



Combining the Clinical and Operational Perspectives in Heterogeneous Treatment Effect Inference in Healthcare Processes

Sam Verboven^{1(✉)} and Niels Martin^{2,3(✉)}

¹ Data Analytics Laboratory, Vrije Universiteit Brussel, Pleinlaan 2,
1050 Brussels, Belgium
sam.verboven@vub.be

² Research Group Business Informatics, Hasselt University, Martelarenlaan 42,
3500 Hasselt, Belgium
niels.martin@uhasselt.be

³ Research Foundation Flanders (FWO), Egmontstraat 5, 1000 Brussels, Belgium

Abstract. Recent developments in causal machine learning open perspectives for new approaches that support decision-making in healthcare processes using causal models. In particular, Heterogeneous Treatment Effect (HTE) inference enables the estimation of causal treatment effects for individual cases, offering great potential in a process mining context. At the same time, HTE literature typically focuses on clinical outcome measures, disregarding process efficiency. This paper shows the potential of jointly considering the clinical and operational effects of treatments in the context of healthcare processes. Moreover, we present a simple pipeline that makes existing HTE machine learning techniques directly applicable to event logs. Besides these conceptual contributions, a proof-of-concept application starting from the publicly available sepsis event log is outlined, forming the basis for a critical reflection regarding HTE estimation in a process mining context.

Keywords: Heterogeneous Treatment Effect · Process Mining · Machine Learning · Event Log

1 Introduction

Process mining techniques aim to extract valuable insights from process execution data captured in an event log [1]. As it starts from data entries representing real-life behaviour, instead of the assumed or ideal behaviour [1], process mining offers evidence-based insights in processes [20]. Within the healthcare domain, process mining techniques have been used for various use cases, such as automatically discovering the order of activities, assessing whether clinical guidelines have been followed, or identifying bottlenecks in a healthcare process [24].

While process mining in healthcare often focuses on conveying process insights to practitioners based on historical data, there is increasing awareness

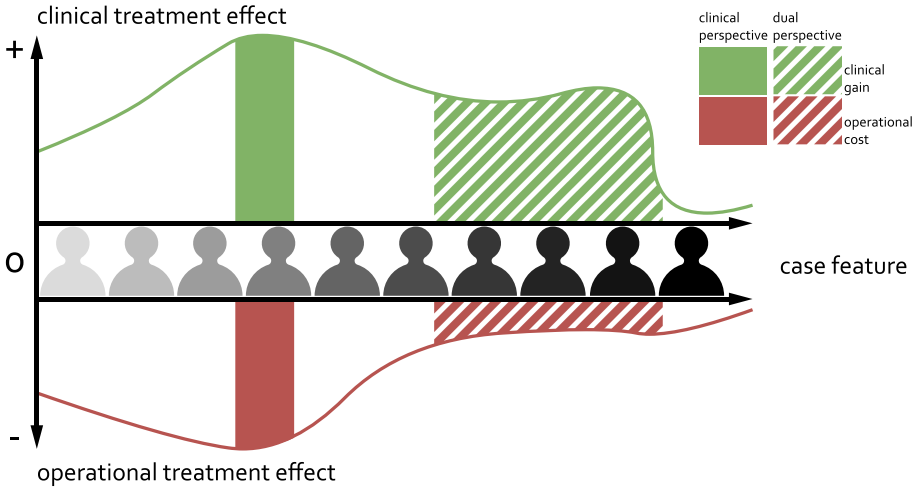


Fig. 1. The top graph y-axis depicts the positive clinical treatment effect. The bottom graph y-axis depicts the negative operational treatment effect (i.e., the operational cost). Along the shared x-axis, a case feature value is varied. Green areas below the treatment effect curve represent the clinical gain for a given policy. Red areas represent the operational cost. Filled areas represent the policy if we only take the clinical treatment effect into account. Dashed areas represent the policy when case-level process efficiency effects are also taken into account. Both clinical and operational effects can be estimated from event logs. In this example, taking individual operational efficiency effects into account more than doubles the total clinical effect. (Color figure online)

of the need for a complementary set of proactive techniques that can instigate actions in active processes [20]. This awareness, combined with recent developments in causal machine learning, opens perspectives for new approaches that support decision-making in healthcare processes using causal models.

