# Prescriptive Analytics for Business Processes Using Causal Inference (Extended Abstract)

Zahra Dasht Bozorgi[1]

[1]University of Melbourne, 700 Swanston St, Carlton, VIC 3053, Australia

**Abstract**

In this Ph.D. project, we aim to address the need for effective recommendations by devising new process monitoring methods which are prescriptive in nature. These methods will recommend the process participants, at tactical and operational levels, what action they need to take to achieve a given process objective. Moreover, we will use causal inference methods to provide recommendations based on *causal relationships* between the proposed actions and the outcome of interest.

**Keywords**

prescriptive process monitoring, causal inference, machine learning

## 1. Introduction

Recent advances in process mining techniques have allowed companies to manage and improve their processes more efficiently. Particularly, process mining has benefited from advances in machine learning techniques to provide accurate predictions of the future state of business processes in an area known as *predictive process monitoring*. Recent predictive monitoring methods can produce highly accurate predictions of the future state of the process. However, having reliable predictions does not always lead to improvement of the process. A study by Dees *et al.* [1], shows that if good predictions are followed by bad recommendations, the desired improvement is not achieved. This study illustrates the need for new techniques that find the best interventions based on the context of each case.

Causal inference is a field in statistics concerned with discovering and quantifying causal relationships. This field has been widely used in other domains such as medicine, social sciences, and marketing. In medicine, causal inference techniques are used to determine the effectiveness of new drugs. In social sciences and marketing, new policies and campaigns are evaluated using state-of-the-art causal inference methods. The success of causal inference in these fields serves as a motivation to use these techniques in the area of process mining.

In this Ph.D. project, we aim to address the need for effective recommendations by devising new prescriptive process monitoring methods which are based on *causal relationships*. These methods will make recommendations, at tactical and operational levels, about what actions should be taken to achieve a given process objective.

*ICPM 2022 Doctoral Consortium and Tool Demonstration Track*
zdashtbozorg@student.unimelb.edu.au (Z. D. Bozorgi)
https://www.linkedin.com/in/zahradbozorgi/ (Z. D. Bozorgi)
0000-0002-1595-3934 (Z. D. Bozorgi)

## 2. Research Problem

Starting from a set of pre-defined treatments (a.k.a. process interventions) and associated costs, my PhD project explores the use of causal inference in prescriptive monitoring, to recommend what treatments are to be used in what context such that a given benefit function is maximised for the organisation. Specifically, the overall aim of this project is to answer the following question:

*How to provide cost-aware recommendations based on causal relations between a proposed intervention and a target of interest?*

Since the area of prescriptive process monitoring remains largely under-explored, we select three main dimensions of exploration for scoping purposes. These are:

- Who to treat, exploring which cases should get the treatment
- When to treat, exploring when the treatment should be applied
- How to treat, exploring which treatment to choose

Accordingly, we define the following research questions to explore these dimensions:

1. RQ1: Can causal inference provide useful recommendations about which business process to treat?
2. RQ2: Can causal inference provide useful recommendations about when to treat a business process?
3. RQ3: Can causal inference provide useful recommendations about multiple parameterizable treatments of a business process?

## 3. Proposed Solutions

To address the identified research problems, we propose three studies each addressing one research question. To the best of our knowledge, this is the first time causal inference has been used in a prescriptive process monitoring context.

### 3.1. RQ1: Who to treat

The main purpose of this study is to investigate the use of causal inference for prescriptive monitoring. We hypothesise that if we can establish a causal relationship between a possible intervention and a performance metric, the recommendations that result from such causal relationships will result in more benefit for the company than prediction-based recommendations. We divided this study into two sub-studies, one addressing outcome improvement and the other cycle time reduction. We chose these two objectives because one is an example of a binary outcome, while the other is described as a real-valued attribute.

Our approach proposed in [2] is a rule-based prescriptive system (Figure 1). First, a set of interventions which are highly correlated with a positive outcome of interest are identified. Then using a causal effect estimation method called Uplift Tree [3], we identify contexts in

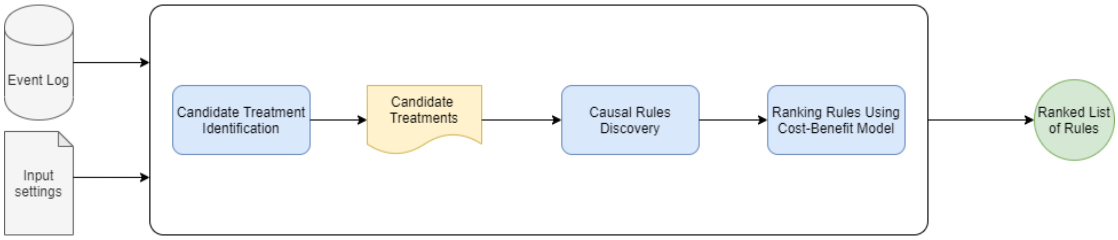


**Figure 1:** Overview of the approach for outcome improvement.

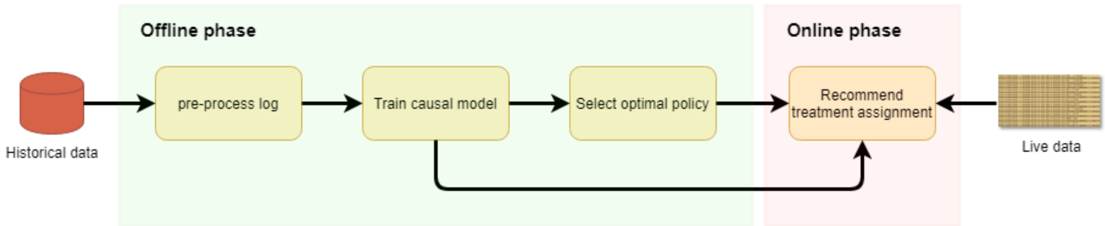


**Figure 2:** Overview of the approach for cycle time reduction.

which those interventions causally influence the outcome. We also propose a cost model that identifies the Return-on-Investment (ROI) of an intervention. The decision to intervene in a process is then made based on the ROI. We show that using this method, we can automatically identify improvement actions that are traditionally done by humans.

