Counterfactual Randomization: Rescuing Experimental Studies from Obscured Confounding

Andrew Forney
Department of Computer Science
Loyola Marymount University
Los Angeles, CA

Elias Bareinboim
Department of Computer Science
Purdue University
West Lafayette, IN

Abstract

Randomized clinical trials (RCTs) like those conducted by the FDA provide medical practitioners with average effects of treatments, and are generally more desirable than observational studies due to their control of unobserved confounders (UCs), viz., latent factors that influence both treatment and recovery. However, recent results from causal inference have shown that randomization results in a subsequent loss of information about the UCs, which may impede treatment efficacy if left uncontrolled in practice (Bareinboim, Forney, and Pearl 2015). Our paper presents a novel experimental design that can be noninvasively layered atop past and future RCTs to not only expose the presence of UCs in a system, but also reveal patient- and practitioner-specific treatment effects in order to improve decision-making. Applications are given to personalized medicine, second opinions in diagnosis, and employing offline results in online recommender systems.

1 Introduction

Randomized Clinical Trials (RCTs) are considered the gold-standard for evidence generation throughout the empirical sciences; annually, the FDA alone spends billions of dollars conducting thousands of RCTs to vet new drugs and medical treatments (Giffin et al. 2010). RCTs are distinctly superior to observational studies in that the assigned treatment is randomized, as opposed to allowing the actors (e.g., physicians, patients) to decide treatment themselves. This randomization provides control for confounding bias (Pearl 2000, Ch. 6), which appears due to the existence of unobserved confounders (UCs) generating uncontrolled variations to the treatment and outcome. Randomization of the treatment allocation constitutes one of the pillars of modern experimental design (Fisher 1951; Wainer 1989) and broadly, within the scientific method itself.

As recent results from causal inference have demonstrated (Bareinboim, Forney, and Pearl 2015), the control for UCs provided by RCTs may yield population-level treatment outcomes, but comes with a cost: information about individual- or unit-level patient and prescriber characteristics is lost, and the treatment selection mechanisms that exist in practice remain unmeasured. Such scenarios are motivated by many instances of confounded decision-making from the medical domain (White 2005; Brookhart et al. 2010; Cormack et al. 2018), in which "despite their attention to evidence, studies repeatedly show marked variability in what healthcare providers actually do in a given situation" (Stead, Starmer, and McClellan 2008). In the case where this variability is due to UCs, better treatment policies than those practiced may exist (Ruberg and Shen 2015).

The focus on more individualized treatment effects (ITEs) has fallen under a variety of headings in the medical and statistical sciences, including "personalized medicine" (Hamburg and Collins 2010), the "effect of treatment on the treated (ETT)" (Pearl 2000; Pearl, Glymour, and Jewell 2016), and "c-specific effects" (Pearl 2017). Yet, for all of these labels, the data sciences still lack an empirical methodology for measuring ITEs without strong assumptions within the model, about the UCs, or about the treatments. Closer attempts (Bareinboim, Forney, and Pearl 2015; Forney, Pearl, and Bareinboim 2017) have come from the online decision-making domain, in which counterfactual treatment effects akin to the ETT are measured, but imply that UCs have already evaded detection following an offline RCT. Related work has examined ITEs and attempted to measure counterfactual outcomes from observational data, but crucially, assume the inexistence of UCs (Papangelou et al. 2018; Shalit, Johansson, and Sontag 2016).

The present work takes a data-scientific perspective to traditional empirical inquiry, acknowledging that the best tools available to machine learning may yield analyses that are only as rich as the data provided. As RCTs are likely to remain the dominant tool for experimentation, the current effort explores approaches that enhance, rather than replace, their findings. Specifically, we make three primary contributions: (1) We begin by formalizing a causal interpretation for the differing treatment policies of actors (e.g., prescribing physicians) in confounded decision-making scenarios, called *heterogeneous-intent* (HI). (2) We then introduce a new experimental procedure for the early detection of UCs and HIs through a technique that can be layered atop a traditional, offline RCT; moreover, this technique can be used to salvage ITEs from *existing* RCT results. (3) Finally, we algorithmitize an online recommender system that exploits the common practice of second-opinions in diagnostics, and can employ the results of (2) to maximize personalized treatment efficacy with minimal learning and in the presence of UCs.

2 Example: The Confounded Physicians

We begin with a motivating example depicting UCs in medical decision-making. In this scenario, physicians regularly prescribe one of two FDA-approved drugs to treat a certain condition. Each of the drugs, denoted $X \in \{0, 1\}$, have been shown to be equally effective at treating the condition in a randomized clinical trial (RCT); specifically, for patient recovery $Y \in \{0, 1\}$ where $Y = 1 = y_1$ indicates recovery, the study found a 70% recovery rate for each drug, i.e., $P(y_1|do(x)) = 0.7 \; \forall \; x \in X$. In reviewing her own patient records, one physician confirms this recovery rate, noting that the recovery rates of each patient she has treated, $P(y_1|x) = 0.7 \; \forall \; x \in X$, are also recovering at the experimentally reported rates. However, consulting with a colleague in her practice, she finds some discrepancy.

Supposing that patient populations between physicians are exchangeable, we will consider only two of many such possible unobserved confounding factors that may affect both treatment and recovery. The first is the patient's socio-economic status (SES), encoded as either low-SES ($S = 0$) or high-SES ($S = 1$). A patient's SES may be heuristically assessed by the physician (for example, through anecdotal indicators or appearance of the patient) and influence their treatment based on differences between the short- or long-term expenses of different therapies (Van Ryn and Burke 2000; Haider et al. 2015). Consider also that SES may covary with certain nutritional quality, such that higher SES patients may have access to better or more diverse meals that interact with the given treatments.