Causal approaches in healthcare processes are confronted with three challenges. First, the effect of the same process intervention (e.g., the execution of a particular activity) can vary widely across patients. Nonetheless, current intervention guidelines are often developed at the population level and, hence, tuned to the *average* case. However, the goal in healthcare process management is evolving towards determining the optimal intervention for any case. Secondly, when causal models consider treatment effects at the patient level, there is a predominant focus on clinical outcome measures, with no regard for process efficiency. In practice, clinical and operational measures are not independent from each other. For instance, while a process intervention may be desirable from a clinical perspective (e.g., reduced likelihood for a particular adverse event), it might have negative implications from an operational point of view (e.g., in terms of ICU length-of-stay). Moreover, increased operational efficiency can also lead to improved clinical outcomes, as more patients receive treatment. Finally, causal models require assumptions to be made based on a priori domain

knowledge [22]. In other words, for models to have a causal interpretation, causal theory needs to be taken into account, preferably before data gathering.

Against the background of these three challenges, this paper explores the potential of Heterogeneous Treatment Effect (HTE) inference within the context of healthcare processes. Recent advances in causal machine learning enable the estimation of the causal treatment effect at the level of an individual using observational data. Consequently, event logs qualify as input for Heterogeneous Treatment Effect (HTE) modelling. In process mining, a treatment can represent any intervention within a healthcare process such as admitting a drug, executing selected activities in a specific order, or letting a particular resource perform an activity. Typical clinical outcomes include general life expectancy measures (e.g., expected days of survival [4]), and disease-specific parameters (e.g., tumor size [6]). Besides being suitable for HTE modelling, event logs also include important clues regarding the operational efficiency of a healthcare process (e.g., the length of stay or the resource involvement). This paper introduces a joint perspective on clinical and operational efficiency. The importance of adopting this joint perspective is illustrated conceptually in Fig. 1. The estimated operational and clinical treatment effects support crucial decisions within resource-constrained healthcare processes. This way, using HTE estimation provides detailed insights into the potential trade-offs between objectives are provided at the case level. A proof-of-concept application is presented using the publicly available sepsis event log [18].

The remainder of this paper is structured as follows. Section 2 introduces HTE inference and discusses the related work. Section 3 presents how HTE inference can be used in healthcare processes. In Sect. 4 a proof-of-concept is presented within the context of the sepsis event log. The paper ends with a discussion in Sect. 5 and a conclusion in Sect. 6.

2 Background

2.1 Heterogeneous Treatment Effects

The goal of HTE estimation is the estimation of the causal effect of a treatment $W \in \{0, 1\}$ on an outcome $Y \in \mathbb{R}$ for an individual i characterised by features $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^n$, where \mathcal{X} denotes the n -dimensional universe of features. We adopt the standard causal effect formulation in line with the standard Rubin/Neyman Potential Outcomes Framework [25]. In the binary setting, there are two potential outcomes (POs), Y_0 and Y_1 , that signify the outcomes when $W = 0$ and $W = 1$, respectively. The HTE can then be specified as:

$$\tau(\mathbf{x}) := \mathbb{E}[Y_1 \mid \mathbf{x}] - \mathbb{E}[Y_0 \mid \mathbf{x}] = \mathbb{E}[Y_1 - Y_0 \mid \mathbf{x}]. \quad (1)$$

From hereon, we will refer to $\tau(\mathbf{x})$ as the HTE.

Methods for HTE Estimation with Observational Data. From a machine learning point of view, two central elements distinguish HTE estimation (sometimes referred to as CATE/ITE) from a standard supervised learning problem. First, the HTE is unobservable for any individual, also referred to as the fundamental problem of causal inference [12]. For instance, when we execute an extra activity (treatment) in the process for an individual, we only observe the throughput time (outcome) with the extra step. For an individual, when $W = 1$ we observe Y_1 , when $W = 0$ we observe Y_0 , never both. Effectively, HTE models estimate something that cannot be observed directly. To still estimate $\tau(\mathbf{x})$, the dominant estimation strategy involves joint modelling of both POs in a multi-task neural network with one output per potential outcome. An estimate of the HTE is then constructed as the difference between PO estimates [26].