In another study [4], we propose a cost-aware prescriptive monitoring method that is designed to reduce cycle time (Figure 2). The core of this recommendation system is the Orthogonal Random Forest (ORF) [5], which is a causal estimation method that works with continuous outcome variables. The approach consists of two phases: an offline phase in which given an event log, a causal effect estimation model is trained, and the best policy for applying the intervention is selected to maximise the gain in applying it. The second phase is the online phase where the causal effect estimator and the selected policy are used to determine which ongoing cases should receive an intervention. Our results show that selecting an intervention policy based on causal models leads to a higher net gain than policies based on traditional machine learning methods for prediction.

## 3.2. RQ2: When to treat

In the second study, we investigate the question of when to apply the treatment. Another important aspect of generating good recommendations is to ascertain when to recommend. This question has been addressed by previous works on prescriptive process monitoring. In the work by Teinemaa *et al* [6], the authors propose empirical thresholding to determine whether an alarm for triggering an intervention should be raised. Metzger *et al.* [7] propose a method based on reinforcement learning (RL) and show that it outperforms empirical thresholding. The main limitation of this method is that they make simplistic assumptions about the effectiveness of the interventions and the focus is more on the prediction of the performance metric being optimised. To address RQ2, we propose to extend the approach by Metzger *et al.* by including causal effect estimates in the reinforcement learning environment. Specifically, we aim to find when-to-treat policies by guiding a reinforcement learning agent to make decisions based on information from causal models in addition to predictive models. In a prescriptive monitoring method based on reinforcement learning, the environment is defined by case prefixes extracted from event logs and ongoing process executions, as well as the agent's estimates of the performance metric and the causal effect of the treatment. The reward is a function based on one or multiple process performance metrics and can include the cost and benefit of the treatment. Our initial results indicate that our approach leads to more net gain than the traditional RL-based method relying purely on predictions.

### 3.3. RQ3: How to treat

In the previous studies, we assume that there is a binary treatment that can potentially influence a process performance metric. However, in practice, many actions can be taken during the execution of the process that influences performance metrics. With the methods proposed in the first study, many causal models need to be trained separately, and this is computationally inefficient. Furthermore, using many standalone causal models will not consider how the different actions influence each other (i.e., the action interplay). Given the success of reinforcement learning in the study by Metzger *et al.* and our own findings in the previous RQ, we plan to extend our RL-based framework to deal with multiple treatments and treatments with continuous values. We will expand the action space to include more values and continuous values and will explore different RL algorithms that are suitable for multi-action and continuous action spaces.

## 4. Methodology

This project will follow a Design Science research method [8]. The rigour of the approaches will be ensured by conducting an extensive literature review and constructing a comprehensive evaluation benchmark, using well-defined selection and assessment criteria. The relevance of the solutions will be ensured via an extensive evaluation of the developed techniques with real-life and simulated data sets, and where possible, through case studies with relevant organisations.

## 5. Challenges and Further Developments

The main challenge in this study is that we assume that all the variables that influence intervention and outcome are observed in the event log. This might lead to biased estimations that cause the prescriptive system to make a sub-optimal recommendation. We try to alleviate this threat to validity by performing sensitivity analysis on our models to measure their robustness to unmeasured confounding. Another challenge in this study is the evaluation of the proposed methods. While we use cost models and other evaluation methods in the causal inference literature to evaluate our prescriptive methods, the best and most rigorous way of evaluating these methods is by conducting an A/B test. Also, we plan to seek domain expert feedback to evaluate our methods qualitatively.

## References


[1] M. Dees, M. de Leoni, W. M. P. van der Aalst, H. A. Reijers, What if process predictions are not followed by good recommendations?, in: Proceedings of the Industry Forum at BPM co-located with 17th International Conference on Business Process Management BPM, 2019.

[2] Z. D. Bozorgi, I. Teinemaa, M. Dumas, M. L. Rosa, A. Polyvyanyy, Process mining meets causal machine learning: Discovering causal rules from event logs, in: 2nd International Conference on Process Mining, ICPM, 2020.

[3] P. Rzepakowski, S. Jaroszewicz, Decision trees for uplift modeling with single and multiple treatments, Knowl. Inf. Syst. (2012).

[4] Z. D. Bozorgi, I. Teinemaa, M. Dumas, M. L. Rosa, A. Polyvyanyy, Prescriptive process monitoring for cost-aware cycle time reduction, in: 3rd International Conference on Process Mining, ICPM, 2021.

[5] M. Oprescu, V. Syrgkanis, Z. S. Wu, Orthogonal random forest for causal inference, in: Proceedings of the 36th International Conference on Machine Learning, ICML, 2019.

[6] I. Teinemaa, N. Tax, M. de Leoni, M. Dumas, F. M. Maggi, Alarm-based prescriptive process monitoring, in: International Conference on Business Process Management, Springer, 2018, pp. 91–107.

[7] A. Metzger, T. Kley, A. Palm, Triggering proactive business process adaptations via online reinforcement learning, in: International Conference on Business Process Management, Springer, 2020, pp. 273–290.

[8] R. H. von Alan, S. T. March, J. Park, S. Ram, Design science in information systems research, MIS quarterly (2004).