Biometrika (2024), **00**, 0, *pp.* 1–20 https://doi.org/10.1093/biomet/asae016
Advance Access publication 19 March 2024

Optimal regimes for algorithm-assisted human decision-making

By M. J. STENSRUD, J. D. LAURENDEAU and A. L. SARVET

Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, Station 8, 1015 Lausanne, Switzerland

mats.stensrud@epfl.ch julien.laurendeau@epfl.ch aaron.sarvet@epfl.ch

Summary

We consider optimal regimes for algorithm-assisted human decision-making. Such regimes are decision functions of measured pre-treatment variables and, by leveraging natural treatment values, enjoy a superoptimality property whereby they are guaranteed to outperform conventional optimal regimes. When there is unmeasured confounding, the benefit of using superoptimal regimes can be considerable. When there is no unmeasured confounding, superoptimal regimes are identical to conventional optimal regimes. Furthermore, identification of the expected outcome under superoptimal regimes in nonexperimental studies requires the same assumptions as identification of value functions under conventional optimal regimes when the treatment is binary. To illustrate the utility of superoptimal regimes, we derive identification and estimation results in a common instrumental variable setting. We use these derivations to analyse examples from the optimal regimes literature, including a case study of the effect of prompt intensive care treatment on survival.

Some key words: Causal inference; Dynamic treatment regime; Instrumental variable; Natural value of treatment; Optimal regime; Single world intervention graph.

1. Introduction

Foundational work on causal inference and dynamic treatment regimes presents a promising pathway towards precision medicine (Robins, 1986; Murphy, 2003; Robins, 2004; Richardson & Robins, 2013; Tsiatis et al., 2019; Kosorok et al., 2021). In a precision-medicine system, decision rules might be algorithmically individualized based on an optimal rule previously learned from nonexperimental or experimental data (Topol, 2019). However, wide-scale implementation of such a system will usually roll out under the supervision of existing medical care providers (Matheny et al., 2019). Indeed, there is some resistance to the notion that implementation of an optimal regime, successfully learned from the data, will result in better expected outcomes compared with existing human decision rules. This resistance stems in part from the belief that existing care providers will have access to relevant information for decision-making that is not recorded in the observed data (Verghese et al., 2018). While this belief does not preclude identification of decision rules that are optimal with respect to a set of measured covariates (Miao et al., 2018; Cui & Tchetgen Tchetgen, 2021b,c; Han, 2021; Kallus & Zhou, 2021; Pu & Zhang, 2021; Qiu et al., 2021; Qi et al., 2023), care providers may be inclined to override the treatment

recommendations provided by the identified optimal regimes, based on their privileged patient observations.

We present a method for leveraging human intuition, encoded in intended treatment values, by identifying a superoptimal regime using data from either nonexperimental or experimental studies, and we clarify when a fusion of such data is beneficial. The superoptimal regime indicates to a care provider, in an algorithm-assisted decision setting, precisely when expected outcomes would be maximized if the care provider overrides the optimal regime recommendation and, importantly, when the optimal regime recommendation should be followed regardless of the care provider's assessment. This superoptimal regime is identical to the conventional optimal regime in settings with no unmeasured confounding. However, when there is unmeasured confounding, the superoptimal regime yields expected outcomes that are as good as or better than both the optimal regime and the implicit factual regime independently implemented by the care provider in the observed data, which have been studied in previous works (Miao et al., 2018; Cui & Tchetgen Tchetgen, 2021b,c; Han, 2021; Kallus & Zhou, 2021; Qi et al., 2023). Furthermore, in many settings identification of the superoptimal regime requires no additional assumptions beyond those used to identify the optimal regime and its expected outcome, i.e., value function. Indeed, we can identify superoptimal regimes by making small modifications to existing results (Cui & Tchetgen Tchetgen, 2021b,c; Qiu et al., 2021).

This article builds on the literature arising from historical interest in the so-called average treatment effect on the treated (Bloom, 1984; Heckman, 1990). One strand of literature expands on the average treatment effect on the treated by identifying and estimating a general class of causal parameters defined by the values of patients' natural treatment choices or intentions in the absence of intervention (Robins et al., 2004, 2007; Haneuse & Rotnitzky, 2013; Richardson & Robins, 2013; Young et al., 2014; Díaz et al., 2021). While this strand of literature is ostensibly interested in the values of parameters like the average treatment effect on the treated per se, a second strand of literature is especially concerned with effect heterogeneity and the implications for transportability of clinical trial results. This second strand focuses on identification of parameters similar to the average treatment effect on the treated by using augmented experimental designs, sometimes referred to as patient preference trials. Unlike conventional two-arm randomized trials, patient preference trials include an additional, third arm where individuals can choose the treatment they receive (Knox et al., 2019). While recruitment to the third arm historically has been done in different ways (Rücker, 1989; Collinge et al., 2009; McLaughlin & Spiess, 2022), modern formulations of patient preference trials require that individuals state their treatment preference before randomization to treatment, control, or taking their stated preferred treatment (Long et al., 2008; Knox et al., 2019). The third arm corresponds to an observational setting in the sense that a representative sample of individuals selects treatment based on their own preferences. If these preferences are associated with the response to treatment, then knowing the preferences provides information that is relevant to making decisions. However, this information is inaccessible in conventional randomized trials. We can view patient preference trials as target trials (Hernán & Robins, 2016) that would allow the identification of our parameters of interest by design.

Finally, an independent collection of works in the machine learning literature has studied the optimal selection of treatment based on a patient's treatment intentions in an online experimental learning setting (Bareinboim et al., 2015; Forney et al., 2017; Forney & Bareinboim, 2019).

The present work unifies and extends these related, but independently developed, literatures and our framing clarifies connections between optimal and superoptimal regimes that are obscured in the extant literature. Furthermore, like the literature characterized by Robins et al. (2004), Richardson & Robins (2013) and others, we do not focus on, or restrict ourselves to, experimental settings: we emphasize the nonexperimental setting and treat the experimental setting as a special case.

2. Preliminaries

2.1. *Nonexperimental data structure*

Consider a treatment $A \in \{0, 1\}$, a pre-treatment vector $L \in \mathcal{L}$ and an outcome $Y \in \mathbb{R}$. Suppose that we have access to n independent and identically distributed observations of (L, A, Y) among patients who received treatment in a nonexperimental setting. An unmeasured variable $U \in \mathcal{U}$ can be a common cause of A and Y. Some of our results, in particular those in the case study in § 7, will further rely on observations of an instrumental variable $Z \in \{0, 1\}$.

2.2. *Potential outcomes and the natural values of treatment*

Let superscripts denote potential outcome variables. In particular, Y^a is the potential outcome when the treatment A is fixed at the value $a \in \{0, 1\}$. More specifically, we let Y^g be the potential outcome under an arbitrary regime g, where the treatment is assigned as a function of measured covariates. Following Richardson & Robins (2013), we use the $+$ symbol to distinguish between the assigned value of treatment under the regime, A^{g+}, and the natural value of treatment under the regime, A^g. The natural value will be important in our arguments, and we state its definition explicitly (Richardson & Robins, 2013).

Definition 1 (Natural value of treatment). *The natural value of treatment A^g is the value of treatment that an individual would choose in the absence of it being assigned by an intervention.*

We use counterfactuals to define the natural values of treatment, like previous authors (Muñoz & Van Der Laan, 2012; Haneuse & Rotnitzky, 2013; Young et al., 2014). However, we could alternatively give the natural values an interventionist interpretation, which does not require conceptualization of counterfactuals: following Robins et al. (2007) and Geneletti & Dawid (2011), the natural value of treatment is a variable that is temporally prior to, but deterministically equal to, the active treatment in nonexperimental data.