Second, standard supervised learning methods cannot handle treatment assignment policies that are not uniformly random, i.e., datasets with assignment bias. Assignment bias thus arises in an observational dataset when the propensity to receive treatment depends on the characteristics of individuals. In reality, this is almost always the case. For example, people with a more advanced stage of cancer will have a higher propensity to receive more radical treatment options. As such, treatment assignment bias induces the treated and untreated distributions to differ. In machine learning literature, this is called covariate shift [27]. Most algorithms for HTE estimation from observational data include some component to counteract such covariate shift. Examples of such components include inverse propensity weighting, propensity score matching [16], PPM [26].

Assumptions for HTE Estimation with Observational Data. Even though machine learning methods have been designed to tackle both aforementioned challenges, not all requirements can be validated or learned directly from the data. To guarantee that the treatment effect can be identified in the Rubin-Neyman PO framework, the following standard assumptions are made:

Assumption 1 (Stable Unit Treatment Value (SUTVA)). *First, there cannot be spillover effects between the potential outcomes of individuals in different treatment groups. Second, each unit is assumed to be presented with identical versions of each treatment. Third, we observe through the factual outcome Y the potential outcome associated with the assigned treatment.*

For example, Frank’s hospital stay length (outcome) should not depend on whether Sarah received antibiotics (treatment), and the antibiotics both would receive are the same. When Frank is assigned treatment ($W = 1$), we observe potential outcome Y_1 . This assumption is usually validated based on expert knowledge.

Assumption 2 (Overlap). *For all individuals $\mathbf{x} \in \mathbb{R}^n$, and all treatments $W \in \{0, 1\}$, the following holds: $0 < p(W|\mathbf{x}) < 1$.*

Overlap implies that for the whole feature support region every instance has a non-zero probability of receiving treatment. Intuitively, if there are no examples

of both potential outcomes for some regions of x , we cannot reliably estimate the causal effect for those x .

Assumption 3 (No hidden confounders). *This assumption implies that all variables that impact both treatment assignment and outcome are observed. As such, $(Y_0, Y_1) \perp\!\!\!\perp W | \mathbf{x}$.*

For example, to assess the effect of regular walking on mortality, a straightforward confounder is health status [11]: individuals with poor health walk less – effect on treatment assignment – and have higher chances of dying – effect on the outcome. Not including health status would lead a model to overestimate the causal effect of walking on health. Hence, in this context, it is crucial to collect health status data to avoid confounding bias.

Assumptions 2 and 3 together constitute *strong ignorability* given a set of covariates. When both SUTVA and strong ignorability hold, estimation of causal effects based on the factual outcomes in observational data is possible [25]. The assumptions regarding hidden confounders and SUTVA are fundamentally untestable based on observational data alone [14]. As such, expert knowledge plays a crucial role in HTE inference. The no hidden confounders assumption is the most difficult to satisfy. But as the dimensionality of X increases, the larger the probability that hidden confounders are observed. Consequently, a practical heuristic would be to gather as many features as possible in future event logs. This guideline facilitates causal learning, but stands in contrast with current process mining practices of narrow data gathering, often limited to which activities have been executed for a patient and when they were executed.

2.2 Related Work

HTE Estimation in Healthcare. Causal effect estimation allows us to address questions such as ‘how effective is a given treatment in curing this person?’ and ‘which treatment is more effective for this specific individual?’. Such questions are of critical importance in clinical decision-making. Moreover, recent availability of electronic healthcare records (EHR) and methodological advances have spurred increased interest in HTE inference as a clinical tool [3, 4, 6, 22].

Previous work for healthcare solely considers purely clinical outcomes of actions. However, it has been shown in a business context that taking into account costs greatly improves total profit [5, 32]. Similarly, it makes sense to account for overarching operational objectives. While the average treatment effect has been studied for multiple clinical outcomes (e.g., [17]), no existing work to our knowledge combines both operational and clinical effects of the same treatment.