The second UC is the patient's treatment request, which can be influenced by Direct-to-Consumer Advertising (DTCA) (Lyles 2002; Ventola 2011). In particular, a patient may request a treatment ($R = 1$) or not ($R = 0$), which may influence a physician's decision if they decide to accommodate such requests. Consider also that an indirect pathway may link the medication requested to certain recovery covariates; e.g., drugs advertised on a sports station will be observed by patients who tend to get better exercise, and thus have better cardiovascular health (which may then interact with the treatments).

Recall that different physicians can have diverse treatment policies. In particular, consider that more accommodating physicians will attempt to honor their patients' requests for one medication over the other, but are also influenced by their perception of each patient's SES. Physicians of this "type" assign treatment by the structural equation, $X \leftarrow f_X^{P_1}(S, R) = XOR(S, R)$. Now, suppose another type of physician is aware of the influences of DTCA, and consciously (though without record) refuses to let patient requests influence their decisions; as such, these physicians' treatments can be modeled by the structural equation $X \leftarrow f_X^{P_2}(S) = S$.

Furthermore, from an omniscient viewpoint, we note that there is an even patient distribution over SES and requesters for each drug i.e., $P(r, s) = 0.25 \; \forall \; r \in R, s \in S$. As such, the *true* probabilities of recovery under each confounder state are listed in Table 1(a); the FDA's experimental study, and the observations of the accommodating (P_1) and stringent physicians (P_2), are shown in Table 1(b).

Table 1: (a) Recovery rates as a function of drug choice X, patient SES status S, and patient treatment request R. (b) Recovery rates according to the FDA experiment, $P(y_1|do(X))$, the observations of physicians 1 $P^{P_1}(y_1|X)$ and 2 $P^{P_2}(y_1|X)$.

(a)	$S = 0$		$S = 1$	
$P(y_1\|X, S, R)$	$R = 0$	$R = 1$	$R = 0$	$R = 1$
$X = 0$	0.70	0.80	0.60	0.70
$X = 1$	0.90	0.70	0.70	0.50

(b)	$P(y_1\|do(X))$	$P^{P_1}(y_1\|X)$	$P^{P_2}(y_1\|X)$
$X = 0$	0.70	0.70	0.75
$X = 1$	0.70	0.70	0.60

Scrutinizing this data, we see that the observational treatment policy of physician P_1 represents a case of *invisible confounding*; viz., $P(Y|do(X)) = P^{P_1}(Y|X) \; \forall \; x, x' \in X$, yet there are indeed confounding factors present in the system that the observational, experimental, and counterfactual distributions over recovery do not reveal alone. The plight of physician 2 is not entirely better; while the recovery rates associated with the ostensibly optimal drug $X = 0$ are superior in two configurations of S, R, and it appears as though P_2 receives more discriminant information about the UCs compared to P_1 (since $P(Y|do(X)) \neq P^{P_2}(Y|X) \; \forall \; x \in X$) we can see from Table 1(a) that there exist conditions under which $X = 1$ is actually the optimal assignment choice.

This scenario highlights two important problems in the domain of personalized medicine: (1) the influence of UCs in the diagnostic system may not be revealed in the FDA's experimental trials, and can affect physicians in practice, and (2) even in juxtaposing observational and experimental data (Tab. 1(b)), UCs can still evade detection (see, for example, P_1 above). These two points assert the need for a mechanism for reconciling differences in subjective physician treatment policies and their outcomes.

3 Background

To address the problems posed in the previous section, we begin by modeling confounded decision-making scenarios using the language of Structural Causal Models (SCM) to distinguish the observational, experimental, and counterfactual outcome distributions and articulate the notions of confounding and intent.

Definition 3.1. ***(Structural Causal Model)*** *(Pearl 2000, pp. 204-207) A Structural Causal Model is a 4-tuple, $M = \langle U, V, F, P(u) \rangle$ where: (1) U is a set of background variables (also called exogenous), that are determined by factors outside the model. (2) V is a set $\{V_1, V_2, ..., V_n\}$ of endogenous variables that are determined by variables in the model, viz. variables in $U \cup V$. (3) F is a set of functions $\{f_1, f_2, ..., f_n\}$ such that each f_i is a mapping from (the respective domains of) $u_i \cup PA_i$ to V_i where $U_i \subseteq U$ and $PA_i \subseteq V \setminus V_i$ and the entire set F forms a mapping from U to V. In other words, each f_i in $v_i = f_i(pa_i, u_i), i = 1, ..., n$ assigns a value to V_i*

that depends on (the values of) a select set of variables. (4) $P(u)$ is a probability function defined over the domain of U.

A key assumption in a confounded decision-making (CDM) scenario is that the deciding agents do *not* possess the fully-specified SCM (i.e., the identities and states of all unobserved variables and structural equations), but rather, have a partially-specified model of causal assumptions. In these settings, the key insight from (Bareinboim, Forney, and Pearl 2015) was that, for every decision variable, actors could condition on their "natural" treatment choice (i.e., the decision that they would have made reactively to the UCs, as realized in the observational setting) as a context in which to make a final choice. This observational treatment choice is known as the actor's intent.

Definition 3.2. ***(Intent)*** *For all variables requiring an actor's decision $\Pi_i \in \Pi$ in a SCM M, let the actor's intended choice $I_{\Pi_i,t} = i_{\Pi_i,t}$ be the choice that the actor would make observationally for unit t and the present unit's configuration of UCs $U_t = u_t$. Formally, for parents of $\Pi_i, pa(\Pi_i)$, let $I_{\Pi_i,t} = f_{\Pi_i}(pa(\Pi_i)_t,\ u_{\Pi_i,t})$.*

By choosing the treatment that maximizes the intent-specific reward (i.e., recovery), actors could control for some of the UC state that is otherwise summarized by an experimental maximization criteria. Intent-specific reward quantities were shown to be counterfactual, corresponding to the known counterfactual Effect of the Treatment on the Treated (ETT), of the format $P(Y_x|x')$. The new maximization criteria for actors in CDM scenarios was deemed the Regret Decision Criteria.