The natural value of treatment under the regime g, A^g, is equal to A in any nonexperimental study that investigates the effect of a point treatment; that is, $A = A^g$ with probability 1. Thus, if A is observed, then A^g is observed. In particular, this is true in nonexperimental studies that identify causal effects in the presence of unmeasured confounding, for instance by using instrumental variables or proxy variables (Miao et al., 2018; Tchetgen Tchetgen et al., 2020).

To fix ideas about natural treatment values, consider a doctor who determines whether a patient will be transferred to an intensive care unit, or ICU; let $A = 1$ denote ICU admission and $A = 0$ no ICU admission. In the observed data, the doctor determined the ICU admission and thus the natural value A^g is equal to A with probability 1. We could, however, conceive of a regime where the assigned ICU admission, A^{g+}, is determined by some

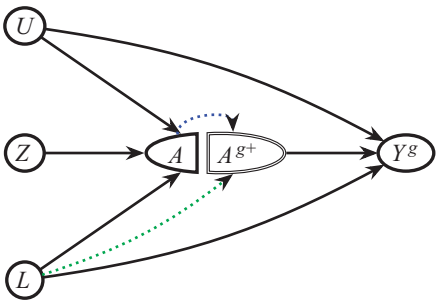

Fig. 1. A dynamic SWIG with instrumental variable Z describing a regime that depends on A and L, consistent with a superoptimal regime.

arbitrary function g of the pre-treatment covariates L. It is possible that the assignment A^{g+} differs from the natural value A.

2.3. *Definitions of treatment regimes*

In this subsection we formally define L-optimal and L-superoptimal regimes in a point-treatment setting, where the prefix L emphasizes their definitional dependence on the elements of the covariate vector L. Throughout, we assume that larger values of Y are desirable. Furthermore, we assume a non-exceptional law, so that there is a unique (super)optimal regime. We elaborate on this assumption in the Supplementary materials.

Definition 2 (L-optimal regimes). *The L-optimal regime g_{opt} assigns treatment $A^{g_{\text{opt}}+} = a$ given a vector $L = l$ by*

$$g_{\text{opt}}(l) \equiv \underset{a \in \{0,1\}}{\arg\max}\, E(Y^a \mid L = l).$$

Definition 3 (L-superoptimal regimes). *The L-superoptimal regime g_{sup} assigns treatment $A^{g_{\text{sup}}+} = a$ given $A = a'$ and $L = l$ by*

$$g_{\text{sup}}(a', l) \equiv \underset{a \in \{0,1\}}{\arg\max}\, E(Y^a \mid A = a', L = l).$$

We refer to the counterfactual expectation $E(Y^a \mid L = l)$ as a conditional value function. In particular, $E(Y^{g_{\text{opt}}} \mid L = l)$ and $E(Y^{g_{\text{sup}}} \mid A = a', L = l)$ are conditional value functions under the L-optimal and L-superoptimal regimes, respectively.

Treatment rules given by L-optimal and L-superoptimal regimes can be presented in single world intervention graphs, or SWIGs (Richardson & Robins, 2013), as illustrated by the instrumental variable setting shown in Fig. 1: the green arrow encodes regime-specific effects of the measured covariates L on the assigned value of treatment under the regime, A^{g+}, a feature of both the L-optimal and the L-superoptimal regimes; the blue arrow further encodes the effect of the natural value of treatment A on A^{g+}, a feature of the L-superoptimal, but not the L-optimal, regime.

Consider again the setting in which a patient might be transferred to an ICU. Suppose we have access to nonexperimental data from a setting where physicians determined ICU admission and thus $A = A^g$. Using these data, an investigator aims to find the dynamic

regime for ICU admission that gives the highest seven-day survival in a future decision setting. To specify this regime, we could assign A^{g+} as a function of measured covariates L describing the patient's age, gender and a collection of clinical measurements. However, beyond using the values of L, we could also ask the treating physician the following question: if you were to choose, would you transfer the patient to an ICU? The answer to this question would encode the natural treatment value A, when assuming that the answer actually agrees with the decision they would have made if we did not intervene. We can indeed use both L and A as input to our decision rule; a superoptimal regime will precisely let A^{g+} be a function of both L and A.

This brief example suggests how natural treatment value interventions can feasibly be implemented; just before intervening we ask the decision-maker about the treatment they intend to provide, and then we record their response to this question as a covariate. Nearly identical measurement strategies for 'patient preference' or 'intent' are employed in the literature on patient preference trials; see, for example, Rücker (1989), Long et al. (2008) and Knox et al. (2019).

We now consider results for superoptimal regimes in settings with observational data, wherein unmeasured confounding between the treatment and the outcome is often expected. Such settings are increasingly studied in the optimal regimes literature (Miao et al., 2018; Cui, 2021; Cui & Tchetgen Tchetgen, 2021b,c; Han, 2021; Kallus & Zhou, 2021; Qi et al., 2023).

3. Superoptimal regimes and their properties

Our first proposition states that L-superoptimal regimes are always better than, or as good as, L-optimal regimes.

Proposition 1 (Superoptimality). *The expected potential outcome under the L-superoptimal regime is better than or equal to that under the L-optimal regime:*

$$E(Y^{g_{\text{opt}}} \mid L = l) \leqslant E(Y^{g_{\text{sup}}} \mid L = l)$$

for all $l \in \mathcal{L}$.

Proof. Using laws of probability and Definitions 2 and 3,

$$\begin{aligned} E(Y^{g_{\text{opt}}} \mid L = l) &= \sum_{a'} E(Y^{g_{\text{opt}}} \mid A = a', L = l) \operatorname{pr}(A = a' \mid L = l) \\ &\leqslant \sum_{a'} E(Y^{g_{\text{sup}}} \mid A = a', L = l) \operatorname{pr}(A = a' \mid L = l) \\ &= E(Y^{g_{\text{sup}}} \mid L = l), \end{aligned}$$

where the inequality follows because, by the definitions of g_{opt} and g_{sup},

$$E(Y^{g_{\text{opt}}} \mid A = a', L = l) \leqslant E(Y^{g_{\text{sup}}} \mid A = a', L = l)$$

for all $a' \in \{0, 1\}$. □

Proposition 1 is not surprising because the regime g_{sup} uses more observed information than g_{opt}; that is, the L-superoptimal regime is optimized, not only with respect to L, but also with respect to A. A similar argument has appeared in Bareinboim et al. (2015) for an online bandit setting with no additional covariates, proposing that rewards are maximized when an agent bases decisions on their natural treatment choice.

In the remainder of this article, we will assume that interventions on the treatment variable A are well-defined, such that the following causal consistency assumption holds.

Assumption 1 (*Consistency*). If $A = a'$, then $Y = Y^{a'}$ for all $a' \in \{0, 1\}$.

Remark 1. Consistency in Assumption 1 can equivalently be formulated as $Y = Y^A$. This formulation highlights that the factual outcome is equivalent to a particular counterfactual outcome under a regime that assigns treatment A^{g+} according to the trivial regime $g(A, L) = A$ for all patients. Thus, the factual regime is a member of the class of regimes that depends on A and L, among which g_{sup} maximizes the expected potential outcome. Hence, under consistency, the expected potential outcome under the L-superoptimal regime is better than or equal to that under the factual regime.

We will also invoke the usual positivity assumption.

Assumption 2 (*Positivity*). We have that $\text{pr}(A = a \mid L) > 0$ with probability 1 for all $a \in \{0, 1\}$.

