Technical University of Munich
TUM School of Computation, Information and Technology

Deciding What to Measure: Methods for Feature Importance and Acquisition

Henrik Quirin von Kleist

Complete reprint of the dissertation approved by the TUM School of Computation, Information and Technology of the Technical University of Munich for the award of the

Doktor der Naturwissenschaften (Dr. rer. nat.)

Chair: Prof. Dr. Mathias Drton

Examiners:

1. Prof. Dr. Daniel Rückert

2. Prof. Dr. Niki Kilbertus

3. Assoc. Prof. Ilya Shpitser, Ph.D.

The dissertation was submitted to the Technical University of Munich on 29 August 2025 and accepted by the TUM School of Computation, Information and Technology on 5 November 2025.

Acknowledgments

I would like to thank everyone who supported and guided me throughout my PhD.

First, I am grateful to Narges Ahmidi, who supervised me during the first three years of my PhD at Helmholtz Munich. From the beginning, she was open to my ideas and gave me the freedom to pursue my own direction. Her thoughtful mentorship played a key role in shaping my development as a researcher.

I am also thankful to Carsten Marr, who welcomed me into his group during the final phase of my PhD. I appreciated his clarity, steady guidance, and the trust he placed in my work. His leadership always kept the people behind the research in focus, which made for a supportive and motivating environment.

I owe special thanks to Ilya Shpitser, whose mentorship had a deep influence on my understanding of causal inference. I am particularly thankful for the opportunity to visit Johns Hopkins University—an experience that was valuable both academically and personally. Our collaboration has continued beyond the visit, and working with him remains one of the most rewarding aspects of my PhD.

I would also like to thank Daniel Rückert for his support as my supervisor at TUM. His feedback and broader perspective helped me stay grounded and focused throughout the process.

It was also a pleasure to supervise and work with my master's students Alonso, Joshua, and Marios. I greatly enjoyed our discussions, and appreciated their enthusiasm, curiosity, and thoughtful perspectives on the topics we explored together.

Many thanks to my colleagues in Narges's group: Alireza, Elly, Octavia, Leopold, and Theresa. In particular, I want to thank Alireza for the many helpful discussions we've had over the years, which often helped me see problems in a new light.

I'm also grateful to my colleagues at Johns Hopkins: Trung, Zixiao, Jaron, and Numair. I especially thank Trung for the many engaging conversations and his curiosity to explore the same questions that motivated me.

This journey was shaped by the input and support of many people, and I'm truly thankful to all of you.

Abstract

In many real-world domains, decisions must be made under uncertainty. Acquiring additional information can improve outcomes—but often comes at a cost. In healthcare, diagnostic tests can reduce uncertainty and support clinical decisions, but may also be invasive, expensive, or time-consuming.

This thesis investigates how machine learning systems can navigate such trade-offs through *active feature acquisition (AFA)*—the task of learning policies that selectively acquire input features in a sequential, cost-sensitive manner to optimize the performance of a downstream prediction model. To bring AFA into practical use, two main challenges must be addressed: how to quantify the value of acquiring information under missing data, and how to evaluate acquisition policies using retrospective data.

The first paper included in this cumulative thesis addresses the question of how to measure the value of features when data is missing. It introduces a formal distinction between two types of feature importance: *observed data feature importance*, which quantifies the value of features under the actual data distribution where missingness may be present, and *full data feature importance*, which considers a hypothetical setting in which all features are fully observed. To complement these metrics, the paper proposes the *feature measurement importance gradient (FMIG)*, which estimates the marginal benefit of measuring a feature more frequently.

The second paper formalizes the *active feature acquisition performance evaluation (AFAPE)* problem: how to reliably evaluate AFA policies using retrospective data collected under a different acquisition policy. We identify key assumptions that make this feasible—most importantly, the *no direct effect (NDE)* assumption, which requires that acquisitions affect only what is observed, not the underlying feature values, and the *no unobserved confounding (NUC)* assumption, which ensures acquisition decisions are explainable by observed variables. By linking AFAPE to frameworks in *missing data analysis* and *offline reinforcement learning*, we clarify when existing tools can be applied. Under both assumptions, we introduce a *semi-offline reinforcement learning* framework and develop novel estimators. A semiparametric efficiency analysis then characterizes the tradeoffs among estimators, clarifying how to balance statistical efficiency and robustness to model misspecification.

These contributions provide a principled foundation for feature acquisition in dynamic, cost-sensitive settings. By clarifying when feature importance and policy evaluation are identifiable and estimable, the thesis lays the groundwork for optimizing active acquisition under diverse assumptions and deployment conditions.

Kurzfassung

In zahlreichen Anwendungsfeldern müssen Entscheidungen unter Unsicherheit getroffen werden, und das gezielte Erheben zusätzlicher Informationen kann die Ergebnisse verbessern – ist jedoch häufig mit Kosten verbunden. In der Medizin beispielsweise können diagnostische Tests zur Reduktion von Unsicherheit und zur Unterstützung klinischer Entscheidungen beitragen, sind jedoch oft invasiv, teuer oder zeitaufwendig.

Diese Dissertation untersucht, wie maschinelles Lernen helfen kann, solche Zielkonflikte zu bewältigen – anhand des Konzepts des *active feature acquisition (AFA)*. Dabei geht es um die Entwicklung von Algorithmen, die selektiv und sequenziell Informationen erheben, um die Vorhersageleistung eines Modells bei begrenztem Ressourceneinsatz zu optimieren. Um AFA praktisch einsetzbar zu machen, müssen zwei zentrale Herausforderungen adressiert werden: Wie lässt sich der Wert von Informationsgewinnung aus echten, unvollständigen Datensätzen quantifizieren? Und wie kann man AFA-Algorithmen anhand retrospektiver Daten evaluieren?

Die erste in dieser kumulativen Dissertation enthaltene Arbeit befasst sich mit der Frage, wie sich der Wert einzelner Informationen in unvollständigen Datensätzen messen lässt. Sie führt eine formale Unterscheidung zwischen zwei Arten von Informationswichtigkeit ein: *Informationswichtigkeit unter beobachteten Daten*, die den Wert von Informationen unter der realen, möglicherweise durch fehlende Werte geprägten Datenverteilung beschreibt, und *Informationswichtigkeit unter vollständigen Daten*, die ein hypothetisches Szenario mit vollständig erhobenen Informationen betrachtet. Ergänzend wird der *feature measurement importance gradient (FMIG)* vorgeschlagen – ein Maß für den marginalen Nutzen einer häufigeren Erhebung bestimmter Informationen.

Die zweite Arbeit formalisiert das *active feature acquisition performance evaluation (AFAPE)*-Problem: wie sich AFA-Algorithmen zuverlässig anhand retrospektiver Daten bewerten lassen, die unter einer anderen Erhebungsstrategie gesammelt wurden. Wir identifizieren zentrale Voraussetzungen, unter denen eine solche Bewertung möglich ist – insbesondere die *no direct effect (NDE)*-Annahme, wonach sich die Erhebung nur auf die Beobachtung, nicht aber auf die zugrundeliegenden Informationswerte auswirkt, sowie die *no unobserved confounding (NUC)*-Annahme, der zufolge Erhebungsentscheidungen vollständig durch beobachtbare Variablen erklärbar sind. Durch die Verknüpfung von AFAPE mit etablierten Sichtweisen aus der *Fehlende-Daten-Analyse* und dem *Offline Reinforcement Learning* zeigen wir auf, wann und wie existierende Methoden genutzt werden können. Unter beiden Annahmen führen wir

ein *semi-offline reinforcement learning*-Framework ein und entwickeln neue Schätzverfahren. Eine semiparametrische Effizienzanalyse beleuchtet die Zielkonflikte zwischen statistischer Effizienz und Robustheit gegenüber Modellfehlspezifikation.

Diese Beiträge schaffen eine fundierte Grundlage für die gezielte Informationsgewinnung in dynamischen, ressourcenbeschränkten Einsatzfeldern. Indem sie darlegt, wann Informationswichtigkeit und die Bewertung von AFA-Algorithmen berechenbar sind, legt die Dissertation das methodische Fundament für die Optimierung von AFA-Algorithmen unter unterschiedlichen Annahmen und Einsatzbedingungen.

List of Publications

This thesis is based on the following peer-reviewed publications:

1. von Kleist, H., Wendland, J., Shpitser, I., and Marr, C. Feature Importance Metrics in the Presence of Missing Data. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267. PMLR, 2025. Accepted for publication. To appear.

2. von Kleist, H., Zamanian, A., Shpitser, I., and Ahmidi, N. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings. *Journal of Machine Learning Research*, 26(60):1–84, 2025.

Contents

1. Introduction

In many real-world settings, access to additional information can improve decision-making, yet acquiring that information is often costly. These acquisition costs—whether financial, ethical, time-consuming, physically invasive, or cognitively and computationally demanding—fundamentally shape how individuals and institutions choose to gather and act on information.

In healthcare, for example, diagnostic tests can significantly enhance clinical decision-making but may be invasive, expensive, or time-consuming [13]. Screening policies similarly involve trade-offs: while early detection through widespread testing can reduce disease burden, it also incurs high costs and risks overdiagnosis or unnecessary treatment [27, 18]. These decisions must weigh the marginal value of additional information against the cost of acquiring it.

The same tension arises in criminal justice. Pretrial risk assessments are used to estimate the likelihood of recidivism and inform bail decisions, but high-quality assessments often require sensitive or costly information, such as detailed interviews or background checks [12]. In online services, recommender systems selectively solicit user feedback to personalize content, aiming to improve relevance while minimizing user burden [44]. In industrial settings, such as machine maintenance for robotics, data acquisition to assess system health can be expensive or disruptive, prompting strategic decisions about when and what to measure [29].

Costly information also plays a central role in economics. Under the framework of rational inattention [25], economists have shown that behaviors which appear suboptimal can often be understood as rational responses once the costs of acquiring or processing information are taken into account. For instance, the well-documented "home bias" in international finance—where investors overweight domestic assets despite diversification benefits—can arise from the high cost of monitoring foreign markets [25]. Similarly, in macroeconomics, firms adjust prices infrequently not due to a lack of information per se, but because tracking and interpreting aggregate shocks can be prohibitively costly relative to expected gains [25]. The persistence of high unemployment after recessions may likewise reflect the costs of staying informed about labor market conditions [25]. When acquiring and processing relevant information requires substantial resources or attention, firms and workers may delay adjustments—leading to sluggish or incomplete responses to economic shocks.

These examples—from healthcare and criminal justice to online services and macroeconomics—reflect a common challenge: deciding what information is worth acquiring under

cost constraints. Among these, healthcare is especially illustrative and is a central focus of this thesis. Diagnostic testing, in particular, offers a concrete lens on how information acquisition decisions are made in practice, often under uncertainty. To examine this more closely, I now turn to empirical evidence on how tests are used in medical settings.

While acquisition costs clearly shape decision-making in real-world settings, studies have shown that diagnostic tests are often overused or underused in practice. A large review of 63 studies across 15 countries found that among 47 diagnostic tests, 17 were underused in more than 50% of cases, while 11 were overused more than half the time [32]. These problems are not trivial: diagnostic testing is a critical part of patient care, and nearly half of all diagnostic errors (44%) in the U.S. occur during the testing phase—through failures to order, report, or follow up on results—compared to 32% due to errors in clinical assessment [43].

These patterns reflect more than just cognitive limitations. In many cases, poor information decisions are driven by competing incentives. A U.S. survey found that 91% of physicians believe unnecessary tests or procedures are frequently ordered to avoid malpractice lawsuits [3]. Patient expectations also play a role—physicians may order tests to satisfy patients rather than explain why they are unnecessary. Financial incentives can further distort test ordering: one study found that the odds of ordering a common laboratory test were up to eight times higher among specialist physicians with a financial stake in an on-site lab [4].

Even in simplified, incentive-free settings, humans often struggle to make optimal information-cost trade-offs. Experimental studies have shown that participants tend to underpurchase information or choose suboptimal sources, even when better options are available at the same cost [10].

Together, these findings underscore the difficulty of deciding what to measure—whether due to institutional incentives or inherent cognitive limitations. This motivates the need for tools that can support such decisions in a systematic and cost-aware way.

1.1. Active Feature Acquisition

One approach to addressing this challenge is to develop algorithmic policies that recommend what to measure and when. In machine learning, this problem is formalized as *active feature acquisition (AFA)* [1, 21, 22, 8, 45, 54]. AFA methods aim to select a subset of features to observe—balancing the cost of acquiring each feature against its expected value for a downstream task such as classification. For example, in medical diagnosis, an AFA system might recommend a personalized sequence of tests, choosing only those that meaningfully improve accuracy relative to their cost or risk. To support the practical use of AFA, this thesis addresses two key methodological challenges.

1. How can we determine features that are most valuable to acquire? In practice, many datasets contain missing values due to selective testing or measurement constraints. Traditional feature importance metrics typically assume fully observed inputs, which limits their relevance in these settings. The first paper of this thesis [51] develops a framework for defining and estimating feature importance under missing data—providing a foundation for identifying for which features an AFA system would be most beneficial.

2. How can we evaluate the performance of an AFA system before deployment? In realistic settings, the data available for evaluation is not collected under the AFA system's policy, but instead reflects the acquisition choices made by humans—for example, a physician deciding which tests to order. As a result, it can be difficult to assess how the AFA system would have performed under its own policy. The second—and main—paper of this thesis [52] formalizes this problem as *active feature acquisition performance evaluation (AFAPE)* and develops new estimators for evaluating AFA agents using retrospective data.

Together, these two studies aim to build a more rigorous foundation for designing, interpreting, and evaluating AFA systems under real-world constraints. Throughout both studies, I developed the core ideas, formalized the problem settings, and carried out the theoretical and empirical work. My co-authors contributed through discussion, feedback, and editorial input. In the following, I provide a more detailed summary of the two papers.

1.2. Paper 1: Feature Importance Metrics in the Presence of Missing Data

This paper [51] examines how to define and interpret feature importance metrics when features are selectively missing. It introduces a distinction between two perspectives: *full data feature importance*, which reflects a feature's importance had there been no missingness, and *observed data feature importance*, which evaluates importance under the actual measurement policy. While both are useful, they target different questions and rely on different assumptions—yet are often conflated in practice. The paper clarifies this distinction and categorizes commonly used estimation approaches based on which perspective they implicitly target and how they behave under different missingness mechanisms.

To complement these metrics, the paper introduces the *feature measurement importance gradient (FMIG)*—a model-agnostic metric that quantifies how prediction performance would improve if a feature were measured more frequently. FMIG avoids strong assumptions and provides actionable guidance for adapting measurement strategies. These contributions offer a foundation for reasoning about feature value under missingness—a necessary first step toward designing effective AFA systems.

1.3. Paper 2: Active Feature Acquisition Performance Evaluation

This paper [52] addresses the challenge of evaluating AFA agents using retrospective data. In many real-world applications, including healthcare, available datasets reflect decisions made by humans—such as clinicians—not the acquisition policies of a proposed AFA system. As a result, evaluating how an AFA agent would perform at deployment requires counterfactual reasoning: estimating the predictive and acquisition costs the agent would incur under its own policy, not the one used to generate the data.

We formalize this as the *active feature acquisition performance evaluation* (AFAPE) problem and analyze the assumptions required to estimate such counterfactual costs. The paper shows that AFAPE sits at the intersection of missing data analysis and offline reinforcement learning, depending on whether one assumes that acquisitions don't affect the underlying feature values (the no direct effect assumption) or that they depend only on observed variables (the no unobserved confounding assumption).

Building on this insight, the paper proposes a new *semi-offline reinforcement learning* framework that unifies both perspectives. Within this setting, we introduce three estimators—direct, inverse probability weighted, and doubly robust—that can consistently estimate agent performance under milder positivity assumptions than standard approaches.

These contributions provide the tools needed to evaluate AFA agents before deployment—an essential prerequisite for training them optimally.

2. Methods

This chapter lays the statistical foundation for the research questions addressed in this thesis. At their core, many of these questions are *counterfactual* in nature. For example: *What would have been the importance of feature X_1, had it been fully observed?* Or: *What would the outcomes have been if an AFA agent, rather than a human expert, had made the feature acquisition decisions?*

To reason about such questions, I draw on tools from both classical statistics and causal inference. To structure the methodological approach, I follow the standard three-step pipeline of statistical analysis:

1. **Estimand Specification**: I begin by formally defining the quantities to be estimated — referred to as *estimands*. This requires a language for formulating counterfactual queries. I rely on the potential outcomes framework [42] and causal graphical models [34, 36], which provide a formal apparatus for expressing what would have happened under alternative actions—for example, if an agent had selected a different set of features than those observed. This step clarifies the target of analysis and distinguishes between different notions of feature importance or predictive performance under varying acquisition strategies.

2. **Identification**: Next, I examine the conditions under which these estimands can be inferred from the observed data. For this, I use causal graphical models to make key assumptions explicit and interpretable. These models help represent the data-generating process, account for mechanisms of missingness, and clarify the role of confounding. By encoding assumptions graphically, I can assess whether the observed data contains sufficient information to identify the estimand.

3. **Estimation**: Finally, I construct estimators to recover the target quantities from observed data. This step draws on techniques from machine learning and semiparametric theory, with an emphasis on developing methods that are not only flexible but also statistically sound. I highlight key properties such as unbiasedness, consistency, efficiency, and robustness, which serve as criteria for evaluating the quality of different estimators throughout the thesis.

To make these ideas more concrete, I interleave this chapter with simplified toy problems that mirror some of the core challenges addressed in the papers. These examples isolate specific estimation or identification difficulties and illustrate, on a smaller scale, the reasoning and techniques used in the papers.

2.1. Estimand Specification

To answer a statistical question using data, one must first clearly and mathematically define what is to be estimated—this is the *estimand*. Estimand specification is a foundational step in any principled analysis, as it determines both the interpretation of results and the assumptions needed for identification and estimation.

Unfortunately, this step is often neglected in the machine learning literature, leading to ambiguity about the target of inference. For instance, in our work on feature importance under missing data [51], we show that many studies fail to distinguish between two fundamentally different questions: (i) how important a feature is in a dataset that contains missing values, and (ii) how important that feature would be if all data were fully observed.

These correspond to different estimands—reflecting the *observed* and *full data* distributions, respectively—yet they are often conflated in both method and interpretation. This leads to misleading conclusions, especially when the missingness mechanism is informative.

A similar ambiguity arises in our work on evaluating active feature acquisition performance [52]. We observe that prior studies often report performance estimates under the implicit assumption that agents can only select from the features acquired in the retrospective dataset. This assumption—effectively limiting the agent to historical test choices—is rarely made explicit and is generally unrealistic. By specifying the estimand precisely, we clarify the evaluation task and avoid this common but rarely acknowledged modeling error.

These examples highlight the central role of estimand specification in both the interpretation of feature importance and the evaluation of AFA policies. Yet one reason this step is so often neglected is that machine learning lacks a standard formalism for expressing counterfactual questions. In the following subsection, I introduce the *potential outcomes framework* [42] —a well-established formalism from causal inference that provides the language needed to express counterfactual questions precisely and unambiguously.

2.1.1. The Potential Outcomes Framework

In the potential outcomes framework [42], the random variable $Y_{(a)}$ denotes the outcome that would be observed if a treatment variable A were set to value a—possibly contrary to fact. This framework provides a formal language for reasoning about counterfactuals and will be used throughout this chapter to define target quantities.

I begin with a simple treatment–outcome setting to illustrate key ideas. For example, A might represent a medical treatment (e.g., a prescribed medication) and Y a health outcome (e.g., recovery from disease). $Y_{(a)}$ then describes what the patient's outcome would have been had they received treatment a, regardless of what actually occurred.

In the context of this thesis, the "treatment" corresponds to a set of feature acquisition decisions—such as which diagnostic tests were performed. The outcome may then be the value of a specific feature or the performance of a predictive model. From this perspective, many missing data problems can be naturally viewed through the lens of causal inference. To make this connection explicit, I briefly introduce the standard notation for representing missing data using potential outcomes [28].

Missing Data as a Causal Problem

Let $X_{(1)}$ denote the fully observed (true) value of a variable, X the possibly missing observed value, and $R \in \{0, 1\}$ a binary *missingness indicator*, where $R = 1$ indicates that X is observed. These variables are linked via the *consistency assumption*:

$$X = \begin{cases} X_{(1)} & \text{if } R = 1, \\ \text{"?"} & \text{if } R = 0. \end{cases} \tag{2.1}$$

Here, $X_{(1)}$ can be interpreted as the potential outcome of X under the hypothetical intervention $R = 1$—that is, had the value been observed. This notation makes explicit the difference between what is actually seen in the dataset and what is assumed to exist in principle, but may not be observed due to missingness.

Returning to the feature importance example, the potential outcomes framework allows one to express precisely what is meant by "feature importance had there been no missing data." By defining feature importance in terms of the counterfactual feature vector $X_{(1)}$, one can isolate the estimand of interest from artifacts of the missingness mechanism. This clarity is essential not only for sound interpretation, but also for ensuring that estimation strategies are aligned with the target quantity.

To visualize and reason more explicitly about the causal structure of the underlying data-generating processes and missingness mechanisms, I complement this framework with *graphical causal models*[34]. These models provide a transparent and intuitive way to encode assumptions, assess identifiability, and guide the choice of estimation strategies.

2.1.2. Graphical Causal Models

Graphical causal models provide a compact and interpretable way to represent assumptions about the data-generating process. In this thesis, I use *directed acyclic graphs (DAGs)* [34] and *single-world intervention graphs (SWIGs)* [36] to encode causal relationships between variables.

Formally, a DAG $\mathcal{G}(V)$ consists of a set of nodes V representing variables, and directed edges indicating direct causal influence. The graph is acyclic, meaning it contains no directed cycles. A DAG defines a statistical model by specifying a factorization of the joint distribution:

$$p(V) = \prod_{V_i \in V} p(V_i \mid \text{pa}_{\mathcal{G}}(V_i)),$$

where $\text{pa}_{\mathcal{G}}(V_i)$ denotes the parents of V_i in $\mathcal{G}$—that is, all nodes with an edge pointing into V_i.

This factorization implies a set of conditional independencies. In particular, the absence of an edge between two variables encodes an assumption of conditional independence given other variables. These relationships can be formally identified using the concept of *d-separation* [34].

Two variables X and Y are said to be d-separated given a set Z—written $X \perp\!\!\!\perp_{d\text{-}sep} Y \mid Z$—if every path between X and Y is blocked when conditioning on Z. Under the *global Markov property*, d-separation implies conditional independence:

$$X \perp\!\!\!\perp Y \mid Z.$$

A path is blocked by Z if at least one of the following conditions holds:

1. The path contains a chain structure $A \to B \to C$ or $A \leftarrow B \to C$, and the middle node B is in Z.

2. The path contains a collider structure $A \to B \leftarrow C$, and neither B nor any of its descendants are in Z.

These graphical criteria allow one to formally express and inspect the assumptions underlying a statistical model. Throughout this thesis, I use DAGs to assess identifiability, characterize missingness mechanisms, and define estimands precisely.

To represent potential outcomes within causal graphs, I use *single-world intervention graphs* (SWIGs) [36]. SWIGs extend standard DAGs by explicitly encoding the consequences of interventions—for example, setting a treatment variable to a fixed value.

Figure 2.1B shows the SWIG corresponding to the DAG in Figure 2.1A, under an intervention that sets A to the value a. In a SWIG, the intervened variable A is split into two nodes: one representing its natural (pre-intervention) version, and one representing the fixed value imposed by the intervention. There is no edge between them. Downstream variables are relabeled as potential outcomes—e.g., $Y_{(a)}$—to reflect their dependence on the intervention.

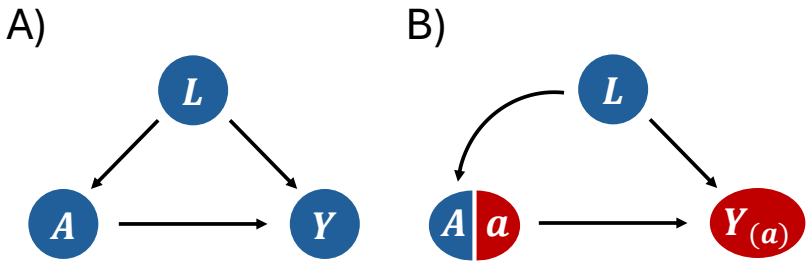


Figure 2.1.: (A) A standard causal DAG illustrating the relationships between variables A (the treatment), L (the confounder), and Y (the outcome). (B) The corresponding single-world intervention graph (SWIG) under the intervention that sets A to a. In the SWIG, the variable A is split into its natural version A and its intervened value (fixed to a), and downstream variables such as Y are relabeled as potential outcomes $Y_{(a)}$. All incoming edges into the intervened variable are removed, reflecting the modified causal structure under intervention.

Missing Data Graphs

Having introduced graphical causal models more generally, I now return to the missing data setting to illustrate how these models can be adapted to represent assumptions about missingness. A common approach is to use *missing data graphs* (or *m-graphs*) [26, 46].

Missing data graphs are designed to represent the structure of missingness problems and differ from conventional causal graphs in that they include variables representing unobserved ground-truth values. Figure 2.2 illustrates examples for a setting where the true features $X_{(1)}$ are multivariate, split into components $X_{(1),1}$ and $X_{(1),2}$. The graphs also depict the missingness indicators R_1 and R_2, as well as the observed proxy variables X_1 and X_2, which are deterministically defined by $X_{(1)}$ and R through the *consistency assumption* (see Eq. 2.1), represented by dotted blue arrows.

The inclusion of counterfactuals in m-graphs allows for the formalization of different assumptions about the missingness mechanism. These are typically categorized into three types [23]:

- *Missing-completely-at-random* (MCAR): The missingness of a variable is statistically independent of both observed and unobserved variables, although it may still be influenced by other missingness indicators. In graphical terms, there are no arrows from any observed or unobserved data variables (e.g., L, $X_{(1),1}$, $X_{(1),2}$) into the missingness indicators. An example of an MCAR graph is shown in Figure 2.2A).
- *Missing-at-random* (MAR): The missingness of a variable may only depend on variables that are observed. For example, in Figure 2.2B), the observed covariate L influences R_1 and R_2. More complex MAR scenarios can occur, such as when $X_{(1),1}$ affects R_2,

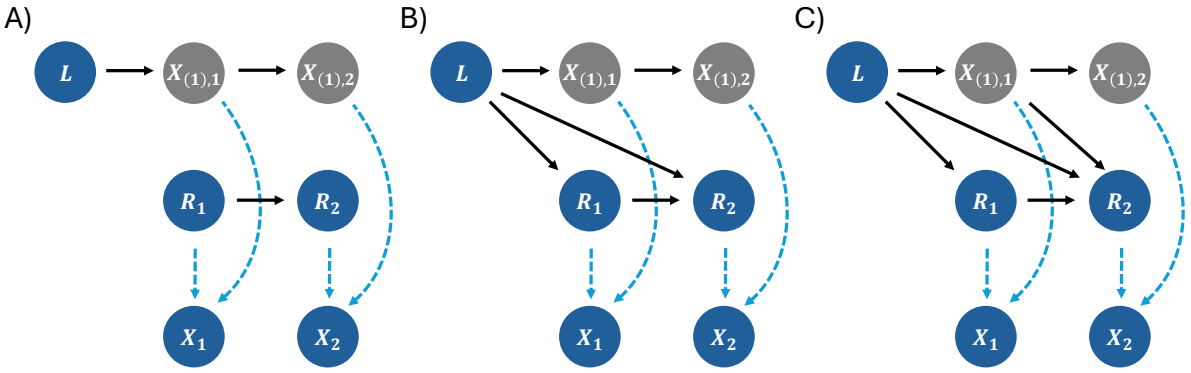


Figure 2.2.: Missing data graphs for a setting with one fully observed covariate L and two partially missing features $X_{(1),1}$ and $X_{(1),2}$. A) Missing-completely-at-random (MCAR): No arrows from data variables (L, $X_{(1),1}$, $X_{(1),2}$) into the missingness indicators. B) Missing-at-random (MAR): Missingness depends on the observed variable L. C) Missing-not-at-random (MNAR): The missingness indicator R_2 depends on the unobserved variable $X_{(1),1}$, violating the MAR assumption. In all graphs, dotted blue arrows represent deterministic relationships imposed by the consistency assumption.

but only when $R_1 = 1$. However, such context-dependent independence assumptions cannot be explicitly represented in standard m-graphs.

- *Missing-not-at-random* (MNAR): All remaining scenarios fall under this category. Here, the missingness may depend on variables that are themselves unobserved. For instance, Figure 2.2C) shows a case where $X_{(1),1} \to R_2$, indicating that the missingness of X_2 depends directly on the (potentially) unobserved value of $X_{(1),1}$.

In our AFAPE paper [52], we illustrate how an m-graph can emerge naturally from a causal graph through the introduction of the *no direct effect (NDE)* assumption [38]. This assumption—commonly implied in missing data problems—states that the act of measuring a variable affects only whether it is observed, not the value it would take. While the NDE assumption is often reasonable, it can be violated in certain domains. For example, in medical settings, an invasive diagnostic test might influence physiological responses such as the heart rate, thereby altering other variables in the system. In such cases, the value that would have been observed if the variable had been measured might differ from the actual, unmeasured value.

2.2. Identification

Once an estimand has been clearly defined, the next step is to determine whether—and under which assumptions—it is possible to estimate this target parameter from observed data. This

step is known as *identification*. Identification typically involves rewriting the estimand—often defined in terms of unobserved or counterfactual quantities—as a functional of the observed data distribution. For example, the counterfactual variable $X_{(1)}$ must not appear directly in the identifying expression, since it is not observed.

2.2.1. Identification Assumptions

Before proceeding, I state the most fundamental assumptions about the structure of the data that are needed for identification. Throughout this work, I assume that individual data points are independent of one another. This assumption is often referred to as the *no interference assumption* [14]:

No Interference Assumption: *The potential outcomes of one individual are not affected by the treatments given to other individuals.*

This means that not only are the observed individual data points statistically independent, but their counterfactual outcomes are as well.

To identify estimands involving potential outcomes, the counterfactual quantities must be connected to the observed data. This is done through the *consistency assumption* [14]:

Consistency Assumption: *If the treatment actually received is $A = a$, then the observed outcome equals the potential outcome under $A = a$:*

$$\text{If } A = a, \text{ then } Y = Y_{(a)}.$$

In the context of missing data, this corresponds to the assumption that $X = X_{(1)}$ when $R = 1$. Together, the no interference and consistency assumptions are often grouped under the *stable unit treatment value assumption (SUTVA)* [41].

However, SUTVA alone is not sufficient for identification. Suppose one aims to identify the counterfactual mean $\theta = \mathbb{E}[Y_{(a)}]$. By the consistency assumption, it is known that $Y = Y_{(a)}$ whenever $A = a$, which might suggest one could estimate θ with the empirical mean $\hat{\mathbb{E}}_n[Y \mid A = a]$ (where $\hat{\mathbb{E}}_n$ denotes the empirical average over a dataset of size n). This strategy works in randomized experiments, where treatment assignment is independent of potential outcomes by design. But in observational settings, treatment decisions may depend on factors that also influence the outcome. In such cases, the treated and untreated groups are not directly comparable, and the conditional mean $\mathbb{E}[Y \mid A = a]$ may not reflect the counterfactual quantity $\mathbb{E}[Y_{(a)}]$.

To address this, an additional assumption is required: *exchangeability*, also known as *ignorability* or *no unmeasured confounding* [14]:

Exchangeability Assumption: *Let L be a set of observed covariates. Then:*

$$Y_{(a)} \perp\!\!\!\perp A \mid L.$$

This assumption states that, conditional on L, the treatment assignment A is independent of the potential outcome $Y_{(a)}$. Intuitively, after adjustment for the relevant covariates, there are no unobserved common causes of both treatment and outcome. In the context of missing data, the same assumption is required: the missingness indicator R must be conditionally independent of the unobserved value $X_{(1)}$, given a set of covariates L.

Graphical models offer a powerful tool to assess the plausibility of exchangeability. In the SWIG shown in Figure 2.1B, $Y_{(a)}$ and A are d-separated given L and thus exchangeability holds. This is because the path $A \leftarrow L \rightarrow Y_{(a)}$ is blocked by L since this corresponds to a chain structure of which the middle node is in the conditioning set.

Finally, for identification to be valid, it must be ensured that all strata defined by the conditioning set L is actually observed in the data. This is guaranteed by the *positivity assumption* [14]:

Positivity Assumption:

$$p(A = a \mid L = \ell) > 0 \quad \text{for all } \ell \text{ such that } p(L = \ell) > 0.$$

Without this condition, certain subgroups would lack sufficient data to estimate conditional expectations—rendering identification impossible.

2.2.2. Identification via the G-formula

With these four assumptions—no interference, consistency, exchangeability, and positivity—it is possible to identify the estimand using the *G-formula* [37, 39]. In my example, the target quantity $\theta = \mathbb{E}[Y_{(a)}]$ can be rewritten as:

$$\begin{aligned}
\theta = \mathbb{E}[Y_{(a)}] &= \sum_l \mathbb{E}[Y_{(a)} \mid L = l]\, p(L = l) \\
&\overset{1*)}{=} \sum_l \mathbb{E}[Y_{(a)} \mid L = l, A = a]\, p(L = l) \\
&\overset{2*)}{=} \sum_l \mathbb{E}[Y \mid L = l, A = a]\, p(L = l). \qquad (2.2)
\end{aligned}$$

Here:

- Step $1*$) uses the *exchangeability assumption*, which allows us to condition on $A = a$ without introducing bias, assuming $Y_{(a)} \perp\!\!\!\perp A \mid L$. To ensure this conditional expectation is well-defined, the *positivity assumption* is also required: one must observe individuals with $A = a$ for all values l of L with positive support.

- Step $2*$) uses the *consistency assumption*, which states that if $A = a$, then the observed outcome Y equals the counterfactual $Y_{(a)}$.

The final expression depends only on observed quantities—namely, the distribution of L and the conditional expectation of Y given $L = l$ and $A = a$. Under these assumptions, the estimand is therefore identified from the observed data.

2.2.3. Reducing the Positivity Assumption

The positivity assumption is especially important—and especially fragile—in real-world applications. It often fails to hold in practice, including in many of the settings considered in this thesis. The assumption becomes particularly unrealistic in high-dimensional contexts or, in my case, when test selection (or more generally, feature acquisition) is highly deterministic.

For example, in medicine, certain diagnostic tests may never be performed for specific patient groups—for instance, an invasive test might not be ordered for patients below a certain risk threshold. In such cases, it is not possible to reliably estimate what would have happened had those tests been conducted, because the necessary counterfactuals are never observed.

Both papers in this thesis confront such violations and propose strategies to mitigate their impact. Broadly, there are two ways to reduce reliance on the positivity assumption:

1. **Leverage alternative modeling assumptions:** In some cases, one can introduce additional assumptions that reduce the need for positivity. For example, if only a subset of covariates $L_1 \subset L$ suffices for exchangeability, the positivity assumption only needs to hold over L_1, which may be more plausible. In our AFAPE paper [52], we adopt a different type of assumption: namely, that feature acquisitions affect what is observed but not the underlying feature values themselves (the *no direct effect (NDE)* assumption). This allows us to identify the counterfactual performance of an AFA system without requiring that every possible acquisition pattern appear in the data.

2. **Redefine the estimand:** Another approach is to revise the question of interest to one that is identifiable under weaker assumptions. For example, rather than estimating counterfactual feature values across the full population, we might restrict attention to subgroups where the test is at least occasionally performed. This idea appears in the feature importance paper [51], where we shift focus from a full data feature importance metric—which relies on strong counterfactual support—to a marginal measure of the value of increasing a feature's measurement frequency slightly.

In the following, I illustrate how each strategy is applied in the context of the two papers.

Leveraging Alternative Assumptions in the AFAPE Problem

In our AFAPE paper [52], we reduce reliance on the positivity assumption by leveraging the no direct effect (NDE) assumption. While the full identification result relies on the richer structure developed in the paper, the basic idea can be illustrated using a simpler toy example. This example shows how the methodological tools introduced earlier—SWIGs, conditional independence, and identification theory—can be combined to support alternative identification strategies when full positivity fails.

I consider the problem of identifying the counterfactual mean $\mathbb{E}[Y_{(a)}]$ using the causal graphs in Figure 2.3. The graphs include two observed covariates, L_1 and L_2, as well as an unobserved confounder U. The goal is to find an adjustment set W such that $Y_{(a)} \perp\!\!\!\perp A \mid W$, thereby satisfying exchangeability.

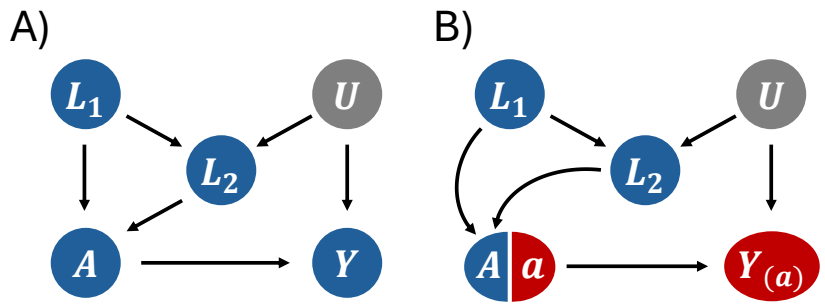


Figure 2.3.: Toy example illustrating positivity reduction via conditional identification. (A) Causal graph of the retrospective distribution. (B) SWIG under an intervention setting $A = a$.

The minimal valid adjustment set among observed variables is $W = \{L_1, L_2\}$. Adjusting only for L_1 leaves the path $A \leftarrow L_2 \leftarrow U \rightarrow Y_{(a)}$ unblocked, while adjusting only for L_2 opens the collider path $A \leftarrow L_1 \rightarrow L_2 \leftarrow U \rightarrow Y_{(a)}$. Although adjusting for U would block all backdoor paths, U is unobserved. These conclusions follow from the d-separation criterion introduced in Section 2.1.2.