Causality in Process Mining. Within the process mining field, there has been growing interest in the identification of causal patterns from an event log. This interest is exemplified by approaches developed to conduct root-cause analysis [10, 28, 31], even though they focus on finding characteristics that are correlated

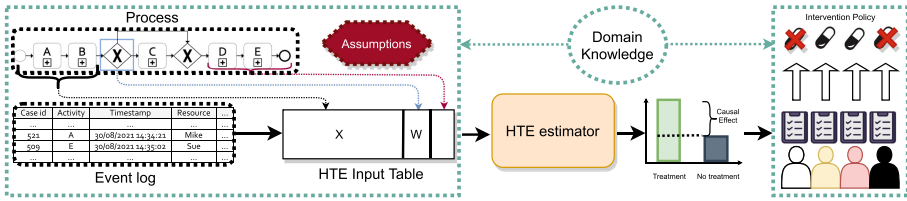


Fig. 2. Basic process flow for using causal models starting from event logs. An HTE input table can be constructed from an event log, allowing the application of standard causal machine learning methods. Domain knowledge plays a vital role in the determination of data collection, the intervention point, the validation of the assumptions, and final policy guidance.

with certain phenomena, without assessing whether the observed correlations are causal in nature. In contrast, Hompes et al. [13] and Narendra et al. [21] identify causal relationships at the process level starting from an event log using the Granger causality test and structural causal models, respectively.

Limited research has considered causal effects at a case level in the process mining field. Qafari and van der Aalst [23] use counterfactual reasoning to detect statements indicating why an undesirable outcome has happened for a particular case. Bozorgi et al. [7] also focus on the case level by proposing a technique that provides case level recommendations of treatments. The technique generates candidate treatments using action rule mining, after which an uplift tree and associated rules are generated for each candidate treatment. They apply their approach within the context of a loan application context [7].

Our work extends existing work on causality in process mining in general and HTE inference in particular, by jointly considering clinical treatment effects and operational treatment effects at the case level. Moreover, we formalise a simple pipeline that makes existing HTE machine learning techniques directly applicable to event logs.

3 Heterogeneous Treatment Effect Inference in Healthcare Processes

Definition 1 (Event, Trace and Event Log). Let \mathcal{A} represent the universe of attributes. An event $e \in \mathcal{A} \rightarrow \mathcal{X}$ is an assignment of values to attributes. Let $\mathcal{E} = \mathcal{A} \rightarrow \mathcal{X}$ represent the universe of events. A trace $t \in \mathcal{E}^*$ is a sequence of events referring to the same case c . Let $\mathcal{T} = \mathcal{E}^*$ represent the universe of traces. An event log L collects the traces of a set of cases, i.e., $L \subset \mathcal{T}$.

Definition 2 (HTE input table). Given an event log L , \mathcal{X} represents the universe of features which can be calculated over L . Let $f \in L \rightarrow \mathcal{X}$ be a feature function assigning values x calculated over L . Then, the HTE input table I consists of a set of entries γ , one entry $\forall c \in L$. Each entry $\gamma = (X_1, X_2, \dots, X_n, W, Y)$, where $X_1, \dots, X_n \in \mathcal{X}$, $W \in \{0, 1\}$ represents whether the treatment has been assigned, and Y represents the value of the outcome measure.

Under Assumptions 1 to 3, the HTE is identifiable and can be estimated using causal machine learning algorithms based on the HTE input table. Then, using Definitions 1 and 2, we can featurise the event log such that it translates to a standard set-up that facilitates use of all state-of-the-art machine learning algorithms for HTE estimation. A visual depiction of this pipeline can be found in Fig. 2.