The following lemma, which exploits positivity and consistency, is similar to arguments that have appeared in work on treatment effects on the treated (Robins et al., 2007; Geneletti & Dawid, 2011; Bareinboim et al., 2015; Dawid & Musio, 2022) and will be used in our derivations of identification results.

Lemma 1. *Under consistency and positivity,* $E(Y^a \mid A = a', L = l)$ *for* $a, a' \in \{0, 1\}$ *and* $l \in \mathcal{L}$ *can be expressed as*

$$E(Y^a \mid A = a', L = l) = \begin{cases} E(Y \mid A = a', L = l), & a = a', \\ \frac{E(Y^a \mid L=l) - E(Y \mid A=a, L=l)\,\text{pr}(A=a \mid L=l)}{\text{pr}(A=a' \mid L=l)}, & a \neq a'. \end{cases} \tag{1}$$

Proof. When $a = a'$, the equation holds by consistency. When $a \neq a'$, the result follows from the law of total probability, positivity and consistency. □

Based on Lemma 1, we can use simple algebra to derive the following result, also leveraged by Bareinboim et al. (2015) in a setting without covariates.

Corollary 1. *Under consistency and positivity, the* L*-superoptimal regime* $g_{\text{sup}}(a', l)$ *for* $a' \in \{0, 1\}$ *and* $l \in \mathcal{L}$ *is equal to*

$$g_{\text{sup}}(a', l) = \begin{cases} a', & E(Y \mid L = l) \geqslant E(Y^{1-a'} \mid L = l), \\ 1 - a', & E(Y \mid L = l) < E(Y^{1-a'} \mid L = l). \end{cases}$$

The next proposition states conditions for identification of L-superoptimal regimes from observed data.

PROPOSITION 2 (IDENTIFICATION OF SUPEROPTIMAL REGIMES). *Under consistency and positivity, the L-superoptimal regime and its value function are identified by the joint distribution of* (L, A, Y) *whenever* $E(Y^a \mid L = l)$ *for all* $a \in \{0, 1\}$ *and* $l \in \mathcal{L}$ *is identified.*

Proof. This follows from Lemma 1 and Corollary 1, because all the terms on the right-hand side of (1) are identified under the two conditions in the proposition. □

Proposition 2 is useful because it justifies a two-step procedure for identification of L-superoptimal regimes using nonexperimental data. First, we use existing approaches to identify conditional outcome means and the conditional densities of the natural treatment values, i.e., the propensities. Second, we apply the result in Lemma 1 to compute counterfactual outcomes conditional on natural treatment values, which allow us to identify L-superoptimal regimes. Furthermore, Proposition 2 shows that the L-superoptimal regime g_{sup} is identified whenever conditional potential outcome means, $E(Y^a \mid L = l)$, are identified in a nonexperimental study, which encompasses studies using instrumental variables or proxy variables as important special cases.

Remark 2 (*Instrumental variables*). Corollary 2 implies that L-superoptimal regimes are identified under assumptions suggested in two recent articles by Qiu et al. (2021) and Cui & Tchetgen Tchetgen (2021c), who developed theory for identification and estimation of optimal regimes in the presence of unmeasured confounding. That is, under assumptions given in the Supplementary Material, the expected outcomes under the regimes of Qiu et al. (2021) and Cui & Tchetgen Tchetgen (2021c) will be worse than or equal to those under the L-superoptimal regimes, and in both cases the L-superoptimal regimes require no extra assumptions for identification of value functions. However, there also exist alternative conditions for identifying optimal treatment rules in instrumental variable settings, which require only identification of the causal effect conditional on L, or its sign, but not $E(Y^a \mid L = l)$ itself, as thoroughly discussed by Cui & Tchetgen Tchetgen (2021b); see also Han (2021).

Remark 3 (*Proximal inference*). Corollary 2 is also valid in proximal learning settings (Miao et al., 2018). Interestingly, heuristic arguments have been used to justify the inclusion of other covariates, but not the natural value A, in the decision function in proximal inference settings. Regarding the inclusion of an instrumental variable Z, Qi et al. (2023) write: 'This may be reasonable since Z may contain some useful information of U, which can help improve the value function.'

We emphasize that the results presented thus far have been agnostic to the absence of unmeasured confounding, which is often equated with the following assumption.

Assumption 3 (*L-exchangeability*). We have that $Y^a \perp\!\!\!\perp A \mid L$ for $a \in \{0, 1\}$.

The next results describe different properties of the L-superoptimal regime that depend on the truth value of L-exchangeability.

COROLLARY 2. *L-exchangeability implies that* $g_{\text{sup}}(A, L) = g_{\text{opt}}(L)$ *with probability* 1.

Proof. Let $a^* = \underset{a \in \{0,1\}}{\arg\max}\, E(Y^a \mid L = l)$. If L-exchangeability holds, then

$$E(Y^{a^*} \mid L = l) = E(Y^{a^*} \mid A = a', L = l)$$

for all $a' \in \{0,1\}$ and $l \in \mathcal{L}$. Thus, $a^* = \underset{a\in\{0,1\}}{\arg\max}\, E(Y^a \mid A = a', L = l)$ for all $a' \in \{0,1\}$. □

Remark 4. Suppose that an L-superoptimal regime yields better outcomes than an L-optimal regime in a given study. Then it follows from Corollary 2 that L-exchangeability fails. This fact can be used to construct tests for unmeasured confounding; see the Supplementary Material for more details. Furthermore, when L-exchangeability fails, an investigator will often assume that there exists a variable U, often called an unmeasured confounder, that exerts effects on A and Y. Then, measuring U in the future will further improve decision-making. Because A often represents a decision made by a human in the course of natural practice, the investigation and measurement of causes of A, such as U, may be feasible.

Corollary 3. *Consistency implies that* $E(Y^{g_{\text{sup}}}) \geqslant E(Y)$. *When, additionally, L-exchangeability holds,* $E(Y^{g_{\text{opt}}}) \geqslant E(Y)$.

Proof. As in Remark 1, $Y = Y^A$ is generated under a special case of a regime that depends on the natural value of treatment, where $A^{g+} = g(A, L) = A$ with probability 1. Because g_{sup} is the optimal such regime, $E(Y^{g_{\text{sup}}}) \geqslant E(Y)$. When L-exchangeability holds, application of Corollary 2 completes the proof. □

Remark 5. Given an identified optimal regime, suppose that a human care provider insists that their own intuition about treatment decisions is superior, owing to their access to privileged observations not used by the regime. Corollary 2 highlights that this insistence is contradicted when the optimal regime is identified under assumptions of no unmeasured confounding. Their claim might be illustrated by paths in the SWIG of Fig. 1: if this privileged information were truly useful for decision-making, i.e., $U \rightarrow Y^g$, and were leveraged by the clinician in the observed data, i.e., $U \rightarrow A$, then we would not usually suppose that $Y^a \perp\!\!\!\perp A \mid L$.

Remark 6. Previous work provides optimality guarantees that exclude L-superoptimal regimes. Kallus & Zhou (2021) considered a setting with unmeasured confounding and identified L-optimal regimes that are guaranteed to be as good as a 'baseline' regime. However, their baseline regime is restricted to be a function solely of measured baseline covariates L. Both the factual regime and the L-superoptimal regime are functions of unmeasured factors U when there is unmeasured confounding and so neither qualifies as a baseline regime. Ben-Michael et al. (2022) gave a similar safety guarantee, requiring the baseline regime to be a deterministic function of L. Thus, these safety guarantees in general cover neither the factual regime nor the superoptimal regime.