Applying the standard identification strategy via the g-formula gives:

$$\begin{aligned}\theta = \mathbb{E}[Y_{(a)}] &= \sum_{l_1,l_2} \mathbb{E}[Y_{(a)} \mid L_1 = l_1, L_2 = l_2]\, p(L_1 = l_1, L_2 = l_2) \\ &\overset{1*)}{=} \sum_{l_1,l_2} \mathbb{E}[Y_{(a)} \mid L_1 = l_1, L_2 = l_2, A = a]\, p(L_1 = l_1, L_2 = l_2) \\ &= \sum_{l_1,l_2} \mathbb{E}[Y \mid L_1 = l_1, L_2 = l_2, A = a]\, p(L_1 = l_1, L_2 = l_2), \end{aligned} \tag{2.3}$$

where step $1*$) uses the exchangeability condition $Y_{(a)} \perp\!\!\!\perp A \mid L_1, L_2$. This identification strategy requires the standard positivity assumption:

Positivity Assumption (Standard):

$$p(A = a \mid L_1 = l_1, L_2 = l_2) > 0 \quad \text{for all } l_1, l_2 \text{ with } p(L_1 = l_1, L_2 = l_2) > 0.$$

This assumption is often too strong to hold in practice. In our paper, we apply a simple identification trick to weaken it. It depends on the specific causal structure in the graph which also applies in this toy problem. From the SWIG, we observe that $Y_{(a)} \perp\!\!\!\perp L_1$, since the path $L_1 \to L_2 \leftarrow U \to Y_{(a)}$ is blocked. This allows us to fix $L_1 = l_1^*$ arbitrarily and identify $\mathbb{E}[Y_{(a)}]$ as follows:

$$\begin{aligned}
\theta = \mathbb{E}[Y_{(a)}] &= \mathbb{E}[Y_{(a)} \mid L_1 = l_1^*] \\
&= \sum_{l_2} \mathbb{E}[Y_{(a)} \mid L_1 = l_1^*, L_2 = l_2]\, p(L_2 = l_2 \mid L_1 = l_1^*) \\
&= \sum_{l_2} \mathbb{E}[Y_{(a)} \mid L_1 = l_1^*, L_2 = l_2, A = a]\, p(L_2 = l_2 \mid L_1 = l_1^*) \\
&= \sum_{l_2} \mathbb{E}[Y \mid L_1 = l_1^*, L_2 = l_2, A = a]\, p(L_2 = l_2 \mid L_1 = l_1^*). \qquad (2.4)
\end{aligned}$$

Since L_1 has been fixed, the required positivity assumption is now considerably weaker:

Positivity Assumption (Reduced):

$$\begin{aligned}
p(A = a \mid L_1 = l_1^*, L_2 = l_2) > 0 \quad &\text{for all } l_2 \text{ with } p(L_2 = l_2 \mid L_1 = l_1^*) > 0, \\
&\text{and for at least one } l_1^* \text{ with } p(L_1 = l_1^*) > 0.
\end{aligned}$$

Changing the Estimand in the Feature Importance Problem

In some cases, the positivity assumption cannot be relaxed without introducing additional assumptions that are implausible or undesirable. If positivity fails entirely, the target parameter is not identified—meaning the data does not contain enough information to answer the research question. In such settings, the analyst must acknowledge this limitation and consider whether an alternative, "answerable" question might still yield useful insight.

This issue arises in our feature importance paper [51]. There, we consider the problem of evaluating a classifier's performance in a hypothetical world where no features are missing. Identifying this estimand requires strong positivity: for each subset of features $X_{(1)}$, there must be a nonzero probability that the full feature vector is observed. In real-world

datasets—especially in clinical or time-series settings—such an assumption is often unrealistic due to systematic sparsity in measurement patterns.

To address this, we propose shifting the estimand. Instead of asking what would happen if we had observed all features, we ask: what would happen if the measurement probability of a given feature were increased slightly? This alternative question still informs the design of acquisition strategies, but can be identified under substantially weaker assumptions.

The resulting metric—the feature measurement importance gradient (FMIG)—is inspired by *incremental propensity score interventions (IPSI)* [17]. IPSI formalizes interventions that slightly shift the propensity of taking an action, rather than enforcing a deterministic treatment assignment $A = a$. By modifying acquisition probabilities via odds ratios rather than fixed values, IPSI avoids positivity violations: if the acquisition probability is zero for a given subgroup, no intervention is made there.

In the case of FMIG, we apply this idea to the acquisition process itself. We estimate the effect of a small increase in the measurement probability of a given feature on predictive performance. Because the intervention does not require full data availability, FMIG remains identifiable even in sparse and selective datasets.

2.3. Estimation

Once identification has been achieved, the next step is estimation—that is, the process of computing an estimate of the target parameter from observed data. In this section, I first review key properties that make an estimator desirable in practice. I then provide a brief overview of semiparametric theory, which serves as the foundation for constructing estimators with these properties. Finally, I present a few simple examples from classical causal inference, which later reappear—often in more complex or adapted forms—in both papers.

2.3.1. Desirable Properties of Estimators

Let $\hat{\theta}_n$ denote an estimator for a target parameter θ, based on a dataset of size n. An estimator maps an observed dataset to a point estimate of the parameter. For instance, a simple estimator for the population mean $\theta = \mathbb{E}[Y]$ is the empirical mean:

$$\hat{\theta}_n = \hat{\mathbb{E}}_n[Y],$$

where $\hat{\mathbb{E}}_n$ denotes the sample average over n observations.

I now outline several desirable properties of estimators that are commonly used to evaluate their performance, both in theory and in practice.

Unbiasedness [7]. The bias of an estimator is defined as $\text{Bias}(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n] - \theta$. An estimator is said to be *unbiased* if its expected value equals the true parameter: $\text{Bias}(\hat{\theta}_n) = 0$. Unbiasedness ensures that, across repeated samples, the estimator does not systematically over- or under-estimate the target.

Unbiasedness is particularly important in safety-critical applications. In our AFAPE paper, we show that biased estimators can lead to severely distorted evaluations of an AFA system's performance. In medical contexts, such distortions could result in deploying AI systems that make suboptimal or even harmful decisions for patients—potentially putting patient outcomes at serious risk.

Consistency [47]. An estimator is *consistent* if it converges in probability to the true value of the parameter as the sample size grows:

$$(\hat{\theta}_n - \theta) \xrightarrow{p} 0.$$

Consistency guarantees that with enough data, the estimator becomes arbitrarily close to the target parameter.

In the healthcare setting, consistency is crucial for the reliable translation of models into real-world practice. For instance, when developing tools to support diagnostic decisions, it is essential that estimators become increasingly accurate as more patient data becomes available.

Efficiency [47]. An estimator is said to be (asymptotically) *efficient* if, in large samples, it achieves the lowest possible variance when estimating the target parameter. In parametric models—where all parts of the data-generating process are specified—a classical result known as the *Cramér–Rao bound* [7] characterizes this minimal variance. The well-known maximum likelihood estimator (MLE), for example, is efficient under appropriate regularity conditions and correct model specification.

Efficiency is particularly important when data is scarce or expensive to collect. In the AFAPE setting, for example, limited dataset sizes make it essential to use estimators that use the available data optimally and thus have a reduced variance.

Robustness. An estimator can be considered more robust if it retains desirable properties—such as consistency—even under weaker or fewer modeling assumptions. Estimators that rely on a reduced positivity assumption, as discussed in Subsection 2.2.3, are one example. Many methods further impose strong parametric assumptions that fully specify parts of the data-generating process. In contrast, more robust estimators avoid such constraints and allow these components to remain unrestricted. Moreover, even when some assumptions

are violated, a robust estimator may still yield reasonable estimates that do not deviate substantially from the true parameter.

In medical applications, robustness is especially desirable, as many assumptions are either untestable or unlikely to hold in the face of complex biological interactions, noisy measurements, and heterogeneous patient behavior. This is particularly important when answering counterfactual questions, which are notoriously difficult to address using retrospective data and require a high degree of trust in the underlying model assumptions.

Each of these properties addresses a different aspect of estimator performance. Achieving all of them simultaneously, however, requires careful modeling choices. In the next section, I introduce the semiparametric theory framework [47], which provides tools for characterizing and navigating trade-offs between estimator properties—such as efficiency and robustness—while maintaining consistency and unbiasedness under appropriate conditions.

2.3.2. Semiparametric Theory

The maximum likelihood estimator (MLE) is an efficient, unbiased, consistent, and often straightforward estimator to derive. It is frequently used for supervised machine learning tasks. Highly parameterized models—with up to billions of parameters in the case of vision-language models [11]—offer enormous flexibility and thus require fewer restrictive modeling assumptions. This flexibility, however, comes at the cost of requiring extremely large datasets. In these settings, performance can be directly evaluated on held-out test data, which helps mitigate concerns about model misspecification.

The estimation problems addressed in this thesis differ in a crucial way. My goal is not to build predictive models for supervised tasks, but to evaluate models under counterfactual conditions where no ground-truth outcomes are observed. In such settings, the standard supervised learning workflow—train on labeled data, evaluate on a test set—no longer applies. Robustness to modeling assumptions thus becomes not just beneficial but essential.

Fully parametric models specify functional forms for every part of the data-generating process, while nonparametric models leave the process entirely unrestricted. Semiparametric models strike a balance: they allow us to encode structure where we have justified knowledge—such as the no direct effect assumption in AFAPE—while leaving other parts of the model flexible [47]. Semiparametric theory provides a principled framework for analyzing such models and for deriving estimators that trade off robustness and statistical efficiency. It therefore forms the theoretical foundation for much of the estimation methodology developed in the AFAPE paper.

I provide a more in-depth review of semiparametric theory in the AFAPE paper, where we develop estimators tailored to that setting. I therefore omit a detailed technical discussion here. Instead, I now present a few simple examples of estimators for the toy problem of estimating

$\theta = \mathbb{E}[Y]$, many of which reappear—often in more complex or adapted forms—throughout both papers.

2.3.3. Examples of Popular Estimators

The following presents three estimators for $\theta = \mathbb{E}[Y]$: the plug-in estimator, the inverse probability weighting (IPW) estimator, and the doubly robust (DR) estimator, which combines elements of the first two.

Plug-in Estimator of the G-formula [14]. A simple and widely used estimator is the plug-in estimator derived from the G-formula. It is constructed by replacing the unknown components of the G-formula with models fitted to the data. This approach is based on the identification result presented in Subsection 2.2.2. In the example, where $\theta = \mathbb{E}[Y_{(a)}] = \sum_l \mathbb{E}[Y \mid L = l, A = a]\, p(L = l)$, the parameter can be estimated by learning a model for the conditional expectation $\hat{\mathbb{E}}[Y \mid L = l, A = a]$ and approximating the distribution $p(L)$. In practice, the empirical distribution of L is typically used instead of modeling $p(L)$ explicitly.

The resulting plug-in estimator is given by:

$$\hat{\theta}_{G_{\text{form}}} \equiv \hat{\mathbb{E}}_n \left[\hat{\mathbb{E}}[Y|L, A = a] \right]$$

Inverse Probability Weighting (IPW) Estimator [14]. The G-formula can also be expressed as:

$$\begin{aligned}
\mathbb{E}[Y_{(a)}] &= \sum_l \mathbb{E}[Y \mid L = l, A = a]\, p(L = l) \\
&= \sum_{y,l,a'} y p(Y = y \mid L = l, A = a)\, \mathbb{I}(a' = a) p(L = l) \\
&= \sum_{y,l,a'} y p(Y = y \mid L = l, A = a)\, \mathbb{I}(a' = a) p(L = l) \cdot \frac{p(A = a'|L = l)}{p(A = a'|L = l)} \\
&= \sum_{y,l,a'} y p(Y = y \mid L = l, A = a')\, p(A = a'|L = l) p(L = l) \cdot \frac{\mathbb{I}(a' = a)}{p(A = a'|L = l)} \\
&= \sum_{y,l,a'} \frac{\mathbb{I}(a' = a) y}{p(A = a'|L = l)} p(Y = y, L = l, A = a') \\
&= \mathbb{E}\left[\frac{\mathbb{I}(A = a) Y}{p(A|L = l)} \right]
\end{aligned}$$

The corresponding IPW estimator is:

$$\hat{\theta}_{IPW} \equiv \hat{\mathbb{E}}_n \left[\frac{\mathbb{I}(A = a) Y}{\hat{p}(A|L)} \right]$$

where $\hat{p}(A|L)$ is the *propensity score model*, estimated from data.

Doubly Robust Estimator [2, 14]. The IPW estimator relies on a model for treatment assignment (propensity scores), while the plug-in estimator depends on an outcome regression model. Semiparametric theory provides a way to combine both into a single estimator that remains consistent if either model is correctly specified.

The doubly robust estimator is given by:

$$\hat{\theta}_{DR} \equiv \hat{\mathbb{E}}_n \left[\hat{\mathbb{E}}[Y|L, A = a] + \frac{\mathbb{I}(A = a)}{\hat{p}(A|L)} (Y - \hat{\mathbb{E}}[Y|L, A = a]) \right]$$

This estimator is termed "doubly robust" because it retains consistency even if only one of the two nuisance models—the propensity score or the outcome regression—is correctly specified.

2.3.4. Nuisance Function Estimation and Cross-fitting

The estimators above require nuisance functions: the propensity score $\hat{p}(A|L)$ and the outcome regression $\hat{\mathbb{E}}[Y|L, A = a]$. These functions must be estimated from data.

To estimate an outcome regression model, all data points for which $A = a$ can be used to regress Y on L. This regression can be performed using simple parametric models or more complex machine learning methods. For the resulting estimators to be unbiased and consistent, however, the model assumptions for these nuisance functions must be satisfied. This requirement introduces an additional assumption that is often not made explicit during the identification stage, where it is typically presumed that nonparametric estimation is feasible.

If flexible machine learning models are used, directly plugging in the estimated nuisance functions into the final estimator may introduce bias. This issue can be addressed using a technique known as *cross-fitting* [9]. In cross-fitting, the dataset is partitioned into folds: one part is used to estimate the nuisance functions, and the other to evaluate the final estimator. This process is then repeated across folds (as in cross-validation), and the results are aggregated. Cross-fitting ensures that desirable asymptotic properties—such as unbiasedness and efficiency—are retained, even when machine learning is used for nuisance estimation.

2.4. Conclusion of the Statistical Pipeline

This chapter has outlined the key components of the statistical pipeline used to address the counterfactual questions posed in the two papers accompanying this thesis. I have

reviewed the three foundational steps of principled statistical analysis—estimand specification, identification, and estimation—and illustrated how they guide the development of rigorous and interpretable methods.

Throughout the chapter, I have used simplified examples to demonstrate the core ideas underlying the more complex methodology employed in the papers. These examples show how precise mathematical notation helps clarify the question of interest and enables the application of appropriate statistical tools. I have highlighted how causal identification techniques make explicit the assumptions needed to estimate counterfactual quantities from observed data. In particular, I have discussed how strong assumptions—such as positivity—can be relaxed, either by leveraging additional structure or by redefining the estimand.

Finally, I have presented basic estimation strategies and shown how concepts like double robustness and cross-fitting allow for flexible, efficient, and reliable inference—even when using modern machine learning tools. Together, these building blocks form the methodological foundation for the two central contributions of this thesis: quantifying feature importance under missing data and evaluating active feature acquisition systems under distribution shift.

3. Publications

In the following, I summarize the main publications of this thesis.

3.1. Feature Importance Metrics in the Presence of Missing Data

This paper [51] addresses the challenge of defining and interpreting feature importance metrics in the presence of missing data—an issue common in domains like healthcare, where features are often selectively measured due to cost, risk, or workflow constraints. The work introduces a conceptual distinction between two perspectives: *full data feature importance*, which assesses a feature's importance had their been no missingness, and *observed data feature importance*, which evaluates importance under the actual measurement policy. These perspectives correspond to different estimands, assumptions, and use cases, yet are frequently conflated in applied work.

To clarify these distinctions, the paper uses the leave-one-covariate-out (LOCO) metric as a case study and demonstrates how each notion of importance requires its own identification strategy and estimation procedure. It highlights how common approaches often fail to specify their estimand and may implicitly rely on strong assumptions about the missingness process and positivity. The paper categorizes a range of estimation methods, mainly focusing on imputation, according to the estimands they target and the assumptions they require.

To complement these two perspectives, the paper introduces the *feature measurement importance gradient (FMIG)*—a novel, model-agnostic metric that quantifies how marginal increases in a feature's measurement frequency would affect downstream predictive performance. FMIG addresses a different question: not how important a feature is under a fixed measurement policy, but how valuable it would be to measure it more often. Crucially, FMIG remains identifiable even in high-missingness regimes and does not rely on strong positivity assumptions, making it particularly relevant for real-world data acquisition settings.

Through a series of motivating examples and synthetic experiments, the paper illustrates how different importance metrics align with different decision-making needs—such as understanding biological mechanisms, training models under existing measurement policies, or guiding the design of future data collection strategies. The proposed framework enables

more rigorous, transparent, and context-aware use of feature importance metrics in settings where missing data is the norm.

Individual Contribution: I developed the main ideas of the paper—including the distinction between full and observed data feature importance, the formulation of FMIG, and the connection to identification theory. I also implemented the methods, carried out the theoretical and empirical analysis, and wrote the majority of the paper. My co-authors provided helpful feedback and supported the refinement of the manuscript through discussion and editorial input.

3.2. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings

This paper [52] introduces and addresses the problem of active feature acquisition performance evaluation (AFAPE) in settings with time-varying features. We formalize the AFAPE problem by specifying an estimand: the expected sum of misclassification costs and feature acquisition costs incurred by a given AFA agent and classifier, estimated from retrospective data. Since the AFA agent would have made different acquisition decisions than those observed in the historical dataset (e.g., collected by physicians), the evaluation task corresponds to a counterfactual inference problem.

We analyze AFAPE under two key assumptions: (i) the no direct effect (NDE) assumption, which posits that feature acquisitions do not affect the underlying feature values; and (ii) the no unobserved confounding (NUC) assumption, which states that acquisition decisions depend only on previously observed features. Under these assumptions, we demonstrate that AFAPE can be mapped to existing statistical frameworks: offline reinforcement learning (under NUC) and missing data analysis (under NDE), each providing tools for identification and estimation.

To overcome the limitations of these individual approaches, we introduce a novel semi-offline reinforcement learning framework that unifies both perspectives. This framework enables more efficient use of retrospective data while significantly relaxing the required positivity assumptions. Within this setting, we develop three new estimators: a direct method (DM), an inverse probability weighting estimator (IPW), and a double reinforcement learning estimator (DRL). The DRL estimator is doubly robust, maintaining consistency even if either the Q-function (a special, sequential version of an outcome regression model) or propensity model is misspecified.

These contributions are further supported by a semiparametric efficiency analysis, which connects the evaluation strategies within a common theoretical framework. Synthetic experiments demonstrate that our estimators outperform standard evaluation heuristics in data

efficiency and robustness, particularly under limited overlap between acquisition policies. Our results highlight the risks of biased evaluation and offer practical solutions for safer deployment of AFA agents.

Individual Contribution: I developed the main ideas of the paper, including the formulation of the AFAPE problem, the integration of missing data and offline RL perspectives, and the design of the semi-offline framework and associated estimators. I also conducted the theoretical analysis, including the application of semiparametric theory to unify the evaluation settings, and implemented the methods and experiments. Co-authors contributed helpful feedback and supported the refinement of the manuscript through discussion and editorial input.

4. Discussion and Outlook

In this thesis, I addressed two central challenges in the development and deployment of AFA systems: understanding when AFA methods are applicable, and evaluating them reliably before deployment.

4.1. Summary of Contributions

My first contribution focuses on assessing feature importance under missing data. The ability to evaluate the value of a feature in the presence of selective measurement is argued to be a crucial first step in identifying settings where AFA methods are both applicable and impactful. A distinction was introduced between *full data* and *observed data* feature importance metrics, clarifying the assumptions required to define and estimate each. To support actionable decision-making, a novel, model-agnostic metric—the *feature measurement importance gradient (FMIG)*—was proposed to guide measurement policies by prioritizing features whose more frequent acquisition would most improve predictive performance.

The second contribution addresses the challenge of evaluating AFA agents using retrospective data. Reliable evaluation is especially critical in high-stakes domains such as healthcare, where flawed assessments prior to deployment can lead to harmful outcomes. To formalize this challenge, the *active feature acquisition performance evaluation (AFAPE)* problem was introduced and analyzed under a range of assumptions. This led to the development of a *semi-offline reinforcement learning* framework that unifies perspectives from missing data and offline RL. Within this framework, new estimators were derived that offer substantially improved data efficiency and require weaker positivity assumptions than existing approaches—enabling more reliable and generalizable evaluation procedures.

4.2. Future Directions

The contributions of this thesis represent an initial step toward making AFA systems viable for real-world deployment. Nonetheless, several important directions remain for future research to fully realize this goal.

4.2.1. Extending AFA Performance Evaluation

While this work included an extensive discussion on the evaluation of active feature acquisition methods, real-world settings can be much more diverse. Future research must broaden the AFAPE framework to accommodate a wider range of data structures, decision contexts, and domain-specific constraints.

Static Feature Settings and Complex Missingness. The AFAPE framework developed in our paper [52] assumes time-varying features, where an agent selects which features to acquire at future time steps. In contrast, static feature settings—where features are fixed—allow agents to defer expensive acquisitions and strategically explore alternatives, making the order of acquisition central to the evaluation problem. In a separate, not-yet-published manuscript [49], I, together with co-authors, analyze AFAPE in this static setting, assuming the acquisition order is unobserved in the retrospective dataset. It is shown that under the no direct effect (NDE) assumption, the AFAPE estimand remains identifiable. The static setting also gives rise to more complex missingness mechanisms, such as the *randomized monotone missingness (RMM)* model [40], which assumes that each acquisition decision may depend on all previously acquired features. While such mechanisms are highly relevant in practice—for instance, when clinicians decide which test to run next based on earlier results—they are particularly challenging to model when the order of acquisition is not observed. Future work should analyze AFAPE under missingness assumptions like the RMM model to further extend its theoretical foundations and improve its practical applicability.

AFA for Treatment Assignment. While this thesis focused on using acquired information for diagnostic prediction, future applications may involve treatment decisions based directly on the acquired features. Extending AFA to such settings requires accounting for the causal effects of both feature acquisition and the treatments that follow. Some prior work has explored these questions [38, 6, 24, 30, 19], but adapting the semi-offline RL framework to model treatment effects remains an open and promising direction.

Toward Fully Efficient Estimators. The estimators developed in the AFAPE paper improve efficiency but are not fully efficient. While nearly efficient estimators—such as those proposed by Liu et al. [24]—can be adapted to the AFA setting discussed here, they rely on strong a positivity assumption. Our paper showed that this assumption can be relaxed to enhance robustness [52], but the construction of fully semiparametrically efficient estimators under the weaker positivity assumption remains an open problem.

Multimodal Data Integration. This thesis focuses on static and time-series tabular datasets, which already capture many important aspects of real-world decision-making problems. However, many practical applications—particularly in healthcare—involve multimodal data sources, such as free-text clinical notes, medical imaging (e.g., MRI or CT scans), and genomic or biochemical profiles. While the theoretical framework developed in our paper applies broadly to such settings, adapting AFA methods to high-dimensional, heterogeneous data will require integrating more expressive models—such as large-scale transformers [48] or domain-specific neural architectures—for tasks like training nuisance models, classifiers, and agents. These models can be plugged into the existing framework, but care must be taken to ensure they behave reliably in counterfactual settings. Applying AFA to large-scale multimodal datasets is thus a promising next step for extending the practical impact of this work.

Assumptions for Identifiability and Efficiency. Static feature settings, structured missingness, and multimodal data are examples of assumptions that affect identifiability and evaluation in AFAPE. Relaxing assumptions can increase robustness and widen applicability, while strengthening them—when justified—can significantly improve statistical efficiency. Below we highlight several additional assumptions or modeling considerations that warrant further investigation:

- **Assumptions about the data distribution**: While the AFAPE paper focused on assumptions about the acquisition process, it did not impose structural constraints on the underlying data-generating process of the features and the label. Introducing such assumptions can improve estimation efficiency [47]. One example is smoothness or continuity in feature evolution—particularly relevant in time-series settings—where it may be reasonable to assume that feature values do not change abruptly over short time intervals, as is often the case in clinical or physiological data.
- **Markov assumptions**: In offline RL, it is common to assume that the next state is conditionally independent of prior states given the current state and action [20]. Similar or slightly weaker assumptions may apply to both the data and acquisition processes in AFA. For example, one may assume that the patient's condition from a month ago is no longer relevant for the progression of a common cold, given their current symptoms and test results. When appropriate, Markov assumptions can substantially reduce model complexity and improve estimation efficiency [16].
- **No interference and budget constraints**: Throughout this thesis, it was assumed that one individual's acquisition decisions and outcomes do not affect those of others. In practice, however, interference can arise due to shared resources—for example, limited diagnostic equipment or physician availability. Budget constraints can further induce cross-individual dependencies, where the agent must operate under global resource limitations. Relaxing the no-interference assumption and formalizing these broader constraints is essential for application in a wider range of real-world settings.

- **Measurement error**: Feature values may be noisy due to biological variability or measurement imprecision. In such cases, the AFA agent might benefit from re-acquiring the same feature to increase confidence. Explicitly modeling measurement error could improve both policy quality and evaluation accuracy.
- **Time delays**: In many domains, especially healthcare, test results are not available immediately after testing. Modeling these delays—e.g., hours or days between test order and result—can influence both optimal acquisition timing and how agent performance is evaluated retrospectively.

Analyzing how these assumptions affect identification and efficiency—and determining when they can be plausibly justified in practice—offers a promising direction for both theoretical development and practical deployment of AFA systems.

4.2.2. Training Optimal AFA Agents

Realizing the full potential of AFA requires not only reliable evaluation but also effective training of agents. Future research should explore learning algorithms specifically adapted to the semi-offline RL framework proposed in this thesis, as well as the integration of foundation models into AFA systems.

Policy Learning in the Semi-Offline RL Setting. One route to training optimal AFA agents could involve extending approaches from the dynamic treatment regimes literature, such as marginal structural models (MSMs) [14] and structural nested mean models (SNMMs) [14], to the AFA setting. In this discussion, however, I adopt an alternative RL perspective.

Several families of RL algorithms are promising candidates for extension to the semi-offline AFA setting:

- **Policy gradient methods:** These methods optimize a parameterized policy by estimating the gradient of the expected loss and updating the policy parameters accordingly. Since AFAPE provides estimators of the expected loss, policy gradient approaches are a natural fit for this setting.
- **Actor-critic algorithms:** These methods train two components: an actor that selects actions and a critic that evaluates them. A natural extension in our setting is to reuse the Q-function from the semi-offline evaluation framework as the critic.
- **Q-learning:** Adapting Q-learning to semi-offline RL would require extending the semi-offline version of the Bellman equation used in AFAPE to a Bellman optimality equation [20] for policy improvement. This direction is promising but more complex: the semi-offline Q-function conditions on features that may not be observable to the agent at decision time.

- **Model-based approaches:** These are especially promising in the AFA setting because, under the no direct effect (NDE) assumption, the transition model reduces to an imputation model—making it easier to learn from data. Planning and simulation based on these imputation models can then support policy optimization.

As shown in the AFAPE paper [52], semi-offline RL occupies an intermediate position between offline and online RL. While offline RL relies entirely on retrospective data and online RL assumes a simulatable environment for exploration, the semi-offline setting allows simulating some actions—those with observed data—while others remain unobservable. These constraints introduce new challenges for adapting RL algorithms. In offline RL, regularization is often used to enforce the positivity assumption by discouraging policies from selecting unsupported actions. In the semi-offline setting, where the positivity assumption is relaxed, regularization strategies must be rethought. At the same time, techniques from online RL may also prove useful. Online RL typically assumes access to a full simulator and focuses on balancing exploration and exploitation. The semi-offline RL approach also uses a sampling policy, raising analogous challenges. Incorporating this sampling policy into learning algorithms and adapting exploration strategies accordingly is a promising direction for future work.

Foundation Models for AFA. Foundation models—such as large language models (LLMs) like GPT-4 [31]—have demonstrated notable generalization capabilities across diverse tasks, particularly in natural language and vision domains. Trained on extensive and varied datasets, these models have achieved state-of-the-art performance in areas including healthcare, supporting applications like medical question answering, clinical documentation, and decision support [53]. Their ability to integrate contextual knowledge and generalize to new tasks with minimal fine-tuning underscores their potential for adaptability [5], positioning them as valuable tools for downstream decision-making.

Given their flexibility and reasoning abilities, one might consider applying a foundation model directly—either as an AFA agent or as a combined AFA agent and classifier. For instance, an LLM could be prompted with a partial feature set to recommend the next most informative test, potentially achieving competitive performance without task-specific training. Such a model could leverage domain knowledge acquired during pretraining, adhere to medical heuristics consistent with clinical reasoning, and balance competing objectives (e.g., recommending tests beneficial for both primary and secondary diagnoses). Additionally, constraints like resource availability or equipment load could be incorporated at inference time through natural language instructions.

However, deploying foundation models in this context presents challenges. First, AFA represents an unseen task for most foundation models, making their reliability under real-world constraints uncertain. This is where the AFAPE framework becomes instrumental: it is designed to evaluate the performance of *any* agent—including foundation models—using

retrospective data. This enables principled assessment of a language model's decision-making capabilities, thereby supporting greater trust and accountability.

Second, while foundation models exhibit impressive generalization, their strongest performance often arises when fine-tuned on tasks closely aligned with their downstream objectives. Since counterfactual reasoning is typically not part of a foundation model's pretraining, fine-tuning these models specifically for AFA tasks or integrating them into hybrid systems with components explicitly trained for counterfactual inference may enhance their effectiveness.

Moreover, while LLMs can process multiple modalities, they often rely on text representations, which may not fully capture the inherent structure of structured data. This reliance can lead to challenges in accurately interpreting and reasoning over such data formats. Studies have highlighted that LLMs may struggle with tasks requiring deep understanding of structured data due to their training predominantly on unstructured text [15, 33].

Overall, integrating foundation models into AFA workflows represents a promising yet underexplored avenue. Future research should investigate methods to adapt or guide these models, compare their outputs to trained AFA agents, and employ the evaluation techniques presented in this thesis to ensure rigorous evaluation.

4.2.3. Translating AFA into Practice

While the core methods developed in this thesis are broadly applicable, translating AFA systems into real-world use always involves domain-specific challenges. In what follows, I focus on deployment in healthcare settings.

Successfully deploying AFA systems in medical environments requires more than algorithmic excellence. It involves navigating sociotechnical complexities, supporting human–machine collaboration, and confronting ethical trade-offs that arise in real-world decision-making. This translation demands not only rigorous evaluation but also thoughtful integration into workflows and values.

Doctor–AI Collaboration. Translating AFA into real-world medical workflows requires more translational research. The assumptions underpinning most AFA models—such as the missingness mechanism or the lack of interference across individuals—are often untestable and may be violated in practice. For example, a test recommended by an AFA model might not be feasible for a particular patient if the same equipment is simultaneously needed elsewhere. In such settings, real-world deployment requires careful rollout strategies to monitor and validate performance under these practical constraints.

Moreover, doctors operate within broader diagnostic and operational contexts that extend beyond the narrow objective modeled by an AFA agent. A recommended test may serve

multiple diagnostic purposes. For instance, an AFA model predicting sepsis in the ICU may suggest a blood test for that specific task, but the same test might also be critical for monitoring kidney function or medication dosage. These multi-objective uses are difficult to encode fully in the AFA formulation and must instead be accounted for by the clinician. Educating clinicians about what an AFA tool is and is not designed to do—its assumptions, scope, and limitations—is therefore essential.

Collaboration frameworks should therefore be designed to support physicians rather than replace them. Prior research has shown that human–AI teams often outperform either alone in diagnostic settings [35]—a dynamic that must be carefully studied for AFA. This includes not only performance but also clinicians' trust in and willingness to adopt such tools, which may depend on interpretability, perceived autonomy, and alignment with clinical goals.

Ethical Considerations. At the heart of the AFA problem lies a trade-off between the cost of acquiring information and the risk of making incorrect decisions with incomplete data. The AFAPE paper formalizes the problem as balancing feature acquisition costs against misclassification costs. While some of these costs are monetary—such as the price of an MRI scan or the downstream healthcare expenses from a misdiagnosis—others are more subtle or intangible. These include patient discomfort, psychological burden, delays in treatment, and, in the most serious cases, physical harm or even death.

Determining how to balance such costs is ultimately an ethical question, one that cannot—and should not—be resolved solely through algorithmic optimization. It demands broader societal deliberation involving patients, practitioners, ethicists, and policymakers. Algorithms may provide transparency in how trade-offs are computed, but not in how they should be valued.

By making decision processes mathematically explicit, AFA systems surface ethical dilemmas that were previously handled informally or subconsciously by clinicians. This formalization offers an opportunity—not to dictate answers—but to frame important questions more clearly for societal and clinical discourse.

List of Figures

Bibliography

[1] An, Q., Li, H., Liao, X., and Carin, L. Active feature acquisition with POMDP models. *Submitted to Pattern Recognition Letters*, 2006.

[2] Bang, H. and Robins, J. M. Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, 61(4):962–973, 2005.

[3] Bishop, T. F., Federman, A. D., and Keyhani, S. Physicians' views on defensive medicine: a national survey. *Archives of Internal Medicine*, 170(12):1081–1083, 2010.

[4] Bishop, T. F., Federman, A. D., and Ross, J. S. Laboratory Test Ordering at Physician Offices With and Without On-Site Laboratories. *Journal of general internal medicine*, 25(10): 1057–1063, 2010.

[5] Budnikov, M., Bykova, A., and Yamshchikov, I. P. Generalization potential of large language models. *Neural Computing and Applications*, 37(4):1973–1997, 2025.

[6] Caniglia, E. C., Robins, J. M., Cain, L. E., Sabin, C., Logan, R., Abgrall, S., Mugavero, M. J., Hernández-Díaz, S., Meyer, L., Seng, R., Drozd, D. R., Seage III, G. R., Bonnet, F., Le Marec, F., Moore, R. D., Reiss, P., van Sighem, A., Mathews, W. C., Jarrín, I., Alejos, B., Deeks, S. G., Muga, R., Boswell, S. L., Ferrer, E., Eron, J. J., Gill, J., Pacheco, A., Grinsztejn, B., Napravnik, S., Jose, S., Phillips, A., Justice, A., Tate, J., Bucher, H. C., Egger, M., Furrer, H., Miro, J. M., Casabona, J., Porter, K., Touloumi, G., Crane, H., Costagliola, D., Saag, M., and Hernán, M. A. Emulating a trial of joint dynamic strategies: An application to monitoring and treatment of HIV-positive individuals. *Statistics in Medicine*, 38(13): 2428–2446, 2019.

[7] Casella, G. and Berger, R. *Statistical inference*. CRC press, 2024.

[8] Chang, C.-H., Mai, M., and Goldenberg, A. Dynamic Measurement Scheduling for Event Forecasting using Deep RL. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 951–960. PMLR, 2019.

[9] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

[10] Connolly, T. and Thorn, B. K. Predecisional information acquisition: Effects of task variables on suboptimal search strategies. *Organizational Behavior and Human Decision Processes*, 39(3):397–416, 1987.

[11] Dehghani, M., Djolonga, J., Mustafa, B., Padlewski, P., Heek, J., Gilmer, J., Steiner, A., Caron, M., Geirhos, R., Alabdulmohsin, I., Jenatton, R., Beyer, L., Tschannen, M., Arnab, A., Wang, X., Riquelme, C., Minderer, M., Puigcerver, J., Evci, U., Kumar, M., Steenkiste, S. v., Elsayed, G. F., Mahendran, A., Yu, F., Oliver, A., Huot, F., Bastings, J., Collier, M. P., Gritsenko, A., Birodkar, V., Vasconcelos, C., Tay, Y., Mensink, T., Kolesnikov, A., Pavetić, F., Tran, D., Kipf, T., Lučić, M., Zhai, X., Keysers, D., Harmsen, J., and Houlsby, N. Scaling Vision Transformers to 22 Billion Parameters. *arXiv:2302.05442 [cs]*, 2023.

[12] Desmarais, S. L., Zottola, S. A., Duhart Clarke, S. E., and Lowder, E. M. Predictive Validity of Pretrial Risk Assessments: A Systematic Review of the Literature. *Criminal Justice and Behavior*, 48(4):398–420, 2021.

[13] Feldman, L. Managing the cost of diagnosis. *Managed care (Langhorne, Pa.)*, 18(5):43–45, 2009.

[14] Hernán, M. A. and Robins, J. M. *Causal Inference: What If*. CRC Boca Raton, FL, 2020.

[15] Huang, S., Gu, Y., Hu, X., Li, Z., Li, Q., and Xu, G. Reasoning Factual Knowledge in Structured Data with Large Language Models. *arXiv:2408.12188 [cs]*, 2024.

[16] Kallus, N. and Uehara, M. Double Reinforcement Learning for Efficient Off-Policy Evaluation in Markov Decision Processes. *Journal of Machine Learning Research*, 21(167): 1–63, 2020.

[17] Kennedy, E. H. Nonparametric Causal Effects Based on Incremental Propensity Score Interventions. *Journal of the American Statistical Association*, 114(526):645–656, 2019.

[18] Krahn, M. D., Mahoney, J. E., Eckman, M. H., Trachtenberg, J., Pauker, S. G., and Detsky, A. S. Screening for Prostate Cancer: A Decision Analytic View. *JAMA*, 272(10):773–780, 1994.

[19] Kreif, N., Sofrygin, O., Schmittdiel, J. A., Adams, A. S., Grant, R. W., Zhu, Z., van der Laan, M. J., and Neugebauer, R. Exploiting nonsystematic covariate monitoring to broaden the scope of evidence about the causal effects of adaptive treatment strategies. *Biometrics*, 77(1):329–342, 2021.

[20] Levine, S., Kumar, A., Tucker, G., and Fu, J. Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems. *arXiv:2005.01643 [cs, stat]*, 2020.

[21] Li, Y. and Oliva, J. Active Feature Acquisition with Generative Surrogate Models. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 6450–6459. PMLR, 2021.

[22] Li, Y., Shan, S., Liu, Q., and Oliva, J. B. Towards Robust Active Feature Acquisition. *arXiv:2107.04163 [cs]*, 2021.