Definition 3.3. ***(Regret Decision Criteria (RDC))*** *(Bareinboim, Forney, and Pearl 2015) The Regret Decision Criteria states that an agent's intended action $I = i \in X$ serves as evidential context for the state of its environment, in which it may then interventionally act. RDC agents thus maximize the reward Y from a counterfactual perspective, such that the optimal action $x^* \in X$, conditioned on the intended action x', is defined as: $x^* = \text{argmax}_{x \in X}\, P(Y_x|x')$*

RDC is a counterfactual optimization criteria because the intended action x' and executed action x need not be equivalent, but was shown to be empirically estimable. The purpose of RDC was to minimize the lost rewards in a CDM setting as a function of each trial's UCs $U_t = u_t$ without knowing the identity nor state of U. RDC was shown to be produce superior policies compared to traditional, experimental techniques, by which the optimal action is defined as $x^* = \text{argmax}_{x \in X}\, P(Y_x)$.

The present work seeks to take the contributions of RDC in the online CDM domain and make several extensions: (1) Note that the above definitions rest upon the assumption that all actors' intent-generating functions are equivalent; in the following section, we relax this assumption to generalize to problems like those faced by the Confounded Physicians. Using this new formalism, we (2) show how RDC's procedure in the online decision-making domain can be generalized to offline experimental design, and finally, (3) how (2) can be used to the benefit of online decision-making.

4 Heterogeneous Intent

Returning to the Confounded Physicians example, suppose each physician collects data on their intent-specific recovery rates of each drug (results displayed in Table 2); this procedure simply requires they acknowledge their intended treatment choice, and then sample the available treatments under that context. Perhaps surprisingly, the intent-specific recovery rates of P_1 appear to be no different than the observational and experimental recovery rates for each drug. Maximizing via RDC, the expected recovery rate of P_1's patients will be 70%, and a marginally better 72.5% for P_2.

Table 2: Results $P(Y_x = 1|x')$ of RDC dynamic experiments conducted by physicians P_1 and P_2.

	$x'^{P_1} = 0$	$x'^{P_1} = 1$	$x'^{P_2} = 0$	$x'^{P_2} = 1$
$x = 0$	0.70	0.70	0.75	0.65
$x = 1$	0.70	0.70	0.80	0.60

The physicians face a perplexing situation in which the results of P_1's RDC experiment suggest that no confounding exists, yet P_2's seems to suggest that there does. They ponder which is the correct interpretation, and more importantly, how they may improve the recovery rates of their patients. The first insight towards addressing this goal is to acknowledge that P_1 and P_2 possess heterogeneous intents for treatment, as defined below.

Definition 4.1. ***(Heterogeneous Intents (HI))*** *Let A_1 and A_2 be two actors within a CDM instance, and M^{A_1} be the SCM associated with the choice policies of A_1 and likewise M^{A_2} be the SCM associated with the choice policies of A_2. For any decision variable $X \in \Pi_M$ and its associated intent $I = f_x$, the actors are said to have heterogeneous intent if $f_I^{A_1} \in F_{M^{A_1}}$ and $f_I^{A_2} \in F_{M^{A_2}}$ are distinct, viz., if $f_I^{A_1} \neq f_I^{A_2}$.*

Acknowledging that actors in the system may experience HI for the same treatment choice allows us to expand an SCM to account for multiple intent functions that, if sensitive to the UCs in even slightly different ways, can improve the accuracy of the learned parameters compared to the true reward distribution. By assumption, the learning agent does not possess the fully-specified model of the task, and as such, will never be able to explicitly estimate $P(U|I^{A_1}, ..., I^{A_a})$. That said, the benefits of conditioning on HIs are tangible in the HI specific reward distribution alone.

In the Confounded Physician scenario, consider the recovery rates associated with each drug in the context of each physicians' intent configuration. Although an extreme case, Tab. 3 shows that we have not only recovered the "true" reward distribution without ever knowing the identities of U, but also, knowing the intents of each IEC in concert, can obtain a superior average recovery rate (77.5%) that is higher than either physician's RDC maximization alone (70.0% and 72.5% for P_1 and P_2, respectively, Tab. 2).

The merit of combining intended treatments from different practitioners is not foreign to medicine; both patient-

Table 3: Recovery rates $P(Y_x = 1|I^{P_1}, I^{P_2})$ for each drug given the intents of physicians P_1 and P_2. Optimal treatments are indicated by asterisks (*).

	$I^{P_1}=0$		$I^{P_1}=1$	
	$I^{P_2}=0$	$I^{P_2}=1$	$I^{P_2}=0$	$I^{P_2}=1$
$X=0$	0.70	*0.70	*0.80	0.60
$X=1$	*0.90	0.50	0.70	*0.70

and practitioner-requested "second opinions" are commonplace and influential in the treatment process, and have been the subjects of recent studies (Vashitz et al. 2012; Meyer, Singh, and Graber 2015). The novel focus of *this* work is to study these concerted opinions at a more systemic level, and then determine if and how physicians with HI may aid the diagnostic process at a smaller scale. To do so, we aggregate the individual actors' SCMs into one that models them together.

Definition 4.2. ***(HI Structural Causal Model (HI-SCM))*** *A Heterogeneous Intent Structural Causal Model (HI-SCM) $M^{\boldsymbol{A}}$ is an SCM that combines the individual SCMs of actors $\boldsymbol{A} = \{A_1, A_2, ..., A_a\}$ such that each decision variable in $M^{\boldsymbol{A}}$ is a function of each actors' individual intents.*

The utility of an HI-SCM can be seen in the distinction between a learning *agent* (i.e., an intelligent system attempting to maximize treatment efficacy) and an *actor* (i.e., a decision-maker like a physician) in a CDM task; the agent's objective is to use its available data to recommend the best treatment to the deciding actors. Fig. 1 depicts an *agent's* prototypical HI-SCM; to reinforce the value of these models, consider the scenario wherein $I^{P_1} = 0, I^{P_2} = 0$. Consulting the HI recovery rates under this condition, an agent using an HI-SCM could suggest treatment $X = 1$ with expected recovery rate of 90% (Tab. 3), compared to the individual actors' 70% (P_1) and 80% (P_2) intent-specific recoveries (Tab. 2), and the FDA experiment's 70% recoveries (Tab. 1(b)).