4. On experimental data

Thus far we have considered only observed data (L, A, Y) generated in a nonexperimental setting. As anticipated in Definition 1, we did so because we leverage the natural value of treatment, i.e., the treatment an individual would choose in the absence of it being assigned by an intervention. In a nonexperimental setting, no intervention is made and the treatment a patient actually receives, A, is indeed equal to this natural value. In experimental settings, however, a patient's natural treatment intentions may be subverted by the experimental design. Therefore, we must introduce additional notation to disambiguate patients'

received and intended treatments in the factual data. To this end, we let A^* denote the treatment a patient actually receives. Formally, we define a setting to be nonexperimental when $A = A^*$ with probability 1, such that the actual treatment value equals the intended treatment value in the preceding results. In contrast, A may not equal A^* in an experimental setting. Here we discuss several consequences of this distinction, including strategies for identifying the L-superoptimal regime with experimental data that differ from those for the nonexperimental setting.

A first consequence of the experimental setting is that Assumption 1, as defined, will almost certainly be violated; in an experiment, a patient actually receives the treatment value corresponding to A^*. To illustrate the argument, consider the assumption that $Y = Y^{A^*}$ in an experimental setting. If $A \neq A^*$ for some individuals, then $Y = Y^{A^*} \neq Y^A$ for those individuals, so we would not expect Assumption 1 to hold. Instead, the following assumption is more reasonable.

Assumption 4 (*Consistency in an experiment*). If $A^* = a$, then $Y = Y^a$ for $a \in \{0, 1\}$.

The results in § 3 all suppose Assumption 1 and not Assumption 4. Therefore, they will not in general apply to experimental data. Similar reasoning has historically motivated alternative trial designs, such as patient preference trials, in which investigators would attempt to measure A and A^* concurrently.

A second consequence of the experimental setting is that A^* is usually allocated such that $Y^a \perp\!\!\!\perp A^*$ by design, and so covariates L will be measured for reasons other than confounding control. Therefore, it is unlikely that L-exchangeability, Assumption 3, will hold in experimental data, as defined. Instead, the following assumption is more reasonable,

Assumption 5 (*L-exchangeability in an experiment*). We have that $Y^a \perp\!\!\!\perp A^* \mid L$ for $a \in \{0, 1\}$.

Despite the irrelevance of the results in § 3, the experimental setting may seem especially appealing for L-superoptimal regime identification: because A^* is randomized by design, we can adopt an even more elaborate exchangeability assumption that includes A as a covariate.

Assumption 6 (*(L, A)-exchangeability in an experiment*). We have that $Y^a \perp\!\!\!\perp A^* \mid L, A$ for $a \in \{0, 1\}$.

Furthermore, we do not have $A = A^*$ by definition, and so the following positivity condition will usually hold.

Assumption 7 (*Positivity in an experiment*). We have that $\text{pr}(A^* = a \mid L, A) > 0$ with probability 1 for $a \in \{0, 1\}$.

Lemma 1 permits identification of the L-superoptimal regime even when Assumption 7 is contradicted, as it is in a nonexperimental setting. With Assumptions 4, 6 and 7 and experimental data (Forney & Bareinboim, 2019), for example in patient preference trial designs, we might trivially identify the L-superoptimal regime without appealing to Lemma 1. In this second approach, the natural treatment value A is simply regarded as an additional covariate and thus effectively subsumed into L.

Lemma 2. *Under consistency, positivity and* (L, A)*-exchangeability in an experiment, i.e., under Assumptions* 4*,* 6 *and* 7*,*

$$E(Y^a \mid A = a', L = l) = E(Y \mid A^* = a, A = a', L = l).$$

Proof. The equality holds through sequential application of Assumptions 4 and 6, where Assumption 7 ensures that the functional remains well-defined for all values of $a, a' \in \{0, 1\}$. □

Unfortunately, the natural treatment value A is not measured in most experimental settings. Therefore, when only experimental data are available and A is unmeasured, Lemma 2 cannot be used to identify the L-superoptimal regime. However, the L-optimal regime can be learned with such data via identification of $E(Y^a \mid L = l)$. The claim and proof are trivial, by considering Lemma 2 without A in the conditioning set and replacing Assumption 6 with Assumption 5.

Remark 7. While not useful on its own for learning the L-superoptimal regime, knowledge of the parameters $E(Y^a \mid L = l)$ from an experiment will be instrumental as a supplement to nonexperimental data, even if L-exchangeability does not hold for those nonexperimental data. Suppose that the nonexperimental data and the experimental data are random draws from the same superpopulation. If the conditions of Lemma 1 are met for the nonexperimental data, then its identification functional can be evaluated using the combination of the parameters $E(Y^a \mid L = l)$ learned in the experiment and those parameters of (L, A, Y) directly observed in the nonexperimental setting. This heuristic for combining experimental and nonexperimental data has been suggested by Bareinboim et al. (2015) for the identification of a $\emptyset$-superoptimal regime. Furthermore, patient preference trials ensure the availability of such data by design, regardless of whether the natural treatment values A are measured in the assigned treatment arms. As an illustration, consider the ICU setting introduced in § 2.2. We could construct a three-arm trial in which we randomly assign the patient to ICU admission, no ICU admission or following the doctor's preference.

5. On algorithm-assisted human decision-making

One vision for optimal regimes is to use them in an algorithmic treatment-assignment paradigm, wherein treatments are assigned completely according to learned algorithms without human intervention. This algorithmic paradigm would replace current paradigms centred on consensus standards-of-care guidelines and human care providers' intuition, which could be fallible. However, the medical community may be resistant to ceding control to such algorithms in the absence of theoretical guarantees that expected outcomes will be better under the targeted optimal regime. We have shown in Corollary 3 that the superiority of the L-optimal regime is indeed guaranteed whenever L-exchangeability holds. Nevertheless, the medical community has historically expressed a deep scepticism of L-exchangeability or any identification strategy that depends on independence conditions in nonexperimental data; see, for example, the *Journal of the American Medical Association's* prohibition on causal language for the results of nonexperimental studies (AMA Manual of Style Committee, 2020). When an L-optimal regime is learned in the absence of L-exchangeability,

for example when the L-optimal regime is learned using data from a conventional two-arm trial, a clinician's scepticism may be justified: it cannot be guaranteed that

$$E(Y^{g_{\text{opt}}}) \geqslant E(Y^{A}).$$

A primary benefit of the superoptimal regime g_{sup} is to provide an algorithm with the guarantee that

$$E(Y^{g_{\text{sup}}}) \geqslant E(Y^{A}).$$

We illustrated in § 4 that the superoptimal regime is estimable from a combination of experimental and nonexperimental data whereby all relevant assumptions are enforced by design; thus such results may be acceptable to a sceptical medical community.

Nevertheless, current formulations of the L-superoptimal regime g_{sup} regard treatment intentions A as simply an additional covariate. Thus, this formulation suggests a paradigm where the algorithm is rhetorically centred. This radical departure from existing treatment assignment paradigms may result in the persistence of scepticism and resistance, despite the guarantees of the L-superoptimal regime. Therefore, we give the following equivalent formulation of g_{sup} that suggests a paradigm where the human care provider remains centred.