[23] Little, R. J. A. and Rubin, D. B. *Statistical analysis with missing data*. Wiley series in probability and statistics. Wiley, third edition edition, 2020.

[24] Liu, L., Shahn, Z., Robins, J. M., and Rotnitzky, A. Efficient Estimation of Optimal Regimes Under a No Direct Effect Assumption. *Journal of the American Statistical Association*, 116(533):224–239, 2021.

[25] Maćkowiak, B., Matějka, F., and Wiederholt, M. Rational Inattention: A Review. *Journal of Economic Literature*, 61(1):226–73, 2023.

[26] Mohan, K., Pearl, J., and Tian, J. Graphical Models for Inference with Missing Data. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

[27] Mushlin, A. I. and Fintor, L. Is screening for breast cancer cost-effective? *Cancer*, 69(S7): 1957–1962, 1992.

[28] Nabi, R., Bhattacharya, R., and Shpitser, I. Full law identification in graphical models of missing data: Completeness results. In *International Conference on Machine Learning*, pp. 7153–7163. PMLR, 2020.

[29] Nentwich, C. and Reinhart, G. Towards Data Acquisition for Predictive Maintenance of Industrial Robots. *Procedia CIRP*, 104:62–67, 2021.

[30] Neugebauer, R., Schmittdiel, J. A., Adams, A. S., Grant, R. W., and van der Laan, M. J. Identification of the Joint Effect of a Dynamic Treatment Intervention and a Stochastic Monitoring Intervention Under the No Direct Effect Assumption. *Journal of Causal Inference*, 5(1):20160015, 2017.

[31] OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.-L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R., Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H. W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus, L., Felix, N., Fishman, S. P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafstein, M., Gray, S., Greene, R., Gross, J., Gu, S. S., Guo, Y., Hallacy, C., Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S., Jain, S., Jang, J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B., Jun, H., Kaftan, T., Kaiser, , Kamali, A., Kanitscheider, I., Keskar, N. S., Khan, T., Kilpatrick, L., Kim, J. W., Kim, C., Kim, Y., Kirchner, J. H., Kiros, J., Knight, M., Kokotajlo, D., Kondraciuk, , Kondrich, A., Konstantinidis, A., Kosic, K., Krueger, G., Kuo, V., Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C. M., Lim, R., Lin, M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A., Malfacini, K., Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne, A., McGrew, B., McKinney, S. M., McLeavey, C., McMillan,

P., McNeil, J., Medina, D., Mehta, A., Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa, E., Mossing, D., Mu, T., Murati, M., Murk, O., Mély, D., Nair, A., Nakano, R., Nayak, R., Neelakantan, A., Ngo, R., Noh, H., Ouyang, L., O'Keefe, C., Pachocki, J., Paino, A., Palermo, J., Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos, A., Pavlov, M., Peng, A., Perelman, A., Peres, F. d. A. B., Petrov, M., Pinto, H. P. d. O., Michael, Pokorny, Pokrass, M., Pong, V. H., Powell, T., Power, A., Power, B., Proehl, E., Puri, R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach, K., Ross, C., Rotsted, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar, S., Sastry, G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K., Sherbakov, T., Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M., Sitkin, J., Slama, K., Sohl, I., Sokolowsky, B., Song, Y., Staudacher, N., Such, F. P., Summers, N., Sutskever, I., Tang, J., Tezak, N., Thompson, M. B., Tillet, P., Tootoonchian, A., Tseng, E., Tuggle, P., Turley, N., Tworek, J., Uribe, J. F. C., Vallone, A., Vijayvergiya, A., Voss, C., Wainwright, C., Wang, J. J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann, C. J., Welihinda, A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter, C., Wolrich, S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T., Yoo, S., Yu, K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S., Zheng, T., Zhuang, J., Zhuk, W., and Zoph, B. GPT-4 Technical Report, 2024.

[32] O'Sullivan, J. W., Albasri, A., Nicholson, B. D., Perera, R., Aronson, J. K., Roberts, N., and Heneghan, C. Overtesting and undertesting in primary care: a systematic review and meta-analysis. *BMJ open*, 8(2):e018557, 2018.

[33] Pang, C., Cao, Y., Yang, C., and Luo, P. Uncovering Limitations of Large Language Models in Information Seeking from Tables. In Ku, L.-W., Martins, A., and Srikumar, V. (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 1388–1409, Bangkok, Thailand, 2024. Association for Computational Linguistics.

[34] Pearl, J. *Causality*. Cambridge university press, 2009.

[35] Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., and Cherubini, A. Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific Reports*, 12(1):14952, 2022.

[36] Richardson, T. S. and Robins, J. M. Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality. 2013.

[37] Robins, J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.

[38] Robins, J., Orellana, L., and Rotnitzky, A. Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine*, 27(23):4678–4721, 2008.

[39] Robins, J. M. Causal Inference from Complex Longitudinal Data. In Bickel, P., Diggle, P., Fienberg, S., Krickeberg, K., Olkin, I., Wermuth, N., Zeger, S., and Berkane, M. (eds.),

Latent Variable Modeling and Applications to Causality, volume 120, pp. 69–117. Springer New York, 1997.

[40] Robins, J. M. and Gill, R. D. Non-Response Models for the Analysis of Non-Monotone Ignorable Missing Data. *Statistics in Medicine*, 16(1):39–56, 1997.

[41] Rubin, D. B. Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593, 1980.

[42] Rubin, D. B. Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[43] Schiff, G. D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B. L., Elstein, A. S., Hasler, S., Kabongo, M. L., and Krosnjar, N. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Archives of internal medicine*, 169(20):1881–1887, 2009.

[44] Schnabel, T., Bennett, P. N., Dumais, S. T., and Joachims, T. Using Shortlists to Support Decision Making and Improve Recommender System Performance. *arXiv:1510.07545 [cs]*, 2016.

[45] Shim, H., Hwang, S. J., and Yang, E. Joint Active Feature Acquisition and Classification with Variable-Size Set Encoding. *Advances in Neural Information Processing Systems*, 31, 2018.

[46] Shpitser, I., Mohan, K., and Pearl, J. Missing data as a causal and probabilistic problem. Technical report, 2015.

[47] Tsiatis, A. A. *Semiparametric theory and missing data*. Springer series in statistics. Springer, 2006.

[48] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, u., and Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[49] von Kleist, H., Zamanian, A., Shpitser, I., and Ahmidi, N. Evaluation of Active Feature Acquisition Methods for Static Feature Settings. *arXiv:2312.03619 [cs, stat]*, 2023.

[50] von Kleist, H., Zamanian, A., Shpitser, I., and Ahmidi, N. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings. *arXiv:2312.01530 [cs, stat]*, 2023.

[51] von Kleist, H., Wendland, J., Shpitser, I., and Marr, C. Feature Importance Metrics in the Presence of Missing Data. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267. PMLR, 2025.

[52] von Kleist, H., Zamanian, A., Shpitser, I., and Ahmidi, N. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings. *Journal of Machine Learning Research*, 26(60):1–84, 2025.

[53] Wang, D. and Zhang, S. Large language models in medical and healthcare fields: applications, advances, and challenges. *Artificial Intelligence Review*, 57(11):299, 2024.

[54] Yin, H., Li, Y., Pan, S. J., Zhang, C., and Tschiatschek, S. Reinforcement Learning with Efficient Active Feature Acquisition. *arXiv:2011.00825 [cs]*, 2020.

A. Original Publications

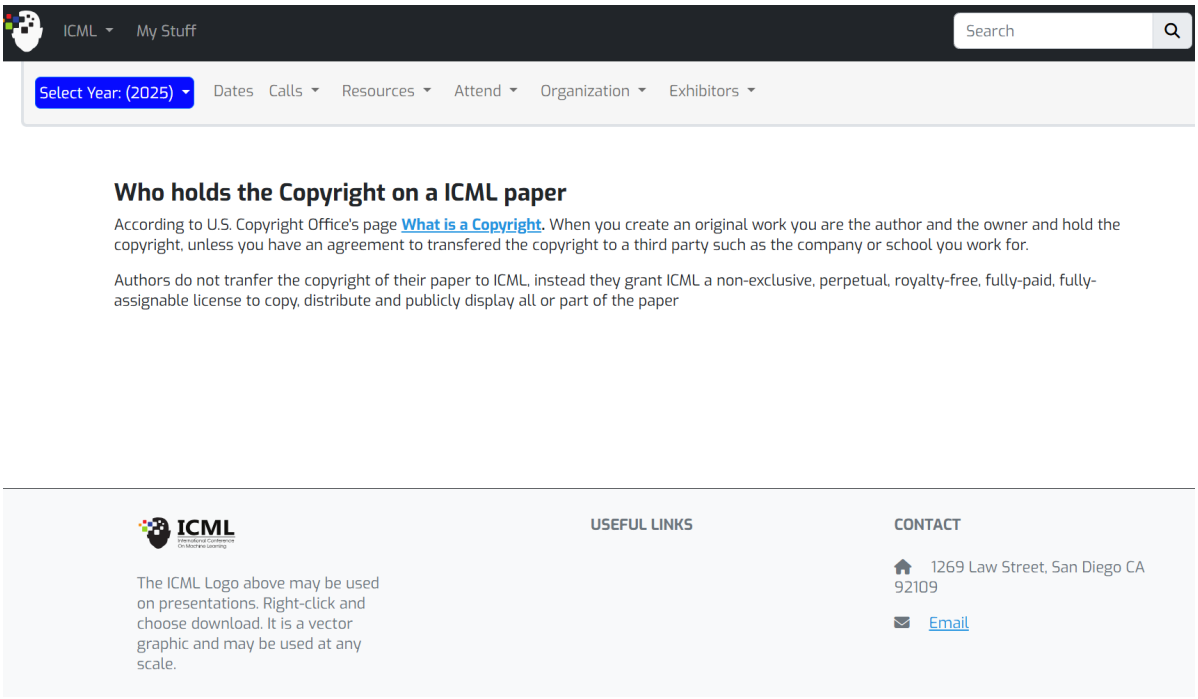
A.1. Feature Importance Metrics in the Presence of Missing Data

This publication was accepted for presentation at the *42nd International Conference on Machine Learning (ICML)*, to be held in Vancouver, Canada, in 2025. It will appear in the *Proceedings of Machine Learning Research (PMLR)*, volume 267. A one-page summary is provided in Section 3.1 of this thesis.

Full citation:
von Kleist, H., Wendland, J., Shpitser, I., and Marr, C. Feature Importance Metrics in the Presence of Missing Data. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267. PMLR, 2025. Accepted for publication. To appear.

License and Reprint Permission:
According to the official ICML copyright policy, authors retain the copyright to their papers. They grant ICML only a non-exclusive, perpetual, royalty-free license to distribute and display the work. Therefore, authors retain the right to reuse their work in dissertations and other non-commercial contexts.

The relevant policy is available on the ICML website at: `https://icml.cc/FAQ/Copyright`. A screenshot is shown below.



Official ICML copyright policy stating that authors retain copyright. This screenshot serves as reprint permission proof.

Feature Importance Metrics in the Presence of Missing Data

Henrik von Kleist [1 2 3] **Joshua Wendland** [4] **Ilya Shpitser** [3] **Carsten Marr** [1]

Abstract

Feature importance metrics are critical for interpreting machine learning models and understanding the relevance of individual features. However, real-world data often exhibit missingness, thereby complicating how feature importance should be evaluated. We introduce the distinction between two evaluation frameworks under missing data: (1) feature importance under the full data, as if every feature had been fully measured, and (2) feature importance under the observed data, where missingness is governed by the current measurement policy. While the full data perspective offers insights into the data generating process, it often relies on unrealistic assumptions and cannot guide decisions when missingness persists at model deployment. Since neither framework directly informs improvements in data collection, we additionally introduce the feature measurement importance gradient (FMIG), a novel, model-agnostic metric that identifies features that should be measured more frequently to enhance predictive performance. Using synthetic data, we illustrate key differences between these metrics and the risks of conflating them.

1. Introduction

Feature importance metrics play a crucial role in supervised machine learning, offering a means to evaluate the impact of individual features on a model's performance. A subset of these metrics, known as global feature importance metrics (Ewald et al., 2024), determine the value of including a feature in the prediction model, typically given a certain baseline of other features. These metrics help clarify the relevance of features in applications like medical decision-making, where they can provide insights into the diagnostic and prognostic value of medical tests (Yoo et al., 2025; Rai et al., 2024; Ma et al., 2023). In practice, however, particularly in clinical contexts, features are often measured inconsistently, leading to missing values (Zamanian et al., 2024). This raises important questions about how to define and interpret feature importance in such scenarios.

To ground our discussion, we present three simplified scenarios in heart attack diagnosis, each illustrating a distinct perspective on how to reason about feature importance in the presence of missing data:

Scenario 1: A biomedical researcher aims to understand the intrinsic relationship between troponin levels, a biomarker typically measured via a blood test, and the occurrence of heart attacks - independent of hospital testing protocols. Using a retrospective dataset with missing values, the researcher is interested in feature importance as if every feature had been fully measured.

Scenario 2: A machine learning engineer develops a heart attack prediction model and evaluates the relevance of troponin levels. Given a retrospective dataset with missingness, the model is to be trained for deployment in a setting where such missingness will persist due to unchanged hospital testing protocols. The engineer is thus interested in **feature importance under the current missingness** where features sometimes assume a special value "?" representing missingness.

Scenario 3: Since troponin tests are costly and time-consuming, but provide valuable diagnostic information, an analyst seeks to optimize their use after deploying the prediction model. The analyst wants to understand how prediction performance would change if troponin was measured more frequently. They are thus interested in **feature importance under an increase in measurement probability**.

In this work, we examine how feature importance metrics can be defined and estimated in such scenarios. To clarify the structure of the problem, we organize our analysis

[1]Institute of AI for Health, Helmholtz Munich - German Research Center for Environmental Health, Neuherberg, Germany [2]TUM School of Computation, Information and Technology, Technical University of Munich, Garching, Germany [3] Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA [4] Faculty of Computer Science, Ruhr University Bochum Bochum, Germany . Correspondence to: Henrik von Kleist <henrik.kleist@tum.de>.

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

around the classical three-step framework of statistical inference: (i) **Estimand definition**, which entails specifying the target parameter to be computed; (ii) **Identification**, which assesses whether the estimand can be expressed as a function of the observed data; and (iii) **Estimation**, which concerns how the estimand is calculated from the observed data.

We find that current practices for reporting feature importance metrics in scenarios with missing data lack rigor across all three steps:

- ***Estimands are not clearly specified:*** When transitioning from a fully observed dataset to one with missing data, the target parameter - i.e. the feature importance metric - must be redefined to reflect the data structure. In current practice, analysts do not report based on which data the feature importance is calculated. For example, is the metric intended to reflect the importance of features in the observed data with missingness, or in the full data that would have been available had there been no missingness?
- ***Identification assumptions are not stated, and identification is not performed:*** Analysts often fail to specify the necessary assumptions required for identification. This includes independence assumptions about the missingness process, such as whether data are assumed to be Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR) (Bhattacharya et al., 2020; Nabi et al., 2020; Mohan & Pearl, 2021). It also includes the positivity assumption (Petersen et al., 2012), which ensures sufficient overlap in feature availability across subpopulations. As a result, it remains unclear under which assumptions the reported feature importance metrics hold true. Examples of this issue can be found in Yoo et al. (2025); Rai et al. (2024); Ma et al. (2023); Beld et al. (2024); Peng et al. (2023); Guan et al. (2023); Zhou et al. (2023); Qi et al. (2022); Xie et al. (2023); Bao et al. (2023); Yang et al. (2022); Zhang et al. (2023).
- ***Estimation methods are often biased:*** A variety of methods are used in practice to handle missing data when computing feature importance metrics, including but not limited to mean/median/zero imputation (Zhuang et al., 2023; Lucini et al., 2023; Ishii et al., 2023; Huang et al., 2023; Danilatou et al., 2022), conditional mean imputation (Vo et al., 2024), complete case analysis (Zhang et al., 2022), and multiple imputation (Peng et al., 2023; Guan et al., 2023; Qi et al., 2022; Xie et al., 2023; Chen et al., 2023; Zeng et al., 2023). It is often unclear to which estimand these methods correspond to and whether they introduce bias. For instance, some methods implicitly assume that the missing feature contains no or excessive information, leading to distorted feature importance estimates. Consequently, reported metrics may misrepresent the true underlying importance of features.

To address the lack of estimand specification, we introduce the distinction between two concepts: **(1) Full data feature importance metrics**, which refer to the ground truth data had there been no missingness, and **(2) Observed data feature importance metrics**, which refer to the observed data and thus analyze feature importance in the context of the current measurement policy. We illustrate the implications of these two perspectives using a simple yet practical feature importance metric: the leave-one-covariate-out (LOCO) metric (Lei et al., 2018).

We then turn to the problem of identification. While observed data feature importance metrics are directly identifiable from the observed data distribution, the identification of full data feature importance metrics relies on strong - and often unrealistic - assumptions. In particular, we examine the role of the positivity assumption, which can critically limit the reliability of estimation in practice.

Next, we address the problem of estimation by reviewing commonly used approaches for computing feature importance metrics under missing data. We classify these methods according to the estimands they target and assess their ability to produce unbiased results under various missingness mechanisms. We demonstrate that estimation results can differ drastically between full data and observed data feature importance metrics. For instance, consider a costly or invasive medical test that provides a highly informative feature: if performed for every patient, the feature would be an excellent predictor, yielding high full data feature importance. However, due to cost constraints, it is rarely conducted in practice - leading to a much lower observed data feature importance.

While full data and observed data feature importance metrics are appropriate for the questions posed in Scenarios 1 and 2, they fall short in addressing Scenario 3, where the goal is to assess a features importance under changes in the measurement process. To fill this gap, we introduce the **feature measurement importance gradient (FMIG)**, a novel, model-agnostic, and computationally efficient estimand. FMIG quantifies the sensitivity of prediction performance to small increases in the measurement probability of a given feature. Unlike full data feature importance metrics, FMIG does not require a positivity assumption, making it more robust and practically applicable in real-world settings with high rates of missingness.

2. Related Methods

In Appendix A, we provide an overview of feature importance metrics. In the following, we discuss related approaches for assessing feature importance and the closely related problem of feature selection in the presence of missing data.

2.1. Estimand Specification

To the best of our knowledge, no prior studies have explicitly differentiated between observed and full data feature importance metrics. However, some papers incidentally report both types of metrics because they use multiple methods - some of which align with observed data feature importance metrics and others with full data feature importance - without noticing the difference in the underlying estimands (Vo et al., 2024; Srinivasan et al., 2016; Rai et al., 2024). The same issue arises in feature selection (Fan et al., 2023; Mera-Gaona et al., 2021; Ergul Aydin & Kamisli Ozturk, 2024). Notably, only Hapfelmeier (2012) acknowledged that multiple imputation addresses a different question than their proposed method, but the difference in estimands remained unexplored. In contrast, our work is the first to formally define these concepts and clarify their implications.

2.2. Identification

Challenges related to estimating feature importance or performing feature selection under different missingness mechanisms, such as MCAR, MAR, and MNAR, have been discussed in some studies (Vo et al., 2024; Hapfelmeier, 2012; Srinivasan et al., 2016; Fan et al., 2023; Beld et al., 2024; Zhao & Long, 2017), but MNAR remains particularly underexplored (Zhao & Long, 2017). Many papers fail to specify the missingness assumptions under which their reported metrics are valid (Yoo et al., 2025; Rai et al., 2024; Ma et al., 2023; Beld et al., 2024; Peng et al., 2023; Guan et al., 2023; Zhou et al., 2023; Qi et al., 2022; Xie et al., 2023; Bao et al., 2023; Yang et al., 2022; Zhang et al., 2023). Moreover, no prior work has, to the best of our knowledge, addressed the role of the positivity assumption in the context of feature importance metrics under missing data.

2.3. Estimation

Various studies have investigated methods for estimating feature importance metrics (Vo et al., 2024; Srinivasan et al., 2016) or conducting feature selection (Gunn et al., 2023; Seijo-Pardo et al., 2018; 2019; Mera-Gaona et al., 2021; Hu et al., 2021; Genossar et al., 2024; Ergul Aydin & Kamisli Ozturk, 2024; Doquire & Verleysen, 2012) in the presence of missing data. These approaches include imputation-based methods (Zhao & Long, 2017; Vo et al., 2024; Seijo-Pardo et al., 2018; 2019) and methods that intrinsically handle missing data (Vo et al., 2024). Among imputation-based methods, conditional mean imputation and similar approaches like KNN-imputation (Smit et al., 2022) or SVD imputation are commonly used (Srinivasan et al., 2016; Hapfelmeier, 2012; Ergul Aydin & Kamisli Ozturk, 2024; Smit et al., 2022), but introduce bias (Seijo-Pardo et al., 2018).

In contrast, multiple imputation (MI) is recognized as an unbiased method for estimating feature importance under missing data (Moons et al., 2006). However, prior work has emphasized that the label must be included in the imputation model to ensure validity (Moons et al., 2006). Despite this recommendation, it is still common in practice to exclude the label (Srinivasan et al., 2016; Hapfelmeier, 2012; Ergul Aydin & Kamisli Ozturk, 2024; Zeng et al., 2023).

While these methods have been analyzed in various studies, no work has systematically categorized them based on the type of feature importance metric they compute - whether observed data or full data feature importance. In addition to this categorization, we identify which estimation methods introduce bias, further clarifying their limitations and appropriate applications.

2.4. Action Importance Metrics in Reinforcement Learning and Causal Inference

The feature measurement importance gradient (FMIG), introduced in this work, quantifies how predictive performance would change if the probability of measuring a given feature were slightly increased. This metric frames the measurement process as a decision-making problem, where the decision is whether to measure a feature in a specific context defined by the already-measured features. This perspective naturally aligns with concepts in reinforcement learning (RL) and causal inference, which study the importance of actions and their effects.

Shapley values, which are widely used in feature importance evaluation (Lundberg & Lee, 2017), have been also applied in RL to assess the importance of individual actions (Beechey et al., 2023). Concepts like the credit assignment problem in RL (Pignatelli et al., 2024), which focus on attributing rewards to specific actions, highlight related challenges in quantifying the importance of actions and suggest potential connections to feature measurement tasks.

FMIG also shares a conceptual link with incremental propensity score interventions introduced by Kennedy (2019), which provide an interpretable framework for analyzing the effects of small changes in action probabilities (propensity scores). We adapt this idea to the setting of feature measurement and, by focusing only on the gradient, simplify estimation while retaining an interpretable metric.

3. Full Data and Observed Data Feature Importance Metrics

In this section, we introduce notation, define the full data and observed data LOCO metrics, and discuss identification and estimation. We consider a classification setting, but our results naturally extend to regression.

3.1. Notation

We define the key variables and functions as follows:

- **Observed features**: $X \in (\mathbb{R} \cup \{\text{"?"}\})^d$, where X_i represents the value of the i-th feature. If $X_i = \text{"?"}$, the feature is missing; otherwise, $X_i \in \mathbb{R}$.
- **Missingness indicators**: $R \in \{0,1\}^d$, where $R_i = 1$ indicates that feature X_i is observed, and $R_i = 0$ indicates it is missing.
- **Ground truth features**: $X_{(1)} \in \mathbb{R}^d$, representing the true values of all features, had they all been observed. Using potential outcome notation (Rubin, 2005), $X_{(1),i}$ denotes the value of feature X_i, potentially contrary to fact, had $R_i = 1$. We obtain:

$$X_i = \begin{cases} X_{(1),i} & \text{if } R_i = 1, \\ \text{"?"} & \text{if } R_i = 0. \end{cases} \tag{1}$$

- **Label**: $Y \in \{0,1,\ldots,K-1\}$, the categorical outcome associated with the feature set X, where K is the number of possible classes. We do not consider label missingness in this work.
- **Missingness process**: $\pi\colon \mathbb{R}^d \times \{0,1,\ldots,K-1\} \to [0,1]^{2^d}$, representing the mechanism that determines the missingness indicator R. In the most general case $\pi(R|X_{(1)}, Y)$ depends on both the ground truth features $X_{(1)}$ and the label Y.
- **Classifier**: $f_{cl}\colon (\mathbb{R} \cup \{\text{"?"}\})^d \times \{0,1\}^d \to [0,1]^K$, defines a mapping from the observed features and missingness indicators to the label probabilities. While X already encodes missingness through the presence of "?" values, we explicitly include the missingness indicators R in the notation to emphasize the dependence on the missingness pattern. Using missingness indicators in the classifier is common practice in many settings (Van Ness et al., 2023; Singh et al., 2021).
- **Loss function**: $L\colon \{0,1,\ldots,K-1\} \times [0,1]^K \to \mathbb{R}$, defines a loss function, for example cross entropy, for the classifier f_{cl} based on the true labels and the predicted label probabilities.

We further let $X_{-j}, R_{-j}, X_{(1),-j}$ denote the reduced dimension observed features, missingness indicators and ground truth features when feature j is excluded. Similarly, we let $f_{cl,-j}\colon (\mathbb{R} \cup \{\text{"?"}\})^{d-1} \times \{0,1\}^{d-1} \to [0,1]^K$ denote the classifier that excludes information from feature j.

3.2. Estimand Definition: Full Data and Observed Data LOCO Metrics

For full data, the LOCO metric for feature j evaluates as follows.

Full data LOCO:

$$\begin{aligned} LOCO_j^{FD} = &\mathbb{E}\left[L\left(Y, f_{cl,-j}(X_{(1),-j}, R_{-j} = \vec{1})\right)\right] \\ &-\mathbb{E}\left[L\left(Y, f_{cl}(X_{(1)}, R = \vec{1})\right)\right] \end{aligned} \tag{2}$$

Here, the full data LOCO metric measures the expected loss difference if feature j is excluded compared to if it is included, with all other features fully observed.

In contrast, the observed data LOCO metric evaluates as follows.

Observed data LOCO:

$$LOCO_j^{OD} = \mathbb{E}\left[L\left(Y, f_{cl,-j}(X_{-j}, R_{-j})\right)\right] - \mathbb{E}\left[L\left(Y, f_{cl}(X, R)\right)\right] \tag{3}$$

It differs from $LOCO_j^{FD}$ in that it replaces the ground truth features $X_{(1)}$ with their observed proxies X and the missingness indicators R, which embeds the feature importance in the context of the measurement policy that was used in the acquisition of the dataset. The observed data LOCO metric $LOCO_j^{OD}$ thus quantifies the loss difference when feature j is excluded compared to if it is included, but only when measured under the current measurement policy, with other features partially observed under their respective measurement policies.

Remark 3.1. Note that the observed data LOCO metric, $LOCO_j^{OD}$, does not capture the contrast in loss that would occur if feature j were no longer measured. We introduce an alternative estimand, termed *leave-one-covariate-unmeasured (LOCU)*, which describes this scenario in detail in Appendix B. The LOCU metric coincides with $LOCO^{OD}$ only under certain restrictive independence assumptions about the missingness process.

3.3. Identification

The observed data LOCO metric is already identified, meaning it is expressed as a function of the observed data. This is not the case for the full data LOCO metric, which depends on the unobserved ground truth features $X_{(1)}$. To identify the full data LOCO metric, missing data identification methods can be applied. Generally, the following assumptions are required:

Stable Unit Treatment Value Assumption (SUTVA): This assumption ensures that the counterfactual feature values are well-defined, as described by Eq. 1. Specifically, it requires that the observed features are equal to the counterfactual

features when $R = 1$. Additionally, it assumes that data points do not interfere with one another - specifically, that missingness in one instance is not influenced by the features or missingness indicators of other instances.

Independence assumptions about the missingness process: These assumptions describe the relationship between the missingness process π and the data:

- Missing Completely at Random (MCAR): The missingness process is independent of both observed and unobserved data, $\pi(R|X_{(1)}, Y) = \pi(R)$.
- Missing at Random (MAR): The missingness process depends only on the observed data.
- Missing Not at Random (MNAR): The missingness process may depend on unobserved data, such as the ground truth features $X_{(1)}$.

Positivity assumption: The positivity assumption requires that $\pi(R = \vec{1} \mid X_{(1)}, Y) > 0$ for all $X_{(1)}, Y$ with $p(X_{(1)}, Y) > 0$. In essence, it ensures that for every ground truth data point, there is a positive probability that it is fully observed. The positivity assumption is particularly critical in many real-world settings, as it is often violated. Consider, for example, a time-series medical prediction setting. The assumption requires that for every patient, there is a chance that all medical tests are being performed jointly and at every time point - a scenario that is rarely, if ever, met in practice. As a result, full data feature importance metric estimates become highly unreliable in such contexts.

3.4. Estimation

In the following, we categorize various estimation methods for both full data and observed data feature importance metrics. Note that feature importance is defined relative to a given classifier f_{cl}, and thus reflects the true importance of a feature only to the extent that the classifier accurately captures the relationship between X and Y.

3.4.1. Estimation of the Full Data LOCO Metric

Inverse probability weighting (Seaman & White, 2013) and multiple imputation (Sterne et al., 2009) are widely used methods from the missing data literature for estimating full data target parameters. They can thus also be used to compute the full data LOCO metric. Both approaches reconstruct samples from the ground truth distribution $p(X_{(1)}, Y)$, enabling the estimation of full data LOCO metrics.

Inverse-Probability Weighting (IPW): The IPW estimator reconstructs samples from the complete cases by applying appropriate weights. Using Bayes' rule, the ground-truth distribution is expressed as:

$$p(X_{(1)}, Y) = \frac{p(R = \vec{1}, X_{(1)}, Y)}{\pi(R = \vec{1} \mid X_{(1)}, Y)}.$$

This formulation is valid if the propensity score $\pi(R = \vec{1} \mid X_{(1)}, Y) = p(R = \vec{1} \mid X_{(1)}, Y)$ is identifiable, as $p(R = \vec{1}, X_{(1)}, Y) = p(R = \vec{1}, X, Y)$ is already a function of the observed data. When the missingness mechanism is MCAR, this estimator simplifies to complete case analysis.

Multiple Imputation (MI): The MI estimator reconstructs $p(X_{(1)}, Y)$ by decomposing $X_{(1)}$ into its observed components X_o[1] and missing components X_m, as follows:

$$p(X_{(1)}, Y) = \sum_R p(X_m \mid X_o, Y, R) p(X_o, Y, R).$$

Here, $p(X_o, Y, R)$ corresponds to the observed data, while $p(X_m \mid X_o, Y, R)$ must be identified and estimated.

To ensure unbiased estimation, the imputation model must thus include the label Y as an input. This was also verified in experiments (Moons et al., 2006). However, this introduces a practical concern: if the imputation model is misspecified, the imputed values may contain excessive information about Y, potentially leading to inflated prediction performance estimates.

To mitigate this risk, practitioners often exclude Y from the imputation model (Srinivasan et al., 2016; Hapfelmeier, 2012; Ergul Aydin & Kamisli Ozturk, 2024). Imputing X_m using $p(X_m \mid X_o, R)$ assumes $X_m \perp\!\!\!\perp Y \mid X_o, R$, implying that missing features contribute no additional information for predicting Y beyond X_o and R. Such a practice effectively dismisses the potential importance of missing features before the computation even begins, rendering the approach fundamentally flawed.

3.4.2. Estimation of the Observed Data LOCO Metric

Estimating observed data feature importance metrics is more straightforward because the observed data LOCO metric is already identified. The primary requirement is ensuring that the classifier can appropriately handle missing values. Crucially, when assessing observed data feature importance metrics, no additional information should be introduced for the missing feature. This can be achieved either by applying the classifier directly (if it supports "?" values) or by using simple imputation methods (e.g., mean or median imputation) that do not inject additional predictive information. In the case of imputation, it is particularly important to include the unchanged missingness indicators R in the classifier's input to preserve the informativeness of the missingness

[1] X_o differs from X because X contains placeholders (e.g., "?" for missing values) that encode information about R.

pattern. While mean imputation is often considered a biased approach (Mera-Gaona et al., 2021), this is true only when a full data feature importance method is of interest. For observed data feature importance, mean imputation remains unbiased. A more rigorous mathematical justification for this result is provided in Appendix C.

Other popular approaches to estimate feature importance metrics can handle missing data intrinsically. These include for example tree-based methods such as XGBoost (Vo et al., 2024; Chen & Guestrin, 2016) and are often based on a principle known as missingness incorporated in attributes (MIA) (Twala et al., 2008). These methods merely handle the missing values within the classification and do not imagine counterfactual, "what if everything had been observed", scenarios. They are therefore also included in the observed data feature importance category.

3.4.3. Biased Estimation Methods

A common but highly problematic approach to computing feature importance metrics is conditional mean imputation. This method replaces missing values with

$$\mathbb{E}[X_m \mid X_o, Y, R],$$

or, more commonly (when Y is excluded), with

$$\mathbb{E}[X_m \mid X_o, R].$$

While this simplification may appear reasonable, it introduces substantial bias when the goal is to estimate either full data or observed data feature importance metrics. We provide a detailed mathematical justification for a range of feature importance metrics in Appendix C and offer some intuition below.

Including Y may create a deterministic relationship between Y and the imputed values, potentially allowing a classifier to exploit this relationship and achieve artificially high prediction accuracy. Excluding Y, on the other hand, can still introduce a deterministic relationship between X_m and X_o, thereby transferring predictive information from X_o to the imputed values. This severely skews feature importance estimates. This method is thus neither suited to estimate full data, nor observed data feature importance metrics. Note that the same holds for all similar imputation methods that do not model the full density for the imputed features. This includes for example KNN-imputation (Zhang, 2012) which is also frequently used to report feature importance under missing data (Zhou et al., 2023).

In conclusion, when using imputation for the estimation of $LOCO^{FD}$ and $LOCO^{OD}$, the missing values must be imputed in a manner consistent with the classifiers respective "working conditions." For $LOCO^{FD}$, the classifier operates under full data availability. Therefore, missing values must be imputed using samples from the full data distribution. When done correctly, $LOCO^{FD}$ is evaluated on a dataset that mirrors the original dataset as if no missingness had occurred. In this case, the classifier can leverage the imputed values to potentially improve performance. For $LOCO^{OD}$, evaluation occurs under conditions where missingness is present. Since the observed data already reflects the classifier's working conditions, no imputation is required. However, one may opt for impute-then-regress classifiers, where imputation serves purely as a means to handle missing values elegantly. In such cases, the imputation step does not introduce additional information beyond what is already present in the observed input features.

Remark 3.2. While the LOCO metrics in Eqs. 2 and 3 are valid for any classifier f_{cl}, the choice of training method is essential for interpretability. Classifiers must be trained for the specific conditions under which they will be deployed. To evaluate observed data feature importance metrics, the classifier must perform optimally in the presence of missing data. In contrast, full data feature importance metrics require a classifier trained for settings where all features are available. The implications of this distinction are discussed in Appendix D.

4. Feature Measurement Importance Gradient

The observed data and full data LOCO metrics assess the impact of completely omitting X_j or $X_{(1),j}$ from the prediction model, providing an "all-or-nothing" perspective on feature importance. However, in practice, improving predictive performance often requires more fine-grained considerations. Rather than asking whether a feature is important in absolute terms, a more actionable question is: which feature, if measured more frequently, would yield the greatest improvement in prediction performance? For instance, if we could increase the measurement frequency of certain features, how should we prioritize them to maximize predictive gains? Answering this question provides practical guidance for optimizing data collection strategies.

To address this, we introduce the feature measurement importance gradient (FMIG), a novel metric that quantifies how prediction performance changes with small increases in the measurement probability of a given feature. We formalize FMIG by defining perturbation interventions on the measurement policy and discuss its identification and estimation.

Specifically, FMIG captures the sensitivity of predictive performance to marginal changes in measurement probability. To this end, we introduce a perturbation δ_j to the current measurement policy π, increasing the probability of measuring feature j. We denote the perturbed policy as π_{δ_j}.

Given this perturbation, we can determine the gradient of the negative loss with respect to the perturbation to rank

features in their importance:

Feature Measurement Importance Gradient (FMIG) G_j**:**

$$G_j = \nabla_{\delta_j} \mathbb{E}\Big[-L\Big(Y, f_{cl}(X_{(\pi_{\delta_j})}, R_{(\pi_{\delta_j})})\Big)\Big]\Big|_{\delta_j=\delta_j^*} \tag{4}$$

where δ_j^* is the value of δ_j that represents no perturbation. Here, $X_{(\pi_{\delta_j})}$ and $R_{(\pi_{\delta_j})}$ denote the counterfactual observed features and missingness indicators, had the perturbed policy π_{δ_j} been applied. We take the gradient with respect to the negative loss so that the FMIG quantifies the expected reduction in loss when the feature is measured slightly more frequently and thus leads to positive feature importances.

4.1. Measurement Process Assumptions

The definition of interventions on the measurement policy relies on the data's measurement process. Here, we consider a special MAR process. Features are measured sequentially, where the measurement of feature j depends on all prior missingness indicators and observed features.

Under these assumptions, the measurement policy factorizes as:

$$\pi(R|X_{(1)}, Y) = \prod_{i=1}^{d} \pi_i(R_i|\underline{R}_{i-1}, \underline{X}_{i-1}), \tag{5}$$

where $\underline{R}_{i-1} \equiv \{R_1, \ldots, R_{i-1}\}$, $\underline{X}_{i-1} \equiv \{X_1, \ldots, X_{i-1}\}$ and we let $\underline{R}_0 \equiv \{\}$ and $\underline{X}_0 \equiv \{\}$.