4 Proof-of-Concept: Sepsis Event Log

4.1 Case Description

Healthcare process management benefits from a joint perspective on operational and clinical objectives of interventions. In many real-world healthcare processes, there is an apparent conflict between operational and clinical objectives. For instance, from a purely clinical point of view, extended hospital stay and extensive treatment with close supervision of clinicians is often optimal. From an operational perspective, typical process efficiency measures (e.g., throughput time) are improved with shorter treatment and earlier discharge. In reality, all healthcare processes are resource-constrained to some extent. Even when not explicitly considered, choices are made on the efficiency – effectiveness plane. HTE modelling allows mapping of the effects on both dimensions at the individual level, improving decision-making.

We empirically illustrate the potential of HTE inference in healthcare processes based on event logs, using the publicly available sepsis event log [18]. This event log contains events related to the trajectory of 1050 patients admitted to the emergency department (ED) of a Dutch hospital with sepsis symptoms [19]. The activities included in the event log relate, amongst others, to the moment when registration and triage took place, when laboratory results were recorded in the system, when antibiotics or liquid were administered and when the patient was discharged from the ED. Moreover, several parameters recorded in the triage document are available as event attributes. Finally, the observation that swift treatment with antibiotics is *always* advised according to the clinical guideline, but not applied in almost half the cases, illustrates the relevance of operational efficiency limits in treatment assignment [19].

4.2 Data Setup

Due to the fundamental problem of causal inference, the ground truth HTE is unobservable, and only one of the potential outcomes – i.e., the factual outcome – is ever observed. Consequently, we cannot directly assess HTE generalization performance based on factual data alone (e.g., using MSE, MAE). Furthermore, the factual outcome distribution reflects biased treatment selection. Hence, a biased model will perform better than a model that successfully corrects against assignment bias.

Table 1. Evaluation axes of Heterogeneous Treatment Effect models.

	Model Capacity	Counterfactual Estimation
Data	Original	Semi-synthetic
Metric type	Standard (e.g., MSE)	Specialised (e.g., PEHE)

These observations are reflected in the standard quantitative evaluation strategies for HTE estimators (Table 1). These strategies separately assess (i) the functional capacity to model the underlying response functions, i.e., whether the model can overfit the factual data, and (ii) its counterfactual estimation capabilities, i.e., whether the model correctly handles assignment bias to yield unbiased estimates of the potential outcomes. While (i) can be evaluated on the factual data, (ii) by definition requires a (semi-)synthetic setup. In such a setup the original features and treatment assignment are retained, but the potential outcomes are simulated [2, 4, 8, 14, 26, 29]. This way, the original assignment bias and feature structure in the dataset stay intact, while allowing quantitative evaluation of the HTE model with the Precision in Estimation of Heterogeneous Effects (PEHE) measure.

$$PEHE = \frac{1}{N} \sum_{i=1}^N (HTE_i - \hat{HTE}_i)^2 \quad (2)$$

In line with Alaa and van der Schaar [2], the data generating model for the clinical potential outcomes is specified by: $f_{c_0}(x) = \epsilon + \exp\left(\left(x + \frac{1}{2}\right)\Omega\right)$, and $f_{c_1}(x) = \epsilon + \Omega x - \omega$, for no treatment and treatment, respectively. The regression coefficients are comprised by Ω , and sampled uniformly from $[0, 0.1, 0.2, 0.3, 0.4]$. $\epsilon \sim \mathcal{N}(0, 1)$ adds i.i.d. sampled zero-centered additive Gaussian Noise to the potential outcomes. Finally, ω is selected such that the average treatment effect of the simulated clinical outcome matches the original sepsis event log data.

Next, we use the same functional form for the data generating model for the potential outcomes of the operational model, f_{o_0} and f_{o_1} . We sample regression coefficients comprised by Ω from $[0, -0.1, 0.2, -0.3, 0.4]$. Furthermore, ω is selected such that the operational cost of treating is always positive. After all, doing nothing should be cheapest. The synthetic operational efficiency effect has a mean of 0.69, a standard deviation of 0.17, and ranges from 0.22–1.28. Note that the original feature structure and treatment assignment bias from the sepsis log are once more retained.