Yet, the Confounded Physicians scenario exemplifies a fortunate case in which the actors' combined intents yields fruitful information about the state of the UCs compared to either intent alone. In general, it would be naïve to assume that every actor contributes new knowledge to the system, and instead, a recommender agent can attempt to filter those that do not. Towards this end, consider that actors with *homogeneous intent* (i.e., that have the same intent structural equations), provide equivalent information about the state of the UCs. Namely, consider two actors with equivalent intent functions in our prototypical HI-SCM (Fig. 1), and their implications for intent-specific treatment outcomes:

$$\begin{aligned} f_I^{A_1}(U) = f_I^{A_2}(U) &\Rightarrow P(U, I^{A_1}) = P(U, I^{A_2}) \\ &\Rightarrow P(Y_x|I^{A_1}) = P(Y_x|I^{A_2}) \\ &= P(Y_x|I^{A_1}, I^{A_2}) \end{aligned}$$

As a consequence, the intents of "like" actors can be clustered, and thus the dimensionality of the conditioning set is reduced. To do so, we define the notion of an Intent Equivalence Class for actors with the same intent functions:

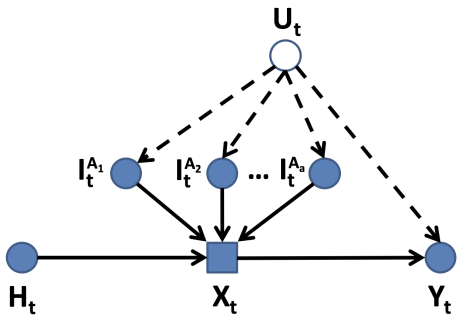

Figure 1: Graphical model of a prototypical HI-SCM M^A for a recommender agent viewing unit t, actor intents I_t^a, decision variable X_t, outcome Y_t, unobserved confounders U_t, and agent history H_t.

Definition 4.3. ***(Intent Equivalence Class (IEC))*** *In a HI-SCM M^A, we say that any two actors $A_i \neq A_j$ belong to separate intent equivalence classes $\boldsymbol{\Phi} = \{\phi_1, \phi_2, ..., \phi_m\}$ of intent functions f_I if $f_I^{A_i} \neq f_I^{A_j}$.*

The intents I^{A_k}, I^{A_p} of two actors $A_k \neq A_p$ in the same IEC, $\phi_r = \{A_k, A_p, ...\}$, are thus exchangeable, and can be instead summarized by annotating with their IEC, ϕ_r; e.g., in our prototypical HI-SCM, we can represent IEC-specific treatment outcomes for two actors in the same IEC as follows: $\phi_1 = \{A_1, A_2\} \Rightarrow P(Y_x|I^{A_1}) = P(Y_x|I^{A_2}) = P(Y_x|I^{A_1}, I^{A_2}) = P(Y_x|I^{\phi_1})$.

Using IECs to cluster equivalent actors reduces sampling requirements for each HI combination, but the primary benefit of considering separate IECs *in concert* is the ability to form superior treatment policies. Just as (Bareinboim, Forney, and Pearl 2015; Forney, Pearl, and Bareinboim 2017) demonstrated that intent-specific treatment maximizations would always be superior to experimental, we show that IEC-specific maximizations are superior to each actors' individually.

Theorem 4.1 (IEC-Specific Reward Superiority)**.** *Let X be a decision variable in a HI-SCM M^A with measured outcome Y, and let I^{ϕ_i} and I^{ϕ_j} be the heterogeneous intents of two distinct IECs ϕ_i, ϕ_j in the set of all IECs in the system, $\boldsymbol{\Phi}$. Maximized HI-specific rewards will always be at least as high as homogeneous, namely:*

$$\max_{x \in X} P(Y_x|I^{\phi_i}) \leq \max_{x \in X} P(Y_x|I^{\phi_i}, I^{\phi_j}) \; \forall \phi_i, \phi_j \in \Phi$$

See appendix for proof.

Thm. 4.1 thus provides a new maximization target for treatment selection in CDM scenarios, which extends RDC:

Definition 4.4. ***(HI Regret Decision Criteria (HI-RDC))*** *In a CDM scenario modeled by an HI-SCM M^A with treatment X, outcome Y, actor intended treatments I^{A_i}, and set of actor IECs $\boldsymbol{\Phi} = \{\phi_1, ..., \phi_m\}$, the optimal treatment $x^* \in X$ is the one that maximizes the IEC-specific treatment outcome, or formally: $x^* = \text{argmax}_{x \in X} P(Y_x|I^{\phi_1}, ..., I^{\phi_m})$*

Vitally, HI-RDC relies on knowing the IECs to which actors belong, which (by Def. 4.3) requires knowledge of each

actors' intent functions. In practice, however, these functions are not known by recommender agents a priori, nor will the actors in the presence of UCs (i.e., when influencing factors are unknown even to them). As such, agents require an empirical means of clustering actors into IECs that befits application of HI-RDC in pursuit of maximizing treatment efficacy. Because intents are observational in nature, they can likewise be observationally sampled over a number of units, and then grouped according to the following criteria:

Theorem 4.2 (Empirical IEC Clustering Criteria)**.** *Let A_i, A_j be two actors modeled by a HI-SCM, and I^{A_i}, I^{A_j} their intents for some decision. Actors A_i, A_j are clustered into the same IEC, $\{A_i, A_j\} \in \phi_r$, whenever their intended actions over the same units correlate, as their intent-specific treatment outcomes will agree. Formally:*

$$\rho(I^{A_i}, I^{A_j}) = 1 \Rightarrow \{A_i, A_j\} \in \phi_r \in \mathbf{\Phi}$$
$$\Rightarrow P(Y_x|I^{A_i}) = P(Y_x|I^{A_i}, I^{A_j})$$

See appendix for proof.