PROPOSITION 3. *There exists a function* $\gamma : \mathcal{L} \to \{0, 1, 2\}$ *such that the following equality holds with probability* 1:

$$g_{\text{sup}}(A, L) = \begin{cases} g_{\text{opt}}(L), & \gamma(L) = 0, \\ A, & \gamma(L) = 1, \\ 1 - A, & \gamma(L) = 2. \end{cases}$$

The function $\gamma(l)$ is identified as

$$\gamma(l) = \begin{cases} 0, & \{\tau_l(1) \geqslant 0, \tau_l(0) \geqslant 0\} \text{ or } \{\tau_l(1) < 0, \tau_l(0) < 0\}, \\ 1, & \{\tau_l(1) \geqslant 0, \tau_l(0) < 0\}, \\ 2, & \{\tau_l(1) < 0, \tau_l(0) \geqslant 0\}, \end{cases}$$

where we let $\tau_l(a') = E(Y^{a=1} \mid A = a', L = l) - E(Y^{a=0} \mid A = a', L = l)$. A proof is provided in the Supplementary Material. Proposition 3 formulates an algorithm that directly negotiates between the L-optimal regime g_{opt} and a human care provider's own privileged intuition, captured by their natural treatment intention A. When a provider encounters a patient, they are given the value of the random variable $\gamma(L)$: if $\gamma(L) = 0$, then the provider is instructed to follow the L-optimal regime's recommendation, $g_{\text{opt}}(L)$; if $\gamma(L) = 1$, then the provider is instructed to override the L-optimal regime's recommendation and provide the treatment according to their natural intention, A; finally, if $\gamma(L) = 2$, then the provider is instructed to override the L-optimal regime's recommendation and provide the treatment opposite to their natural intention, $1 - A$. With this formulation, the superoptimal regime approach can be described as a strategy for optimally negotiating between a typical L-optimal regime and a provider's privileged intuition: when the L-optimal regime is already known, the function γ can be learned to indicate to a care provider when the L-optimal

regime should be followed, or else should be overridden as a function of their natural treatment intention A. Because this formulation is equivalent to the L-superoptimal regime g_{sup}, the provider has guarantees that the algorithm will outperform the status quo. Thus, the use of L-superoptimal regimes can be accurately described as algorithm-assisted human decision-making.

However, the term algorithm-assisted human decision-making has also been used to describe settings where the decision-maker receives information from an algorithm and subsequently makes their decision. For example, consider an experiment that randomly assigned judges to receive no information or output from a public safety assessment algorithm (Imai et al., 2023). The algorithmic output included a recommended decision and the predicted risks of certain adverse outcomes. The judges made their autonomous decisions after receiving this information. The motivation for providing the algorithmic output matches the motivation for L-optimal regimes, at least when adhering to classical decision-making criteria (Sarvet & Stensrud, 2023a,b; Stensrud et al., 2023), that is, finding a decision rule 'that minimizes the prevalence of negative outcomes while avoiding unnecessarily harsh decisions' (Imai et al., 2023). However, this type of algorithm-assisted decision-making could easily be augmented with algorithmic output of L-superoptimal regimes. Specifically, the algorithm could output $g_{\text{sup}}(a', L)$ for both $a' \in \{0, 1\}$. Suppose that the decision-maker receives this information before they state their intended treatment value; the L-superoptimal regimes are provided for each intended treatment value. Then, the decision-maker can use the algorithmic output even without uncovering their intended treatment A. For example, an algorithm might inform an ICU doctor that for a patient with covariates l the recommended treatment is $g_{\text{sup}}(a', l) = a'$ for both $a' \in \{0, 1\}$. This means that the algorithm supports the doctor's intended decision for such a patient, whatever it may be. So, if an ICU doctor plans to give treatment a', they receive confirmation that this aligns with the algorithm's recommendation. However, if instead the algorithm suggests $g_{\text{sup}}(a', l) = a$ for any $a' \neq a$, then a doctor who intended to give such a treatment a' would be alerted to a discrepancy with the algorithm's recommendation. They then have the option to reconsider their decision. Therefore, knowing $g_{\text{sup}}(a', l)$ for $a' \in \{0, 1\}$ gives the ICU doctor additional information to support their decisions, even without the doctor disclosing their intentions.

We emphasize that any override of the original algorithm, g_{opt} or g_{sup}, will in general forgo the optimality guarantees of that original algorithm. In particular, we cannot formally guarantee that the decision-maker, based on their free will, would make better decisions using algorithmic information. To study such decision settings, we would need different experimental data, where a decision-maker's intended treatments are measured both before and after they are given algorithmic information.

The effect of algorithms on natural human decisions themselves is related to a more general complication that may arise with the use of L-superoptimal regimes in practice (McLaughlin & Spiess, 2022). While an investigator might expect the conditional distribution of the potential outcomes given covariates, $f_{Y^a|L}$, to remain stable across time, e.g., for biological reasons, they might not expect the same stability in the conditional distribution of the natural treatment given covariates, $f_{A|L}$. For example, this conditional distribution may change when individuals know that an algorithm will use A as input. When this generalizability problem is present, we cannot guarantee the performance of L-superoptimal regimes in the future decision setting, as the L-superoptimal regime is a function of $f_{A|L}$. In contrast, we might still have guarantees for the L-optimal regime, because this regime is not a function of $f_{A|L}$. However, to guarantee that g_{opt} in the observed data setting corresponds

to the L-optimal regime in a future decision setting, we still need to make assumptions about shared probabilistic structure across these settings; see, e.g., Bareinboim & Pearl (2013) and Dahabreh et al. (2019). Thus, the conventional L-optimal regimes also require strong assumptions for generalizability. To have analogous guarantees for g_{sup}, we would further need to make assumptions on the probabilistic structure of the natural value. For example, it would be sufficient, but not necessary, that individuals make their intended decision A in the same way as they would in the observed data. The plausibility of this condition is context-dependent. Suppose that a decision-maker receives the results of a study showing that the superoptimal regime agrees with the intended treatment value for individuals with a particular covariate value l. Then we would not expect the decision-maker to change their natural decisions for such individuals in a future decision setting; the study just confirmed that their intended treatment in this context is the best treatment option, given the observed data, and we might believe that $f_{A|l}$ is stable. However, suppose instead that a decision-maker receives the results from a study showing that their natural decisions for individuals with covariate value l are opposite to those recommended by the superoptimal regime. If the decision-maker changes their natural decision process after seeing these study results, then we would not expect $f_{A|l}$ to be stable. To mitigate the problem of an unstable distribution, we could, in the future decision setting, attempt to retrieve an intended treatment value that is representative of the study result. For example, we could instruct the decision-maker to provide the previously learned algorithm with their intended treatment had they not received the recent study results. We leave to future work the elaboration of weaker conditions for the stability of superoptimal regimes.

Finally, consider a future decision setting where the L-optimal and L-superoptimal regimes both maintain their nominal guarantees. As in classical settings without unmeasured confounding, the estimated L-optimal and L-superoptimal regimes, learned with finite data, might differ from the true L-optimal and L-superoptimal regimes (Hubbard et al., 2016). Thus, due to sampling variability, the estimated L-superoptimal and L-optimal regimes might perform worse than the true L-superoptimal and L-optimal regimes, respectively. Similarly, the estimated L-superoptimal regime might also perform worse than the observed regime. In future work, we will study strategies to control the error of deviating from the natural regime when the natural regime actually is optimal, based on familywise error rates and false discovery rates.

6. On the nonprescriptive use of superoptimal regimes

The formulation of g_{sup} in Proposition 3 highlights a counterintuitive possibility of an L-superoptimal regime: when $\text{pr}\{\gamma(L) = 2\} > 0$, the L-superoptimal regime indicates that a decision-maker should assign precisely the treatment value that is the opposite of their natural intentions, $1 - A$, for some patients. This could be the case when humans currently use outcome-predicting variables in precisely the opposite way to that which would optimize outcomes.