4.2. Measurement Policy Interventions

To define perturbation interventions, we leverage the factorization from Eq. 5:

$$\pi_{\delta_j}(R|X_{(1)}, Y) = \\ = \pi_{j,\delta_j}(R_j|\underline{R}_{j-1}, \underline{X}_{i-1}) \prod_{i=1; i\neq j}^{d} \pi_i(R_i|\underline{R}_{i-1}, \underline{X}_{i-1}).$$

Here, the intervention modifies the measurement policy only for feature j. We adapt the perturbation from incremental propensity score interventions (Kennedy, 2019), defined as:

$$\pi_{j,\delta_j} = \frac{\delta_j \, \pi_j}{\delta_j \, \pi_j + 1 - \pi_j}, \quad \text{for } 0 < \delta_j < \infty,$$

where $\pi_j \equiv \pi_j(R_j = 1|\underline{R}_{j-1}, \underline{X}_{j-1})$. The parameter δ_j is the *odds ratio* of the intervention:

$$\delta_j = \frac{\pi_{j,\delta_j}/(1-\pi_{j,\delta_j})}{\pi_j/(1-\pi_j)} = \frac{\text{odds}_{\pi_{j,\delta_j}}(R_j = 1|\underline{R}_{j-1}, \underline{X}_{j-1})}{\text{odds}_{\pi_j}(R_j = 1|\underline{R}_{j-1}, \underline{X}_{j-1})}.$$

Values of $\delta_j < 1$ correspond to a decrease in the probability of measuring feature j, while $\delta_j > 1$ increases the probability. When $\delta_j = \delta_j^* = 1$, the original measurement policy is recovered: $\pi_{j,\delta_j=1} = \pi_j$.

4.3. Identification

Using the perturbation interventions, the feature measurement importance gradient G_j can be identified as follows:

Lemma 4.1. ***(Identification of the Feature Measurement Importance Gradient (FMIG))***
Under SUTVA and the assumption about the missingness process from Eq. 5, the feature measurement importance gradient (FMIG) for feature j is identified as:

$$G_j = \nabla_{\delta_j} \mathbb{E}\Big[-L\Big(Y, f_{cl}(X_{(\pi_{\delta_j})}, R_{(\pi_{\delta_j})})\Big)\Big]\Big|_{\delta_j=1} \\ = \mathbb{E}\Big[-L\left(Y, f_{cl}(X,R)\right)(-1)^{1-R_j}(1-\pi_j(R_j|\underline{R}_{j-1}, \underline{X}_{j-1}))\Big].$$

The proof is provided in Appendix E. It demonstrates why, unlike full data feature importance metrics, the FMIG does not require a positivity assumption due to the chosen perturbation. This allows for far more reliable estimation in real-world settings with high fractions of missing values.

4.4. Estimation

The functional G_j, as identified in Lemma 4.1, is straightforward to estimate. It requires only the measurement policy of the feature of interest to be learned. An estimator is given by:

$$\hat{G}_j = \hat{\mathbb{E}}_n\Big[-L\left(Y, f_{cl}(X,R)\right)(-1)^{1-R_j}(1-\hat{\pi}_j(R_j|\underline{R}_{j-1}, \underline{X}_{j-1}))\Big],$$

where $\hat{\mathbb{E}}_n$ is the empirical mean over a dataset of size n, and $\hat{\pi}_j$ is a model learned to approximate π_j. When using flexible machine learning models for $\hat{\pi}_j$, these should be trained on a separate data split. To restore full data efficiency, a cross-fitting scheme can be employed (Chernozhukov et al., 2018).

We extend the theory of the feature measurement importance gradient to time-series settings in Appendix F. In this setting, each feature is assumed to be measured at fixed time points, with predictions made for a time-series label at each time point.

5. Experiments

Missing data estimation methods cannot be tested on real-world data with real missingness due to the unavailability of ground truth features $X_{(1)}$. To address this limitation, we perform a series of synthetic experiments to illustrate the differences between feature importance metrics, the impact of positivity violations, and the significance of appropriate estimation methods.

5.1. Experiment Design

We consider a time-series prediction task where features $X^t \in (\mathbb{R} \cup \{\text{"?"}\})^d$ $(t \in \{1, 2, 3\})$ are used to predict labels

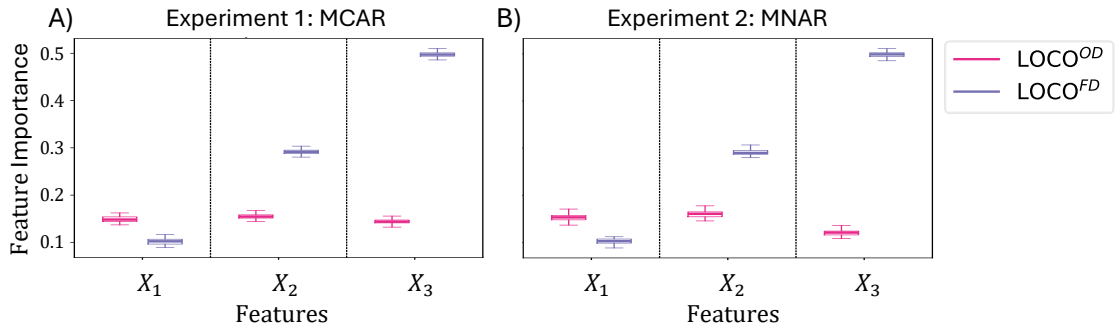


Figure 1. Comparison of feature importance metrics (ground truth estimates). A) Experiment 1: The full data LOCO metric increases from X_1 to X_3, reflecting the features' growing predictive importance. However, the observed data LOCO remains flat because the increasing predictive value is offset by higher missingness rates from X_1 to X_3. B) Experiment 2: Due to MNAR missingness ($X_{(1),3} \to R_1, R_2$), the observed data LOCO of X_3 is further reduced because the missingness mechanism induces correlations between $X_{(1),3}$ and X_1, X_2 via R_1, R_2. Confidence intervals were computed using the non-parametric bootstrap with 50 samples.

Y^t at each time step. We use an "impute-then-regress" classifier with zero imputation and a temporal convolutional network (TCN) (Bai et al., 2018). Missingness mechanisms (MCAR, MNAR) are introduced at varying missingness rates across experiments. Full experimental details are provided in Appendix G. Additionally, Appendix H presents an experiment highlighting the importance of training classifiers appropriately for the metric under evaluation.

5.2. Synthetic Experiments and Results

Experiment 1: Rarely measured variables have reduced observed data feature importance

Design: We evaluate three variables with increasing predictive importance but decreasing measurement probabilities under an MCAR scenario: $p(R_1^t = 1) = 0.75$, $p(R_2^t = 1) = 0.5$, $p(R_3^t = 1) = 0.3$.

Results: While the full data LOCO reflects the increasing predictive importance from X_1 to X_3, the observed data LOCO remains flat (Figure 1A). This occurs because the gains in predictive performance from more informative features are counteracted by their higher missingness rates. The experiment illustrates how observed data feature importance metrics lead to different conclusions than full data feature importance metrics.

Experiment 2: Informative missingness affects observed feature importance

Design: Using the same dataset as Experiment 1, we simulate MNAR missingness where $X_{(1),3}^t$ influences missingness of features at subsequent time points. Average missingness probabilities remain equal to Experiment 1 with $p(R_1^t = 1) \approx 0.75$, $p(R_2^t = 1) \approx 0.5$, $p(R_3^t = 1) = 0.3$.

Results: Figure 1B) shows that the full data LOCO remains unchanged. However, the observed data LOCO of $X_{(1),3}$ decreases as its information is "leaked" into correlated variables (R_1, R_2). The experiment highlights how informative missingness redistributes feature importance in observed data.

Experiment 3: Positivity violations introduce bias in estimation

Design: Using the same dataset as Experiment 1, we simulate stronger MCAR missingness: $p(R_1^t = 1) = 0.4$, $p(R_2^t = 1) = 0.2$, $p(R_3^t = 1) = 0.1$. This results in a near-complete absence of fully observed cases, effectively violating the positivity assumption.

Results: Figure 2 illustrates biases in estimating components of the full data LOCO metric. The complete case analysis, which is generally unbiased under MCAR, produces unbiased estimates in Figure 2A) for Experiment 1. However, in Experiment 3, where the positivity assumption is violated, the complete case estimator exhibits strong bias (Figure 2B). These results highlight the limitations of standard estimation methods for the full data LOCO metric when positivity is violated.

Experiment 4: FMIG provides additional insights

Design: We evaluate three variables with equal full data feature importance, using an MCAR mechanism. We have $p(R_1^t = 1) = 1$, $p(R_2^t = 1) = 0.5$ and a constant feature $X_{(1),3}$ with $p(R_3^1 = 1) = 1$ and $p(R_3^t = 1) = 0.5$ for $t \geq 2$.

Results: Figure 3 illustrates the FMIG metric is zero for X_1 as it is already fully observed. FMIG is positive for X_2 which could, if measured more often, improve prediction performance. Measuring the constant $X_{(1),3}$ multiple times provides no predictive improvement. Since the additional feature measurement is completely uninformative, the FMIG even becomes slightly negative. The insight that measuring $X_{(1),3}$ more often will not improve prediction per-

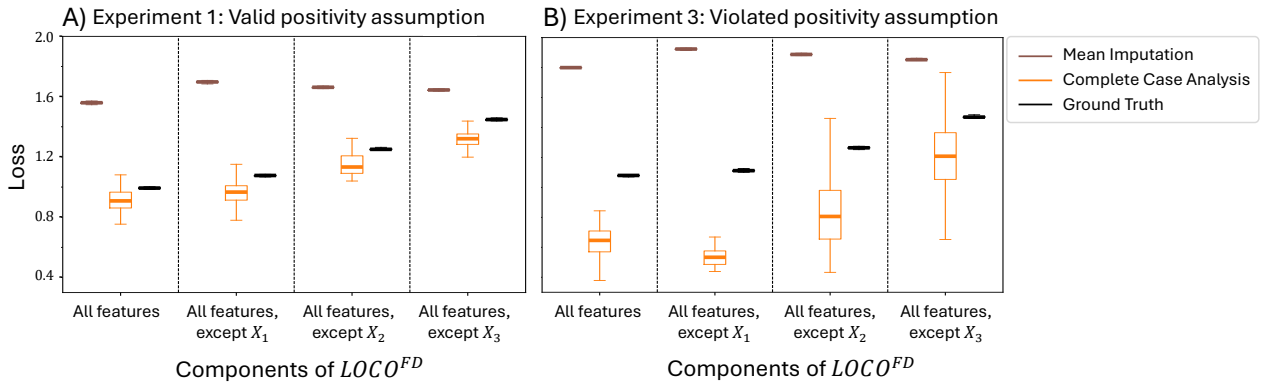


Figure 2. Effects of the positivity assumption violations on the estimation of full data LOCO components. The plot shows estimates using different estimators for $\mathbb{E}\left[-\sum_{t=1}^{T} L\left(Y^t, f_{cl}(\underline{X}^t_{(1)}, \underline{R}^t = \vec{1})\right)\right]$ (denoted as "All features") and $\mathbb{E}\left[-\sum_{t=1}^{T} L\left(Y^t, f_{cl,-j}(\underline{X}^t_{(1),-j}, \underline{R}^t_{-j} = \vec{1})\right)\right]$ for each feature (denoted as "All features, except X_1", "All features, except X_2", and "All features, except X_3"). A) Experiment 1: Under a valid positivity assumption, the complete case estimator provides an acceptable estimate of the ground truth. B) Experiment 3: Under violations of the positivity assumption, the complete cases analysis provides a biased estimate of the ground truth. Note that the classifiers used to evaluate feature importance metrics were trained on the observed data from each experiment. As a result, even the ground truth estimates differ slightly, with a lower loss in Experiment 1 due to less missingness during classifier training. Confidence intervals were computed using the non-parametric bootstrap with 50 samples.

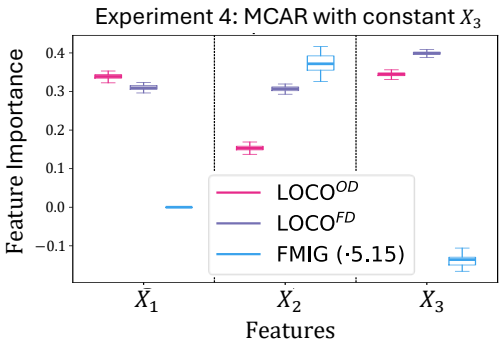


Figure 3. $X_{(1),3}$ is constant and fully observed at $t = 1$, but less frequently at later times. The feature measurement importance gradient FMIG is low for this feature, even though it has high observed data and full data feature importance.

formance is not apparent from observed or full data LOCO metrics, but only from the FMIG.

6. Discussion & Conclusion

In this work, we examined feature importance in the presence of missing data. Full data feature importance metrics, which assume no missingness, offer insights into the data-generating process but are often difficult to estimate and impractical when missingness persists at deployment. To address this, we introduced observed data feature importance metrics, which assess importance under a given measurement policy and highlighted key differences between these metrics, categorizing estimation methods based on their target estimands.

Neither observed nor full data feature importance directly informs data collection. To address this, we introduce the feature measurement importance gradient (FMIG), a novel metric that identifies which features, if measured more frequently, would yield the greatest improvement in predictive performance. Synthetic experiments illustrate the distinct insights these metrics provide, emphasizing the need for tailored approaches in missing data settings. FMIG is thus not an alternative to the observed data or full data LOCO metrics, but an additional tool to assess the importance of a feature under a change in measurement probability.

While FMIG provides information on which features should be measured more frequently, it does not prescribe an optimal measurement policy. Active feature acquisition (AFA) (von Kleist et al., 2025; 2023), on the other hand, explicitly balances measurement costs against predictive value to derive optimal policies. FMIG thus complements AFA by identifying features worth prioritizing for optimization.

Our experimental evaluation is limited to synthetic data. Demonstrating the usefulness of feature importance metrics in real-world settings particularly where prior knowledge or separately collected, fully observed datasets are available is an important direction for future work. Such demonstrations could highlight the practical relevance of these metrics and inform their application in domains with varying measurement constraints.

Acknowledgements

Joshua Wendland was supported by the European Research Council (ERC) Starting Grant DEUCE, grant no. 101077178. Carsten Marr acknowledges funding from the ERC under the European Unions Horizon 2020 research and innovation program (grant agreement no. 866411) and support from the Hightech Agenda Bayern.

Impact Statement

Our method enhances the explainability of AI models by providing a principled way to assess feature importance under missing data, thereby improving trust and reliabilityespecially in high-stakes applications such as medical decision-making. By clarifying biases in existing methods, our work contributes to more transparent, interpretable, and fair AI systems, ultimately fostering more informed and responsible deployment of machine learning models.

References

Bai, S., Kolter, J. Z., and Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv:1803.01271 [cs]*, 2018.

Bao, C., Deng, F., and Zhao, S. Machine-learning models for prediction of sepsis patients mortality. *Medicina intensiva*, 47(6):315–325, 2023.

Beechey, D., Smith, T. M. S., and imek, Explaining Reinforcement Learning with Shapley Values. *arXiv:2306.05810 [cs]*, 2023.

Beld, J.-J. v. d., Pathak, S., Geerdink, J., Hegeman, J. H., and Seifert, C. Feature importance to explain multimodal prediction models. A clinical use case. *arXiv:2404.18631 [cs]*, 2024.

Bhattacharya, R., Nabi, R., Shpitser, I., and Robins, J. M. Identification In Missing Data Models Represented By Directed Acyclic Graphs. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, pp. 1149–1158. PMLR, 2020.

Chen, T. and Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco California USA, 2016. ACM.

Chen, Z., Li, T., Guo, S., Zeng, D., and Wang, K. Machine learning-based in-hospital mortality risk prediction tool for intensive care unit patients with heart failure. *Frontiers in cardiovascular medicine*, 10(101653388): 1119699, 2023.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Danilatou, V., Nikolakakis, S., Antonakaki, D., Tzagkarakis, C., Mavroidis, D., Kostoulas, T., and Ioannidis, S. Outcome Prediction in Critically-Ill Patients with Venous Thromboembolism and/or Cancer Using Machine Learning Algorithms: External Validation and Comparison with Scoring Systems. *International journal of molecular sciences*, 23(13), 2022.

Doquire, G. and Verleysen, M. Feature selection with missing data using mutual information estimators. *Neurocomputing*, 90:3–11, 2012.

Ergul Aydin, Z. and Kamisli Ozturk, Z. Filter-based feature selection methods in the presence of missing data for medical prediction models. *Multimedia Tools and Applications*, 83(8):24187–24216, 2024.

Ewald, F. K., Bothmann, L., Wright, M. N., Bischl, B., Casalicchio, G., and König, G. A Guide to Feature Importance Methods for Scientific Inference. *arXiv:2404.12862*, 2024.

Fan, M., Peng, X., Niu, X., Cui, T., and He, Q. Missing data imputation, prediction, and feature selection in diagnosis of vaginal prolapse. *BMC Medical Research Methodology*, 23(1):259, 2023.

Genossar, B., On, T., Islam, M. M., Eliav, B., Roy, S. B., and Gal, A. MISFEAT: Feature Selection for Subgroups with Systematic Missing Data. *arXiv:2412.06711 [cs]*, 2024.

Guan, C., Ma, F., Chang, S., and Zhang, J. Interpretable machine learning models for predicting venous thromboembolism in the intensive care unit: an analysis based on data from 207 centers. *Critical care (London, England)*, 27(1):406, 2023.

Gunn, H. J., Hayati Rezvan, P., Fernández, M. I., and Comulada, W. S. How to apply variable selection machine learning algorithms with multiply imputed data: A missing discussion. *Psychological Methods*, 28(2):452–471, 2023.

Hapfelmeier, A. Random Forest variable importance with missing data. 2012.

Hu, L., Joyce Lin, J.-Y., and Ji, J. Variable selection with missing data in both covariates and outcomes: Imputation and machine learning. *Statistical Methods in Medical Research*, 30(12):2651–2671, 2021.

Huang, C.-T., Wang, T.-J., Kuo, L.-K., Tsai, M.-J., Cia, C.-T., Chiang, D.-H., Chang, P.-J., Chong, I.-W., Tsai, Y.-S., Chu, Y.-C., Liu, C.-J., Chen, C.-H., Pai, K.-C., and Wu, C.-L. Federated machine learning for predicting acute kidney injury in critically ill patients: a multicenter study in Taiwan. *Health information science and systems*, 11 (1):48, 2023.

Ishii, E., Nawa, N., Hashimoto, S., Shigemitsu, H., and Fujiwara, T. Development, validation, and feature extraction of a deep learning model predicting in-hospital mortality using Japan's largest national ICU database: a validation framework for transparent clinical Artificial Intelligence (cAI) development. *Anaesthesia, critical care & pain medicine*, 42(2):101167, 2023.

Kennedy, E. H. Nonparametric Causal Effects Based on Incremental Propensity Score Interventions. *Journal of the American Statistical Association*, 114(526):645–656, 2019.

Le Morvan, M., Josse, J., Scornet, E., and Varoquaux, G. Whats a good imputation to predict with missing values? In *Advances in Neural Information Processing Systems*, volume 34, pp. 11530–11540, 2021.

Lei, J., GSell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

Lucini, F. R., Stelfox, H. T., and Lee, J. Deep Learning-Based Recurrent Delirium Prediction in Critically Ill Patients. *Critical care medicine*, 51(4):492–502, 2023.

Lundberg, S. M. and Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Ma, L., Zhang, C., Gao, J., Jiao, X., Yu, Z., Ma, X., Wang, Y., Tang, W., Zhao, X., Ruan, W., and Wang, T. Mortality Prediction with Adaptive Feature Importance Recalibration for Peritoneal Dialysis Patients: a deep-learning-based study on a real-world longitudinal follow-up dataset. *arXiv:2301.07107 [cs]*, 2023.

Mera-Gaona, M., Neumann, U., Vargas-Canas, R., and López, D. M. Evaluating the impact of multivariate imputation by MICE in feature selection. *PLOS ONE*, 16 (7):e0254720, 2021.

Mohan, K. and Pearl, J. Graphical Models for Processing Missing Data. *Journal of the American Statistical Association*, 116(534):1023–1037, 2021.

Moons, K. G. M., Donders, R. A. R. T., Stijnen, T., and Harrell, F. E. Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology*, 59(10):1092–1101, 2006.

Nabi, R., Bhattacharya, R., and Shpitser, I. Full law identification in graphical models of missing data: Completeness results. In *International Conference on Machine Learning*, pp. 7153–7163. PMLR, 2020.

Peng, C., Yang, F., Li, L., Peng, L., Yu, J., Wang, P., and Jin, Z. A Machine Learning Approach for the Prediction of Severe Acute Kidney Injury Following Traumatic Brain Injury. *Neurocritical care*, 38(2):335–344, 2023.

Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., and van der Laan, M. J. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, 21(1):31–54, 2012.

Pignatelli, E., Ferret, J., Geist, M., Mesnard, T., Hasselt, H. v., Pietquin, O., and Toni, L. A Survey of Temporal Credit Assignment in Deep Reinforcement Learning, 2024.

Qi, J., Lei, J., Li, N., Huang, D., Liu, H., Zhou, K., Dai, Z., and Sun, C. Machine learning models to predict in-hospital mortality in septic patients with diabetes. *Frontiers in endocrinology*, 13(101555782):1034251, 2022.

Rai, T., Shen, Y., He, J., Mahmud, M., Brown, D. J., Kaur, J., ODowd, E., Baldwin, D. R., and Hubbard, R. Understanding Feature Importance of Prediction Models Based on Lung Cancer Primary Care Data. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2024.

Robins, J. A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.

Rubin, D. B. Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

Seaman, S. R. and White, I. R. Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, 22(3):278–295, 2013.

Seijo-Pardo, B., Alonso-Betanzos, A., Bennett, K., Bolon-Canedo, V., Guyon, I., Josse, J., and Saeed, M. Analysis of imputation bias for feature selection with missing data. *Computational Intelligence*, 2018.

Seijo-Pardo, B., Alonso-Betanzos, A., Bennett, K. P., Bolón-Canedo, V., Josse, J., Saeed, M., and Guyon, I. Biases in feature selection with missing data. *Neurocomputing*, 342:97–112, 2019.

Singh, J., Sato, M., and Ohkuma, T. On missingness features in machine learning models for critical care: observational study. *JMIR Medical Informatics*, 9(12):e25022, 2021.

Smit, J. M., Krijthe, J. H., Endeman, H., Tintu, A. N., de Rijke, Y. B., Gommers, D. A. M. P. J., Cremer, O. L., Bosman, R. J., Rigter, S., Wils, E.-J., Frenzel, T., Dongelmans, D. A., De Jong, R., Peters, M. A. A., Kamps, M. J. A., Ramnarain, D., Nowitzky, R., Nooteboom, F. G. C. A., De Ruijter, W., Urlings-Strop, L. C., Smit, E. G. M., Mehagnoul-Schipper, D. J., Dormans, T., De Jager, C. P. C., Hendriks, S. H. A., Achterberg, S., Oostdijk, E., Reidinga, A. C., Festen-Spanjer, B., Brunnekreef, G. B., Cornet, A. D., Van den Tempel, W., Boelens, A. D., Koetsier, P., Lens, J. A., Faber, H. J., Karakus, A., Entjes, R., De Jong, P., Rettig, T. C. D., Arbous, M. S., Lalisang, R. C. A., Tonutti, M., De Bruin, D. P., Elbers, P. W. G., Van Bommel, J., and Reinders, M. J. T. Dynamic prediction of mortality in COVID-19 patients in the intensive care unit: A retrospective multi-center cohort study. *Intelligence-based medicine*, 6(101771737): 100071, 2022.

Srinivasan, K., Currim, F., Ram, S., Lindberg, C., Sternberg, E., Skeath, P., Najafi, B., Razjouyan, J., Lee, H.-K., Foe-Parker, C., Goebel, N., Herzl, R., Mehl, M. R., Gilligan, B., Heerwagen, J., Kampschroer, K., and Canada, K. Feature Importance and Predictive Modeling for Multi-source Healthcare Data with Missing Values. In *Proceedings of the 6th International Conference on Digital Health Conference*, DH '16, pp. 47–54, New York, NY, USA, 2016. Association for Computing Machinery.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., and Carpenter, J. R. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338:b2393, 2009.

Twala, B. E. T. H., Jones, M. C., and Hand, D. J. Good methods for coping with missing data in decision trees. *Pattern Recognition Letters*, 29(7):950–956, 2008.

Van Ness, M., Bosschieter, T. M., Halpin-Gregorio, R., and Udell, M. The Missing Indicator Method: From Low to High Dimensions. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, pp. 5004–5015, New York, NY, USA, 2023. Association for Computing Machinery.

Verdinelli, I. and Wasserman, L. Feature Importance: A Closer Look at Shapley Values and LOCO. *arXiv:2303.05981 [stat]*, 2023.

Vo, T. L., Nguyen, T., Hammer, H. L., Riegler, M. A., and Halvorsen, P. Explainability of Machine Learning Models under Missing Data. *arXiv:2407.00411 [cs]*, 2024.

von Kleist, H., Zamanian, A., Shpitser, I., and Ahmidi, N. Evaluation of Active Feature Acquisition Methods for Static Feature Settings. *arXiv:2312.03619 [cs, stat]*, 2023.

von Kleist, H., Zamanian, A., Shpitser, I., and Ahmidi, N. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings. *Journal of Machine Learning Research*, 26(60):1–84, 2025.

Xie, W., Li, Y., Meng, X., and Zhao, M. Machine learning prediction models and nomogram to predict the risk of in-hospital death for severe DKA: A clinical study based on MIMIC-IV, eICU databases, and a college hospital ICU. *International journal of medical informatics*, 174 (ct4, 9711057):105049, 2023.

Yang, B., Zhu, Y., Lu, X., and Shen, C. A Novel Composite Indicator of Predicting Mortality Risk for Heart Failure Patients With Diabetes Admitted to Intensive Care Unit Based on Machine Learning. *Frontiers in endocrinology*, 13(101555782):917838, 2022.

Yoo, S.-K., Fitzgerald, C. W., Cho, B. A., Fitzgerald, B. G., Han, C., Koh, E. S., Pandey, A., Sfreddo, H., Crowley, F., Korostin, M. R., Debnath, N., Leyfman, Y., Valero, C., Lee, M., Vos, J. L., Lee, A. S., Zhao, K., Lam, S., Olumuyide, E., Kuo, F., Wilson, E. A., Hamon, P., Hennequin, C., Saffern, M., Vuong, L., Hakimi, A. A., Brown, B., Merad, M., Gnjatic, S., Bhardwaj, N., Galsky, M. D., Schadt, E. E., Samstein, R. M., Marron, T. U., Gönen, M., Morris, L. G. T., and Chowell, D. Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data. *Nature Medicine*, pp. 1–12, 2025.

Zamanian, A., von Kleist, H., Ciora, O.-A., Piperno, M., Lancho, G., and Ahmidi, N. Analysis of Missingness Scenarios for Observational Health Data. *Journal of Personalized Medicine*, 14(5):514, 2024.

Zeng, Z., Liu, Y., Yao, S., Liu, J., Xiao, B., Liu, C., and Gong, X. Neural networks based on attention architecture are robust to data missingness for early predicting hospital mortality in intensive care unit patients. *Digital health*, 9 (101690863):20552076231171482, 2023.

Zhang, L., Wang, Z., Zhou, Z., Li, S., Huang, T., Yin, H., and Lyu, J. Developing an ensemble machine learning model for early prediction of sepsis-associated acute kidney injury. *iScience*, 25(9):104932, 2022.

Zhang, S. Nearest neighbor selection for iteratively *k*NN imputation. *Journal of Systems and Software*, 85(11): 2541–2552, 2012.

Zhang, Y., Hu, J., Hua, T., Zhang, J., Zhang, Z., and Yang, M. Development of a machine learning-based prediction model for sepsis-associated delirium in the intensive care unit. *Scientific reports*, 13(1):12697, 2023.

Zhao, Y. and Long, Q. Variable selection in the presence of missing data: imputation-based methods. *WIREs Computational Statistics*, 9(5):e1402, 2017.

Zhou, H., Liu, L., Zhao, Q., Jin, X., Peng, Z., Wang, W., Huang, L., Xie, Y., Xu, H., Tao, L., Xiao, X., Nie, W., Liu, F., Li, L., and Yuan, Q. Machine learning for the prediction of all-cause mortality in patients with sepsis-associated acute kidney injury during hospitalization. *Frontiers in immunology*, 14(101560960):1140755, 2023.

Zhuang, J., Huang, H., Jiang, S., Liang, J., Liu, Y., and Yu, X. A generalizable and interpretable model for mortality risk stratification of sepsis patients in intensive care unit. *BMC medical informatics and decision making*, 23(1): 185, 2023.

A. Background on Feature Importance Metrics

Feature importance metrics can be categorized into local and global approaches (Ewald et al., 2024). Local metrics focus on explaining individual predictions, while global metrics provide insights into the aggregate importance of features in the prediction process. This work centers on global, model-agnostic methods that can be applied to any predictive model.

Feature importance metrics assign an importance value to each feature, reflecting its contribution to prediction performance. Beyond their role in explaining model behavior, feature importance metrics are often used to derive reliable, population-level insights about the data-generating process. They are also closely linked to feature selection methods, as both seek to evaluate the relevance of features within a predictive model. In scenarios involving missing data, the challenges faced in assessing feature importance and selecting features often overlap, allowing similar methodologies to be applied across these tasks.

The field of feature importance metrics is dominated by three main approaches (Ewald et al., 2024). Permutation feature importance evaluates the impact of randomly permuting a features values on model performance. It is model-agnostic and widely used for its simplicity but may underestimate feature importance when features are highly correlated. Marginalization-based methods assess importance by marginalizing out the feature of interest, often requiring strong assumptions about the data-generating process to be valid. Finally, model refitting methods compare model performance before and after removing specific features. Techniques like LOCO (Lei et al., 2018) fall into this category, directly assessing the dependency of predictions on a given feature.

B. The Leave-One-Covariate-Unmeasured (LOCU) Metric

In this appendix, we introduce an alternative estimand to the observed data LOCO metric, $LOCO^{OD}$. Unlike $LOCO^{OD}$, which evaluates the impact of leaving one covariate out of the input to the classifier, this new metric considers the scenario where the measurement of a covariate is entirely omitted.

We denote this alternative metric as *leave-one-covariate-unmeasured (LOCU)*. It is defined as follows:

LOCU:

$$LOCU = \mathbb{E}\left[L\left(Y, f_{cl}(X_{(-j)}, R_{(-j)})\right)\right] - \mathbb{E}\left[L\left(Y, f_{cl}(X, R)\right)\right],$$

where $X_{(-j)}$ denotes the counterfactual value X would have taken if $R_j = 0$. This represents a scenario where all features are measured according to the current measurement policy, except for feature j, which is no longer measured. LOCU thus quantifies the change in prediction loss if feature j is omitted from measurement. This can provide valuable insights for data collection decisions. For instance, in medical settings, if a particular test is costly, the metric can help assess the impact of discontinuing the test on prediction performance.

LOCU differs from the observed data LOCO metric $LOCO^{OD}$ in cases where the measurement process of other features depends on the measurement of feature j (i.e., there are causal arrows from R_j to other missingness indicators). For example, suppose feature i is measured only if a test on feature j yields a positive result. In this situation, omitting the measurement of feature j would also prevent the measurement of feature i, potentially leading to a more substantial change in prediction performance.

While causal inference methods can be used to identify and estimate LOCU in such scenarios, this aspect is considered outside the scope of this work.

C. Proofs for Unbiased Estimation Results for the LOCO Metrics

In this Appendix, we add mathematical proofs that complement the discussion about estimation for $LOCO^{OD}$ and $LOCO^{FD}$. We define θ as a general estimand and $\hat{\theta}$ as its corresponding estimator. The bias of the estimator is given by:

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta.$$

Our results extend to a broader class of full data feature importance metrics of the form:

$$\theta^{FD} = \mathbb{E}\left[g(\{f_{cl,s}(X_{(1),s}, R_s = \vec{1}) : s \in \mathscr{S}\}, Y)\right],$$

where g can be any function and $\mathscr{S}$ represents the set of all feature subsets. This definition encompasses a wide range of metrics, including LOCO and Shapley values. Notably, Shapley values can be expressed as a weighted average of LOCO values across submodels (Verdinelli & Wasserman, 2023). Additionally, we consider observed data feature importance metrics θ^{OD} of the same form, with $X_{(1),s}$ and $R_s = \vec{1}$ replaced by X_s and R_s.

C.1. Conditional Mean Imputation for Full Data Feature Importance Metrics

Firstly, we demonstrate that conditional mean imputation results in biased estimation of full data feature importance metrics. Our analysis is based on the formulation used for the unbiased multiple imputation estimator:

$$\begin{aligned}\theta^{FD} &= \sum_{X_{(1)},Y} g(\{f_{cl,s}(X_{(1),s}, R_s = \vec{1}) : s \in \mathscr{S}\}, Y) p(X_{(1)}, Y) \\ &= \sum_{X_m,X_o,Y,R} g(\{f_{cl,s}(X_{m\wedge s}, X_{o\wedge s}, R_s = \vec{1}) : s \in \mathscr{S}\}, Y) p(X_m|X_o, Y, R) p(X_o, Y, R),\end{aligned}$$

where $p(X_m|X_o, Y, R)$ represents the imputation density. When the imputation model is learned, one can apply Monte Carlo integration to obtain an unbiased estimator $\hat{\theta}^{FD}$ for θ^{FD}.

Conditional mean imputation, however, simplifies the above expression to:

$$\theta^{FD} \approx \sum_{X_m,X_o,Y,R} g(\{f_{cl,s}(\mathbb{E}[X_{m\wedge s}|X_o, Y, R], X_{o\wedge s}, R_s = \vec{1}) : s \in \mathscr{S}\}, Y) p(X_o, Y, R),$$

which assumes that the expectation operator can be pulled inside the functions g and f_{cl}. This assumption is only valid if both functions are linear, which is generally not the case. Consequently, the use of conditional mean imputation introduces bias.

C.2. Mean and Conditional Mean Imputation for Observed Data Feature Importance Metrics

Next, we demonstrate that mean imputation (or a broader class of imputation methods) does not introduce bias for observed data feature importance metrics θ^{OD}. Since θ^{OD} is defined as a function of the classifier f_{cl}, the reported feature importance metric reflects classifier-specific importance rather than a general global measure.

A commonly used class of classifiers, referred to as impute-then-regress classifiers (Le Morvan et al., 2021), first impute the missing values and subsequently classify. If the classifier is sufficiently flexible and the dataset is large enough, the choice of imputation method becomes inconsequential, as no new information is introduced. Thus, any classifier that maps X_s and R_s to Y can be used, including those employing mean imputation, without inducing bias.

However, this conclusion holds only if imputation is performed within the classifier itself using only its input features. If conditional mean imputation is applied to the entire dataset before choosing the subset for the classifier, bias arises. Let the imputed features be:

$$X_i' = \begin{cases} X_i & \text{if } R_i = 1, \\ \hat{\mathbb{E}}_n[X_i|X_o, R = \vec{1}] & \text{if } R_i = 0. \end{cases}$$

In this case, the imputed variable X_i' effectively becomes a function of other features, $X_i' \equiv f(X_o, R = \vec{1})$. The resulting bias in the estimator is given by:

$$\begin{aligned}\text{Bias}(\hat{\theta}^{OD}) &= \mathbb{E}[\hat{\theta}^{OD}] - \theta^{OD} \\ &= \mathbb{E}\left[\hat{\mathbb{E}}_n[g(\{f_{cl,s}(X_s', R_s) : s \in \mathscr{S}\}, Y)]\right] - \mathbb{E}\left[g(\{f_{cl,s}(X_s, R_s) : s \in \mathscr{S}\}, Y)\right].\end{aligned}$$

This bias is generally nonzero because X_s' depends on all features rather than just X_s. Consequently, applying conditional mean imputation before classification results in biased estimates.

D. Impact of Classifier Training on LOCO Metrics

The LOCO metrics in Eqs. 2 and 3 highlight the importance of aligning classifier training with the conditions of deployment. Specifically:

Full data feature importance metrics require a classifier f_{cl}^{FD} designed to handle fully available features:

$$f_{cl}^{FD}(X_{(1)}, R = \vec{1}) = p(Y \mid X_{(1)}, R = \vec{1}).$$

Observed data feature importance metrics require a classifier f_{cl}^{OD} capable of handling missing data:

$$f_{cl}^{OD}(X, R) = p(Y \mid X, R).$$

Training these classifiers introduces unique challenges. For f_{cl}^{FD}, the absence of $X_{(1)}$ in the training data creates a missing data problem, complicating the training process. In contrast, training f_{cl}^{OD} involves dealing with missing values ("?") in X. While some models inherently handle missing data (Twala et al., 2008), a common alternative is to impute missing values. However, this approach can introduce discontinuities, which hinder the learning process.

To address these challenges, "impute-then-regress" classifiers (Le Morvan et al., 2021) provide an effective solution. These classifiers leverage the following decomposition:

$$f_{cl}^{OD}(X, R) = p(Y \mid X_m = E[X_m \mid X_o, R], X_o, R).$$

Here, missing values (X_m) are replaced with their conditional expectations based on observed values X_o (and missingness indicators R). This imputation is intrinsic to the classifier and differs from post hoc imputation used during feature importance evaluation. Conditional mean imputation is often a practical and effective choice in this context.

The improper training of classifiers can introduce additional bias in LOCO metric estimations, particularly for full data feature importance metrics when f_{cl}^{OD} is used instead of f_{cl}^{FD}. These biases are exacerbated in missing data scenarios such as MAR and MNAR, compared to MCAR. Awareness of these limitations is crucial for ensuring robust and interpretable feature importance assessments.

E. Proof of Lemma 4.1

In this appendix, we prove Lemma 4.1 which states the form of the feature measurement importance gradient.