The requirement of Thm. 4.2 that actor intents perfectly correlate for admission into the same IEC is done to support the theoretical development of HI-RDC, but is typically too stringent for application; in a future section, we demonstrate that this criteria can be softened according to some agent-defined tolerance to account for actor-specific noise. Because Thm. 4.2 shifts the actor IEC clustering from a functional comparison to an empirically sampled one, we require strategies for collecting and then exploiting these samples.

The new treatment-efficacy maximization target provided by HI-RDC (Def. 4.4), coupled with the Empirical IEC Clustering Criteria (Thm. 4.2) can be applied in a similar fashion to online recommender systems as originally presented by (Bareinboim, Forney, and Pearl 2015). However, there are two problems with leaping directly into the online domain: (1) in the absence of any prior knowledge about the actors the agent would be advising, the ethics of exploratory recommendations (i.e., treatment suggestions that are intended to adequately sample IECs and HI-specific treatment effects) are somewhat dubious, and may take time to converge to the optimal treatment policy; and (2) if UCs are indeed present in the system (at which point HI-RDC becomes applicable to deconfound in the online domain), and the confounded treatments have already undergone experimental vetting, this implies that they were approved for use without knowledge of possible confounding.

Motivated by these problems, we next detail techniques in both offline and online experimental design that can detect the influence of UCs with simple additions to the traditional RCT procedure, and then use the enriched data that results to inform an online recommender system.

HI-Randomized Clinical Trials

Consider now an RCT in a CDM scenario modeled by a HI-SCM. For a single participant (or unit) t and randomized treatment X_t, this means that all observational causal influences of treatment assignment are severed, as by the $do(X_t = x_t)$ operation, and the influence of any UCs summarized across

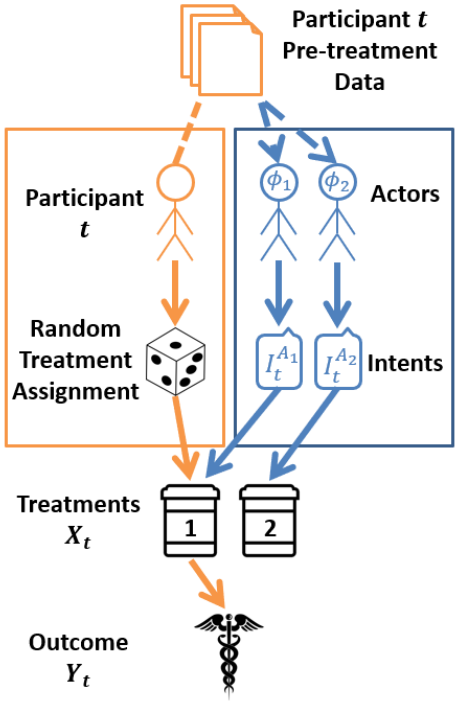

Figure 2: Depiction of an HI-RCT, with (left) the traditional RCT procedure, and (right) the additional HI collection layered on top.

experimental conditions. Assuming that individuals in the experimental population are representative of those treated in practice, any physician's *intended* treatments for each participant in the RCT can be compared to the treatment that was *randomly assigned*. Before examining the merit of this comparison, we define this procedure as a HI-RCT:

Definition 4.5. ***(HI Randomized Clinical Trial (HI-RCT))*** *Let X be the treatment of a Randomized Clinical Trial (RCT) in which all participants are randomly assigned to some experimental condition via intervention $do(X = x)$ with measured outcome Y. Furthermore, let $\mathbf{\Phi} = \{\phi_1, ..., \phi_m\}$ be the set of all IECs for actors in the HI-SCM M^A for which the RCT is meant to apply. A Heterogeneous Intent RCT (HI-RCT) is an RCT wherein treatments are randomly assigned to each participant, but in addition, the intended treatments I^{Φ} of sampled actors are collected for each participant.*

Fig. 2 depicts the HI-RCT procedure, and demonstrates how the added component of actors' HI collection (right) requires no changes to the traditional RCT procedure (left). In the standard RCT paradigm, (1) a participant's pre-treatment demographics, health history, and other relevant factors are collected, after which (2) they are randomly assigned to some experimental condition, and (3) the results of that treatment by some dependent measure (e.g., recovery) are observed. In the enhanced HI-RCT paradigm, (1) a participant's pre-treatment data is shared with actors who would typically assign treatment (e.g., prescribing physicians), (2) those actors

provide their intended[1] treatment for this participant which, crucially, may be different from that which was randomly assigned, and (3) the results of the randomly assigned treatment can be compared in meaningful ways to those intended by practitioners, yielding the following data:

(1) *Actor IECs* ($\mathbf{\Phi}$), by examining $\rho(A_i, A_j)$ between actors A_i, A_j, IECs can be determined using Thm. 4.2. (2) *Experimental Treatment Effects* (Y_x), the interventional data recorded in a traditional RCT, ignoring actor HIs, which is useful for predicting population-level treatment effects. (3) *Observational Treatment Effects* ($Y_x|I^{A_i},\ x = I^{A_i}$), when the intent of an actor matches the randomly assigned treatment, which can be used to predict the influence of UCs on treatment efficacy if left uncontrolled. (4) *Counterfactual Treatment Effects* ($Y_x|I^{A_i},\ x \neq I^{A_i}$), when the intent of an actor disagrees with the randomly assigned treatment, which can be used to identify actors with superior / inferior treatment policies and detect UCs. (5) *HI-Specific Treatment Effects* ($Y_x|I^{\phi_1}, ..., I^{\phi_m}$), when multiple IECs exist in the system, HI-specific treatment effects can be determined.