An algorithm-driven health-care system might dismiss this occurrence as an ancillary curiosity; if $\gamma(L) = 2$, then providing treatment $1 - A$ would simply be the optimal choice, given covariates A and L. However, g_{sup} is more than simply a prescriptive treatment policy; a positive probability of $\gamma(L) = 2$ might indicate an opportunity to radically adjust existing theories or systems of patient care for some groups, which were apparently grossly misformulated. The history of the study of human behaviour offers many examples of

fallacies where humans systematically, but unintentionally, undermine their own objectives, and iatrogenic harm is one well-documented subclass of this phenomenon. Detecting such occurrences is surely an important scientific aim, as major paradigm shifts in medical history have been portended by the scientific community's attention to such paradoxes (Kuhn, 1970).

Example 1 (*Semmelweis*). Consider the case of Ignaz Semmelweis, a 19th-century Hungarian physician. Semmelweis famously observed that it was precisely the women who were admitted to elite teaching hospital wards in anticipation of obstetric complications, $A = 1$, who were experiencing increased mortality from puerperal fever. This observation was ostensibly paradoxical: the elite venues, $A = 1$, purported to offer the best possible care. If Semmelweis had used data to learn the L-superoptimal regime, he would have observed that $\text{pr}\{\gamma(L) = 2\} > 0$; that is, there exist subgroups for which the best thing to do is to not admit to the elite teaching hospital, i.e., $A^{g_{\text{sup}}+} = 0$, precisely those patients who would otherwise be admitted to such a ward, $A = 1$, and to admit those patients, $A^{g_{\text{sup}}+} = 1$, who would otherwise be treated in a less prestigious venue, $A = 0$. Semmelweis ultimately uncovered an explanation: women sent to the prestigious hospitals were the most likely to need surgical intervention, which was then often provided by physicians returning from autopsy procedures with hands unwashed (Semmelweis, 1983). Semmelweis's observations helped initiate a hygiene and hand-washing revolution in medicine.

Semmelweis did not need the formalisms of superoptimal regimes to make his discovery. Instead, he relied on savvy intuition and large effect sizes. Superoptimal regime methodology provides a tool for systematic surveillance of iatrogenic harm, even when effect sizes are modest, or human intuition would otherwise fail.

7. Case study: instrumental variables

7.1. *Motivation*

The results we have derived so far are general. They can be used in any setting where value functions and the joint distribution of the factuals (L, A, Y) are identified. Thus, these results could be of interest in a range of settings where investigators would otherwise aim to find L-optimal regimes in the presence of unmeasured confounding. In each particular setting, an investigator can derive explicit identification formulae, which in turn motivate estimators.

In our first case study, we revisit an example from the seminal paper by Balke & Pearl (1997), illustrating that L-superoptimal regimes for certain values of A can be point identified even if L-optimal regimes are not.

Example 2 (*Vitamin A supplementation and mortality*). Balke & Pearl (1997) derived bounds for average causal effects in instrumental variable settings. These bounds are sharp under an individual-level exclusion restriction, $Y^{a,z} = Y^a$ for all a and z, and the exchangeability assumption $Z \perp\!\!\!\perp \{Y^{a=1}, Y^{a=0}, A^{z=0}, A^{z=1}\}$, which holds by design when Z is randomly assigned. The sharpness guarantees also hold under weaker assumptions (Swanson et al., 2018). To illustrate the practical relevance of these bounds, Balke & Pearl (1997) analysed data from a randomized experiment in northern Sumatra, where 450 villages were randomly offered oral doses of vitamin A supplementation, $Z = 1$, or no treatment, $Z = 0$. Villages receiving vitamin A supplementation were encouraged to provide them to preschool

children aged 12–71 months. The dataset contained 10 231 individuals from villages assigned to vitamin A, $Z = 1$, and 10 919 untreated individuals, $Z = 0$. Balke & Pearl (1997) studied the effect of consuming vitamin A supplementation, $A = 1$, versus no treatment, $A = 0$, on survival, $Y = 1$, after 12 months. Leveraging that Z is an instrument, Balke & Pearl (1997) reported bounds on the average treatment effect,

$$-0.1946 \leqslant E(Y^{a=1} - Y^{a=0}) \leqslant 0.0054.$$

They concluded that the 'vitamin A supplement, if uniformly administered, is seen as capable of increasing mortality rate by much as 19.46% and is incapable of reducing mortality rate by more than 5.4%'. Balke & Pearl (1997) did not consider further covariates, and thus we define $L = \emptyset$. It follows that neither the L-optimal regime,

$$g_{\text{opt}} = \arg\max_{a \in \{0,1\}} E(Y^a),$$

nor the value function $E(Y^a)$ are point identified, but both functionals are nontrivially bounded. However, consider now the regime

$$g_{\text{sup}}(a') = \arg\max_{a \in \{0,1\}} E(Y^a | A = a'),$$

which uses the intended value of vitamin A consumption as input to the decision function. Using Lemma 1 in this particular example, we have point identification of the superoptimal regimes for the treated, which is an analogous parameter to the average treatment effect on the treated,

$$E(Y^{a=1} - Y^{a=0} \mid A = 1) = 0.0032.$$

This example illustrates the point that, under conditions that do not identify an L-optimal regime recommendation, the L-superoptimal regime recommendation for a particular value of A is identifiable. Indeed, we conclude that among children who would consume vitamin A supplementation when offered, the supplementation does have a beneficial effect. In this example, the point identification of the superoptimal conditional average treatment effect follows because vitamin A treatment was inaccessible to those randomly assigned to no treatment, $Z = 0$ (see Balke & Pearl, 1997, Table 1), that is, there is one-sided compliance.

Whereas the L-superoptimal regime recommendation given $A = 1$ is point identified, the L-superoptimal regime recommendation given $A = 0$ is not, that is,

$$-0.33 \leqslant E(Y^{a=1} - Y^{a=0} \mid A = 0) \leqslant 0.0069.$$

7.2. *Point identification and estimation*

To illustrate how explicit identification formulae and estimators can be derived, we further build on recent work on optimal regimes (Cui, 2021; Cui & Tchetgen Tchetgen, 2021b; Qiu et al., 2021). These results are presented in detail in the Supplementary Material, and we provide an overview in this section. Specifically, we give new identification results for L-superoptimal regimes. We derive the nonparametric influence function of corresponding identification functionals and thereby motivate new estimators. Moreover, we suggest a strategy to further improve efficiency when the investigator aims only to identify the

L-superoptimal regime, given by the sign of the value function, as opposed to the value function itself. We apply these new methods to study the effects of ICU admission on survival in § 8.

To illustrate the practical benefits of superoptimal versus optimal regimes in this setting, in the Supplementary Material we revisit an example from Qiu et al. (2021, Remark 5), who emphasized that an L-optimal regime can be worse than the regime that was implemented in the observed data. We show that the L-superoptimal regime is strictly better than the L-optimal regime in their example. We also give another example with a minor change to the setting in Qiu et al. (2021), where the L-superoptimal regime outperforms both the L-optimal and the observed regime.

7.3. *Augmenting regimes with instruments*

Instruments play an important role in L-superoptimal regimes which is different from their role in L-optimal regimes. Let an (L, Z)-optimal regime be defined as

$$g_{\text{z-opt}}(l, z) \equiv \underset{a \in \{0,1\}}{\arg\max}\, E(Y^a \mid L = l, Z = z),$$

which is isomorphic to an L-optimal regime, but further uses the instrument Z. An instrument Z satisfies $Y^a \perp\!\!\!\perp Z \mid L$, which can be read off the SWIG in Fig. 1. Using arguments isomorphic to those of Corollary 2, the (L, Z)-optimal regime is always equal to the L-optimal regime. However, interestingly, an (L, Z)-superoptimal regime is not necessarily equal to an L-superoptimal regime. This follows because in many cases

$$Y^a \not\perp\!\!\!\perp Z \mid L, A;$$

see the SWIG in Fig. 1 as an example. Therein, it can be seen that Z would be d-connected to Y^a given L and A via the path $Z \to A \leftarrow U \to Y^a$, which would be open conditional on A, a collider. A similarly open path would remain if Z were alternatively associated with A via an unmeasured common cause, which is often assumed in some instrumental variable models. The practical implication is that using an instrument Z can further improve superoptimal, but not optimal, regimes; see the Supplementary Material for more details. We give intuition to this result in our ICU example in § 8.