Proof. The gradient simplifies as follows:

$$\begin{aligned}
G_j &= \nabla_{\delta_j} \mathbb{E}\Big[-L\Big(Y, f_{cl}(X_{(\pi_{\delta_j})}, R_{(\pi_{\delta_j})})\Big)\Big]\Big|_{\delta_j=1} = \\
&\overset{*_1}{=} \nabla_{\delta_j} \sum_{R,Y,X_{(1)}} -L\left(Y, f_{cl}(X,R)\right) \pi_{\delta_j}(R|X_{(1)}, Y) p(X_{(1)}, Y)\Big|_{\delta_j=1} = \\
&= \sum_{R,Y,X_{(1)}} -L\left(Y, f_{cl}(X,R)\right) \prod_{i=1}^{d} \pi_i(R_i|\underline{R}_{i-1}, \underline{X}_{i-1}) \nabla_{\delta_j} \pi_{j,\delta_j}(R_j|\underline{R}_{j-1}, \underline{X}_{j-1})\Big|_{\delta_j=1} p(X_{(1)}, Y) = \\
&\overset{*_2}{=} \sum_{R,Y,X_{(1)}} -L\left(Y, f_{cl}(X,R)\right) \prod_{i=1}^{d} \pi_i(R_i|\underline{R}_{i-1}, \underline{X}_{i-1}) \pi_j(R_j|\underline{R}_{j-1}, \underline{X}_{j-1}) (-1)^{1-R_j} (1 - \pi_j(R_j|\underline{R}_{j-1}, \underline{X}_{j-1})) p(X_{(1)}, Y) = \\
&= \mathbb{E}\Big[-L\left(Y, f_{cl}(X,R)\right) \cdot (-1)^{1-R_j} (1 - \pi_j(R_j|\underline{R}_{j-1}, \underline{X}_{j-1}))\Big].
\end{aligned}$$

with the following explanations:

In $*1$) we use the g-formula (Robins, 1986), which requires the SUTVA assumption, exchangeability (ensured by the MAR assumption) and in general positivity. The last assumption of positivity is, however, always guaranteed due to the form of our intervention. Positivity in general necessitates that $\forall r_j, \underline{r}_{j-1}, \underline{x}_{j-1}$ with $p(\underline{R}_{j-1} = \underline{r}_{j-1}, \underline{X}_{j-1} = \underline{x}_{j-1}) > 0$ and $\pi_{j,\delta_j}(R_j = r_j | \underline{R}_{j-1} = \underline{r}_{j-1}, \underline{X}_{j-1} = \underline{x}_{j-1}) > 0$, that we must also have $\pi_j(R_j = r_j | \underline{R}_{j-1} = \underline{r}_{j-1}, \underline{X}_{j-1} = \underline{x}_{j-1}) > 0$. This always holds, as setting with $\pi_j = 0$ yield

$$\pi_{j,\delta_j} = \frac{\delta_j \cdot 0}{\delta_j \cdot 0 + 1 - 0} = 0,$$

settings with $\pi_j = 1$ give

$$\pi_{j,\delta_j} = \frac{\delta_j \cdot 1}{\delta_j \cdot 1 + 1 - 1} = 1.$$

and settings with $\pi_j \in (0,1)$ imply

$$\pi_{j,\delta_j} = \frac{\delta_j \cdot \pi_j}{\delta_j \cdot \pi_j + 1 - \pi_j} > 0$$

since $\delta_j \cdot \pi_j > 0$, and $\delta_j \cdot \pi_j + 1 - \pi_j > 0$ holds.

In $*1$), we also use sums for the integrals to keep notation simple, but for continuous X one can replace the sums with proper integrals.

In $*2$), we used the following simplifications of the gradient of the perturbed measurement policy. We first look at the case $R_j = 1$:

$$\begin{aligned}\nabla_{\delta_j} \pi_{j,\delta_j}(R_j = 1 | \underline{R}_{j-1}, \underline{X}_{j-1})\Big|_{\delta_j=1} &= \nabla_{\delta_j} \frac{\delta_j \cdot \pi_j}{\delta_j \cdot \pi_j + 1 - \pi_j}\Big|_{\delta_j=1} \overset{*_{2.1}}{=} \frac{\pi_j(\delta_j \cdot \pi_j + 1 - \pi_j) - (\delta_j \cdot \pi_j) \cdot (\pi_j)}{(\delta_j \cdot \pi_j + 1 - \pi_j)^2}\Big|_{\delta_j=1} = \\ &= \frac{\pi_j(1-\pi_j)}{(\delta_j \cdot \pi_j + 1 - \pi_j)^2}\Big|_{\delta_j=1} = \pi_j(1-\pi_j)\end{aligned}$$

where $*2.1$) follows from the quotion rule of differentiation.

For $R_j = 0$, we find:

$$\begin{aligned}\nabla_{\delta_j} \pi_{j,\delta_j}(R_j = 0 | \underline{R}_{j-1}, \underline{X}_{j-1})\Big|_{\delta_j=1} &= \nabla_{\delta_j}(1 - \pi_{j,\delta_j}(R_j = 1 | \underline{R}_{j-1}, \underline{X}_{j-1}))\Big|_{\delta_j=1} = -\nabla_{\delta_j} \pi_{j,\delta_j}(R_j = 1 | \underline{R}_{j-1}, \underline{X}_{j-1})\Big|_{\delta_j=1} = \\ &= -\pi_j(1-\pi_j) = -\Big(1 - \pi_j(R_j = 0 | \underline{R}_{j-1}, \underline{X}_{j-1})\Big)\pi_j(R_j = 0 | \underline{R}_{j-1}, \underline{X}_{j-1})\end{aligned}$$

Bringing both together, we find the transformation in $*2$) and thus completes the proof.

□

F. Feature Measurement Importance Gradient for Time-series Settings

In this appendix, we extend the feature measurement importance gradient to time-series settings. For time-step t, we denote the observed features as $X^t \in (\mathbb{R} \cup \{\text{"?"}\})^d$ and all features up to time t as $\underline{X}^t = \{X^1, \ldots, X^t\}$. The time-dependent label is $Y^t \in \{0,1\}$. A classifier predicts Y^t at each time step. This classifier is $f_{cl}^t(\underline{X}^t, \underline{R}^t)$, where $\underline{R}^t$ denotes the missingness indicators up to t.

We redefine the FMIG for the time-series settings as:

Feature Measurement Importance Gradient (FMIG) G_j**:**

$$G_j = \nabla_{\delta_j} \mathbb{E}\left[\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t_{(\pi_{\delta_j})}, \underline{R}^t_{(\pi_{\delta_j})})\right)\right]\Bigg|_{\delta_j = \delta_j^*} \tag{6}$$

where δ_j^* is the value of δ_j that represents no perturbation.

F.1. Measurement Process Assumptions

We make the following assumptions about the factorization of the missingness process:

$$\pi(R|X_{(1)}, Y) = \prod_{t=1}^{T} \pi^t(R^t|\underline{R}^{t-1}, \underline{X}^{t-1}), \tag{7}$$

where we assume the multivariate missingness indicators further factorize into conditionally independent terms:

$$\pi^t(R^t|\underline{R}^{t-1}, \underline{X}^{t-1}) = \prod_{i=1}^{d} \pi_i^t(R_i^t|\underline{R}^{t-1}, \underline{X}^{t-1}). \tag{8}$$

We thus assume that at each time-point, the measurement of feature i will depend on all observed features in the past.

F.2. Measurement Policy Interventions

To redefine perturbation interventions for the time-series setting by now jointly intervening on a feature at all time-steps. We obtain:

$$\pi_{\delta_j}(R|X_{(1)}, Y) = \prod_{t=1}^{T} \pi_{j,\delta_j}^t(R_j^t|\underline{R}^{t-1}, \underline{X}^{t-1}) \prod_{i \neq j}^{d} \pi_i^t(R_i^t|\underline{R}^{t-1}, \underline{X}^{t-1}). \tag{9}$$

where each individual intervention follows the same odds ratio parameterization as in the main body.

F.3. Identification

Using the perturbation interventions, the feature measurement importance gradient G_j can be identified for the time-series setting as the following corollary states:

Corollary F.1. ***(Identification of the Feature Measurement Importance Gradient (FMIG) for Time-series Settings)*** *Under SUTVA and the assumptions about the missingness process from Eqs. 7 and 8, the feature measurement importance gradient (FMIG) for feature j is identified as:*

$$\begin{aligned} G_j &= \nabla_{\delta_j} \mathbb{E}\left[\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}_{(\pi_{\delta_j})}^t, \underline{R}_{(\pi_{\delta_j})}^t)\right)\right]\Bigg|_{\delta_j=1} \\ &= \mathbb{E}\left[\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t, \underline{R}^t)\right) \cdot \left(\sum_{k=1}^{t} (-1)^{1-R_j^k}(1 - \pi_j^k(R_j^k|\underline{R}^{k-1}, \underline{X}^{k-1})\right)\right] \end{aligned} \tag{10}$$

Proof. The gradient simplifies as follows:

$$
\begin{aligned}
G_j &= \nabla_{\delta_j}\mathbb{E}\left[\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t_{(\pi_{\delta_j})}, \underline{R}^t_{(\pi_{\delta_j})})\right)\right]\Bigg|_{\delta_j=1} = \\
&= \nabla_{\delta_j}\sum_{R,Y,X_{(1)}}\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t_{(\pi_{\delta_j})}, \underline{R}^t_{(\pi_{\delta_j})})\right)\pi_{\delta_j}(R|X_{(1)},Y)p(X_{(1)},Y)\Bigg|_{\delta_j=1} = \\
&\overset{*_1}{=} \sum_{R,Y,X_{(1)}}\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t, \underline{R}^t)\right)\left(\nabla_{\delta_j}\prod_{\tau=1}^{t}\pi^\tau_{j,\delta_j}(R_j^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})\Bigg|_{\delta_j=1}\right)\prod_{\tau=1}^{t}\prod_{i=1}^{d}\pi_i^\tau(R_i^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})p(X_{(1)},Y) = \\
&\overset{*_2}{=} \sum_{R,Y,X_{(1)}}\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t, \underline{R}^t)\right)\sum_{k=1}^{t}\left(\prod_{\tau=1,\tau\neq k}^{t}\pi^\tau_{j,\delta_j}(R_j^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})\nabla_{\delta_j}\pi^k_{j,\delta_j}(R_j^k|\underline{R}^{k-1},\underline{X}^{\tau-1})\Bigg|_{\delta_j=1}\right) \\
&\quad\cdot\prod_{\tau=1}^{t}\prod_{i=1}^{d}\pi_i^\tau(R_i^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})p(X_{(1)},Y) = \\
&\overset{*_3}{=} \sum_{R,Y,X_{(1)}}\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t, \underline{R}^t)\right)\sum_{k=1}^{t}\left(\prod_{\tau=1}^{t}\pi^\tau_j(R_j^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})(-1)^{1-R_j^k}\left(1-\pi_j^k(R_j^k|\underline{R}^{k-1},\underline{X}^{k-1})\right)\right) \\
&\quad\cdot\prod_{\tau=1}^{t}\prod_{i=1}^{d}\pi_i^\tau(R_i^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})p(X_{(1)},Y) = \\
&= \sum_{R,Y,X_{(1)}}\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t, \underline{R}^t)\right)\sum_{k=1}^{t}\left((-1)^{1-R_j^k}\left(1-\pi_j^k(R_j^k|\underline{R}^{k-1},\underline{X}^{k-1})\right)\right) \\
&\quad\cdot\prod_{\tau=1}^{t}\pi^\tau_j(R_j^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})\prod_{i=1}^{d}\pi_i^\tau(R_i^\tau|\underline{R}^{\tau-1},\underline{X}^{\tau-1})p(X_{(1)},Y) = \\
&= \mathbb{E}\left[\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t, \underline{R}^t)\right)\cdot\left(\sum_{k=1}^{t}(-1)^{1-R_j^k}(1-\pi_j^k(R_j^k|\underline{R}^{k-1},\underline{X}^{k-1})\right)\right].
\end{aligned}
$$

with the following explanations:

In $*1$), we used the fact that only the part of π that models missingness for feature j depends on the gradient. We further use the fact that $\sum_{t=1}^{T} -L\left(Y^t, f_{cl}^t(\underline{X}^t, \underline{R}^t)\right)$ is independent of any missingness indicators R^τ with $\tau > t$.

In $*2$), we apply the product rule of differentiation.

In $*3$), we leverage the same simplification of $\nabla_{\delta_j}\pi^k_{j,\delta_j}(R_j^k|\underline{R}^{k-1},\underline{X}^{\tau-1})$ as was used in Appendix E for the simpler static feature setting. □

G. Experiment Details

In this appendix, we provide a detailed description of the experimental setup, including the data generation process, missingness mechanisms, training configurations, and the methods used to estimate feature importance metrics. These details align with the experiments presented in Section 5 and Figures 1, 2, 3 and 4.

G.1. Experiment Setup

The experiments are based on a synthetic time-series dataset with three features ($d = 3$) and three time-points ($T = 3$). For time-step t, we denote the observed features as $X^t \in (\mathbb{R} \cup \{\text{"?"}\})^d$ and all features up to time t as $\underline{X}^t = \{X^1, \ldots, X^t\}$. The time-dependent label is $Y^t \in \{0, 1\}$. A classifier predicts Y^t at each time step, and for the observed data, this classifier is $f_{cl}^{OD}(\underline{X}^t, \underline{R}^t)$, where $\underline{R}^t$ denotes the missingness indicators up to t.

The data generation process is defined as follows. Features $X^t_{(1),i}$ are independent across dimensions and generated recursively:

$$X^t_{(1),i} = \begin{cases} \gamma_i X^{t-1}_{(1),i} + (1-\gamma_i)\epsilon^t_i, & \text{if } t > 1, \\ \epsilon^t_i, & \text{if } t = 1, \end{cases}$$

where $\epsilon^t_i \sim \mathcal{N}(0, \sigma = 1)$ and γ_i controls the temporal dependence for feature i. We let $X^0_{(1)} \equiv \vec{0}$. Labels Y^t are binary and determined as:

$$p(Y^t = 1) = \begin{cases} 1, & \text{if } \sum_i W_i X^t_{(1),i} + \sum_i W_i X^{t-1}_{(1),i} > 0, \\ 0.2, & \text{otherwise.} \end{cases}$$

This label distribution simulates a scenario where not all data points are equally easy to classify. Parameters W represent feature importances, which vary across experiments. We generate 100,000 data points and split them into 30% for training the classifier, 30% for training the measurement policy, and 40% for testing.

G.2. Missingness Mechanisms

The missingness process in these experiments follows an MCAR or MNAR scenario. For the MNAR case, the missingness probabilities are modeled as:

$$\pi(R^t_i = 1 | X^{t-1}_{(1)}) = \sigma(\alpha_1 + \alpha_2 X^{t-1}_{(1)}),$$

where σ denotes the logistic function. In this example, $X^{t-1}_{(1)}$ directly influences the missingness of R^t_i, making the process MNAR. The exact missingness probabilities used in each experiment are summarized in Table 1.

G.3. Training and Evaluation

We used an "impute-then-regress" classifier (Le Morvan et al., 2021) with zero imputation and a temporal convolutional network (TCN) (Bai et al., 2018) to classify labels Y^t. The classifier uses four layers, with 32 channels per layer, a batch size of 2,000, dropout rate of 0.2, and a learning rate of 0.001.

Feature importance is evaluated using the observed and full data LOCO and the feature measurement importance gradient (FMIG). When we report the full data LOCO metric and do not mention the estimation method, it refers to the ground truth feature importance, as computed from the complete dataset without missingness. Binary cross-entropy is used as the loss function when calculating the LOCO metrics and the FMIG. We report confidence intervals using the nonparametric bootstrap with 50 samples. However, due to computational cost, we do not retrain the classifier for each bootstrap iteration. As a result, the confidence intervals may be overly optimistic.

G.4. Experiment Configurations

The detailed configurations for each experiment, including the data-generating process parameters (W, γ) and missingness mechanisms (π), are provided in Table 1.

H. Experiment Regarding Classifier Training Procedures

To illustrate the impact of proper classifier training, we modified the design of Experiment 3 by training the classifier on both the observed dataset (denoted as f^{OD}_{cl}) and the full dataset (denoted as f^{FD}_{cl}). This experiment, conducted under a violation of the positivity assumption and a high percentage of missingness, demonstrates how improper classifier training affects feature importance metrics. The results, shown in Figure 4, reveal that classifiers trained on observed data yield lower full-data LOCO metrics compared to those trained on the full dataset. Importantly, both LOCO metrics were evaluated on the ground truth dataset without missingness, highlighting how improper training introduces additional bias on top of the bias already inherent in estimating full data LOCO metrics from observed data. The consequences are expected to be even stronger for MAR and MNAR missingness scenarios, compared to this MCAR experiment.

These findings underscore the critical importance of aligning classifier training with the target metric to ensure accurate and meaningful feature importance evaluations. In practice, however, full alignment is often unfeasible due to the absence of ground truth data. Nonetheless, being aware of this limitation is crucial, as it allows for adjustments of training approaches to mitigate its consequences.

Experiment	Data-Generating Process	Missingness Mechanisms
Exp 1	$W_1 = \frac{1}{6}; \gamma_1 = 0.2$ $W_2 = \frac{2}{6}; \gamma_2 = 0.2$ $W_3 = \frac{3}{6}; \gamma_3 = 0.2$	$\pi(R_1^t = 1) = 0.75$ $\pi(R_2^t = 1) = 0.5$ $\pi(R_3^t = 1) = 0.3$
Exp 2	$W_1 = \frac{1}{6}; \gamma_1 = 0.2$ $W_2 = \frac{2}{6}; \gamma_2 = 0.2$ $W_3 = \frac{3}{6}; \gamma_3 = 0.2$	$\pi(R_1^t = 1 \mid X_{(1),2}^{t-1}) = \sigma(1.7 - 3.0 X_{(1),2}^{t-1})$ $\pi(R_2^t = 1 \mid X_{(1),2}^{t-1}) = \sigma(-3.0 X_{(1),2}^{t-1})$ $\pi(R_3^t = 1) = 0.3$
Exp 3	$W_1 = \frac{1}{6}; \gamma_1 = 0.2$ $W_2 = \frac{2}{6}; \gamma_2 = 0.2$ $W_3 = \frac{3}{6}; \gamma_3 = 0.2$	$\pi(R_1^t = 1) = 0.4$ $\pi(R_2^t = 1) = 0.2$ $\pi(R_3^t = 1) = 0.1$
Exp 4	$W_1 = \frac{1}{3}; \gamma_1 = 0.2$ $W_2 = \frac{1}{3}; \gamma_2 = 0.2$ $W_3 = \frac{1}{3}; \gamma_3 = 1.0$	$\pi(R_1^t = 1) = 1.0$ $\pi(R_2^t = 1) = 0.5$ $\pi(R_3^1 = 1) = 1.0; \pi(R_3^t = 1) = 0.5$ (for $t > 1$)

Table 1. Experiment-specific details of the data-generating processes and missingness mechanisms.

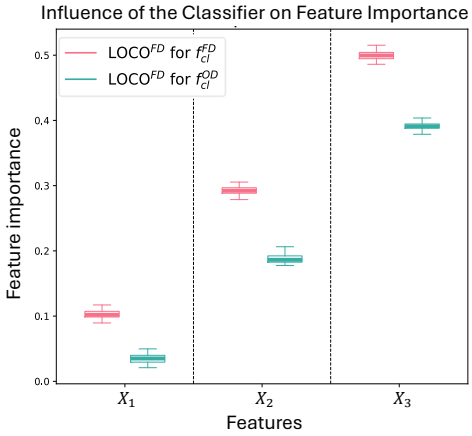


Figure 4. Impact of classifier training on feature importance: The full data LOCO metric derived from a classifier trained on observed data underestimates feature importance compared to a classifier trained on ground-truth data without missingness. Both LOCO metrics are estimated on the ground truth data.

A.2. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings

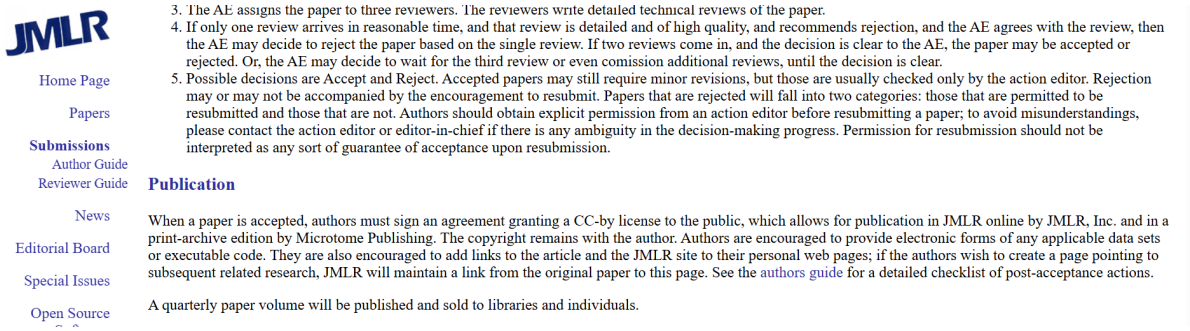
This publication appeared in the *Journal of Machine Learning Research (JMLR)*, Volume 26, Article 60, in 2025. A one-page summary is provided in Section 3.2 of this thesis.

An earlier version of this article was made publicly available as a preprint on arXiv [50], and has since been updated to reflect the final version published in JMLR.

Full citation:
von Kleist, H., Zamanian, A., Shpitser, I., and Ahmidi, N. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings. *Journal of Machine Learning Research*, 26(60):1–84, 2025.

License and Reprint Permission:
JMLR publishes all accepted articles under the Creative Commons Attribution 4.0 International License (CC BY 4.0), and the copyright remains with the authors. This license allows unrestricted reuse, distribution, and reproduction, including incorporation in dissertations.

A statement confirming this policy is available on the JMLR website under the "Submissions" section of the Author Information page: `https://www.jmlr.org/author-info.html`. A screenshot of the relevant section is included below as proof of reprint permission. Furthermore, the details of the CC License are shown below.


JMLR

Home Page
Papers
Submissions
Author Guide
Reviewer Guide
News
Editorial Board
Special Issues
Open Source

3. The AE assigns the paper to three reviewers. The reviewers write detailed technical reviews of the paper.
4. If only one review arrives in reasonable time, and that review is detailed and of high quality, and recommends rejection, and the AE agrees with the review, then the AE may decide to reject the paper based on the single review. If two reviews come in, and the decision is clear to the AE, the paper may be accepted or rejected. Or, the AE may decide to wait for the third review or even comission additional reviews, until the decision is clear.
5. Possible decisions are Accept and Reject. Accepted papers may still require minor revisions, but those are usually checked only by the action editor. Rejection may or may not be accompanied by the encouragement to resubmit. Papers that are rejected will fall into two categories: those that are permitted to be resubmitted and those that are not. Authors should obtain explicit permission from an action editor before resubmitting a paper; to avoid misunderstandings, please contact the action editor or editor-in-chief if there is any ambiguity in the decision-making progress. Permission for resubmission should not be interpreted as any sort of guarantee of acceptance upon resubmission.

Publication

When a paper is accepted, authors must sign an agreement granting a CC-by license to the public, which allows for publication in JMLR online by JMLR, Inc. and in a print-archive edition by Microtome Publishing. The copyright remains with the author. Authors are encouraged to provide electronic forms of any applicable data sets or executable code. They are also encouraged to add links to the article and the JMLR site to their personal web pages; if the authors wish to create a page pointing to subsequent related research, JMLR will maintain a link from the original paper to this page. See the authors guide for a detailed checklist of post-acceptance actions.

A quarterly paper volume will be published and sold to libraries and individuals.

Screenshot from the official JMLR submission policy page. This serves as documentation of reprint permission.

CC BY 4.0

Attribution 4.0 International

Deed

Canonical URL : https://creativecommons.org/licenses/by/4.0/

See the legal code

You are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation .

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Notice

This deed highlights only some of the key features and terms of the actual license. It is not a license and has no legal value. You should carefully review all of the terms and conditions of the actual license before using the licensed material.

Creative Commons is not a law firm and does not provide legal services. Distributing, displaying, or linking to this deed or the license that it summarizes does not create a lawyer-client or any other relationship.

Creative Commons is the nonprofit behind the open licenses and other legal tools that allow creators to share their work. Our legal tools are free to use.

- Learn more about our work
- **Learn more about CC Licensing**
- Support our work
- Use the license for your own material.
- Licenses List
- Public Domain List

Footnotes

appropriate credit — If supplied, you must provide the name of the creator and attribution parties, a copyright notice, a license notice, a disclaimer notice, and a link to the material. CC licenses prior to Version 4.0 also require you to provide the title of the material if supplied, and may have other slight differences.

- More info

indicate if changes were made — In 4.0, you must indicate if you modified the material and retain an indication of previous modifications. In 3.0 and earlier license versions, the indication of changes is only required if you create a derivative.

- Marking guide
- More info

technological measures — The license prohibits application of effective technological measures, defined with reference to Article 11 of the WIPO Copyright Treaty.

- More info

exception or limitation — The rights of users under exceptions and limitations, such as fair use and fair dealing, are not affected by the CC licenses.

- More info

publicity, privacy, or moral rights — You may need to get additional permissions before using the material as you intend.

- More info

Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings

Henrik von Kleist[1,2,3] HENRIK.VONKLEIST@HELMHOLTZ-MUNICH.DE
Alireza Zamanian[2,4] ALIREZA.ZAMANIAN@IKS.FRAUNHOFER.DE
Ilya Shpitser[3] ISHPITS1@JHU.EDU
Narges Ahmidi[1,3,4] NARGES.AHMIDI@HELMHOLTZ-MUNICH.DE
[1] *Institute of AI for Health, Helmholtz Munich - German Research Center for Environmental Health, 85764 Neuherberg, Germany*
[2] *TUM School of Computation, Information and Technology, Technical University of Munich, 85748 Garching, Germany*
[3] *Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA*
[4] *Fraunhofer Institute for Cognitive Systems IKS, 80686 Munich, Germany*

Editor: Jin Tian

Abstract

Machine learning methods often assume that input features are available at no cost. However, in domains like healthcare, where acquiring features could be expensive or harmful, it is necessary to balance a feature's acquisition cost against its predictive value. The task of training an AI agent to decide which features to acquire is called active feature acquisition (AFA). By deploying an AFA agent, we effectively alter the acquisition strategy and trigger a distribution shift. To safely deploy AFA agents under this distribution shift, we present the problem of active feature acquisition performance evaluation (AFAPE). We examine AFAPE under i) a no direct effect (NDE) assumption, stating that acquisitions do not affect the underlying feature values; and ii) a no unobserved confounding (NUC) assumption, stating that retrospective feature acquisition decisions were only based on observed features. We show that one can apply missing data methods under the NDE assumption and offline reinforcement learning under the NUC assumption. When NUC and NDE hold, we propose a novel semi-offline reinforcement learning framework. This framework requires a weaker positivity assumption and introduces three new estimators: A direct method (DM), an inverse probability weighting (IPW), and a double reinforcement learning (DRL) estimator.

Keywords: active feature acquisition, semi-offline reinforcement learning, dynamic testing regimes, missing data, causal inference, semiparametric theory

1. Introduction

Machine learning methods typically assume that the full set of input features will be readily available after deployment, with little to no cost. This is, however, not always the case, as acquiring features may impose a significant cost. In such situations, the predictive value of a feature should be balanced against its acquisition cost. In the medical diagnostics context, the cost of feature acquisition (e.g., for a biopsy test) may include not only monetary cost but also the potential adverse harm to patients. This is why physicians acquire certain features,

e.g., via biopsies, MRI scans, or lab tests, only when their diagnostic values outweigh their costs or risks. The challenge is exacerbated when prediction must be made regarding a large number of diverse outcomes with different sets of informative features. Going back to the medical example, a typical emergency department is able to diagnose thousands of different diseases based on a large set of possible observations. For every new emergency patient with ambiguous symptoms, clinicians must narrow down their search for a proper diagnosis via step-by-step feature acquisitions.

Active feature acquisition (AFA) addresses this problem by designing two AI systems: i) a so-called *AFA agent*, deciding which features must be observed while balancing information gain vs. feature acquisition cost; ii) an ML prediction model, often a classifier, that solves the prediction task based on the acquired set of features. To elucidate the AFA process, we present a hypothetical and simplified scenario of diagnosing heart attacks.

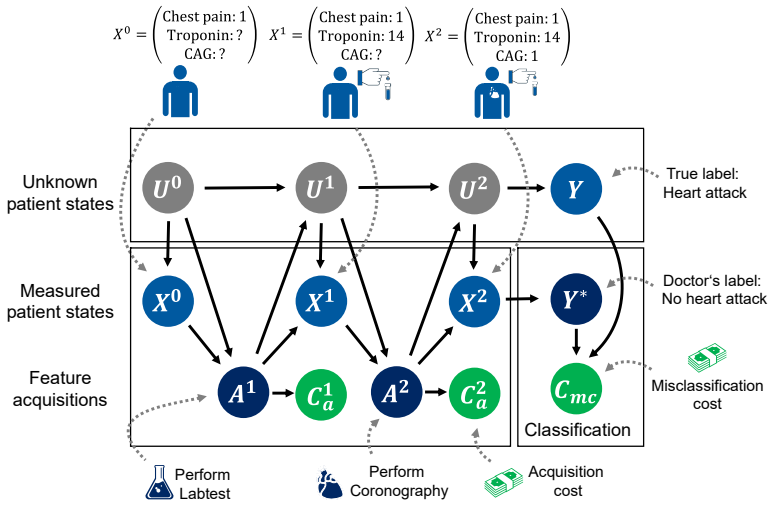


Figure 1: AFA process for a simplified hypothetical heart attack diagnosis example. A patient with chest pain (X^0) prompts the doctor to first order a troponin lab test (A^1) and, upon reviewing the result (X^1), to also order a coronography (CAG) (A^2). The feature acquisitions A^1 and A^2 produce feature acquisition costs C_a^1 and C_a^2. After the acquisition process concludes, the doctor makes a diagnosis Y^*, which, if different from the true underlying condition Y, produces a misclassification cost C_{mc}.

1.1 Heart Attack Diagnosis Example

Figure 1 presents the partially observable decision process that encapsulates the sequential decision-making aspect of the AFA problem for a heart attack diagnosis example. Upon arrival at the hospital, a patient with an unknown health state (U^0) exhibits the symptom of chest pain (X^0 = "chest pain"). At this stage, no additional information is available. The attending doctor decides to order a troponin lab test (A^1 = "acquire troponin") as

part of the feature acquisition process. The laboratory test incurs a feature acquisition cost (C_a^1 = "\$100"). Subsequently, upon reviewing the results of the lab test (X^2), the doctor decides that a coronography (A^2), an invasive imaging procedure, is necessary. Notably, the feature acquisition cost (C_a^2) for this procedure may be substantially higher due to the potential harm to the patient. After the completion of the feature acquisition process, a diagnosis of whether the patient is experiencing a heart attack is performed. A (hypothetical) misclassification cost C_{mc} arises if the diagnosis Y^* and the true condition Y differ.

In general, medical tests may also impact the patient's health (illustrated by the edges $A^t \to U^t$). The invasive coronography from our example may, for example, cause bleeding, infections, hypotension or other problems (Tavakol et al., 2012). Such an effect is denoted as a *direct effect*, and we refer to its absence as the *no direct effect (NDE)* assumption.

Furthermore, the decision to perform a clinical test A^t may not solely rely on past observed variables X^τ ($\tau < t$) but can also depend on past unobserved variables U^τ (illustrated by the edges $U^{t-1} \to A^t$) or even other factors. For example, a doctor might base acquisition decisions on information that is partially not recorded in the medical database. We refer to the assumption that acquisitions are only determined by past observed variables as the *no unobserved confounding (NUC)* assumption.

1.2 Paper Goal

We investigate the evaluation of AFA agents under the distribution shift that occurs since the AFA agent makes different acquisition decisions than the doctors who were responsible for collecting the retrospective data set. The focus of the paper is thus not to design new AFA agents and classifiers but to estimate the performance of *any* AFA agent and classifier at deployment. This means the doctor should be informed, for example, how many wrong diagnoses are to be expected or how much acquisition cost will be incurred on average if an AFA system is deployed.

The paper has two primary objectives: i) Identification, which involves determining the assumptions that enable the unbiased representation of costs from the retrospective data distribution and ii) Estimation, which focuses on turning the obtained identification strategy into point and interval estimators following well-grounded statistical principles. We specifically analyze scenarios that involve both adherence to and violation of the NDE and NUC assumptions. We formulate this problem of *active feature acquisition performance evaluation (AFAPE)* as the problem of estimating the expected counterfactual acquisition and misclassification costs using retrospective data.

1.3 Paper Outline and Contributions

The remainder of this paper is organized as follows. After reviewing the necessary background and related methods in Section 2, we formulate the AFAPE problem in Section 3. The general AFAPE problem is not identified—that is, it is not possible to estimate the counterfactual acquisition and misclassification costs from retrospective data when both the NDE and NUC assumptions are violated.

Therefore, we begin Section 4 by employing the NUC assumption and show that this makes AFAPE amenable to an offline reinforcement learning (RL) / dynamic treatment

regimes (DTR) view. This allows the application of known identification and estimation theory from the offline RL / DTR literature.

In Section 5, we make instead the NDE assumption and assume the NUC assumption can be violated. We demonstrate that under the NDE assumption, the AFA decision process depicted in Figure 1 transforms into a missing data graph (m-graph) (Mohan et al., 2013; Shpitser et al., 2015), a recognized graphical framework in the missing data literature. This enables us to apply established identification and estimation theory from the missing data literature. After solving the missing data problem, the AFAPE problem is transformed into an online RL setting where one can simulate different acquisition trajectories, leading to a trivial solution for AFAPE.

In Section 6, we assume both the NUC and the NDE assumptions hold. In this setting, one can apply either offline RL or missing data methods to solve AFAPE, but both require strong positivity assumptions and do not utilize the data optimally. Therefore, we propose a new viewpoint on AFA, which we denote as *semi-offline reinforcement learning*. Under the semi-offline RL viewpoint, the AFA agent engages with the environment in an online manner, but certain actions (where the underlying feature values are missing in the retrospective data) cannot be explored. The positivity assumption required for identification is drastically reduced under the new semi-offline RL viewpoint. We derive three novel estimators that can be denoted as semi-offline RL versions of known offline RL estimators, including the Q-function based direct method (DM) (Levine et al., 2020), inverse probability weighting (IPW) (Levine et al., 2020), and the double reinforcement learning (DRL) estimator (Kallus and Uehara, 2020). Notably, our DRL estimator is doubly robust, exhibiting consistency even if either the underlying Q-function or the propensity score model is misspecified.

In Section 7, we explore the estimation of AFAPE under the NUC and NDE assumptions using semiparametric theory. We demonstrate how all three viewpoints—the offline RL view (under the NDE assumption), the missing data view (under the NUC assumption), and the semi-offline RL view—interconnect within this theoretical framework. Unfortunately, there is no closed-form solution for an efficient estimator. However, we can enhance the efficiency of all estimators by applying established semiparametric techniques from related fields, such as standard missing data problems (Tsiatis, 2006) and dynamic testing and treatment regimes (Liu et al., 2021). These methods, though, come with significant computational costs, are challenging to implement, and require strong positivity assumptions.

In Section 8, we present synthetic data experiments that exemplify the improved data efficiency and reduced positivity requirements of the semi-offline RL estimators. Our experiments also show that biased evaluation methods commonly used in the AFA literature can lead to detrimental conclusions regarding the performance of AFA agents. Deploying such methods without caution may pose significant risks to patients' lives. We end the paper with a Discussion (Section 9) and Conclusion (Section 10).

2. Background and Related Methods

In the following, we review some of the literature about AFA and provide some background on offline RL/ DTR, missing data, and semi-parametric theory.

2.1 Active Feature Acquisition (AFA)

Research on active feature acquisition (AFA) and related problem formulations have been published under various different names and in different, largely disjoint, research communities. Early research in economics and decision science literature addressed the problem of "Value of Information" (VoI) (LaValle, 1968a,b; Gould, 1974; Hilton, 1979; Hess, 1982; Keisler et al., 2014). Similar methods have also been applied in the medical field, often in terms of cost-effectiveness analysis of screening policies (Mushlin and Fintor, 1992; Krahn et al., 1994; Botteman et al., 2003; Force*, 2009). AFA has further been studied under the name of "dynamic testing regimes" (Liu et al., 2021; Robins et al., 2008) or "dynamic monitoring regimes" (Neugebauer et al., 2017; Kreif et al., 2021) in the causal inference literature, often in combination with dynamic treatment regimes. In these settings, the goal of the feature acquisitions is not to enable better predictions/diagnoses but to enable better treatment decisions.

The name "active feature acquisition" (AFA) (An et al., 2006; Li and Oliva, 2021a; Li et al., 2021; Chang et al., 2019; Shim et al., 2018; Yin et al., 2020) is common in the machine learning literature, but other names are also frequently used. These include, but are not limited to, "active sensing" (Yoon et al., 2019, 2018; Tang et al., 2020; Jarrett and van der Schaar, 2020), "active feature elicitation" (Natarajan et al., 2018; Das et al., 2021), "dynamic feature acquisition" (Li and Oliva, 2021b), "dynamic active feature selection" (Zhang, 2019), "element-wise efficient information acquisition" (Gong et al., 2019), "classification with costly features" (Janisch et al., 2020) and "test-cost sensitive classification" (Xiaoyong Chai et al., 2004).

These diverse research fields share a common characteristic, which involves designing an agent to selectively acquire a subset of features to balance acquisition cost and information gain. The approaches used to design such agents range from simple greedy acquisition strategies to more complex RL-based strategies. However, the focus of this work is not on any specific AFA method but rather on evaluating the performance of *any* AFA method under the acquisition distribution shift. For a more comprehensive literature review of existing AFA methods and a distinction between AFA and other related fields, we direct interested readers to Appendix A.

2.2 (Offline) Reinforcement Learning (RL) / Dynamic Treatment Regimes (DTR)

We show in Section 4 that AFA can be analyzed from an offline RL/ DTR viewpoint. In Section 5, we show that AFAPE can also be analyzed from an online RL viewpoint (if NDE holds and after missingness has been resolved). Online RL allows the interaction of an agent with the environment and thus the simulation of outcomes under any desired policy, thereby leading to a trivial solution for the AFAPE problem. In offline RL, however, such a simulation is not possible due to missing knowledge about the environment. The AFAPE problem then becomes equivalent to the problem of off-policy policy evaluation (OPE) (Dudik et al., 2011; Thomas and Brunskill, 2016; Kallus and Uehara, 2020), in which the goal is to evaluate the performance of a "target" policy (here the AFA policy) from data collected under a different "behavior" policy (here the retrospective acquisition policy of, for example, a doctor). Several estimators have been developed for the OPE problem. These

include the plug-in based on the G-formula (Robins, 1986) (also referred to as model-based evaluation (Levine et al., 2020)), inverse probability weighting (IPW) (Levine et al., 2020) (also known as importance sampling or the Horvitz-Thompson estimator (Horvitz and Thompson, 1952)), the direct method (DM) (Levine et al., 2020), and double reinforcement learning (DRL) (Kallus and Uehara, 2020).