4.3 Model Setup

We use `cfrnet` [26], a popular neural network-based HTE estimator that uses an integral probability metric in its loss function to explicitly balance the covariate distribution of the treated and untreated group within a learned shared representation. After deletion of observations with missing values, 642 patient

observations are retained. We hold out 100 observations for validation, 200 for testing, and use the rest for training. `Cfrnet` is run twice, once for the clinical outcome, and once for the operational outcome, yielding estimates of $H\hat{T}E_C$ and $H\hat{T}E_O$, respectively. We use the same hyperparameters as reported in Shalit et al. [26].

We validated that `cfrnet` performs well for both model capacity and counterfactual estimation by assessing validation set MSE and PEHE on the factual and synthetic data, respectively. Note that outside of proof-of-concept demonstrations, one should rigorously select an appropriate functional class among multiple benchmarks to avoid model misspecification [30].

As a benchmark treatment assignment policy, we rank all individuals by their clinical effect $H\hat{T}E_C$. This is the standard assignment policy in HTE literature. Specific to our setting is that we introduce an operational cost that, if exceeded, prohibits further treatment of individuals. Our synthetic setup reflects that treating each individual has a unique cost that depends on its features x . For the second policy, reflecting the joint clinical-operational perspective, we take into account the impact on the operational budget and treat based on the estimated clinical effect per unit of operational effect, or $H\hat{T}E_C/H\hat{T}E_O$.

4.4 Results

On the test set, taking into account the joint perspective, we treat 170% more patients and achieve a total clinical effect increase of 57.83%, compared to the clinical-only baseline, using the same operational resources. The results highlight the synergy between the process and clinical views. Even with maximisation of clinical effect in mind, it is thus helpful to adopt a joint perspective. Post-deployment, a model can be further evaluated by assessing whether following the model’s recommendations has improved patient outcomes.

5 Discussion

The interplay between causal learning and process mining is a promising frontier for the management of healthcare processes. However, to empirically validate this promise, awareness of the practical requirements of HTE inference is required. Based on the conducted analyses, we enumerate three main lessons. Moreover, we reflect upon two broader perspectives on HTE inference for process mining.

Lesson 1: The HTE input table enables the use of state-of-the-art causal machine learning algorithms. The transformation of an event log to the HTE input table, the standard HTE modelling setup, can be performed with minimal overhead. Hence, state-of-the-art machine learning methods are available to the process mining community to develop causal models using event logs. We believe this simple formalization significantly lowers the threshold for coalescence between the HTE inference and process mining communities.

Lesson 2: To take more effective actions in healthcare, effects on both clinical and operational outcomes need to be modeled. Currently, many process mining works do not give explicit consideration to clinical process outcomes, while the machine learning for healthcare community does not model operational outcomes. In healthcare, the impact of an action on process efficiency is often not uniformly distributed across every case or intervention. We have intuitively illustrated how total clinical gain can be achieved when jointly modelling the clinical and operational effect of actions. However, to apply this in a real-life setting, we need the right combination of data and domain knowledge.

Lesson 3: A paradigm-shift for event log building is needed to fully capitalise on causal learning. To facilitate causal process interventions based on HTE inference, more a priori planning is required than is the current practice in the process mining field. First, the causal assumptions (Sect. 2.1) need to be validated together with domain experts. Second, these assumptions also translate to explicit data requirements. Currently, event logs mainly highlight when particular activities were executed on a patient. However, for HTE inference, confounders also need to be included, which will often require broader data extraction when building an event log. Finally, to enable jointly modelling operational and clinical outcomes, representative outcome measures need to be defined for the application at hand.

Perspective 1: Methodological extensions towards methods that can learn directly from event logs and capitalise on time dependencies are on the horizon. While the definition of the HTE input table offers a simple solution to enable causal learning using event logs, featurising the event log comes at the cost of losing information. For example, the time-series nature of the data is often lost when translating to a tabular data structure. Hence, methods that can learn from the original event logs offer opportunities to learn from richer data and the time dependencies in the log. Possible solutions could originate from time-series compatible models, such as RNN, LSTM or Transformer-based architectures.