8. Application: intensive care unit admissions

Following Keele et al. (2020), we study the effect of prompt ICU admission on seven-day survival. We use resampled data from a cohort study of patients with deteriorating health who were referred for assessment for ICU admission at 48 UK National Health Service hospitals in 2010–2011 (Harris et al., 2015).

Our treatment of interest, $A = 1$, is ICU admission within four hours of arrival at the hospital, referred to as 'prompt ICU admission'. An individual is considered untreated, $A = 0$, if they were not admitted within four hours. Our sample consists of 13 011 patients, of whom 10 478 were treated. One reason for being untreated could be resource constraints, e.g., lack of available ICU beds or insufficient staffing. As in Keele et al. (2020), we use an indicator of the ICU bed occupancy being below or above the median as our instrument, Z, which should affect the outcome Y only through its effect on A; see Fig. 1. We further

Table 1. *Marginal value functions under different regimes, where the percentile* 95% *confidence intervals are estimated by nonparametric bootstrap in* 500 *samples*

Parameter	Estimate (95% confidence interval)
$E(Y)$	0.86 (0.85, 0.86)
$E(Y^{\hat{g}_{\text{opt}}})$	0.93 (0.40, 1.00)
$E(Y^{\hat{g}_{\text{sup}}})$	0.97 (0.77, 1.00)
$E(Y^{\hat{g}_{z\text{-sup}}})$	0.98 (0.77, 1.00)

consider an individual's age, recorded sex and sequential organ failure assessment score as baseline variables, L.

In these nonexperimental data, the individual's natural value of treatment is directly recorded. In a future decision setting, we could measure the natural treatment variable by asking the following question of a doctor treating a patient: would you promptly admit this patient to an ICU? An answer of yes would correspond to $A = 1$, and a no would correspond to $A = 0$, assuming a deterministic relation between the doctor's response to this question and what they actually would have done. Informally, the doctor's response, A, serves as a proxy for factors U that might indicate the risk of seven-day mortality. Furthermore, using the current bed occupancy Z jointly with the doctor's response A could provide a better proxy for factors U not recorded in the observed data, even when ICU bed occupancy Z is independent of U marginally. For example, suppose that U represents a physician's judgement of a patient's underlying mortality risk based on unrecorded injury features or other implicit judgements of patient frailty, encoded as 'moderate' or 'severe', and that a doctor will admit all patients on low-occupancy days, but will admit only 'severe'-risk patients on high-occupancy days. Occupancy has little predictive capacity for a patient's underlying mortality risk marginally, but if it is known that a patient was admitted on a high-occupancy day, then we can deduce that the patient must have been at 'severe' risk.

We estimated observed, L-optimal, L-superoptimal and (L, Z)-superoptimal regimes based on the estimation algorithm in the Supplementary Material, where we also used 60-40 sample splitting to avoid the bias that would result from estimating and evaluating a (super)optimal decision rule in the same sample (Zhang et al., 2012; Qiu et al., 2021). The point estimates of the marginal value functions suggest that the estimated L-superoptimal, i.e., $\hat{g}_{\text{sup}}$, and (L, Z)-superoptimal, $\hat{g}_{z\text{-sup}}$, regimes outperform the alternatives; see Table 1. The fact that the confidence intervals are wide is not surprising, despite the large sample size, because of the reliance on an instrumental variable. However, the imprecision also requires us to caution against making strong conclusions about the estimated (super)optimal regimes, because they could deviate from the true (super)optimal regimes due to finite-sample uncertainty, as discussed in § 5, see also the Supplementary Material.

9. Future directions

An interesting problem is generalizing the results to longitudinal settings with time-varying treatments. A complicating factor is that the non-baseline natural treatment values will not in general correspond with observed treatment values, even when the data arise from a nonexperimental setting. Nevertheless, their distributions may be identified under assumptions commonly invoked to identify dynamic regimes that depend on the natural

value of treatment in time-varying treatment settings, as described in Richardson & Robins (2013) and Young et al. (2014). Generalizations to nonbinary treatments would also be of interest in some settings.

Our identification results motivate estimators of L-superoptimal regimes. We specifically provided semiparametric estimators in an instrumental variable setting. However, alternative estimators of superoptimal regimes can also be developed, and the properties of these estimators must be evaluated on a case-by-case basis. Relatedly, we aim to construct estimators of superoptimal regimes with error control, e.g., control of erroneously deviating from the observed regime.

Finally, there exist results on L-optimal regime identification when conditional outcome means are only partially identified (Cui, 2021; Cui & Tchetgen Tchetgen, 2021a; Pu & Zhang, 2021). The partial identification results can be derived, for example, under conditions that are guaranteed by design. In contrast, point identification in settings with unmeasured confounding, for example in the instrumental variable setting considered here, requires homogeneity assumptions that are not guaranteed to hold. Using L-superoptimal regimes under partial identification conditions is a topic for future investigations.

Acknowledgement

We would like to thank Prof. Vanessa Didelez for insightful comments on a previous version of the manuscript. The data were kindly provided by Prof. Luke Keele. The authors were supported by the Swiss National Science Foundation (200021_207436).

Supplementary material

The Supplementary Material contains additional results and a review of existing conditions for identification and estimation of regimes in instrumental variable settings.

References

AMA Manual of Style Committee (2020). *AMA Manual of Style: A Guide for Authors and Editors*. Oxford: Oxford University Press, 11th ed.

Balke, A. & Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *J. Am. Statist. Assoc.* **92**, 1171–6.

Bareinboim, E., Forney, A. & Pearl, J. (2015). Bandits with unobserved confounders: A causal approach. In *Proc. 28th Int. Conf. Neural Information Processing Systems (NIPS 2015)*, vol. 1. Cambridge, Massachusetts: MIT Press, pp. 1342–50.

Bareinboim, E. & Pearl, J. (2013). A general algorithm for deciding transportability of experimental results. *J. Causal Infer.* **1**, 107–34.

Ben-Michael, E., Greiner, D. J., Imai, K. & Jiang, Z. (2022). Safe policy learning through extrapolation: Application to pre-trial risk assessment. *arXiv:* 2109.11679v3.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Eval. Rev.* **8**, 225–46.

Collinge, J., Gorham, M., Hudson, F., Kennedy, A., Keogh, G., Pal, S., Rossor, M., Rudge, P., Siddique, D., Spyer, M. et al. (2009). Safety and efficacy of quinacrine in human prion disease (PRION-1 study): A patient-preference trial. *Lancet Neurol.* **8**, 334–44.

Cui, Y. (2021). Individualized decision-making under partial identification: Three perspectives, two optimality results, and one paradox. *Harvard Data Science Review*, 22 October 2021; DOI: 10.1162/99608f92.d07b8d16.

Cui, Y. & Tchetgen Tchetgen, E. (2021a). Machine intelligence for individualized decision making under a counterfactual world: A rejoinder. *J. Am. Statist. Assoc.* **116**, 200–6.