2.3 Missing Data

In this paper, we show that AFAPE can be viewed as a missing data (+ a trivial online RL) problem. Thus, known identification and estimation techniques from the missing data literature can be employed. We show that the NUC assumption described in this paper corresponds under NDE to a *missing-at-random (MAR)* assumption. Violations of the NUC assumption correspond, in our setting, to a special, identified *missing-not-at-random (MNAR)* scenario. Estimation strategies generally include inverse probability weighting (IPW) (Seaman and White, 2013), and multiple imputation (MI) (Sterne et al., 2009) (a special case of the plug-in of the G-formula).

2.4 Semiparametric Theory

The goal of AFAPE is to estimate the expected acquisition and misclassification costs that would arise when following the AFA system's decisions. In more general terms, this corresponds to estimating a target parameter $J = J(p)$ of some unknown distribution p given a set of observed samples from p (the retrospective data set). The goal in semiparametric theory is to find suitable estimators for such a target parameter J while leaving at least part of the data-generating process p unrestricted/ unspecified, thereby imposing fewer assumptions, which can lead to more credible estimates. Assumptions that can be taken with a reasonable level of confidence (such as in many AFA settings the NUC or NDE assumptions), can, however, be leveraged to derive more efficient estimators.

A key focus of semiparametric theory is the identification of *influence functions*, which are used to construct estimators with desirable properties, such as a consistently estimable asymptotic variance. The estimator associated with the influence function that has the smallest asymptotic variance is the most efficient one. Often, these influence functions include nuisance functions—unknown components of the model that must be estimated from the data. For example, a nuisance function might model the probability of a doctor acquiring a particular feature. Estimating these nuisance functions typically involves parametric assumptions (e.g., using a logistic regression model). Consequently, the resulting estimator is only *locally efficient*, meaning it achieves efficiency only if the parametric assumptions for the nuisance function hold true.

Even if the parametric assumptions are incorrect, many influence function-based estimators remain consistent due to a property known as *multiple robustness* (Rotnitzky et al., 2017). For instance, the DRL estimators in this paper exhibit a form of double robustness (Scharfstein et al., 1999; Chernozhukov et al., 2018). These estimators rely on learning two nuisance functions and remain consistent as long as the parametric assumption holds for at least one of these functions.

For a more detailed review, please see Appendix B, which covers both the general principles of semiparametric theory and specific insights related to missing data problems.

A thorough understanding of semiparametric theory is primarily necessary for Section 7, which is intended for interested readers.

2.5 Active Feature Acquisition Performance Evaluation (AFAPE)

Although we believe to be the first to explicitly formulate and analyze the AFAPE problem, other AFA papers have reported performance metrics that can be seen as attempts to address it. The reported results, however, often lack assumption statements and justification for the chosen evaluation framework and are, in general, biased or inefficiently estimated. We categorize these results based on the viewpoints analyzed in this paper:

Offline RL view: The offline RL view has been utilized in the AFA context (Chang et al., 2019; Cheng et al., 2018). As we show in this paper, this approach is only valid under the NUC and strong positivity assumptions.

Missing data + online RL view: We show in this paper that one can apply, under the NDE assumption, a missing data + online RL viewpoint to solve AFAPE. While this viewpoint has been taken in the AFA literature before, the missing data part of it has, to our knowledge, only been solved using (conditional) mean imputation (An et al., 2022; Erion et al., 2021; Janisch et al., 2020). (Conditional) mean imputation leads, however, to biased estimation results, as we illustrate in Section 5.

Semi-offline RL view: Some AFA papers (Janisch et al., 2020; Yoon et al., 2018) have addressed the problem of missing data during the online RL simulations by simply blocking the corresponding feature acquisitions. This approach is similar to our proposed semi-offline RL view. However, unlike our approach, these papers did not correct for the distribution shift caused by blocking feature acquisitions, resulting in biased estimation results.

2.6 No Direct Effect (NDE) Assumption

The only work that, to the best of our knowledge, leverages the NDE assumption in a similar way to our semi-offline RL viewpoint is a series of publications from the causal inference literature around the slightly different problem of evaluation of joint dynamic testing and treatment regimes (Robins et al., 2008; Caniglia et al., 2019; Liu et al., 2021; Neugebauer et al., 2017; Kreif et al., 2021). In this setting, the agent is not only tasked with deciding which features to acquire, but also which treatments to give to the patient. The treatment assignment replaces the need for classification/diagnosis in these settings. Robins et al. (2008) introduced within this setting for the first time the term "no direct effect" (NDE) assumption. NDE stated that the feature acquisition decisions have no direct effect (or no long-term direct effect (Liu et al., 2021)) on the health status of the patient, except through their effect on the treatment decisions.

Caniglia et al. (2019) derived an IPW estimator for this context, which demonstrated a 50-fold increase in data efficiency compared to the offline RL IPW estimator, signaling the enormous benefits that can be achieved by leveraging the NDE assumption. We adapt this estimator to the AFA setting and show that it is equivalent to our proposed IPW estimator for a simple setting and a special positivity assumption. However, our IPW estimator can be applied in more general settings under weaker positivity assumptions and combined with our DM method to form the novel DRL estimator for semi-offline RL.

Liu et al. (2021) developed nearly semiparametrically efficient estimators for this problem by modifying the DRL estimator from the offline RL perspective to enhance efficiency under the NDE assumption. In Section 7, we demonstrate that this approach can be adapted for AFA, where it becomes a specialized method within established semiparametric techniques for missing data problems. Our proposed semi-offline RL estimators can also achieve greater data efficiency using this framework. However, this approach has limitations: it requires strong positivity assumptions, is only applicable in very simple settings with one acquisition action per time point, and the resulting augmentation necessitates complex approximations of function spaces, making implementation challenging and resulting in estimators that are only "nearly" efficient. Additionally, it has only been tested in an extremely simplified context with one acquisition action and one time step (Liu et al., 2021).

2.7 Distribution Shift Robust ML Models

Lastly, this work also relates to the general literature on distribution shift-robust ML models. A common problem with the deployment of ML models occurs if the model is trained, for example, on data from hospital 1 but should be deployed to hospital 2. The related literature aims at building robust models that retain their performances across deployment environments (Rockenschaub et al., 2023a,b). One part of the distribution that might change between hospital 1 and hospital 2 is the feature acquisition policy. If this is the case, and if the acquisition policy at hospital 2 is known, one may directly apply our methods to this scenario and treat the acquisition policy at hospital 2 as the AFA policy that is to be evaluated. However, we will not go into more detail about this scenario and will focus on the AFA setting.

3. Active Feature Acquisition Performance Evaluation (AFAPE) Problem Definition

We begin the section by introducing the mathematical notation for the AFA setting and AFAPE problem. A glossary containing all the variables and important terms can be found in Appendix C.

3.1 Feature Acquisition Process

The feature acquisition process, as illustrated in Figure 2, is modeled using the following variables:

- **Unobserved underlying features**: $U^t \in \mathbb{R}^d$ for $t \in \{0, \dots, T\}$. The features are dynamic, meaning they change over time, so generally, $U_i^t \neq U_i^{t+1}$.
- **Feature acquisition decisions**: $A^t \in \mathcal{A}^t = \{0, 1\}^d$ for $t \in \{1, \dots, T\}$, where $A_i^t = 1$ indicates that feature U_i^t will be acquired.
- **Observed feature values**: $X^t \in (\mathbb{R} \cup \{\text{"?"}\})^d$ for $t \in \{0, \dots, T\}$, where "?" denotes a missing feature value that was not acquired. We assume no measurement error, so

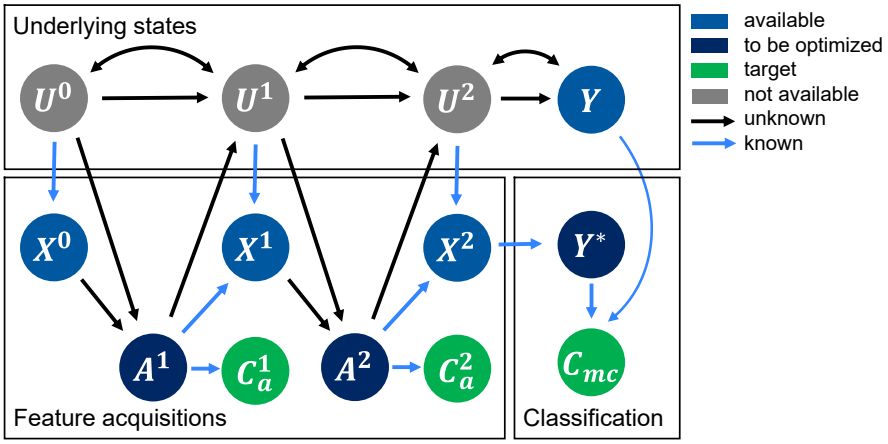


Figure 2: The causal graph depicting the AFA setting as a partially observable decision process consisting of unobserved underlying features U^t, feature acquisition actions A^t, feature measurements $X^t = G_{A^t}(U^t)$, and associated acquisition costs C_a^t. After a number of acquisition steps T (here $T = 2$), a classification Y^* is to be performed. In the case of misclassification (Y^* is not equal to the true label Y), a misclassification cost C_{mc} is produced. Edges showing long-term dependencies are omitted from the graph for visual clarity. These include: $\underline{U}^{t-1}, \underline{X}^{t-1}, \underline{A}^{t-1} \to A^t$; $\underline{X}^T, \underline{A}^T \to Y^*$; $A^t \to \overline{U}^t$; $\underline{U}^{t-1} \leftrightarrow U^t$; $\underline{U}^{t-1} \to U^t$; $\underline{U}^T \leftrightarrow Y$ and $\underline{U}^T \to Y$ (where $\leftrightarrow$ denotes unobserved confounding).

X^t is deterministically determined by A^t and U^t according to:

$$X_i^t = G_{A_i^t}(U_i^t) = \begin{cases} U_i^t & \text{if } A_i^t = 1, \\ \text{"?"} & \text{if } A_i^t = 0. \end{cases} \tag{1}$$

G_A denotes the observation function for a given feature acquisition decision A. Further, we assume for simplicity $X^0 = U^0$.

- **Acquisition costs**: $C_a^t \in \mathbb{R}$ represents the known feature acquisition cost associated with A^t.

Additionally, let $\underline{U}^t$ and $\overline{U}^t$ denote the complete past and complete future of U^t, respectively, where $\underline{U}^t = \{U^0, \ldots, U^t\}$ and $\overline{U}^t = \{U^t, \ldots, U^T\}$. Similarly, define $\underline{A}^t$, $\overline{A}^t$, $\underline{X}^t = G_{\underline{A}^t}(\underline{U}^t)$, and $\overline{X}^t = G_{\overline{A}^t}(\overline{U}^t)$ for the variables A^t and X^t. Let $U = \overline{U}^0$, $A = \overline{A}^1$, and $X = \overline{X}^0$, and denote the space of A by $\mathcal{A}$. The retrospective acquisition policy is given by $\pi_\beta^t(A^t \mid \underline{X}^{t-1}, \underline{U}^{t-1}, \underline{A}^{t-1})$.

3.2 Classification Process

At time T, the feature acquisition process concludes, and the classification of an underlying label is performed based on the acquired information. The classification process includes:

- **Underlying categorical label**: Y, the true label to be classified.
- **Predicted label**: Y^*, obtained using a deterministic classifier $Y^* = f_{cl}(\underline{A}^T, \underline{X}^T)$.
- **Misclassification cost**: $C_{mc} \in \mathbb{R}$, the cost incurred when Y^* differs from Y, defined by a predefined cost function $C_{mc} = f_C(Y^*, Y)$. We alternatively write $C_{mc} = f_C(f_{cl}(\underline{A}^T, \underline{X}^T), Y) \equiv f_C(\underline{A}^T, \underline{X}^T, Y)$ directly as a function of $\underline{A}^T, \underline{X}^T$ and Y.

We let g represent known deterministic distributions or densities, distinct from p used for other distributions or densities. For example, we let $g_{cl}(Y^* \mid \underline{A}^T, \underline{X}^T)$ denote the deterministic distribution of the classifier f_{cl} and $g_C(C_{mc}|Y^*, Y)$ denote the deterministic distribution of the cost function f_C.

We also assume Y is always available in the retrospective data set and allow for unobserved confounding among the unobserved underlying features and the label (represented by edges $\underline{U}^t \leftrightarrow U^{t+1}, Y$), but no additional confounding with A^t. We denote the retrospective data set consisting of the variables A, X, and Y as $\mathcal{D}$.

3.3 Problem Definition: Active Feature Acquisition Performance Evaluation (AFAPE)

Given a target AFA policy $\pi_\alpha^t(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})$ (which is not allowed to depend on the unobserved underlying features $\underline{U}^{t-1}$) and a target classifier $f_{cl}(\underline{X}^T, \underline{A}^T)$, the goal of AFAPE is to estimate the expected acquisition and misclassification costs that would arise, had the target policy π_α and classifier f_{cl} been deployed. The estimation problem for this expected counterfactual cost can be expressed as estimating

$$J_a = \mathbb{E}\left[\sum_{t=1}^{T} C^t_{a,(\pi_\alpha)}\right], \text{ and } J_{mc} = \mathbb{E}\left[C_{mc,(\pi_\alpha)}\right], \tag{2}$$

where $C^t_{a,(\pi_\alpha)}$ and $C_{mc,(\pi_\alpha)}$ denote the potential outcomes of the acquisition and misclassification costs under the AFA policy π_α. Therefore, J_a and J_{mc} represent the expected acquisition and misclassification costs under a distribution induced by π_α rather than by the retrospective acquisition policy π_β. Note that this assumes π_α can be followed perfectly, which may not always hold in practice, for example, if patients refuse certain medical tests or miss appointments.

The goal of this paper is to i) perform identification, i.e., to determine under which assumptions it is possible to resolve this distribution shift and to obtain an unbiased estimate of J_a and J_{mc}; and ii) to derive such unbiased estimators.

As the AFAPE problem is similar for J_a and J_{mc}, we will focus on J_{mc} throughout the main part of the paper. We abbreviate $J_{mc} \equiv J$ and $C_{mc} \equiv C$. We provide the estimation formulas for J_a and for J_{mc} when a prediction is to be performed at each time step in the relevant appendices.

3.4 Problem Definition: Optimization of Active Feature Acquisition Methods

While the focus of the paper is on the AFAPE problem, we provide the definition of the AFA optimization problem for completeness. The goal in AFA is to find the optimal AFA policy $\pi_\alpha^t(A^t|\underline{X}^{t-1}, \underline{A}^{t-1}; \phi_1^*)$ parameterized by ϕ_1^*, and the optimal classifier $f_{cl}(\underline{X}^T, \underline{A}^T; \phi_2^*)$

parameterized by ϕ_2^*, such that their joint application minimizes the expected sum of counterfactual acquisition and misclassification costs:

$$\phi_1^*, \phi_2^* = \arg\min_{\phi_1,\phi_2} J_{\text{total}}(\phi_1, \phi_2) = \arg\min_{\phi_1,\phi_2} \mathbb{E}\left[\sum_{t=1}^{T} C^t_{a,(\pi_\alpha)} + C_{mc,(\pi_\alpha)} \middle| \phi_1, \phi_2\right].$$

3.5 Assumptions

Here, we provide an overview of the key assumptions in this paper. We start by stating the fixed assumptions that hold throughout the paper before stating assumptions that we vary within different sections.

3.5.1 Fixed Assumptions

We make the following assumptions throughout the paper:

Assumption 1 (No measurement noise) There is no noise in feature measurements, as expressed by Eq. 1.

Assumption 2 (Consistency) If $\underline{A}^t = \underline{a}^t$, then $U^t_{(\underline{a}^t)} = U^t$.

This standard consistency assumption from the causal inference literature states that an individual's observed outcomes align with their potential outcomes under the observed acquisition decisions. Here, $U^t_{(\underline{A}^t=\underline{a}^t)}$ represents the potential outcome of U^t under potential acquisition decisions $\underline{A}^t = \underline{a}^t$.

Assumption 3 (No interference) The acquisition decisions for one individual do not affect other individuals.

One prominent example of interference in medical settings is allocation interference, which can occur when a hospital's resources or staff are overwhelmed by a high volume of medical test requests for multiple patients simultaneously, resulting in the inability to fulfill all feature acquisition requests.

3.5.2 Investigated Assumptions

In this paper, we analyze how the following assumptions affect identification and estimation of the target J in the AFAPE problem.

Assumption 4 (No direct effect (NDE)) The unobserved underlying features are not influenced by feature acquisitions (i.e., $A^t \not\to \overline{U}^t$). Equivalently: $U^t \perp\!\!\!\perp \underline{A}^t \mid \underline{U}^{t-1}$.

This is a standard assumption in missing data problems, but it may not hold in all medical settings, as some medical tests can alter certain features of the patient. The NDE assumption is relaxed in Section 4, and made in Sections 5, 6 and 7.

Assumption 5 (No unobserved confounding (NUC)) Acquisition decisions are independent of the unobserved underlying features given past acquisition decisions and measured features: $A^t \perp\!\!\!\perp \underline{U}^{t-1} \mid \underline{X}^{t-1}, \underline{A}^{t-1}$. This is graphically expressed by the missing arrow $\underline{U}^{t-1} \not\to A^t$.

The no unobserved confounding assumption may, for example, be violated in medical settings if certain feature values are seen by the physician and influence their decision-making for further tests but are not recorded in the database. We assume NUC in Sections 4, 6 and 7 and allow certain violations in Section 5. Note that when referring to NUC, we only assume no unobserved confounding of the acquisition actions. Potential unobserved confounding within U and between U and Y is allowed throughout the paper.

Assumption 6 (Positivity/ experimental treatment assignment/ overlap) Certain feature sets have a positive probability of being acquired under the retrospective acquisition policy π_β.

The positivity assumption (also known as experimental treatment assignment assumption or overlap assumption) relates to how much exploration was done under the retrospective acquisition policy π_β. Positivity requirements are crucial for identification and vary between the discussed views. Hence, we derive and discuss them separately for each view.

4. Offline Reinforcement Learning View

Assumptions in this section: Assumption 5 (NUC)
Firstly, we consider the scenario where the NUC assumption (Assumption 5) holds (i.e. $\underline{U}^{t-1} \not\to A^t$), but the NDE assumption (Assumption 4) does not necessarily hold (i.e. $A^t \to \overline{U}^t$). This scenario can be addressed using the offline reinforcement learning (RL) view. The NUC assumption allows us to perform a latent projection (Verma and Pearl, 1990) to project out the unknown variables U^t (along with Y and Y^*) from the causal graph in Figure 2 and obtain the graph in Figure 3 which contains only observed variables. The projected graph allows us to apply established identification and estimation methods from the offline RL literature.

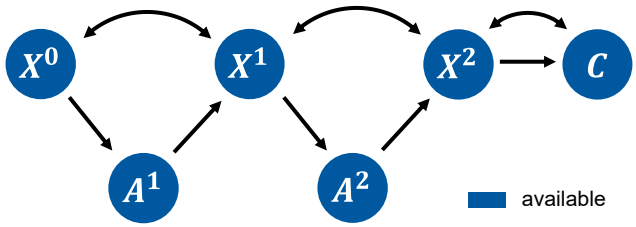


Figure 3: Updated causal graph of the AFA setting under the NUC assumption (Assumption 5) and a latent projection. The graph depicts a standard, identified offline RL setting. Long-term dependencies are omitted from the graph for visual clarity. These include edges $\underline{X}^{t-1}, \underline{A}^{t-1} \to A^t$; $\underline{X}^T, \underline{A}^T \to C$; $\underline{X}^{t-1} \leftrightarrow X^t$ and $\underline{X}^T \leftrightarrow C$.

4.1 Identification

Under the offline RL view, solving the AFAPE problem is equivalent to performing off-policy policy evaluation (OPE). Identification for OPE requires sequential exchangeability (also known as sequential ignorability), which implies that adjusting for $\underline{A}^{t-1}$ and $\underline{X}^{t-1}$ eliminates any confounding factors affecting A^t. The graph in Figure 3 satisfies this requirement. Note

that sequential exchangeability would not hold under violations of the NUC assumption (Assumption 5) because the latent projection under the presence of edges $\underline{U}^{t-1} \to A^t$ would produce confounding edges $\overline{X}^t \leftrightarrow A^t$ and $C \leftrightarrow A^t$.

Identification of J further requires Assumptions 1-3 (no measurement noise, consistency, no interference) and the following (sequential) positivity assumption:

Assumption 6.1 (Positivity for offline RL)

$$\begin{aligned} &\text{if} && p(\underline{X}^{t-1}_{(\pi_\alpha)} = \underline{x}^{t-1}, \underline{A}^{t-1}_{(\pi_\alpha)} = \underline{a}^{t-1})\pi^t_\alpha(a^t|\underline{x}^{t-1}, \underline{a}^{t-1}) > 0, \\ &\text{then} && p(\underline{x}^{t-1}, \underline{a}^{t-1})\pi^t_\beta(a^t|\underline{x}^{t-1}, \underline{a}^{t-1}) \geq \mathcal{O} \\ & && \forall t, a^t, \underline{x}^{t-1}, \underline{a}^{t-1}, \text{ and some constant } \mathcal{O} > 0 \end{aligned}$$

where we introduced the following notation: $\pi^t_\alpha(a^t|\underline{x}^{t-1}, \underline{a}^{t-1}) \equiv \pi^t_\alpha(A^t = a^t|\underline{X}^{t-1} = \underline{x}^{t-1}, \underline{A}^{t-1} = \underline{a}^{t-1})$.

The positivity assumption states that, for every set of actions and observations $\underline{a}^{t-1}, \underline{x}^{t-1}$ reachable under π_α and a desired next action a^t (i.e., an action with positive support under π_α), we require also positive support for a^t under π_β. A violation of this assumption may occur if the acquisition decisions under the AFA policy π_α differ significantly from the decisions made by doctors (π_β).

4.2 Estimation

Estimation can be performed using well-known techniques from the offline RL / DTR literature. The following are common estimators:

1) Model-based evaluation:
The model-based estimator, also known in the causal inference literature as the noniterative conditional expectation (NICE) estimator of the G-formula (Rein et al., 2024; Wen et al., 2021), estimates the target cost J as follows:

$$\hat{J}_{MB\text{-}Off} = \sum_{X,A} \hat{\mathbb{E}}[C|\underline{X}^T, \underline{A}^T] \prod_{t=1}^{T} \hat{p}(X^t|\underline{X}^{t-1}, \underline{A}^t)\pi^t_\alpha(A^t|\underline{X}^{t-1}, \underline{A}^{t-1}) \tag{3}$$

where the integration over X and A can be solved using Monte Carlo integration. Note that we use sums to denote the integration over X. All results in this paper do, however, also hold for continuous X, by replacing the sums with proper integrals.

This estimator requires learning the state transition function $p(X^t|\underline{X}^{t-1}, \underline{A}^t)$ and the expected cost $\mathbb{E}[C|\underline{X}^T, \underline{A}^T]$. We denote the learned nuisance functions as $\hat{p}(X^t|\underline{X}^{t-1}, \underline{A}^t)$ and $\hat{\mathbb{E}}[C|\underline{X}^T, \underline{A}^T]$.

2) Inverse probability weighting (IPW):
The target cost that is estimated by IPW (Levine et al., 2020) is

$$\hat{J}_{IPW\text{-}Off} = \hat{\mathbb{E}}_n \left[\hat{\rho}^T_{Off}\, C\right], \text{where } \hat{\rho}^T_{Off} = \prod_{t=1}^{T} \frac{\pi^t_\alpha(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})}{\hat{\pi}^t_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})}. \tag{4}$$

where $\hat{\mathbb{E}}_n[.]$ denotes the empirical average over the data set $\mathcal{D}$. This estimator requires learning the retrospective acquisition policy/propensity score model $\pi_\beta^t(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})$.

Remark 1 (Cross-fitting estimators) When nuisance functions are estimated using flexible machine learning methods, the training of these functions and the evaluation of the estimator must be conducted on separate data splits to avoid introducing bias. To enhance efficiency, this process can be performed by alternating the training and evaluation splits, a method known as cross-fitting (Chernozhukov et al., 2018). Throughout this paper, we assume that cross-fitting is used, though this is not explicitly noted in the proposed estimators' notation, leading to a slight abuse of notation.

3) Direct method (DM):
The DM estimator, also known as the iterative conditional expectation (ICE) estimator of the G-formula (Wen et al., 2021), estimates the target cost J as:

$$\hat{J}_{DM\text{-}Off} = \hat{\mathbb{E}}_n[\hat{V}^0_{Off}]. \tag{5}$$

This estimator relies on learning a state-action value function Q^t_{Off} or state value function V^t_{Off}:

$$Q^t_{Off} \equiv Q_{Off}(\underline{X}^{t-1}, \underline{A}^t) \equiv \mathbb{E}[C_{(\overline{\pi}^{t+1}_\alpha)}|\underline{X}^{t-1}, \underline{A}^t],$$
$$V^t_{Off} \equiv V_{Off}(\underline{X}^t, \underline{A}^t) \equiv \mathbb{E}[C_{(\overline{\pi}^{t+1}_\alpha)}|\underline{X}^t, \underline{A}^t].$$

where $C_{(\overline{\pi}^{t+1}_\alpha)}$ denotes the potential outcome of C under a policy intervention π_α applied only from time step $t+1$ onwards. Q_{Off} and V_{Off} can be learned using, for example, the dynamic programming (DP) algorithm, which is based on the recursive property of the Bellman equation (Bertsekas, 2012):

$$Q_{Off}(\underline{X}^{t-1}, \underline{A}^t) = \sum_{X^t} V_{Off}(\underline{X}^t, \underline{A}^t) p(X^t|\underline{X}^{t-1}, \underline{A}^t) \tag{6}$$

$$V_{Off}(\underline{X}^t, \underline{A}^t) = \sum_{A^{t+1}} Q_{Off}(\underline{X}^t, \underline{A}^{t+1}) \pi^{t+1}_\alpha(A^{t+1}|\underline{X}^t, \underline{A}^t) \tag{7}$$

In practice, one only needs to learn Q_{Off} such that V_{Off} can be simply computed as $V^t_{Off} = \mathbb{E}_{\pi_\alpha}[Q^{t+1}_{Off}]$ using, for example, Monte Carlo integration over the known AFA policy π_α.

4) Double reinforcement learning (DRL):
The target cost that is estimated by DRL (Kallus and Uehara, 2020) is

$$\hat{J}_{DRL\text{-}Off} = \hat{\mathbb{E}}_n \left[\hat{\rho}^T_{Off} C + \sum_{t=1}^{T} \left(-\hat{\rho}^t_{Off} \hat{Q}^t_{Off} + \hat{\rho}^{t-1}_{Off} \hat{V}^{t-1}_{Off} \right) \right]. \tag{8}$$

The DRL estimator combines approaches 2) and 3) by using both the learned propensity score $\hat{\pi}_\beta$ and the state action value function $\hat{Q}_{Off}$ (and the derived $\hat{V}_{Off}$). This estimator is (locally) efficient and doubly robust, in the sense that it is consistent if either the propensity score model $\hat{\pi}_\beta$, or the state action value function $\hat{Q}_{Off}$ is correctly specified (Kallus and Uehara, 2020).

5. Missing Data (+ Online Reinforcement Learning) View

Assumptions in this section: Assumption 4 (NDE)

In this section, we assume that the NDE assumption holds (i.e., $A^t \not\to \overline{U}^t$), but do not require the NUC assumption (Assumption 5). We observe that the general AFA graph from Figure 2 transforms under NDE into the graph shown in Figure 4A). This new graph represents a temporal missing data graph (m-graph) (Mohan et al., 2013; Shpitser et al., 2015) from the missing data literature. The unobserved underlying feature values U^t are replaced by counterfactuals of the measured feature values $X^t_{(1)}$ since $U^t_{(\underline{a}^t)} = U^t_{(a^t)} = U^t_{(1)} = X^t_{(1)}$ for all potential acquisitions $\underline{a}^t$ and thus also $a^t = \vec{1}$. Due to the temporal restrictions $\overline{X}^t_{(1)} \not\to A^t$, the shown graph can be more precisely specified as the known block-conditional missing data model (Zhou et al., 2010). The graph depicting the counterfactual distribution is shown in Figure 4B).

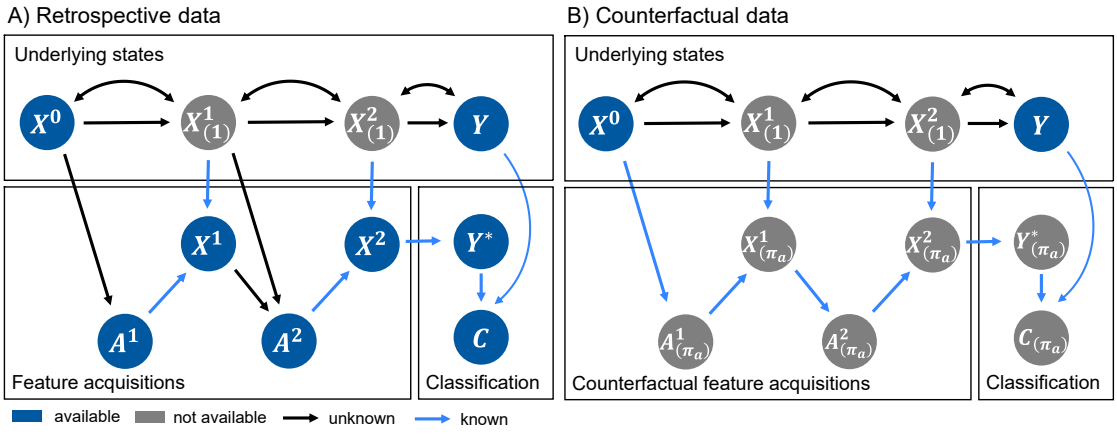


Figure 4: A) Updated causal graph of the AFA process under the NDE assumption (Assumption 4). Unknown state variables U^t are replaced with the counterfactual feature values $X^t_{(1)}$, which represent the values X^t would have taken if A^t was $\vec{1}$ (i.e., the decision to observe all feature values). This graph describing the feature acquisition process is known as a missing data graph (m-graph). B) Graph showing the counterfactual distribution under π_α. Edges showing long-term dependencies are omitted for visual clarity. These include for both graphs $\underline{X}^{t-1}_{(1)} \leftrightarrow X^t_{(1)}$ and $\underline{X}^T_{(1)} \leftrightarrow Y$; for A) $\underline{X}^{t-1}, \underline{X}^{t-1}_{(1)}, \underline{A}^{t-1} \to A^t$, and $\underline{X}^T, \underline{A}^T \to Y^*$; and for B) $X^0, \underline{X}^{t-1}_{(\pi_\alpha)}, \underline{A}^{t-1}_{(\pi_\alpha)} \to A^t_{(\pi_\alpha)}$ and $X^0, \underline{X}^T_{(\pi_\alpha)}, \underline{A}^T_{(\pi_\alpha)} \to Y^*_{(\pi_\alpha)}$.

5.1 Problem Reformulation

Now, we can establish the following theorem. Under the NDE assumption, the AFAPE problem becomes equivalent to a standard missing data problem, for which one can apply known identification and estimation theory.

Theorem 1. *(AFAPE problem reformulation and identification under the missing data view). The AFAPE problem of estimating J (Equation 2) is under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), and Assumption 4 (NDE) equivalent to estimating*

$$J = \sum_{X_{(1)},Y} \underbrace{\mathbb{E}\left[C_{(\pi_\alpha)}|X_{(1)},Y\right]}_{\textit{online RL}} \underbrace{p(X_{(1)},Y)}_{\textit{missing data}}. \tag{9}$$

Furthermore, J is identified if $p(X_{(1)},Y)$ is identified.

Proof The decomposition of J into the two expected values follows from the law of iterated expectations and the independence of $X_{(1)},Y$ from a policy intervention π_α. The fact that the inner expected value is identified can be easily verified by examining the graph representing the counterfactual distribution shown in Figure 4B). The graph shows that all functional relationships between variables that are part of the feature acquisition and classification processes are known (represented as blue edges). In particular, this implies the following factorization:

$$\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)},Y] = \sum_{\underline{X}^T_{(\pi_\alpha)},\underline{A}^T_{(\pi_\alpha)},Y^*_{(\pi_\alpha)},C_{(\pi_\alpha)}} C_{(\pi_\alpha)} q(C_{(\pi_\alpha)},Y^*_{(\pi_\alpha)},X_{(\pi_\alpha)},A_{(\pi_\alpha)}|X_{(1)},Y) \tag{10}$$

with the identifying distribution

$$\begin{aligned} &q(C_{(\pi_\alpha)},Y^*_{(\pi_\alpha)},X_{(\pi_\alpha)},A_{(\pi_\alpha)}|X_{(1)},Y) = \\ &= \prod_{t=0}^{T} \underbrace{g(X^t_{(\pi_\alpha)}|A^t_{(\pi_\alpha)},X^t_{(1)})}_{\text{feature revelations}} \prod_{t=1}^{T} \underbrace{\pi^t_\alpha(A^t_{(\pi_\alpha)}|\underline{X}^{t-1}_{(\pi_\alpha)},\underline{A}^{t-1}_{(\pi_\alpha)})}_{\text{acquisition decisions}} \underbrace{g_{cl}(Y^*_{(\pi_\alpha)}|\underline{X}^T_{(\pi_\alpha)},\underline{A}^T_{(\pi_\alpha)})}_{\text{label prediction}} \underbrace{g_C(C_{(\pi_\alpha)}|Y,Y^*_{(\pi_\alpha)})}_{\text{cost computation}} \end{aligned}$$

which is identified since all (deterministic) distributions $g(.)$ and π_α are known functions. ■

Since all densities $g()$ are deterministic and known, we can further simplify Eq. 10 using simpler notation. We use $X = G_A(X_{(1)})$, $\pi_\alpha(a|G_a(X_{(1)})) \equiv \prod_{t=0}^{T} \pi^t_\alpha(a^t|G_{\underline{a}^{t-1}}(X_{(1)}),\underline{a}^{t-1})$ and $C = f_C(A,G_A(X_{(1)}),Y)$ to obtain:

$$\begin{aligned} \mathbb{E}[C_{(\pi_\alpha)}|X_{(1)},Y] &= \sum_{a\in\mathcal{A}} f_C(\underline{a}^T,G_{\underline{a}^T}(X_{(1)}),Y) \prod_{t=0}^{T} \pi^t_\alpha(a^t|G_{\underline{a}^{t-1}}(X_{(1)}),\underline{a}^{t-1}) \\ &= \sum_{a\in\mathcal{A}} f_C(a,G_a(X_{(1)}),Y)\pi_\alpha(a|G_a(X_{(1)})). \end{aligned} \tag{11}$$

We denote $\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)},Y]$ as the online RL part of the problem because it involves the evaluation of a policy in a known environment. We refer to the outer expected value of $p(X_{(1)},Y)$ as the missing data problem, as it requires the identification of the counterfactual feature distribution.

5.2 Identification

As established in Theorem 1, the AFAPE problem is identified if the missing data problem (i.e., $p(X_{(1)}, Y)$) is identified. The following positivity assumption is required to allow identification of $p(X_{(1)}, Y)$ and therefore for the target parameter J from Eq. 9:

Assumption 6.2 (Positivity for missing data)

if	$p(\underline{X}_{(1)}^{t-1} = \underline{x}^{t-1}, \underline{A}^{t-1} = \vec{1}) > 0,$
then	$\pi_\beta^t(A^t = \vec{1} \mid \underline{X}_{(1)}^{t-1} = \underline{x}^{t-1}, \underline{A}^{t-1} = \vec{1}) \geq \mathcal{O}$
	$\forall t, \underline{x}^{t-1},$ and some constant $\mathcal{O} > 0$

This positivity assumption is very different from the positivity assumption assumed under the offline RL view (Assumption 6.1). It requires the "acquire everything" action trajectory $A^t = \vec{1}\ \forall t$ to have positive support for all possible feature values. In other words, this is a requirement for complete cases among all subpopulations.

Given the positivity assumption, the block-conditional model describing $p(X_{(1)}, Y)$ is identified. We show how identification can be achieved in Appendix D.

Note that Theorem 1 holds even in the more general case without the temporal restriction $\overline{X}_{(1)}^t \not\to A^t$. In this case, $p(X_{(1)}, Y)$ may or may not be identified, depending on what assumptions can be made. There exists a vast literature on identification theory for missing data problems (Bhattacharya et al., 2020; Nabi et al., 2020; Mohan and Pearl, 2021) that can be applied.

Here, we briefly discuss how the common missing data scenarios of missing-completely-at-random (MCAR), missing-at-random (MAR), and missing-not-at-random (MNAR) apply within the AFA framework. MNAR scenarios arise when the missingness of a feature is influenced by the value of another feature which may itself be missing. This situation occurs when the NUC assumption is violated, indicated by the presence of arrows $X_{(1)} \to A$. On the other hand, MAR scenarios occur when the missingness of a feature is dependent only on observed features. This scenario aligns with our setting if the NUC assumption (Assumption 5) holds. Lastly, MCAR represents the simplest case, where the missingness of a feature is independent of all other feature values. In our framework, this corresponds to the absence of any edges $X \to A$.

5.3 Estimation

An estimate of $\mathbb{E}\left[C_{(\pi_\alpha)} | X_{(1)}, Y\right]$, denoted as $\hat{\mathbb{E}}\left[C_{(\pi_\alpha)} | X_{(1)}, Y\right]$, can be readily computed from Eq. 11 using Monte Carlo integration:

$$\hat{\mathbb{E}}[C | X_{(1)}, Y] = \sum_{i}^{n_{MC}} f_C(a_i, G_{a_i}(X_{(1)}), Y) \tag{12}$$

with n_{MC} samples $a_i \sim \prod_{t=1}^{T} \pi_\alpha(A^t | G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^{t-1})$.