Apart from the individual utility of each of the datasets mentioned above, a new comparative criteria emerges that can detect confounding beyond the traditional juxtaposition of observational and experimental data:

Theorem 4.3. ***(HI-RCT Confounding Criteria)*** *In a CDM scenario modeled by an HI-SCM M^A with treatment X, outcome Y, actor intended treatments I^{A_i}, and set of actor IECs $\mathbf{\Phi} = \{\phi_1, ..., \phi_m\}$, there exists some unobserved[2] U such that $X \leftarrow U \rightarrow Y$ whenever $\exists\ x \in X, i^{\Phi} \in I^{\Phi}$: $P(Y_x) \neq P(Y_x|i^{\mathbf{\Phi}})$. See appendix for proof.*

Note that the traditional test for confounding, $P(Y|X) \neq P(Y|do(X))$ (Pearl 2000, Ch. 3), is subsumed by Thm. 4.3 since $P(Y|x) = P(Y_x|x')\ \forall x = x'$. However, in the Confounded Physicians scenario, the traditional test fails from the perspective of P_1, who may have misled agents to conclude no confounding; a comparison of Tabs. 1(b) and 3, however, reveals the presence of UCs.

HI-Online Recommender Systems

If an offline, HI-RCT study detects the presence of UCs in treatment selection, we now examine how a recommender agent can attempt to repair for their harmful influence in the online setting. Returning to the Confounded Physicians example, suppose an HI-RCT has revealed confounding for the treatments under consideration by P_1, P_2; since the UCs cannot be directly controlled for (as they are latent), the prescription of HI-RDC (Def. 4.4) is to learn the IECs for actors in the practice, and then consult their intents for each patient to maximize the chance of recovery. This task is known in the reinforcement learning community as the Multi-armed Bandit (MAB) problem, and specific to the confounded decision-making domain, a MAB problem with UCs (MABUC):

[1]The actors do not know the intents of other actors, nor the randomly assigned treatment or outcome. HIs can be collected *before* or *after* execution of the RCT component if these assumptions are met since intent is a pre-treatment variable, $I = I_x$.

[2]Assuming that any observed confounders have been controlled for (see back-door criterion, (Pearl 2000, Ch. 3)).

Definition 4.6. ***(Multi-armed Bandits with Unobserved Confounders (MABUC))*** *(Bareinboim, Forney, and Pearl 2015) A MABUC problem is characterized as a sequential decision problem over T trials / units, in which a learning agent attempts to maximize rewards Y through choice between k treatments (or arms) $x \in \{x_1, ..., x_k\}$ as decided by the agent's policy Π, history H, and UC state of advised actors U. A metric of the agent's success in a MABUC problem is its cumulative u-regret, namely, the difference (over all trials) between the optimal arm $x^*(u_t)$ under that trial's configuration of UCs u_t, and the arm chosen by the agent according to Π, x_t, or formally: $R^u = \sum_{t=1}^{T} P(y_{x^*(u_t)}|u_t) - y_{x_t^\pi}$*

In a MABUC in which actors experience HI, the task is slightly complicated and decomposes into two learning problems: (1) learning the IECs of actors in the system, and (2) learning the optimal treatment in any UC context, using the actors' intents as proxies for the UC state. Suppose the online agent had access to the results of an HI-RCT in the same domain, which had revealed a diverse, but possibly inexhaustive, set of IECs Φ_{off}. In the case where the IECs of the actors in the online setting are a subset of those in the offline, viz. $\Phi_{on} \subseteq \Phi_{off}$, the agent need only answer problem (1); if a mapping from $\Phi_{on} \rightarrow \Phi_{off}$ can be established, the optimal treatment is immediately available via HI-RDC from the HI-RCT's data (assuming, as is typically the case for FDA experiments, the HI-RCT has sufficient power to reveal these effects). In the case where this mapping does not exist (i.e., $\Phi_{on} \not\subseteq \Phi_{off}$), it is still possible to use the HI-RCT data to accelerate learning in the online domain.

To investigate the relationship between Φ_{off}, Φ_{on}, actors in the online setting can be subjected to a *calibration unit set*, a small questionnaire composed of participant data from the HI-RCT on which they are then asked to provide their intended treatment. By undergoing this calibration, actor IECs can be learned before the agent is required to make recommendations for live patients, and in the case where the online actors' intents correspond to any of the HI-RCT actors' intents, a mapping between $\Phi_{on} \rightarrow \Phi_{off}$ can be made. That said, units chosen from the HI-RCT to compose the calibration set are not best chosen randomly, but can be heuristically selected to improve discriminance over IECs.

Definition 4.7. ***(Actor Calibration-Set Heuristic)*** *Selection of some $n > 0$ calibration units from an offline HI-RCT dataset $\mathcal{D}$ can be used to learn the IECs of agents in an online domain before commencing recommendation. Selection can be guided by three heuristic scores $h(t) = h_c(t) + h_d(t) + h_o(t)$ for the quality of each unit t in the HI-RCT dataset:*

1. *Consistency: how consistent actors of the same IEC $\phi_r = \{A_1, ..., A_i\}$ intended to treat a unit,*
   $h_c(t) = (\#I_t^A \in \phi_r$ agreeing with majority$)/|\phi_r|$
2. *Diversity: how often a configuration of I^{Φ} has been chosen, favoring a diverse set of IEC intent combinations,*
   $h_d(t) = 1/(\#$ of times I^{Φ} appears in calibration set$)$
3. *Optimism: a bias towards choosing units in which the randomly assigned treatment x_t was optimal and succeeded, or suboptimal and failed, $h_o(t) =$*
   $1(P(Y_{x_t}|I_t^{\Phi}) > P(Y_{x'}|I_t^{\Phi})\ \forall\ x' \in X \setminus x_t)1(Y_t = 1) +$
   $1(P(Y_{x_t}|I_t^{\Phi}) < P(Y_{x'}|I_t^{\Phi})\ \exists\ x' \in X \setminus x_t)1(Y_t = 0)$

Algorithm 1 HI-RDC-RCT agent, parameterized by HI-RCT data $\mathcal{D}$, number of samples in the calibration set n, and IEC clustering tolerance τ such that Thm. 4.2 allows for noisy correlation between IEC actor intents, $\rho(I^{A_i}, I^{A_j}) \geq 1 - \tau$.