Perspective 2: More discussion is needed to establish a consensus on policy standards and ethics. While opportunities arise due to novel technologies based on observational data, the adoption of decision support systems in healthcare needs to be soundly motivated. Adoption standards have not yet been established for learning HTEs from observational data of healthcare processes. Existing evaluation standards have mainly evolved from the machine learning field and not from a consensus of requirements from governing bodies (e.g., EMA, FDA) and healthcare organizations. Although stronger theoretical underpinnings can increase trust in HTE model predictions, uncertainty estimates offer an explicit measure of model confidence. Ultimately, the level of uncertainty also influences healthcare process decision making [15]. Finally, we refer to Eichler et al. [9] for a detailed discussion on the requirements of algorithmic decision-support in healthcare based on observational data.

6 Conclusion

In this paper, we introduce a joint approach to HTE inference, combining the clinical and operational perspective of healthcare processes. Despite its potential, careful consideration is required to incorporate HTE inference in the toolbox of healthcare organisations. When the prevailing assumptions are not accounted for when building the event logs, estimates of causal effects will not be identifiable and, hence, biased. Most importantly, strong cooperation with domain experts is needed to check for hidden confounders as violations of this assumption cannot be deduced from the data itself. To the best of our knowledge, no publicly available event logs have been collected with the HTE assumptions in mind, which hampers the development and testability of causal learning for process mining.

References

1. van der Aalst, W.M.P.: *Process Mining - Data Science in Action*. Springer, Heidelberg (2016). <https://doi.org/10.1007/978-3-662-49851-4>
2. Alaa, A.M., van der Schaar, M.: Bayesian inference of individualized treatment effects using multi-task Gaussian processes. arXiv preprint [arXiv:1704.02801](https://arxiv.org/abs/1704.02801) (2017)
3. Berrevoets, J., Alaa, A., Qian, Z., Jordon, J., Gimson, A.E., Van Der Schaar, M.: Learning queueing policies for organ transplantation allocation using interpretable counterfactual survival analysis. In: *Proceedings of the International Conference on Machine Learning*, pp. 792–802 (2021)
4. Berrevoets, J., Jordon, J., Bica, I., Gimson, A., van der Schaar, M.: OrganITE: optimal transplant donor organ offering using an individual treatment effect. In: *Proceedings of the 2020 Annual Conference on Neural Information Processing Systems* (2020)
5. Berrevoets, J., Verboven, S., Verbeke, W.: Optimising individual-treatment-effect using bandits. arXiv preprint [arXiv:1910.07265](https://arxiv.org/abs/1910.07265) (2019)
6. Bica, I., Alaa, A.M., Lambert, C., Van Der Schaar, M.: From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. *Clin. Pharmacol. Theor.* **109**(1), 87–100 (2021)
7. Bozorgi, Z.D., Teinmaa, I., Dumas, M., Rosa, M.L., Polyvyanyy, A.: Process mining meets causal machine learning: Discovering causal rules from event logs. In: van Dongen, B.F., Montali, M., Wynn, M.T. (eds.) *Proceedings of the 2nd International Conference on Process Mining*, pp. 129–136 (2020)
8. Curth, A., van der Schaar, M.: Doing great at estimating CATE? On the neglected assumptions in benchmark comparisons of treatment effect estimators. arXiv preprint [arXiv:2107.13346](https://arxiv.org/abs/2107.13346) (2021)
9. Eichler, H.G., et al.: Are novel, nonrandomized analytic methods fit for decision making? the need for prospective, controlled, and transparent validation. *Clin. Pharmacol. Ther.* **107**(4), 773–779 (2020)
10. Gupta, N., Anand, K., Sureka, A.: Pariket: mining business process logs for root cause analysis of anomalous incidents. In: Chu, W., Kikuchi, S., Bhalla, S. (eds.) *DNIS 2015. LNCS*, vol. 8999, pp. 244–263. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16313-0_19

