2024.
- [15] Tanayott Thaweethai, David E. Arterburn, Karen J. Coleman, and Sebastien Haneuse. Robust inference when combining inverse-probability weighting and multiple imputation to address missing data with application to an electronic health records-based study of bariatric surgery. Ann. Appl. Stat., 15(1):126–147, 2021.