procedure HI-RDC-RCT-INIT($\mathcal{D}, n, \tau$)
  $calSet \leftarrow h(\mathcal{D}, n, \tau)$ ▷ (Def. 4.7)
  $H \leftarrow calSet$ ▷ Agent history gets calSet I^A, X, Y
procedure HI-RDC-RCT-RECOMMEND(t, n, τ)
  $i_t^A \leftarrow f_{I^A}(u_t)$ ▷ Unit's actor intents from UCs
  $\Phi \leftarrow IECs(I_t^A, H, \tau)$ ▷ IEC clustering (Thm. 4.2)
  $x_t \leftarrow \pi(H, I_t^A, \Phi)$ ▷ Policy selects arm (Def. 4.4)
  $y_t \leftarrow f_Y(u_t, x_t)$ ▷ Observe outcome
  $H \leftarrow \{i_t^A, x_t, y_t\}$ ▷ Update history

The calibration set is thus composed:
$h(\mathcal{D}, n) = \{t \in \mathcal{D} : n \text{ highest } h(t)\}$

In the CDM domain, Def. 4.7 serves as a curator for this initial questionnaire, with scores (1, 2) ensuring adequate discriminance to detect a new actor's IEC, and score (3) leveraging knowledge about IEC-optimal treatments gathered from the HI-RCT.[3]

5 Experimental Results

To validate the efficacy of the methodologies detailed in the previous sections, we first simulated an offline HI-RCT and then used the resulting data to inform an online HI-RDC agent in a HI-MABUC setting.[4]

HI-RCT Simulation. The Confounded Physicians scenario was simulated by pairing the intents of 3 actors belonging to one of two IECs $\Phi = \{\phi_1, \phi_2\}$ corresponding to noisy versions of the lenient and stringent physicians in Section 2 (the same structural equations as physicians P_1, P_2, but with 4% error to simulate random noise in human decision-making).[5] To demonstrate ideal conditions with large datasets (akin to an FDA RCT), the sample was comprised of $10,000$ units $\mathcal{D} = \{t \in [1, 10,000] : I_t^A, X_t, Y_t\}$. IEC clusters were correctly established using Thm. 4.2 with tolerance $\tau = 0.1$.

HI-RDC Simulation. With the results of the HI-RCT in hand, we next examined online recommender agents in the MABUC domain within the same Confounded Physicians scenario. The simulations consisted of $N = 1000$ Monte Carlo (MC) repetitions each composed of $T = 10,000$ units / trials. At each trial, the state of the UCs was sampled $u_t \sim P(U)$, the intents of 10 actors belonging to the same two IECs as in the HI-RCT were instantiated $i_t^A \leftarrow f_{I^A}(u_t)$, the agent's policy π selects an arm x_t, and an outcome y_t to that choice is observed. After all MC repetitions were completed, the average, cumulative u-regret (Def. 4.6) was assessed as a metric for each agent's performance.

[3]These scores assume the offline and online populations are exchangeable, as can be formalized via definitions of transportability.

[4]Simulation source code available at:
https://github.com/Forns/hi-mabuc-aaai19

[5]To account for this noise, HI-RDC agents selected a trial's IEC intent by plurality vote of individual actors in each IEC, though future study can be invested in more sophisticated approaches.

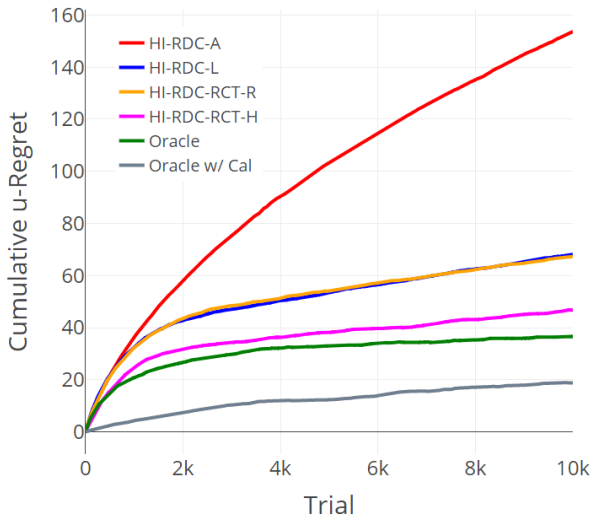

Figure 3: Comparison of agents in Confounded Physicians problem: cumulative u-regret as a function of trial t.

Agent Policies. To make fair comparisons to the results of (Bareinboim, Forney, and Pearl 2015; Forney, Pearl, and Bareinboim 2017) in the MABUC domain, all compared agents explore and exploit treatments using the Thompson Sampling procedure (Ortega and Braun 2014). Performance (Fig. 3) was compared between the following agents (from worst to best performance):

1. *HI-RDC-A*, maximized HI-specific rewards as an actor-intent RDC agent but *without* attempting to cluster actors by their IEC (thus, no calibration from an HI-RCT).
2. *HI-RDC-L*, maximized HI-specific rewards as an IEC-intent HI-RDC agent by *learning* actor IECs (but without calibration from an HI-RCT).
3. *HI-RDC-RCT-R*, maximized HI-specific rewards as an HI-RDC agent by clustering actors into IECs; this agent also started with a calibration set of size $n = 20$, but where datum in the calibration set were chosen at random from the simulated HI-RCT.
4. *HI-RDC-RCT-H*, same as (3) but with calibration samples chosen by Def. 4.7 (agent described in Alg. 1 w/ calibration set size of $n = 20$).
5. *Oracle*, a contextual bandit learner (Langford and Zhang 2008) that (unfairly) treats the state of the UCs as observed factors at each trial and treats them as a context (no calibration from an HI-RCT).
6. *Oracle w/ Calibration*, same as (4) but with a curated calibration set (Def. 4.7) of size $n = 20$ in which the values of all UCs are also known in each sample.