This approach is common in online RL settings, where the agent interacts with the environment. Monte Carlo integration is performed in the following way: First, one takes

a sample x, y from $p(X_{(1)}, Y)$ (obtained through either of the methods shown below) and reveals the initial features x^0 to the agent. The agent's first action, denoted by a'^1, is then sampled from $\pi_\alpha(A^1|x^0)$. Depending on a'^1, the corresponding feature set amongst x^1 is revealed to the agent and the next action a'^2 is sampled. This continues until step T, when the classifier is applied (using the acquired subset of features from the simulation as input) and a resulting misclassification cost c' is computed. This process is repeated multiple times, and the costs are averaged to obtain $\hat{\mathbb{E}}\left[C_{(\pi_\alpha)}|X_{(1)}, Y\right]$.

This online RL process requires samples from $p(X_{(1)}, Y)$. These samples can be obtained using standard missing data estimation methods, which result in the following estimators:

1) Inverse probability weighting (IPW):
The target cost that is estimated by IPW (Seaman and White, 2013) is

$$\hat{J}_{IPW\text{-}Miss} = \hat{\mathbb{E}}_n\left[\hat{\rho}_{Miss}\,\hat{\mathbb{E}}\left[C_{(\pi_\alpha)}|X_{(1)}, Y\right]\right], \text{ where } \hat{\rho}_{Miss} = \prod_{t=1}^{T} \frac{\mathbb{I}(A^t = \vec{1})}{\hat{\pi}^t_\beta(A^t = \vec{1}|\underline{X}^{t-1}_{(1)}, \underline{A}^{t-1} = \vec{1})}, \tag{13}$$

and where $\mathbb{I}(.)$ denotes the indicator function. This IPW approach requires learning the propensity score π_β, but only for the scenario of full data acquisition (where $A = \vec{1}$). Because of the indicator function $\mathbb{I}(A^t = \vec{1})$, only the complete cases are selected for reweighting.

2) Multiple Imputation (MI):
The target cost that is estimated by the multiple imputation (MI) (Sterne et al., 2009) estimator is:

$$\hat{J}_{MI\text{-}Miss} = \hat{\mathbb{E}}_n\left[\sum_{X_m} \hat{\mathbb{E}}\left[C_{(\pi_\alpha)}|X_{(1)}, Y\right] \hat{p}(\underset{\uparrow}{X_m}|\underset{\downarrow}{X_o}, Y)\right]. \tag{14}$$

Missing part of $X_{(1)}$ (↓)
Observed part of $X_{(1)}$ (↑)

This estimator is based on the G-formula, which requires the counterfactual data distribution $p(X_{(1)}, Y)$. In the MI estimator, this density is not modeled fully, but the empirical distribution of the available data is augmented with samples (i.e., imputations) from a model for the missing data. The MI estimator is based on the decomposition $\hat{p}(X_{(1)}, Y) = \hat{p}(X_m|X_o, Y)p(X_o, Y)$, where X_o denotes the observed part and X_m the missing part of $X_{(1)}$. The sampling of the missing part is then usually repeated multiple times to increase the precision of the estimate, hence the name "multiple imputation".

In certain scenarios, MI can outperform the estimators from the semi-offline RL view, which will be described next. Appendix E discusses the advantages and disadvantages of the MI estimator in more detail. The appendix also highlights why using (conditional) mean imputation, which has been previously employed in AFA settings (An et al., 2022; Erion et al., 2021; Janisch et al., 2020) generally leads to biased estimation results.

6. Semi-offline Reinforcement Learning View

Assumptions in this section: Assumption 4 (NDE), Assumption 5 (NUC)
In this section, we assume both the NDE and NUC assumptions hold. In this context,

either the offline RL or the missing data view can be applied to solve AFAPE, but both have limitations, as illustrated in the following scenarios.

We assume an available data point without $X^0_{(1)}$, but with two time-steps and univariate $X^1_{(1)} = 0.6$ and $X^2_{(1)} = 0.9$. Initially, we assume the retrospective acquisition decisions were $A^1 = 1$ and $A^2 = 1$ (denoted as scenario 1). The data point in scenario 1 corresponds to a complete case where both feature values $X^1_{(1)}$ and $X^2_{(1)}$ are known.

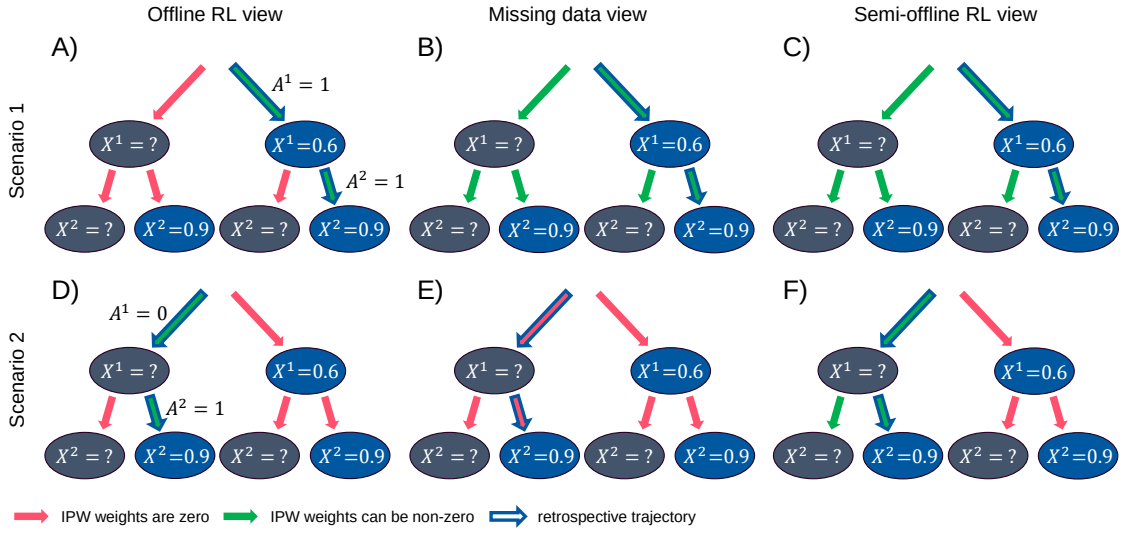


Figure 5: Visualization of data utilization by IPW estimators under different views. Each graph shows the four possible target acquisition trajectories for two exemplary retrospective acquisition scenarios and highlights which target trajectories can receive non-zero IPW weights under the respective views. A), D) The IPW estimator from the offline RL viewpoint: only target trajectories that match the retrospective trajectory can be evaluated. B), E) The IPW estimator from the missing data viewpoint: all trajectories can be evaluated if the datapoint is a complete case; otherwise, no evaluation is possible. C), F) The IPW estimator from the semi-offline RL viewpoint: all trajectories with equal or fewer acquisitions than the retrospective trajectory can be evaluated.

We analyze how this data point is utilized by the IPW estimators within different views. Figures 5A) and 5B) show the four possible target trajectories: $a^1 = 0, a^2 = 0$; $a^1 = 1, a^2 = 0$; $a^1 = 0, a^2 = 1$; and $a^1 = 1, a^2 = 1$ under the offline RL and missing data views, respectively. In the offline RL view, this data point can only be used to evaluate the target trajectory $a^1 = 1$ and $a^2 = 1$, matching the retrospective trajectory. The missing data IPW estimator, however, can evaluate any of the four target trajectories using this data point since it is a complete case.

Now consider scenario 2 with hypothetical retrospective acquisition decisions $A^1 = 0$ and $A^2 = 1$. Figures 5D) and 5E) depict the corresponding trajectories for both views. The

offline RL IPW estimator can use this datapoint to evaluate the matching trajectory $a^1 = 0$ and $a^2 = 1$ but none others. The missing data IPW estimator cannot evaluate any target trajectories as this is not a complete case.

In scenario 2, we only know the value of $X^2_{(1)}$, not $X^1_{(1)}$, since $A^1 = 0$. However, the target trajectories $a^1 = 0$, $a^2 = 0$ and $a^1 = 0$, $a^2 = 1$ do not require the value of $X^1_{(1)}$ and can still be simulated. This motivates our novel semi-offline RL viewpoint, where we consider all target trajectories with the same or fewer acquisitions, compared with the retrospective data point, as simulatable.

Figures 5C) and 5F) show the simulatable trajectories under the semi-offline RL view for both scenarios. Similar to online RL, a policy can sample different trajectories, but under the semi-offline RL view, only among the simulatable ones, while trajectories involving non-simulatable actions (like $a^1 = 1$ in scenario 2) should not be sampled.

Since not all trajectories are simulatable, we restrict the simulation policy to block actions that would result in non-simulatable trajectories. A simple Monte Carlo estimator averaging the costs within the simulated trajectories would be biased due to the blocking of actions, necessitating post-simulation bias correction.

The remainder of this section formalizes the semi-offline RL viewpoint and is organized as follows. First, we introduce the simulation policy, referred to as the semi-offline sampling policy, and illustrate that simulations using this policy do not require information about $X_{(1)}$ not already contained in $X = G_A(X_{(1)})$. Next, we explain how the AFAPE target J can be equivalently formulated based on the semi-offline sampling distribution. We then prove that this reformulated J is identified under a new, weaker positivity assumption, and finally, we derive novel estimators for J.

6.1 The Semi-offline Sampling Policy

To introduce the semi-offline RL view, we first revisit the problem formulation from the missing data view (Eqs. 9 and 11):

$$J = \sum_{X_{(1)},Y} \underbrace{\sum_{a \in \mathcal{A}} f_C(a, G_a(X_{(1)}), Y)\pi_\alpha(a|G_a(X_{(1)}))}_{\text{online RL}} \underbrace{p(X_{(1)}, Y)}_{\text{missing data}}. \tag{15}$$

The online RL part involves integrating over all possible trajectories $a \in \mathcal{A}$, representing subsets of the features. Notably, many terms in this inner integral do not require complete knowledge of $X_{(1)}$. Specifically, any summand with $a \leq A$ (where we let $\leq$ denote an element-wise comparison) does not need information from $X_{(1)}$ beyond what is already in $X = G_A(X_{(1)})$.

This missing data + online RL approach effectively says: "Solve all missingness, then integrate over all possible feature subsets, including many that didn't require addressing the missingness to begin with." This approach is clearly suboptimal. Instead, we propose a novel semi-offline RL viewpoint: "Integrate over all subsets of the observed data and adjust for the bias introduced by excluding subsets where data was missing." We call this the semi-offline RL viewpoint because some subsets/trajectories can be simulated (the online part) while others cannot (the offline part).

We define the semi-offline sampling distribution to include the observed data trajectories. To avoid integrating over unobserved feature subsets, we replace π_α in the distribution of Eq. 15 with a policy π'_{sim} that has no support for trajectories where the corresponding $X_{(1)}$ are missing. A policy π' that enforces this exclusion is defined formally as a blocked policy:

Definition 2. *(Blocked Policy). A policy* $\pi'^t(A'^t|G_{\underline{A}'^{t-1}}(X_{(1)}), \underline{A}'^{t-1}, A^t)$ *is called a 'blocked policy' of the policy* $\pi^t(A'^t|G_{\underline{A}'^{t-1}}(X_{(1)}), \underline{A}'^{t-1})$ *if it satisfies the following conditions:*
1) Blocking of acquisitions of non-available features:

$$\text{if } a'^t \not\leq a^t, \qquad \text{then } \pi'^t(a'^t|G_{\underline{a}'^{t-1}}(X_{(1)}) = \underline{x}'^{t-1}, \underline{a}'^{t-1}, a^t) = 0 \qquad \forall t, a'^t, a^t, \underline{x}'^{t-1}, \underline{a}'^{t-1}$$

2) No blocking of acquisitions of available features:

$$\text{if } a'^t \leq a^t \text{ and } \pi^t(a'^t|\underline{x}'^{t-1}, \underline{a}'^{t-1}) > 0, \quad \text{then } \pi'^t(a'^t|\underline{x}'^{t-1}, \underline{a}'^{t-1}, a^t) > 0 \quad \forall t, a'^t, a^t, \underline{x}'^{t-1}, \underline{a}'^{t-1}$$

Condition 1 ensures that sampling does not depend on values of $X_{(1)}$ that are not contained in $X = G_A(X_{(1)})$. Condition 2 ensures that the online exploration part is utilized by forcing the blocked policy π' to have positive support whenever π has positive support and the desired features are available. A practical choice for π' just sets $\pi'(A'^t = a'^t|G_{\underline{A}'^{t-1}}(X_{(1)}), \underline{A}'^{t-1}, A^t = a^t) = 0$ if $a'^t \not\leq a^t$ and rescales the other probabilities accordingly.

Having defined the blocked policy, we can now introduce the semi-offline sampling distribution:

$$\begin{aligned} p'(A', G_A(X_{(1)}), Y, A) &= \prod_{t=1}^{T} \pi'^t_{sim}(A'^t|G_{\underline{A}'^{t-1}}(X_{(1)}), \underline{A}'^{t-1}, A^t) p(A, G_A(X_{(1)}), Y) \qquad (16) \\ &\equiv \underbrace{\pi'_{sim}(A'|G_{A'}(X_{(1)}), A)}_{\text{semi-offline RL}} \underbrace{p(A, G_A(X_{(1)}), Y)}_{\text{observed data}} \end{aligned}$$

which only involves observed data because $A' \leq A$ due to the blocking.

Similar to online RL, we can sample from this distribution to construct a simulated data set $\mathcal{D}'$, consisting of $X = G_A(X_{(1)}), Y, A, A'$, and $C' = f_C(A', G_{A'}(X_{(1)}), Y)$. In Figure 6, we show the full causal graph of how such sampling is performed.

Note that although the number of observed data samples is limited by the data set size n, the semi-offline RL variables A' and C' can be sampled multiple times beyond this constraint. While π'_{sim} can be chosen to be a blocked AFA policy π'_α, this is not required; different choices of π'_{sim} introduce an off-policy aspect. The only requirement is that π_{sim} (the unblocked version of π'_{sim}) meets the positivity assumption from the offline RL view (Assumption 6.1).

Since we replace π_α with π'_{sim}, the resulting cost samples C' cannot simply be averaged to estimate J. Instead, we reformulate the AFAPE problem as a causal inference problem by intervening on the semi-offline sampling distribution p', reversing π'_{sim} back to π_α. This reformulation is formalized as follows.

6.2 Problem Reformulation

The AFAPE problem can be reformulated under the semi-offline RL view (i.e., under the proposed distribution p') as the following theorem states.

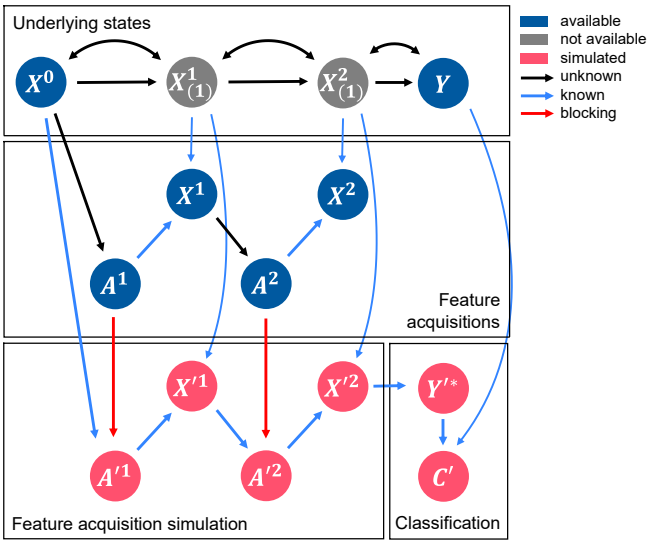


Figure 6: Causal graph for the semi-offline sampling distribution p'. Simulated acquisition actions A'^t and observations $X'^t = G_{A'^t}(X^t_{(1)})$ follow a blocked simulation policy π'_{sim}. The simulation policy is restricted by A^t such that actions A'^t are blocked if $A'^t_i > A^t_i$ for any i. The simulated cost C' can be computed from A', X' and Y: $C' = f_C(Y'^*, Y)$ with $Y'^* = f_{cl}(X', A')$ being the predicted label under the simulated acquisitions. Edges showing long-term dependencies are omitted from the graph for visual clarity. These include $\underline{X}^{t-1}, \underline{A}^{t-1} \to A^t$; $\underline{X}'^{t-1}, \underline{A}'^{t-1} \to A'^t$; $\underline{X}'^T, \underline{A}'^T \to Y^*$; $\underline{X}^{t-1}_{(1)} \leftrightarrow X^t_{(1)}$; and $\underline{X}^T_{(1)} \leftrightarrow Y$.

Theorem 3. *(AFAPE problem reformulation under the semi-offline RL view). The AFAPE problem of estimating J (Eq. 2 or Eq. 9) is under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE) and Assumption 5 (NUC) equivalent to estimating*

$$J = \mathbb{E}_{p'}[C'_{(\pi_\alpha)}]. \tag{17}$$

$C'_{(\pi_\alpha)}$ denotes the potential outcome of C', had, instead of the blocked simulation policy π'_{sim}, the AFA policy π_α been employed.

Proof Starting from Eq. 15, we find:

$$\begin{aligned}
J &= \sum_{X_{(1)},Y} \sum_{a \in \mathcal{A}} f_C(a, G_a(X_{(1)}), Y) \pi_\alpha(a | G_a(X_{(1)})) p(X_{(1)}, Y) \\
&= \sum_{X_{(1)},Y,A_{(\pi_\alpha)}} f_C(A_{(\pi_\alpha)}, G_{A_{(\pi_\alpha)}}(X_{(1)}), Y) p(A_{(\pi_\alpha)}, X_{(1)}, Y) \\
&= \sum_{X_{(1)},Y,A_{(\pi_\alpha)},A} f_C(A_{(\pi_\alpha)}, G_{A_{(\pi_\alpha)}}(X_{(1)}), Y) p(A_{(\pi_\alpha)}, X_{(1)}, Y, A) \\
&\overset{*1}{=} \sum_{X_{(1)},Y,A'_{(\pi_\alpha)},A} f_C(A'_{(\pi_\alpha)}, G_{A'_{(\pi_\alpha)}}(X_{(1)}), Y) p'(A'_{(\pi_\alpha)}, X_{(1)}, Y, A) = \mathbb{E}_{p'}[C'_{(\pi_\alpha)}].
\end{aligned}$$

where the change in $*1$) from $A_{(\pi_\alpha)}$ to $A'_{(\pi_\alpha)}$ is possible since an intervention on the simulated actions leads to the same distribution as an intervention on the observed actions.

■

Remark 2 (Comparison of AFAPE under offline RL and semi-offline RL) The AFAPE problem formulation under the semi-offline RL view (Eq. 17) closely resembles the original AFAPE formulation under the offline RL view (Eq. 2): $J = \mathbb{E}[C_{(\pi_\alpha)}]$. This similarity might raise the question of what is gained by this reformulation if one still needs to adjust for the intervention π_α. The key difference is that, in the offline RL view, one must account for a significant distribution shift from π_β to π_α. In contrast, under the semi-offline RL view, the distribution shift is much smaller, from a blocked π'_{sim} to π_α.

6.3 Identification

In the following section, we focus on identifying the reformulated target J from Eq. 17. Similar to the other views, identification under the semi-offline RL view requires a positivity assumption. However, under the semi-offline RL view, this assumption is considerably less stringent compared to the positivity requirements in the offline RL (Assumption 6.1) and missing data views (Assumption 6.2).

In the offline RL view, the positivity assumption requires support for retrospective trajectories (i.e., action sequences under π_β) that match any target trajectory (i.e., action sequences under π_α). The missing data view demands support for "acquire all" retrospective trajectories, meaning complete cases. In contrast, the semi-offline RL view only necessitates that for any target trajectory, there is positive support for at least one retrospective trajectory with equal or more acquisitions. We now formalize this positivity assumption rigorously. Due to its inherent complexity, we divide the formalization into multiple definitions:

Definition 4. *(Local positivity assumption and local admissible set $\mathcal{A}_{adm}$ for semi-offline RL). Let the local admissible set $\mathcal{A}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t)$, defined for all a'^t and all $\underline{x}^{t-1}, \underline{a}^{t-1}$ s.t. $p(G_{\underline{a}^{t-1}}(X_{(1)}) \equiv \underline{X}^{t-1} = \underline{x}^{t-1}, \underline{A}^{t-1} = \underline{a}^{t-1}) > 0$, be the non-empty set containing all values of a^t for which*

$$(1) \qquad a^t \geq a'^t$$
$$(2) \qquad \pi^t_\beta(a^t|\underline{x}^{t-1}, \underline{a}^{t-1}) \geq \mathcal{O}$$

for some constant $\mathcal{O} > 0$. We further say that the local positivity assumption holds at $\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t$ if $\mathcal{A}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t)$ exists.

The local positivity assumption states that the observed data allows the simulation of a desired action a'^t (i.e., there is positive support for at least one value a^t s.t. $a^t \geq a'^t$). The local positivity assumption is, however, not enough, which leads to the following definition of regional positivity.

Definition 5. *(Regional positivity assumption and regional admissible set $\tilde{\mathcal{A}}_{adm}$ for semi-offline RL). Let the regional admissible set $\tilde{\mathcal{A}}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, \underline{a}'^t) \subseteq \mathcal{A}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t)$,*

defined for all $\underline{x}^{t-1}, \underline{a}^{t-1}, \underline{a}'^{t}$ *s.t.* $p'(G_{\underline{a}^{t-1}}(X_{(1)}) = \underline{x}^{t-1}, \underline{A}^{t-1} = \underline{a}^{t-1}, \underline{A}'^{t} = \underline{a}'^{t}) > 0$*, be the non-empty set containing all values of* a^t *such that* $\tilde{\mathcal{A}}^{t+1}_{adm}(\underline{x}^t, \underline{a}^t, \underline{a}'^{t+1})$ *exists for all* $\underline{x}^t, \underline{a}^t, \underline{a}'^{t+1}$ *such that the following holds for* a'^{t+1}*, and* x^t*:*

$$p(x^t|\underline{x}^{t-1}, \underline{a}^t)\pi^t_\alpha(a'^{t+1}|\underline{x}'^{t-1}, \underline{a}'^t) > 0 \tag{18}$$

where $\underline{x}'^{t-1}$ *denotes the acquired features under* $\underline{a}'^{t-1}$ *and thus a subset of* $\underline{x}^{t-1}$*. We further say that the regional positivity assumption holds at* $\underline{x}^{t-1}, \underline{a}^{t-1}, \underline{a}'^{t}$ *if* $\tilde{\mathcal{A}}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, \underline{a}'^t)$ *exists.*

Regional positivity states that there is not only a value a^t with positive support in the observed data that allows the simulation of a desired action a'^t (i.e., local positivity), but it also ensures for such an a^t, that the simulations of all possible future desired actions are also possible. As regional positivity is still limited to a given data point, we also make the following global positivity assumption.

Assumption 6.3 (Global positivity assumption for semi-offline RL) We say that the global positivity assumption holds if the regional positivity assumption holds for all datapoints a'^1, x^0 s.t. $p(x^0)\pi_\alpha(a'^1|x^0) > 0$.

The following lemma (proved in Appendix F) establishes that the global positivity assumption for semi-offline RL is weaker than both the positivity assumptions required by the offline RL and the missing data views:

Lemma 6. *(Sufficiency conditions for global positivity). The global positivity assumption for semi-offline RL (Assumption 6.3) holds if the positivity assumption from offline RL (Assumption 6.1) or from missing data (Assumption 6.2) holds.*

After having defined positivity for semi-offline RL, we can now perform identification for J:

Theorem 7. *(Identification of J for the semi-offline RL view). The reformulated AFAPE problem of estimating J under the semi-offline RL view (Eq. 17) is under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and Assumption 6.3 (global positivity) identified by*

$$J = \mathbb{E}_{p'}[C'_{(\pi_\alpha)}] = \sum_{A', A, G_A(X_{(1)}), Y} f_C(A', X', Y) q'(A', A, X, Y) \tag{19}$$

with the distribution

$$q'(A', A, X, Y) = \prod_{t=1}^{T} \underbrace{\pi^t_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})}_{\text{distr. subject to constraints}} \underbrace{\pi^t_\alpha(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1})}_{\text{target policy}} \prod_{t=0}^{T} p(X^t|\underline{X}^{t-1}, \underline{A}^t, Y) p(Y) \tag{20}$$

where

$$\pi^t_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \underbrace{\mathbb{I}(A^t \in \tilde{\mathcal{A}}^t_{adm}(\underline{X}^{t-1}, \underline{A}^{t-1}, \underline{A}'^t))}_{\text{support restriction}} f^t_{id}(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) \tag{21}$$

for any function f^t_{id} *s.t.* π^t_{id} *is a valid density.*

Here, we used again $X \equiv G_A(X_{(1)})$ and $X' \equiv G_{A'}(X_{(1)})$ to simplify notation. In fact, since $A' \leq A$, we can also write $X' = G_{A'}(X)$. Additionally, we can define semi-offline RL versions of the Bellman equation:

Theorem 8. *(Bellman equation for semi-offline RL). The semi-offline RL view admits under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and the local positivity assumption at datapoint $\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t$ (from Definition 4), the following semi-offline RL version of the Bellman equation:*

$$Q_{Semi}(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi) = \sum_{X^t} V_{Semi}(\underline{A}'^t, \underline{X}^t, \underline{A}^{t-1}, A^t = a^t, \Xi) p(X^t | \underline{X}^{t-1}, \underline{A}^{t-1}, A^t = a^t, \Xi) \tag{22}$$

$$\text{for any } a^t \in \mathcal{A}^t_{adm}(\underline{X}^{t-1}, \underline{A}^{t-1}, A'^t)$$

$$V_{Semi}(\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi) = \sum_{A'^{t+1}} Q_{Semi}(\underline{A}'^{t+1}, \underline{X}^t, \underline{A}^t, \Xi) \pi^{t+1}_\alpha(A'^{t+1} | \underline{X}'^t, \underline{A}'^t) \tag{23}$$

with semi-offline RL versions of the state-action value function Q_{Semi} and state value function V_{Semi}:

$$Q^t_{Semi} \equiv Q_{Semi}(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi) \equiv \mathbb{E}_{p'}[C'_{(\overline{\pi}^{t+1}_\alpha)} | \underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi]$$
$$V^t_{Semi} \equiv V_{Semi}(\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi) \equiv \mathbb{E}_{p'}[C'_{(\overline{\pi}^{t+1}_\alpha)} | \underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi]$$

where $C'_{(\overline{\pi}^{t+1}_\alpha)}$ denotes the potential outcome of C' under interventions from time step $t+1$ onwards. $\Xi \subseteq \{Y, O\}$, with O denoting all features that are always available, denotes an optional subset of additional variables that can be conditioned on. Furthermore, Q^t_{Semi} and V^t_{Semi} are identified if the regional positivity assumption (from Definition 5) holds at $\underline{X}^{t-1}, \underline{A}^{t-1}, \underline{A}'^t$ and $a^t \in \tilde{\mathcal{A}}^t_{adm}(\underline{X}^{t-1}, \underline{A}^{t-1}, \underline{A}'^t)$.

The proofs for Theorems 7 and 8 are shown in Appendix G.

The functions Q_{Semi} and V_{Semi} are very similar to their counterparts from the offline RL view (Q_{Off} and V_{Off}), with two differences: i) they are learned from a curated data set $\mathcal{D}'$ which arises from sampling p'; and ii) they contain the simulated acquisitions $\underline{A}'^t$, but also the real features and actions $\underline{X}^t, \underline{A}^t$ which are needed to adjust for confounding of the blocking operation.

The identification steps so far have been very specific to knowledge about $\tilde{\mathcal{A}}^t_{adm}$ that needs to be assessed from the data. We now look more closely at a specific, stronger positivity assumption (where $\mathcal{A}_{adm} = \tilde{\mathcal{A}}_{adm}$) that allows the use of the maximum amount of datapoints and therefore leads to easier to use estimators.

Definition 9. *(Maximal regional positivity assumption for semi-offline RL): We say that the maximal regional positivity assumption holds for a datapoint $\underline{x}^{t-1}, \underline{a}^{t-1}, \underline{a}'^t$ if $\tilde{\mathcal{A}}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, \underline{a}'^t) = \mathcal{A}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t)$ and the maximal regional positivity assumption further holds for all $\underline{x}^t, \underline{a}^t, \underline{a}'^{t+1}$ such that $a^t \in \mathcal{A}^t_{adm}(\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t)$ and the following holds for a'^{t+1}, and x^t:*

$$p(x^t | \underline{x}^{t-1}, \underline{a}^t) \pi_\alpha(a'^{t+1} | \underline{x}'^t, \underline{a}'^t) > 0.$$

Assumption 6.4 (Maximal global positivity assumption for semi-offline RL) We say that the maximal global positivity assumption holds if the maximal regional positivity assumption holds for all datapoints x^0, a'^1 s.t. $p(x^0)\pi_\alpha(a'^1|\underline{x}'^0) > 0$.

The maximal regional positivity and maximal global positivity assumptions ensure that we can use all available data points where $A^t \geq A'^t$ without running into positivity problems in later time steps. This makes the identification and estimation significantly easier, as shown next.

We can now propose the following corollary of Theorem 7, which states identification under the maximal global positivity assumption.

Corollary 10. *(Identification of J for the semi-offline RL view under maximal global positivity). The reformulated AFAPE problem of estimating J under the semi-offline RL view (Eq. 17) is under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and Assumption 6.4 (maximum global positivity) identified by Eqs. 19 and 20 where*

$$\pi^t_{id}(A^t|\underline{A}'^{t-1}, A'^t = a'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \\ = \mathbb{I}(A^t \geq a'^t)\pi^t_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1}) f^t_{id}(\underline{A}'^{t-1}, A'^t = a'^t, \underline{X}^{t-1}, \underline{A}^{t-1})$$

for any function f^t_{id} s.t. π^t_{id} is a valid density. This holds in particular for the choice of a truncated π_β:

$$\pi^t_{id}(A^t|\underline{A}'^{t-1}, A'^t = a'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \pi^t_\beta(A^t|A^t \geq a'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \\ = \frac{\mathbb{I}(A^t \geq a'^t)\pi^t_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})}{\pi^t_\beta(A^t \geq a'^t|\underline{X}^{t-1}, \underline{A}^{t-1})}.$$

The proof for Corollary 10 is shown in Appendix H.

Lastly, we also provide the following Remark that brings the factorization of the observational distribution p' from Eq. 16 into a comparable form to the identifying distribution q'.

Remark 3 (Factorization of the semi-offline sampling distribution) The semi-offline sampling distribution p' can be alternatively written as:

$$p'(A', G_A(X_{(1)}), Y, A) = \prod_{t=1}^{T} \pi'^t_{sim}(A'^t|G_{\underline{A}'^{t-1}}(X_{(1)}), \underline{A}'^{t-1}, A^t) p(A, G_A(X_{(1)}), Y) \\ = \prod_{t=1}^{T} \underbrace{\pi'^t_{sim}(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t)}_{\text{known simulation policy}} \underbrace{\pi_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})}_{\text{retro. acquisition policy}} \prod_{t=0}^{T} p(X^t|\underline{X}^{t-1}, \underline{A}^t, Y) p(Y) \quad (24)$$

6.4 Estimation

We propose the following novel estimators for J, which arise from the semi-offline RL viewpoint. We differentiate between estimators derived under the global positivity assumption (Assumption 6.3) and under the (stronger) maximal global positivity assumption (Assumption 6.4).

1) Inverse probability weighting (IPW):
The target cost that is estimated by the semi-offline IPW estimator is

$$\hat{J}_{IPW\text{-}Semi} = \hat{\mathbb{E}}_n \left[\hat{\mathbb{E}}_{n'} \left[\hat{\rho}^T_{Semi}\, C' \middle| A, X, Y \right] \right]. \tag{25}$$

The inner expectation, denoted by $\hat{\mathbb{E}}_{n'}[.]$, represents the empirical average over the simulated values of A', which can involve many more samples compared to the outer expectation, taken over the observed data. The inverse probability weights are under the global positivity assumption:

$$\hat{\rho}^T_{Semi} = \hat{\rho}^T_{Semi}(\pi_{id}) = \prod_{t=1}^{T} \frac{\pi^t_\alpha(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1})}{\pi'^t_{sim}(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t)} \frac{\pi^t_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})}{\hat{\pi}^t_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})} \tag{26}$$

or under the maximal global positivity assumption (by choosing $\pi^t_{id} = \pi^t_\beta(A^t|A^t \geq a'^t, \underline{X}^{t-1}, \underline{A}^{t-1})$:

$$\hat{\rho}^T_{Semi} = \prod_{t=1}^{T} \frac{\pi^t_\alpha(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1})}{\pi'^t_{sim}(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t)} \frac{\mathbb{I}(A^t \geq a'^t)}{\hat{\pi}^t_\beta(A^t \geq a'^t|\underline{X}^{t-1}, \underline{A}^{t-1})}. \tag{27}$$

The following remarks state that the IPW estimators from the offline RL and missing data viewpoints are special cases of the proposed estimator:

Remark 4 (Offline RL IPW estimator as a special version of the semi-offline RL IPW estimator) The IPW estimator from the offline RL view, $\hat{J}_{IPW\text{-}Off}$, is, for the choice $\pi'_{sim} = \pi'_\alpha$, equal to $\hat{J}_{IPW\text{-}Semi}$ with:

$$\pi^t_{id}(A^t|\underline{A}'^{t-1}, A'^t = a'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \mathbb{I}(A^t = a'^t).$$

Remark 5 (Missing data IPW estimator as a special version of the semi-offline RL IPW estimator) The IPW estimator from the missing data view, $\hat{J}_{IPW\text{-}Miss}$, is, for the choice $\pi'_{sim} = \pi'_\alpha$, equal to $\hat{J}_{IPW\text{-}Semi}$ with:

$$\pi_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \mathbb{I}(A^t = \vec{1}).$$

The IPW estimator $\hat{J}_{IPW\text{-}Semi}$ under the maximal global positivity assumption demonstrates the large benefits of the semi-offline RL view over both the offline RL and missing data views. Its second fraction shows that not only datapoints where $A^t = A'^t$ are used (i.e., have positive weight), as in the offline RL view, neither only datapoints where $A^t = \vec{1}$ are used, as in the missing data view, but all datapoints where $A^t \geq A'^t$ can be used. However, this benefit diminishes as the target policy becomes more "data-hungry"—that is, as it acquires more features. In fact, there is no difference between all three estimators for an "acquire all features" policy:

Remark 6 (Equality of IPW estimators for an "acquire all features" policy) The IPW estimators from the missing data view, $\hat{J}_{IPW\text{-}Miss}$, from the offline RL view $\hat{J}_{IPW\text{-}Miss}$, and from the semi-offline RL view $\hat{J}_{IPW\text{-}Semi}$ are identical if $\pi^t_\alpha(A^t = \vec{1}|\underline{X}^{t-1} = \underline{x}^{t-1}, \underline{A}^{t-1} = \vec{1}) = 1\ \forall t, \underline{x}^{t-1}$.

We show in Appendix I that $\hat{J}_{IPW\text{-}Semi}$ (under maximal global positivity) is equivalent in simple AFA settings to an adapted version of the IPW estimator by Caniglia et al. (2019). Our IPW estimators $\hat{J}_{IPW\text{-}Semi}$ can, however, be applied in more general AFA settings.

2) Direct method (DM):
The target cost that is estimated by the semi-offline DM estimator is

$$\hat{J}_{DM\text{-}Semi} = \hat{\mathbb{E}}_n[\hat{V}^0_{Semi}] \tag{28}$$

This estimator is based on learning a semi-offline RL version of the state-action value function Q_{Semi} using the semi-offline version of the Bellman equation (Eqs. 22 and 23). Using Q_{Semi}, one can compute the state value function: $V^t_{Semi} = \mathbb{E}_{\pi_\alpha}[Q^{t+1}_{Semi}]$.

The training process of Q_{Semi} can benefit from using the off-policy aspect of the proposed semi-offline sampling distribution p' (i.e., from using a simulation policy π'_{sim} that is different from π'_α). This is because a deterministic AFA policy, for example, will only generate one exact trajectory of simulated actions A' and costs C' per datapoint X, Y, A. A stochastic simulation policy π'_{sim} can instead be used to generate multiple such trajectories. Usually, a parametric working model, for example a multi-layer perceptron (MLP), is assumed for the nuisance function $\hat{Q}_{Semi}$. The estimation under such a working model will benefit from the additional datapoints generated under a policy $\pi'_{sim} \neq \pi'_\alpha$.

3) Double reinforcement learning (DRL):
The target cost that is estimated by the semi-offline DRL estimator is

$$\hat{J}_{DRL\text{-}Semi} = \hat{\mathbb{E}}_n \left[\hat{\mathbb{E}}_{n'} \left[\hat{\rho}^T_{Semi} C' + \sum_{t=1}^{T} \left(-\hat{\rho}^t_{Semi} \hat{Q}^t_{Semi} + \hat{\rho}^{t-1}_{Semi} \hat{V}^{t-1}_{Semi} \right) \middle| A, X, Y \right] \right]. \tag{29}$$

which holds for both choices for ρ_{Semi}, given that the respective positivity assumption holds. Similar to the DRL estimator from the offline RL view, this approach combines the other two estimators (Eqs. 25 and 28).

The following theorems state some notable facts about these estimators.

Theorem 11. *(Consistency of $\hat{J}_{IPW\text{-}Semi}$). The estimator $\hat{J}_{IPW\text{-}Semi}$ is consistent if the propensity score model $\hat{\pi}_\beta$ is correctly specified.*

Proof We apply the standard inverse probability weighting approach $\mathbb{E}_{q'}[C'] = \mathbb{E}_{p'}[\frac{q'}{p'} C']$ and use the factorizations for q' and p' from Eqs. 20 (Theorem 7) and 24 (Remark 3), respectively, to obtain Eq. 26 for the weights. The weights from Eq. 27 arise from inserting the special choice for π_{id}. ■

Theorem 12. *(Consistency of $\hat{J}_{DM\text{-}Semi}$). The estimator $\hat{J}_{DM\text{-}Semi}$ is consistent if the Q-function $\hat{Q}_{Semi}$ is correctly specified.*

Proof The proof of the consistency of $\hat{J}_{DM\text{-}Semi}$ follows simply from the semi-offline Bellman equation (Theorem 8) and the law of total expectation. ■

Theorem 13. *(Double robustness of $\hat{J}_{DRL\text{-}Semi}$). The estimator $\hat{J}_{DRL\text{-}Semi}$ is doubly robust, in the sense that it is consistent if either the Q-function $\hat{Q}_{Semi}$ or the propensity score model $\hat{\pi}_\beta$ is correctly specified.*

The proof is shown in Appendix J. The estimator $\hat{J}_{DRL\text{-}Semi}$ is a 1-step estimator based on an influence function derived for J under p'. Therefore, the DRL estimator is regular and asymptotically linear (RAL). The influence function is given by the following theorem:

Theorem 14. *(An influence function under the semi-offline RL view). An influence function of J is:*

$$\varphi_{Semi} = -J + \mathbb{E}\left[\rho^T_{Semi} C' + \sum_{t=1}^{T} \left(-\rho^t_{Semi} Q^t_{Semi} + \rho^{t-1}_{Semi} V^{t-1}_{Semi}\right) \middle| A, X, Y\right]. \tag{30}$$

We will prove this theorem in the next section, addressing semiparametric estimation from all three views.