11. Hakim, A.A., et al.: Effects of walking on mortality among nonsmoking retired men. *New Engl. J. Med.* **338**(2), 94–99 (1998)
12. Holland, P.W.: Statistics and causal inference. *J. Am. Stat. Assoc.* **81**(396), 945–960 (1986)
13. Hompes, B.F.A., Maaradji, A., La Rosa, M., Dumas, M., Buijs, J.C.A.M., van der Aalst, W.M.P.: Discovering causal factors explaining business process performance variation. In: Dubois, E., Pohl, K. (eds.) *CAiSE 2017. LNCS*, vol. 10253, pp. 177–192. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59536-8_12
14. Johansson, F.D., Shalit, U., Kallus, N., Sontag, D.: Generalization bounds and representation learning for estimation of potential outcomes and causal effects. *arXiv preprint arXiv:2001.07426* (2020)
15. Kochenderfer, M.J.: *Decision Making Under Uncertainty: Theory and Application*. MIT Press, Cambridge (2015)
16. Kurth, T., et al.: Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am. J. Epidemiol.* **163**(3), 262–270 (2006)
17. Li, D., et al.: Assessment of treatment effect with multiple outcomes in 2 clinical trials of patients with Duchenne muscular dystrophy. *JAMA Network Open* **3**(2), e1921306 (2020)
18. Mannhardt, F.: Sepsis cases - event log, 4TU.ResearchData. dataset (2016). <https://doi.org/10.4121/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460>
19. Mannhardt, F., Blinde, D.: Analyzing the trajectories of patients with sepsis using process mining. In: *CEUR Workshop Proceedings*, vol. 1859, pp. 72–80 (2017)
20. Martin, N., et al.: Recommendations for enhancing the usability and understandability of process mining in healthcare. *Artif. Intell. Med.* **109**, 101962 (2020)
21. Narendra, T., Agarwal, P., Gupta, M., Dechu, S.: Counterfactual reasoning for process optimization using structural causal models. In: Hildebrandt, T., van Dongen, B.F., Röglinger, M., Mendling, J. (eds.) *BPM 2019. LNBIP*, vol. 360, pp. 91–106. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26643-1_6
22. Proserpi, M., et al.: Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat. Mach. Intell.* **2**(7), 369–375 (2020)
23. Qafari, M.S., van der Aalst, W.M.P.: Case level counterfactual reasoning in process mining. In: Nurcan, S., Korthaus, A. (eds.) *CAiSE 2021. LNBIP*, vol. 424, pp. 55–63. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-79108-7_7
24. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: a literature review. *J. Biomed. Inform.* **61**, 224–236 (2016)
25. Rubin, D.B.: Causal inference using potential outcomes: design, modeling, decisions. *J. Am. Stat. Assoc.* **100**(469), 322–331 (2005)
26. Shalit, U., Johansson, F.D., Sontag, D.: Estimating individual treatment effect: generalization bounds and algorithms. In: *International Conference on Machine Learning*, pp. 3076–3085 (2017)
27. Sugiyama, M., Krauledat, M., Müller, K.R.: Covariate shift adaptation by importance weighted cross validation. *J. Mach. Learn. Res.* **8**(5), 985–1005 (2007)
28. Suriadi, S., Ouyang, C., van der Aalst, W.M.P., ter Hofstede, A.H.M.: Root cause analysis with enriched process logs. In: La Rosa, M., Soffer, P. (eds.) *BPM 2012. LNBIP*, vol. 132, pp. 174–186. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36285-9_18
29. Tian, Y., Schuemie, M.J., Suchard, M.A.: Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int. J. Epidemiol.* **47**(6), 2005–2014 (2018)

30. Vansteelandt, S., Bekaert, M., Claeskens, G.: On model selection and model misspecification in causal inference. *Stat. Methods Med. Res.* **21**(1), 7–30 (2012)
31. Vasilyev, E., Ferreira, D.R., Iijima, J.: Using inductive reasoning to find the cause of process delays. In: *Proceedings of the 15th Conference on Business Informatics*, pp. 242–249 (2013)
32. Verbeke, W., Olaya, D., Berrevoets, J., Verboven, S., Maldonado, S.: The foundations of cost-sensitive causal classification. arXiv preprint [arXiv:2007.12582](https://arxiv.org/abs/2007.12582) (2020)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