Note: the oracles are not realizable agents as they rely on observations about the unobserved; they are intended as yardsticks against which to compare the other agents.

Results. HI-RDC-RCT-H significantly outperforms its competitors on metrics of u-regret. In general, HI-RDC may

only partially recover the state of the UCs, in which case sublinear u-regret is not possible. That said, between each non-Oracle agent, noteworthy comparisons demonstrate the benefit of clustering actors into IECs (significant difference in u-regret between HI-RDC-A ($M = 153.52, SD = 49.48$) vs. HI-RDC-L ($M = 68.13, SD = 69.27$), $t(1998) = 31.71, p < .001$), and merit of calibration set selection heuristics over random selection (significant difference in u-regret between HI-RDC-RCT-R ($M = 67.28, SD = 68.26$) vs. HI-RDC-RCT-H ($M = 46.78, SD = 52.94$), $t(1998) = 7.50, p < .001$).

6 Conclusion

The historical reliance on RCTs as the chief means of scientific inquiry suggests that there are rich opportunities to use the formalization of heterogeneous intent (HI) in order to expand their findings; not only do HI-RCTs mingle the findings of observational and experimental studies, but provide new, counterfactual and HI-specific outcomes that can be used to improve personalized diagnostic and treatment policies. While HI-RCTs are useful for detecting previously invisible UCs, we detail a new online recommender agent (employing HI-RDC) to correct for the influence of UCs already manifest in practice. This new agent serves as a "driver-assist" for treatment selection, can benefit from the commonplace practice of diagnostic second-opinions, and can use the results of an HI-RCT to minimize learning.

Acknwledgements

Bareinboim's research is supported in parts by grants from IBM Research, Adobe Research, NSF IIS-1704352, and IIS-1750807 (CAREER).

References

Bareinboim, E.; Forney, A.; and Pearl, J. 2015. Bandits with unobserved confounders: A causal approach. In *Advances in Neural Information Processing Systems*, 1342–1350.

Brookhart, M. A.; Stürmer, T.; Glynn, R. J.; Rassen, J.; and Schneeweiss, S. 2010. Confounding control in healthcare database research: challenges and potential approaches. *Medical care* 48(6 0):S114–S120.

Cormack, D.; Harris, R.; Stanley, J.; Lacey, C.; Jones, R.; and Curtis, E. 2018. Ethnic bias amongst medical students in aotearoa/new zealand: Findings from the bias and decision making in medicine (bdmm) study. *PloS one* 13(8):e0201168.

Fisher, R. 1951. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 6th edition.

Forney, A.; Pearl, J.; and Bareinboim, E. 2017. Counterfactual data-fusion for online reinforcement learners. In *International Conference on Machine Learning*, 1156–1164.

Giffin, R. B.; Lebovitz, Y.; English, R. A.; et al. 2010. *Transforming clinical research in the United States: challenges and opportunities: workshop summary*. National Academies Press.

Haider, A. H.; Schneider, E. B.; Sriram, N.; Dossick, D. S.; Scott, V. K.; Swoboda, S. M.; Losonczy, L.; Haut, E. R.; Efron, D. T.; Pronovost, P. J.; et al. 2015. Unconscious race and social class bias among acute care surgical clinicians and clinical treatment decisions. *JAMA surgery* 150(5):457–464.

Hamburg, M. A., and Collins, F. S. 2010. The path to personalized medicine. *New England Journal of Medicine* 363(4):301–304.

Langford, J., and Zhang, T. 2008. The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in neural information processing systems*, 817–824.

Lyles, A. 2002. Direct Marketing of Pharmaceuticals to Consumers. *Annual Review of Public Health* 23(1):73–91.

Meyer, A. N.; Singh, H.; and Graber, M. L. 2015. Evaluation of Outcomes From a National Patient-initiated Second-opinion Program. *The American Journal of Medicine* 128(10):1138.e25–1138.e33.

Ortega, P. A., and Braun, D. A. 2014. Generalized thompson sampling for sequential decision-making and causal inference. *Complex Adaptive Systems Modeling* 2(1):2.

Papangelou, K.; Sechidis, K.; Weatherall, J.; and Brown, G. 2018. Toward an understanding of adversarial examples in clinical trials. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*.

Pearl, J.; Glymour, M.; and Jewell, N. P. 2016. *Causal Inference in Statistics: A Primer*. Wiley.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press. Second ed., 2009.

Pearl, J. 2017. Detecting latent heterogeneity. *Sociological Methods & Research* 46(3):370–389.

Ruberg, S. J., and Shen, L. 2015. Personalized medicine: four perspectives of tailored medicine. *Statistics in Biopharmaceutical Research* 7(3):214–229.

Shalit, U.; Johansson, F. D.; and Sontag, D. 2016. Estimating individual treatment effect: generalization bounds and algorithms. *arXiv preprint arXiv:1606.03976*.

Stead, W. W.; Starmer, J. M.; and McClellan, M. 2008. Beyond expert based practice. In *Evidence-Based Medicine and the Changing Nature of Healthcare: 2007 IOM Annual Meeting Summary*. National Academies Press.

Van Ryn, M., and Burke, J. 2000. The effect of patient race and socio-economic status on physicians' perceptions of patients. *Social science & medicine* 50(6):813–828.

Vashitz, G.; Pliskin, J. S.; Parmet, Y.; Kosashvili, Y.; Ifergane, G.; Wientroub, S.; and Davidovitch, N. 2012. Do first opinions affect second opinions? *Journal of general internal medicine* 27(10):1265–71.

Ventola, C. L. 2011. Direct-to-Consumer Pharmaceutical Advertising: Therapeutic or Toxic? *P & T : a peer-reviewed journal for formulary management* 36(10):669–84.

Wainer, H. 1989. Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics* 14:121–140.

White, H. D. 2005. Adherence and outcomes: it's more than taking the pills. *The Lancet* 366(9502):1989–1991.