Cui, Y. & Tchetgen Tchetgen, E. (2021b). On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable. *Statist. Prob. Lett.* **178**, 109180.

Cui, Y. & Tchetgen Tchetgen, E. (2021c). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *J. Am. Statist. Assoc.* **116**, 162–73.

Dahabreh, I. J., Robins, J. M., Haneuse, S. J. & Hernán, M. A. (2019). Generalizing causal inferences from randomized trials: Counterfactual and graphical identification. *arXiv:* 1906.10792.

Dawid, A. P. & Musio, M. (2022). What can group-level data tell us about individual causality? In *Statist. Publ. Interest*. Springer, pp. 235–56.

Díaz, I., Williams, N., Hoffman, K. L. & Schenck, E. J. (2021). Nonparametric causal effects based on longitudinal modified treatment policies. *J. Am. Statist. Assoc.* **118**, 1–16.

Forney, A. & Bareinboim, E. (2019). Counterfactual randomization: Rescuing experimental studies from obscured confounding. In *Proc. 33rd AAAI Conf. Artificial Intelligence*. Cambridge, Massachusetts: AAAI Press, pp. 2454–61.

Forney, A., Pearl, J. & Bareinboim, E. (2017). Counterfactual data-fusion for online reinforcement learners. In *Proc. 34th Int. Conf. Machine Learning*. JMLR, pp. 1156–64.

Geneletti, S. & Dawid, A. P. (2011). Defining and identifying the effect of treatment on the treated. In *Causality in the Sciences*. Oxford: Oxford University Press, pp. 728–49.

Han, S. (2021). Comment: Individualized treatment rules under endogeneity. *J. Am. Statist. Assoc.* **116**, 192–5.

Haneuse, S. & Rotnitzky, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statist. Med.* **32**, 5260–77.

Harris, S., Singer, M., Rowan, K. & Sanderson, C. (2015). Delay to admission to critical care and mortality among deteriorating ward patients in UK hospitals: A multicentre, prospective, observational cohort study. *Lancet* **385**, S40.

Heckman, J. (1990). Varieties of selection bias. *Am. Econ. Rev.* **80**, 313–8.

Hernán, M. A. & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–64.

Hubbard, A. E., Kherad-Pajouh, S. & Van Der Laan, M. J. (2016). Statistical inference for data adaptive target parameters. *Int. J. Biostatist.* **12**, 3–19.

Imai, K., Jiang, Z., Greiner, D. J., Halen, R. & Shin, S. (2023). Experimental evaluation of algorithm-assisted human decision-making: Application to pretrial public safety assessment. *J. R. Statist. Soc.* A **186**, 167–89.

Kallus, N. & Zhou, A. (2021). Minimax-optimal policy learning under unobserved confounding. *Manag. Sci.* **67**, 2870–90.

Keele, L., Harris, S., Pimentel, S. D. & Grieve, R. (2020). Stronger instruments and refined covariate balance in an observational study of the effectiveness of prompt admission to intensive care units. *J. R. Statist. Soc.* A **183**, 1501–21.

Knox, D., Yamamoto, T., Baum, M. A. & Berinsky, A. J. (2019). Design, identification, and sensitivity analysis for patient preference trials. *J. Am. Statist. Assoc.* **114**, 1532–46.

Kosorok, M. R., Laber, E. B., Small, D. S. & Zeng, D. (2021). Introduction to the theory and methods special issue on precision medicine and individualized policy discovery. *J. Am. Statist. Assoc.* **116**, 159–61.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago, Illinois: University of Chicago Press.

Long, Q., Little, R. J. & Lin, X. (2008). Causal inference in hybrid intervention trials involving treatment choice. *J. Am. Statist. Assoc.* **103**, 474–84.

Matheny, M., Israni, S. T., Ahmed, M. & Whicher, D., eds. (2019). *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. Washington, DC: The National Academies Press.

McLaughlin, B. & Spiess, J. (2022). Algorithmic assistance with recommendation-dependent preferences. *arXiv:* 2208.07626v3.

Miao, W., Geng, Z. & Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika* **105**, 987–93.

Muñoz, I. D. & Van Der Laan, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics* **68**, 541–9.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Statist. Soc.* B **65**, 331–55.

Pu, H. & Zhang, B. (2021). Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *J. R. Statist. Soc.* B **83**, 318–45.

Qi, Z., Miao, R. & Zhang, X. (2023). Proximal learning for individualized treatment regimes under unmeasured confounding. *J. Am. Statist. Assoc.* to appear; DOI: 10.1080/01621459.2022.2147841.

Qiu, H., Carone, M., Sadikova, E., Petukhova, M., Kessler, R. C. & Luedtke, A. (2021). Optimal individualized decision rules using instrumental variable methods. *J. Am. Statist. Assoc.* **116**, 174–91.

Richardson, T. S. & Robins, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. CSSS Working Paper no. 128, Center for Statistics and the Social Sciences, University of Washington.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Mod.* **7**, 1393–512.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proc. 2nd Seattle Symposium in Biostatistics*, vol. 179 of *Lecture Notes in Statistics*. New York: Springer, pp. 189–326.

Robins, J. M., Hernán, M. A. & Siebert, U. (2004). Effects of multiple interventions. In *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, vol. 1. Geneva, Switzerland: World Health Organization, pp. 2191–230.

Robins, J. M., Vanderweele, T. J. & Richardson, T. S. (2007). Comment on 'Causal effects in the presence of non compliance: A latent variable interpretation' by Antonio Forcina. *Metron* **LXIV**, 288–98.

Rücker, G. (1989). A two-stage trial design for testing treatment, self-selection and treatment preference effects. *Statist. Med.* **8**, 477–85.

Sarvet, A. L. & Stensrud, M. J. (2023a). Aaron Sarvet and Mats Stensrud's contribution to the Discussion of 'Experimental evaluation of algorithm-assisted human decision-making: Application to pretrial public safety assessment' by Imai et al., part 1. *J. R. Statist. Soc.* A **186**, 199–200.

Sarvet, A. L. & Stensrud, M. J. (2023b). Perspectives on harm in personalized medicine. *arXiv:* 2302.01371v2.

Semmelweis, I. P. (1983). *Etiology, Concept and Prophylaxis of Childbed Fever*. Madison, Wisconsin: University of Wisconsin Press. Translated from the 1861 German original.

Stensrud, M. J., Didelez, V. & Sarvet, A. L. (2023). Discussion of 'Experimental evaluation of algorithm-assisted human decision-making: Application to pretrial public safety assessment', part 2. *J. R. Statist. Soc.* A **186**, 201–2.

Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M. & Richardson, T. S. (2018). Partial identification of the average treatment effect using instrumental variables: Review of methods for binary instruments, treatments, and outcomes. *J. Am. Statist. Assoc.* **113**, 933–47.

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X. & Miao, W. (2020). An introduction to proximal causal learning. *arXiv:* 2009.10982.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Med.* **25**, 44–56.

Tsiatis, A. A., Davidian, M., Holloway, S. T. & Laber, E. B. (2019). *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. Boca Raton, Florida: Chapman and Hall/CRC.

Verghese, A., Shah, N. H. & Harrington, R. A. (2018). What this computer needs is a physician: Humanism and artificial intelligence. *J. Am. Med. Assoc.* **319**, 19–20.

Young, J. G., Hernán, M. A. & Robins, J. M. (2014). Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiol. Meth.* **3**, 1–19.

Zhang, B., Tsiatis, A. A., Laber, E. B. & Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68**, 1010–8.

[*Received on* 2 *March* 2022. *Editorial decision on* 27 *February* 2024]