Remark 7 (Efficiency of φ_{Semi} as a function of Ξ) In Theorem 8, we showed that the semi-offline RL version of the Bellman equation holds for any subset $\Xi \subseteq \{Y, O\}$, and the same applies to Theorem 14. The choice of Ξ affects the efficiency of the DRL estimator: a bigger set Ξ corresponds to a higher efficiency of the corresponding DRL estimator. However, the class of influence functions presented here does not include the efficient influence function. In fact, as we will discuss in the next section, the efficient influence function lacks a closed-form expression.

In Appendix K, we extend the estimators discussed in this section to other settings. These include: i) the estimation of J_a, and ii) scenarios where a prediction Y^{*t} is required at each time step t.

7. Semiparametric Theory under NUC and NDE

Assumptions in this section: Assumption 4 (NDE), Assumption 5 (NUC)
In this section, we explore semiparametric estimation approaches for J under both NDE and NUC assumptions. Readers not interested in the detailed semiparametric theory can skip this section. We demonstrate that all three views can be unified within an established semiparametric theory framework for MAR missing data problems. Using this framework, we prove Theorem 14, which defines a class of influence functions derived under the semi-offline RL view. Although no closed-form efficient influence function exists in this setting, efficiency improvements from Tsiatis (2006) and Liu et al. (2021) can be applied. However, these methods pose significant challenges, including strong positivity assumptions, applicability to a limited set of problems, and implementation complexity.

To discuss semiparametric approaches to AFAPE, we first remind the reader of the semiparametric theory review in Appendix B. The following section draws extensively on the foundational framework of Tsiatis (2006) to establish essential context.

Let the observed data influence function be denoted as $\varphi \equiv \varphi(A, G_A(X_{(1)}), Y)$, and the observed data tangent space as Λ, with the corresponding nuisance tangent space denoted by Λ_{nuis}. The full data influence function - an influence function given the counterfactual

variables $X_{(1)}$ - is denoted $\varphi^F \equiv \varphi^F(X_{(1)}, Y)$, with full data tangent space Λ^F and nuisance tangent space Λ^F_{nuis}. The two key relationships between these spaces and the influence functions are as follows:

- An observed data influence function must lie in the orthocomplement of the observed data nuisance tangent space: $\varphi \in \Lambda^{\perp}_{nuis}$.
- The observed data efficient influence function must be in the observed data tangent space: $\varphi_{eff} \in \Lambda$.

Known semiparametric theory for MAR missing data methods can be applied to the AFAPE problem under the NUC assumption. We begin by defining the space of full data influence functions. Since we assume no restrictions on the full data, the full data influence function is unique, efficient, and given by the online RL part of the missing data + online RL view:

$$\varphi^F(X_{(1)}, Y) = \underbrace{\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)}, Y]}_{\text{online RL}} - J = \sum_{a \in \mathcal{A}} f_C(a, G_a(X_{(1)}), Y)\pi_\alpha(a|G_a(X_{(1)})) - J. \quad (31)$$

While the full data influence function is unique, the space of observed data influence functions is generally not. To find it, we first define the orthogonal complement of the nuisance tangent space, which can be expressed as (Theorem 8.3 from Tsiatis (2006)):

$$\Lambda^{\perp}_{nuis} = \left\{ \left[h^*(A, G_A(X_{(1)}), Y) \oplus \Lambda_2\right] - \Pi\left(\left[h^*(A, G_A(X_{(1)}), Y) \oplus \Lambda_2\right] \Big| \Lambda_{nuis,\psi} \right) \right\}, \quad (32)$$

where h^* belongs to the inverse probability weighting (IPW) space Λ^*_{IPW}, Λ_2 is the augmentation space, and $\Lambda_{nuis,\psi} = \Lambda_\psi$ is the nuisance tangent space, or equivalently the tangent space of the acquisition process. We explain these three spaces now in more detail.

The inverse probability weighting space Λ^*_{IPW}:
The function $h^*(A, G_A(X_{(1)}), Y)$ in Eq. 32 can be any function in the IPW space Λ^*_{IPW} which is defined as:

$$\Lambda^*_{IPW} \equiv \left\{ h^*(A, G_A(X_{(1)}), Y) \in \mathcal{H} : \mathbb{E}[h^*(A, G_A(X_{(1)}), Y)|X_{(1)}, Y] = \varphi^{*F}(X_{(1)}, Y) \right\}.$$

where $\mathcal{H}$ denotes the space of random functions with zero mean and finite variance and $\varphi^{*F}(X_{(1)}, Y)$ denotes an element of the orthocomp of the full data nuisance tangent space: $\varphi^{*F} \in \Lambda^{\perp}_{nuis}$.

In fact, if we further restrict the IPW space such that we don't allow any element of the orthocomp of the full data nuisance tangent space $\varphi^{*F}(X_{(1)}, Y)$, but only the full data influence function $\varphi^F(X_{(1)}, Y)$, then we also obtain only observed data influence functions by Eq. 32 (Theorem 8.3 from Tsiatis (2006)). In the following, we thus restrict the IPW space to:

$$\Lambda_{IPW} \equiv \left\{ h(A, G_A(X_{(1)}), Y) \in \mathcal{H} : \mathbb{E}[h(A, G_A(X_{(1)}), Y)|X_{(1)}, Y] = \varphi^F(X_{(1)}, Y) \right\}.$$

In that case, the IPW space contains functions that, when taking the conditional expected value with respect to the full data, equal the full data influence function. The space is denoted as the IPW space because IPW-based identifying functions can be chosen to construct

elements in this space. In fact, as will be shown in this section, all the IPW estimators - from the offline RL, missing data, and semioffline RL views - are applicable and form valid elements in Λ_{IPW}.

The augmentation space Λ_2:
The augmentation space Λ_2 is defined as follows (Lemma 7.4 from Tsiatis (2006)):

$$\Lambda_2 = \Big\{ b(A, G_A(X_{(1)}), Y) \in \mathcal{H} : \mathbb{E}\left[b(A, G_A(X_{(1)}), Y) | X_{(1)}, Y \right] = 0 \Big\}. \tag{33}$$

This space contains functions that, when taken in conditional expectation with respect to the full data, equal zero. This provides intuition behind the decomposition of the space of observed data influence functions: the space consists of one function that, in conditional expectation, equals the full data influence function, plus all functions that become zero in conditional expectation. One must, however, still obtain the residual of the projection of these functions onto the nuisance tangent space of the acquisition process, $\Lambda_{nuis,\psi}$, introduced below.

The nuisance tangent space of the acquisition process $\Lambda_{nuis,\psi}$:
The space $\Lambda_{nuis,\psi}$ corresponds to the observed data nuisance tangent space of the acquisition process and is a subspace of Λ_2 (Theorem 8.1 from Tsiatis (2006)). In our AFA setting, future feature values do not influence past acquisition decisions, resulting in the following conditional independences: $A^t \perp\!\!\!\perp \overline{X}^t_{(1)}, Y | G_{\underline{A}^{t-1}}(\overline{X}^{t-1}_{(1)}), \underline{A}^{t-1}$ which can be translated into tangent space restrictions such that:

$$\Lambda_{nuis,\psi} = \Lambda^1_{nuis,\psi} \oplus \Lambda^2_{nuis,\psi} \oplus ... \oplus \Lambda^T_{nuis,\psi} \tag{34}$$

where each subspace $\Lambda^t_{nuis,\psi}$ is defined as:

$$\Lambda^t_{nuis,\psi} \equiv \Big\{ \gamma^t(A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)})) \in \mathcal{H} : \\ \mathbb{E}\left[\gamma^t(A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)})) | \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}) \right] = 0 \Big\}.$$

As these subspaces are orthogonal, projections onto them are available in closed form and are known to be (as derived in Appendix B):

$$\Pi([.]|\Lambda^t_{nuis,\psi}) = \mathbb{E}\left[[.] | A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}) \right] - \mathbb{E}\left[[.] | \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}) \right].$$

Amongst the class of observed data influence functions, one may further be interested in finding the one with the smallest asymptotic variance, i.e., the efficient observed data influence function φ_{eff}. It can be found via the following projection of h (Theorem 10.1 from (Tsiatis, 2006)):

$$\varphi_{eff} = h(A, G_A(X_{(1)}), Y) - \Pi\left(h(A, G_A(X_{(1)}), Y) | \Lambda_2 \right).$$

Hence, to construct an efficient influence function, one needs to find an element h, construct the space Λ_2, and project onto it. We now embed the three viewpoints of this work—missing data view, offline RL view, and semi-offline RL view—into this framework. We begin with the traditional missing data approach based on complete cases.

7.1 Missing data view

Now, we discuss the standard, traditional missing data approach to choosing an element h of the IPW space Λ_{IPW} and constructing the augmentation space Λ_2. Traditional semiparametric estimators for missing data problems rely on the missing data positivity assumption (Assumption 6.2), which requires complete cases. The corresponding estimators are referred to as augmented inverse probability weighting complete case (AIPWCC) estimators.

These estimators are called complete case estimators because they choose the missing data IPW estimator for h (see Tsiatis (2006) for more details):

$$h_{Miss}(A, G_A(X_{(1)}), Y) = \rho_{Miss}\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)}, Y] - J \in \Lambda_{IPW}. \tag{35}$$

Furthermore, they also require the missing data positivity assumption (Assumption 6.2) for the construction of Λ_2, given as:

$$\Lambda_2 = \Bigg\{ \sum_{a \in \mathcal{A}\setminus\vec{1}} \left[\mathbb{I}(A = a) - \prod_{t=1}^{T} \frac{\mathbb{I}(A^t = \vec{1})\pi_\beta^t(A^t = a^t | G_{\underline{a}^{t-1}}(X_{(1)}), \underline{a}^{t-1}))}{\pi_\beta^t(A^t = \vec{1} | G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^{t-1} = \vec{1})} \right] b_a(G_A(X_{(1)}), Y) : \\ b_a(G_A(X_{(1)}), Y) \in \mathcal{H} \Bigg\}. \tag{36}$$

For the interested reader, we show how both of these choices are derived under the missing data positivity assumption in Appendix L.

In addition to the strong positivity requirements, a key challenge is that general projections onto Λ_2 are not available in closed form (Tsiatis, 2006). Alternatives to still perform such projections and achieve full efficiency include iterative numerical methods. However, these methods are difficult to implement and involve significant computational challenges, which have hindered their practical application (Tsiatis, 2006). Therefore, such methods are beyond the scope of this work, but we direct interested readers to Tsiatis (2006) for further details.

7.2 Offline RL view

An alternative to applying traditional semiparametric theory for missing data problems is to start from the offline RL view. In this approach, one projects the influence function φ_{Off}—associated with the DRL estimator—onto the restricted tangent space under the NDE assumption, denoted by Λ_{NDE}:

$$\varphi_{eff} = \Pi\left(\varphi_{Off} | \Lambda_{NDE}\right) = \varphi_{Off} - \Pi\left(\varphi_{Off} | \Lambda_{NDE}^{\perp}\right).$$

This approach was first introduced by Liu et al. (2021) in the context of dynamic testing and treatment regimes, though it can be adapted to the AFAPE setting. However, their derivation was limited to a single acquirable feature per time step ($A^t \in \{0, 1\}$).

The space $\Lambda_{NDE}^{\perp}$ is adapted to AFAPE, given as:

$$\Lambda_{NDE}^{\perp} = \Lambda_* - \Pi(\Lambda_* | \Lambda_\psi)$$

with

$$\Lambda_* \equiv \Bigg\{ b^t_{\underline{i}^{t-1},\overline{i}^{t+1}}(\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), X_{(1)}^{\underline{i}^{t-1}}, X_{(1)}^{\overline{i}^{t+1}}, Y) \left(\frac{A^t}{\pi^t_\beta} - 1\right) \prod_{t'=1}^{t-1} \left(\frac{A^{t'}}{\pi^{t'}_\beta}\right)^{i^{t'}} \prod_{t'=t+1}^{T} \left(\frac{A^{t'}}{\pi^{t'}_\beta}\right)^{i^{t'}} : \\ b^t_{\underline{i}^{t-1},\overline{i}^{t+1}}(\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), X_{(1)}, Y) \in \mathcal{H} \Bigg\}$$

where $\underline{i}^{t-1} \equiv \{i^1, .., i^{t-1}\}$ and $\overline{i}^{t+1} \equiv \{i^{t+1}, .., i^T\}$ index subsets of $\{0,1\}^{t-1}$ and $\{0,1\}^{T-t}$ respectively and $X_{(1)}^{\underline{i}^{t-1}} \equiv \{X_{(1)}^{t'} : 1 \leq t' \leq t-1, i^{t'} = 1\}$ and $X_{(1)}^{\overline{i}^{t+1}} \equiv \{X_{(1)}^{t'} : t+1 \leq t' \leq T, i^{t'} = 1\}$. Furthermore, we denote $\pi^t_\beta \equiv \pi^t_\beta(A^t = 1 | G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^{t-1})$ for this one acquisition per time-point setting.

The space Λ_* includes a term for each combination of observed features (indexed by the subsets i). This construction also relies on a stringent positivity assumption, which is even stronger than what is required for identification under the offline RL view (Assumption 6.1). Specifically, it demands that $\pi^t_\beta > 0$ for all t, $G_{\underline{A}^{t-1}}(X_{(1)})$, and $\underline{A}^{t-1}$, irrespective of the target policy. Furthermore, since the target parameter doesn't depend on the acquisition process, we have $\Lambda_\psi = \Lambda_{nuis,\psi}$.

In the following lemmas, we demonstrate the equivalence between this offline RL approach and the semiparametric theory for MAR missing data problems:

Lemma 15. *(Relating the offline RL IPW estimator to the IPW space). The functional $h_{Off} \equiv \rho_{Off} C - J$, based on the IPW estimator from the offline RL view, is a valid element of the IPW space: $h_{Off} \in \Lambda_{IPW}$.*

Lemma 16. *(Λ_* is equal to the augmentation space). The augmentation space Λ_2 is equal to Λ_*.*

Both lemmas are proven in Appendix M.

We now demonstrate that both approaches—whether derived from the offline RL or the missing data view—yield the same influence function (as expected):

$$\begin{aligned}
\varphi_{eff} &= \varphi_{Off} - \Pi(\varphi_{Off} | \Lambda^\perp_{NDE}) \\
&\overset{*1}{=} h_{Off} - \Pi(h_{Off} | \Lambda_{nuis,\psi}) - \Pi(h_{Off} - \Pi(h_{Off} | \Lambda_{nuis,\psi}) | \Lambda_* - \Pi(\Lambda_* | \Lambda_{nuis,\psi})) \\
&= h_{Off} - \Pi(h_{Off} | \Lambda_{nuis,\psi}) - \Pi(h_{Off} | \Lambda_2 - \Pi(\Lambda_2 | \Lambda_{nuis,\psi})) \\
&\overset{*2}{=} h_{Off} - \Pi(h_{Off} | \Lambda_2)
\end{aligned}$$

where we use in $*1$) that the influence function can be decomposed: $\varphi_{Off} = h_{Off} - \Pi(h_{Off} | \Lambda_{nuis,\psi})$. In $*2$), we used that $\Lambda_{nuis,\psi} \subset \Lambda_2$.

This shows that a separate projection onto $\Lambda_{nuis,\psi}$ is unnecessary. However, as noted earlier, a projection onto $\Lambda^\perp_{NDE}$ (or Λ_2) is not available in closed form (Liu et al., 2021), as previously discussed in the missing data approach. Liu et al. (2021) suggest instead constructing an arbitrarily large subspace Ω onto which a projection is feasible. In this case, it becomes helpful to first project onto $\Lambda_{nuis,\psi}$ (where closed-form projections are available), ensuring that the resulting functional remains a valid influence function, even when the second projection is onto the approximated space Ω.

The resulting estimator by Liu et al. (2021) is termed "nearly efficient." It is, however, still difficult to implement and has so far only been tested only for a one time point, one acquisition setting.

7.3 Semi-offline RL view

In this section, we explore how the semi-offline RL view integrates with semiparametric theory and derive the corresponding influence function for the semi-offline DRL estimator given in Theorem 14.

We begin by establishing that the IPW estimator derived from the semi-offline RL framework can be used to construct an element of the IPW space Λ_{IPW}:

Lemma 17. *(Relating the semi-offline RL IPW estimator to the IPW space).* $h_{Semi} \equiv h_{Semi}(A, G_A(X_{(1)}), Y) = \hat{\mathbb{E}}_{n'}[\rho^T_{Semi} C' | A, G_A(X_{(1)}), Y] - J$ *is an element of the IPW space* Λ_{IPW}.

The proof of this lemma can be found in Appendix N.

Since constructing the augmentation space Λ_2 often involves strong positivity assumptions and projections onto Λ_2 are typically not available in closed form, we propose an alternative approach. Specifically, we suggest projecting onto subspaces of Λ_2 where closed-form projections are feasible. One such subspace is $\Lambda_{2,Semi}$, a tractable subspace of Λ_2, defined as:

$$\Lambda_{2,Semi}(\Xi) = \Lambda^1_{2,Semi}(\Xi) \oplus \Lambda^2_{2,Semi}(\Xi) \oplus ... \oplus \Lambda^T_{2,Semi}(\Xi)$$

with each

$$\Lambda^t_{2,Semi}(\Xi) = \Big\{ b^t(A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi) \in \mathcal{H} : \\ \mathbb{E}[b(A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi) | \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi] = 0 \Big\}.$$

Notably, we have $\Lambda_\psi \subseteq \Lambda_{2,Semi}(\Xi) \subset \Lambda_2$ with equality $\Lambda_\psi = \Lambda_{2,Semi}$ if $\Xi = \emptyset$.

This yields the following class of influence functions, proving Theorem 14:

$$\begin{aligned}
\varphi_{Semi}(A, G_A(X_{(1)}), Y; \Xi) &= h_{Semi} - \Pi(h_{Semi} | \Lambda_{2,Semi}(\Xi)) \\
&= h_{Semi} - \sum_{t=1}^{T} \mathbb{E}\left[h_{Semi} \middle| A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi \right] + \sum_{t=1}^{T} \mathbb{E}\left[h_{Semi} \middle| \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi \right] \\
&= \mathbb{E}\left[\rho^T_{Semi} f_C(A', G_{A'}(X_{(1)}), Y) | Y, G_A(X_{(1)}), A \right] \\
&- \sum_{t=1}^{T} \mathbb{E}\left[\rho^t_{Semi} Q^t_{Semi} \middle| A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi \right] \\
&+ \sum_{t=1}^{T} \mathbb{E}\left[\rho^{t-1}_{Semi} V^{t-1}_{Semi} \middle| \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi \right] - J \\
&= \mathbb{E}\left[\rho^T_{Semi} f_C(A', G_{A'}(X_{(1)}), Y) - \sum_{t=1}^{T} \rho^t_{Semi} Q^t_{Semi} + \sum_{t=1}^{T} \rho^{t-1}_{Semi} V^{t-1}_{Semi} \middle| A, G_A(X_{(1)}), Y \right] - J.
\end{aligned}$$

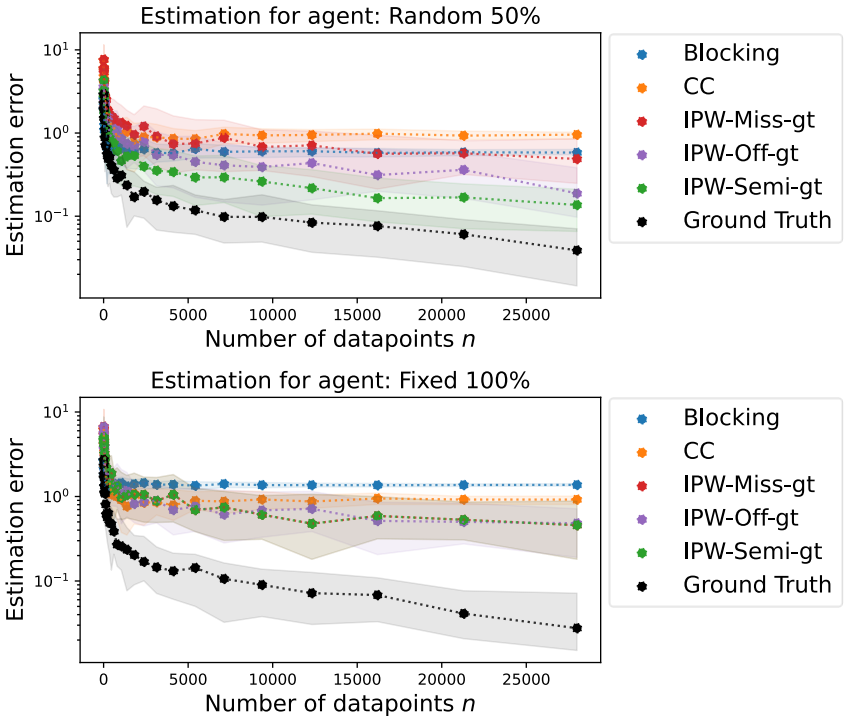


Figure 7: The plots depict convergence as a function of data set size n for sampling-based estimators in Experiment 1. Two agents are shown: one that acquires each costly feature with a probability of 50% and another with 100%. The estimation error is measured as the absolute difference between the estimate and the ground truth, which is calculated on the full data set ($n = 40,000$). The semi-offline RL IPW estimator converges the fastest for the 'Random 50%' agent, while for the 'Fixed 100%' agent, all IPW estimators perform identically, converging at a slower rate.

We provide full details of this derivation in Appendix O.

From the semiparametric viewpoint, it holds that a larger set Ξ will increase the efficiency of the corresponding influence function as mentioned in Remark 7. This is the case since employing a larger set Ξ will result in a larger subspace $\Lambda_{2,Semi}(\Xi) \subset \Lambda_2$, which in turn implies a higher efficiency.

8. Experiments

We evaluate the different estimators on synthetic data sets where the missingness is artificially induced to allow the comparison with the ground truth.

8.1 Experiment Design

We perform five experiments with different violations of the identifying assumptions:

- **Experiment 1:** Assumptions 4 (NDE), 5 (NUC), 6.1 (offline RL positivity), 6.2 (missing data positivity), and 6.4 (maximal global positivity) all hold.

- **Experiment 2:** Assumption 6.2 (missing data positivity) is violated
- **Experiment 3:** Assumption 6.1 (offline RL positivity) is violated for some agents.
- **Experiment 4:** Assumptions 5 (NUC) is violated.
- **Experiment 5:** Assumptions 4 (NDE) is violated.

We evaluate random AFA policies and a proximal policy optimization (PPO) RL agent (Schulman et al., 2017) as AFA agents and use impute-then-regress classifiers (Le Morvan et al., 2021) with unconditional mean imputation and a logistic regression classifier. Nuisance functions ($\hat{Q}_{Semi}$ and $\hat{\pi}_\beta$) are fitted using multi-layer perceptrons and logistic regression models, respectively. The assumed logistic regression model for the propensity score correctly matches the ground truth. We compare the following estimators:

- *Imp-Mean:* Mean imputation (biased estimator)
- *Blocking:* Blocks the acquisitions of not available features but offers no correction. This corresponds to the estimate $\hat{\mathbb{E}}_{n'}[C']$ (with $\pi'_{sim} = \pi'_\alpha$) which is biased.
- *CC:* Complete case analysis (only unbiased under MCAR).
- *IPW-Miss/IPW-Miss-gt:* The IPW estimator from the missing data view. The weights were normalized to reduce the variance of the estimator. *IPW-Miss-gt* uses the ground truth propensity score model π_β instead of its estimate $\hat{\pi}_\beta$.
- *IPW-Off/IPW-Off-gt:* The IPW estimator from the offline RL view with normalized weights and with and without the ground truth propensity score model.
- *IPW-Semi/IPW-Semi-gt:* The IPW estimator (for the maximal global positivity assumption) from the semi-offline RL view with normalized weights and with and without the ground truth propensity score model.
- *DM-Semi:* The semi-offline RL version of the direct method.
- *DRL-Semi/DRL-Semi-gt:* The semi-offline RL version of the double reinforcement learning estimator with normalized weights under the maximal global positivity assumption, with and without the ground truth propensity score model.
- *Ground Truth:* In the experiments where the NDE assumptions hold, the agent is evaluated on the fully observed data set. This corresponds to estimating J using a Monte Carlo estimate, $\hat{\mathbb{E}}\left[C_{(\pi_\alpha)}|X_{(1)}, Y\right]$, derived from samples of the ground truth data without any missingness (i.e., samples from $p(X_{(1)}, Y)$). Conversely, in experiments where the NDE assumption is violated, the ground truth is obtained by running the agent in an environment that is continuously sampled from the true data-generating process.

Complete experiment details are given in Appendix P.

8.2 Results

Figure 7 shows the convergence plots of sampling-based estimators for Experiment 1, highlighting the data efficiency of all estimators when the identifying assumptions hold. As expected, the blocking and complete case estimators are biased and do not converge to the true value of J. Among the unbiased IPW estimators, the semi-offline RL IPW estimator

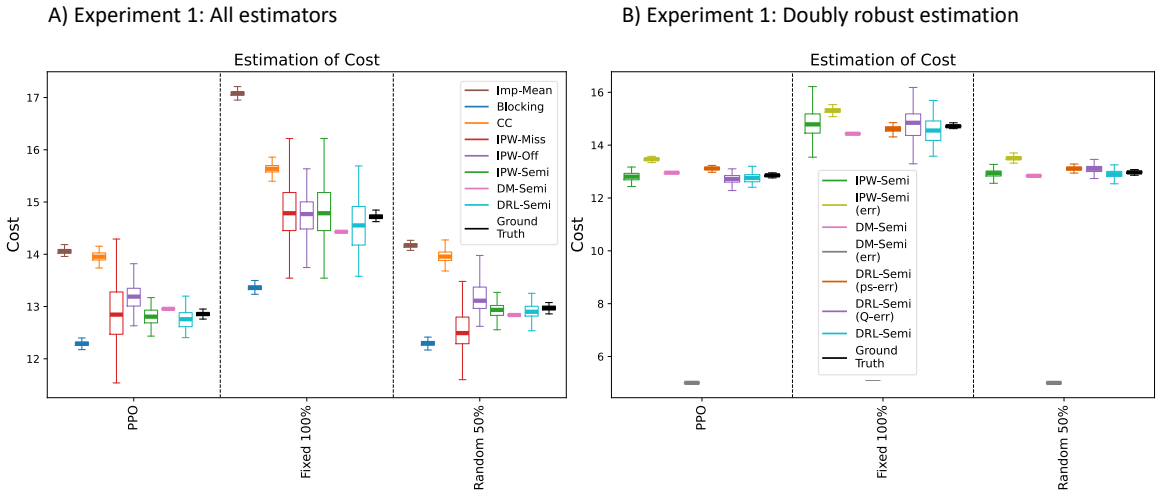


Figure 8: A) General estimation results for Experiment 1. The *Imp-Mean*, *Blocking*, and *CC* estimators show highly biased estimates, while the IPW, semi-offline DM, and DRL estimators align closely with the true target J. B) Estimation results highlighting the double robustness property of the DRL estimator. The *DRL-Semi* estimator continues to provide accurate estimates even when either the propensity score model $\hat{\pi}_\beta$ (*DRL-Semi (ps-err)*) or the Q-model $\hat{Q}_{Semi}$ (*DRL-Semi (Q-err)*) is misspecified.

is the most efficient, achieving the fastest convergence for the 'Random 50%' agent, which acquires each feature with a 50% probability. However, this data efficiency disappears when evaluating the 'Fixed 100%' agent, which acquires all features every time. As noted in Remark 3, all three IPW estimators perform identically in this scenario, as reflected by their equal convergence speeds.

Figure 8A) displays the overall performance of various estimators in Experiment 1. Confidence intervals were computed using the non-parametric bootstrap, excluding the retraining of nuisance functions due to high computational complexity. As a result, the confidence intervals are overly narrow, particularly for the semi-offline DM estimator. The experiment demonstrates that all semi-offline RL estimators accurately approximate the true target parameter J. However, the DM estimator may exhibit slight bias due to potential misspecification of the Q-function. In contrast, the biased mean imputation, blocking, and complete case analysis estimators fail to consistently estimate J.

Figure 8B) illustrates the double robustness property of the semi-offline RL version of the DRL estimator. Even when one of the nuisance functions is misspecified, the DRL estimator still provides estimates that closely approximate the true value of J.

Figures 9A) and B) underscore the importance of the positivity assumption. In Figure 9A), Experiment 2 shows the consequences of violating the missing data positivity assumption (Assumption 6.2)—the fraction of complete cases is only 0.007%. As a result, J for the 'Fixed 100%' agent cannot be identified from any of the views, and all IPW estimators fail

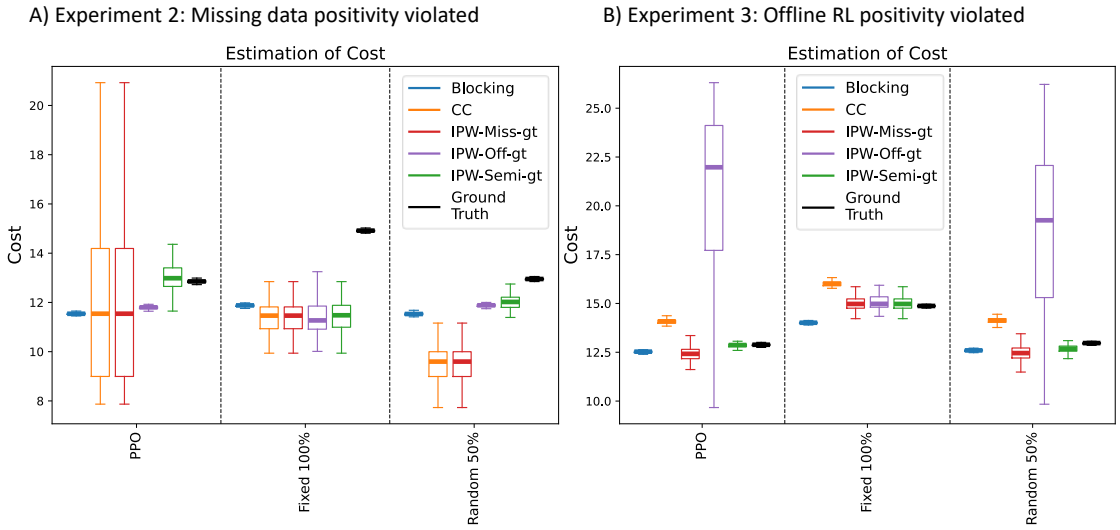


Figure 9: A) Experiment 2: AFA setting with an extremely low fraction of complete cases (0.007%), leading to violations of the missing data positivity assumption. For the 'Fixed 100%' agent, the target J is not identified from any view. However, the semi-offline IPW estimator is less impacted by the lack of complete cases when evaluating the other agents. B) Experiment 3: AFA setting with positivity violations specific to the offline RL view. While other estimators continue to produce accurate estimates, the offline RL IPW estimator fails entirely for the PPO and 'Random 50%' agents.

to provide accurate estimates. The positivity assumption is also violated for the 'Random 50%' agent, though the impact is less severe for the offline and semi-offline RL estimators.

In Figure 9B, Experiment 3 shows the failure of the positivity assumption required by the offline RL view (Assumption 6.1) for two agents. In this experiment, $A_2 = \vec{1}$ for all data points, meaning trajectories where $A_{(\pi_\alpha)} = 0$ have no support. While the other IPW estimators still produce accurate estimates for J, the estimates from the offline RL IPW estimators fail completely.

Finally, Figures 10A) and B) explore the AFA setting when either the NUC or NDE assumptions are violated. In Figure 10A, which depicts an MNAR scenario, all IPW estimators appear to perform well despite the violation of the NUC assumption. For the 'Fixed 100%' agent, this is expected, as all estimators reduce to the missing data IPW estimator, which is identified. However, for the 'Random 50%' agent, the AFAPE target J is not identified from the offline or semi-offline RL views, though the estimation from the semi-offline RL IPW estimator remains relatively robust. It is worth cautioning that violating the NUC assumption may have more severe consequences in real-world applications.

Figure 10B demonstrates the effect of violating the NDE assumption. In this case, only the offline RL view yields consistent estimators, as reflected in the experiment. All other estimators show significant deviations from the ground truth. However, the offline RL IPW estimator shows slight biases, too, possibly due to minor positivity violations.

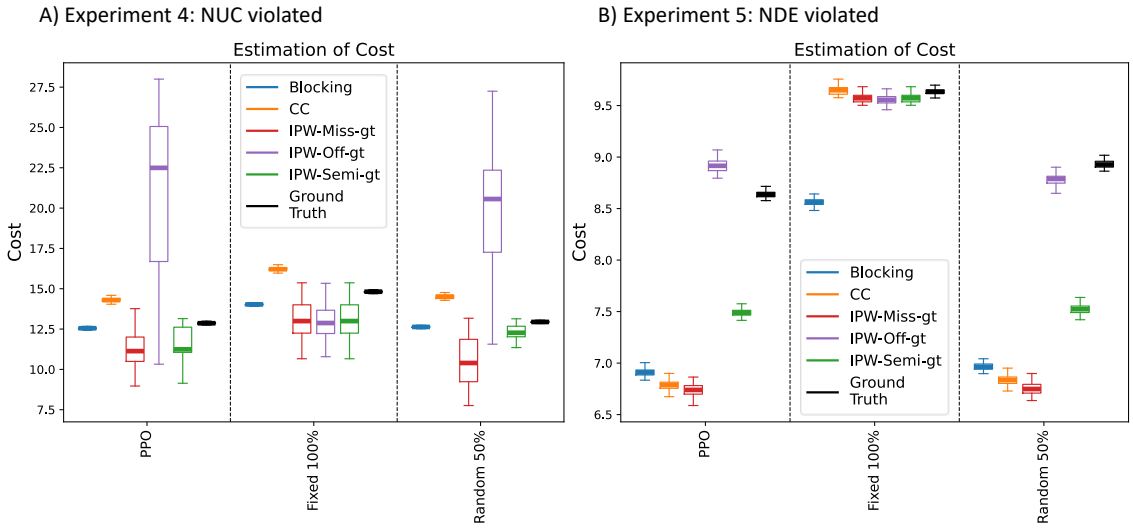


Figure 10: A) Experiment 4: AFA setting with an MNAR acquisition process, resulting in violations of the NUC assumption. Despite this, the semi-offline RL IPW estimator still provides accurate estimates of the target J. B) Experiment 5: AFA setting with a violation of the NDE assumption. All estimators, except the offline RL IPW estimator, produce highly biased estimates. The offline RL IPW estimator also shows some slight deviation from the true J, too, potentially caused by minor positivity violations.

9. Discussion and Future Work

In this study, we explored the various aspects of solving the AFAPE problem. We acknowledge that there is no one-size-fits-all solution, as the choice of assumptions can vary across AFA settings. To facilitate this discussion, we propose a set of questions that data scientists should ask themselves when tackling the AFAPE problem before choosing a viewpoint and estimator.

1) What (conditional) independences hold in the data? The choice of conditional independence assumptions directly influences identifiability and the selection of optimal viewpoints and estimators. When both the NDE and NUC assumptions are violated, the target parameter becomes unidentifiable, making estimation infeasible. If only the NDE assumption fails, the offline RL view is still applicable, while violations of only the NUC assumption allow for the use of the missing data (+online RL) view. In each scenario, leveraging the absence of an edge in the causal graph can effectively eliminate estimation bias that would otherwise persist. When both assumptions are satisfied, one can select from the offline RL, missing data (+online RL), or the novel semi-offline RL view. In such cases, leveraging both the NDE and NUC assumptions through the semi-offline RL view can be advantageous, as it allows for relaxed positivity assumptions and a reduction in estimation variance.

Conclusion: Under NUC, one can apply offline RL methods. Under NDE, one can apply missing data methods. Under both NUC and NDE, one can apply semi-offline RL methods.

2) How much exploration was performed by the retrospective acquisition policy π_β? Positivity requirements, crucial for all viewpoints, demand certain action sequences to be present in the retrospective data. However, real-world data sets—especially in fields like medicine—often breach these assumptions due to the tendency of professionals to follow similar paths with minimal deviation. The semi-offline RL view, fortunately, imposes weaker positivity constraints than both the offline RL and missing data (+online RL) views. Nevertheless, for "data-hungry" AFA policies, the benefits may be less pronounced compared to the missing data view.

Additionally, the choice of the identifying policy π_{id} depends on which positivity assumptions hold in the data. We leave the adaptation of known positivity assessment methods (Petersen et al., 2012) to the semi-offline RL setting as future work.

Conclusion: The semi-offline RL view requires significantly weaker positivity assumptions than the offline RL and missing data (+ online RL) viewpoints.

3) Can the nuisance models be correctly specified and trained? The accuracy of different estimators hinges on the proper specification and training of nuisance functions. Despite the double robustness property of DRL estimators, achieving unbiased estimates still depends on how well these functions are modeled. Estimators like the multiple imputation (MI) method can outperform others in certain contexts, particularly when feature smoothness assumptions are reasonable and easily modeled over time. While machine learning techniques such as deep learning offer flexibility, they demand large data sets, which may not always be available.

Conclusion: No single viewpoint or estimator is superior across all settings. The choice between MI and semi-offline RL estimators depends on prior knowledge and the feasibility of training the nuisance models.

4) Is the available data set size sufficient? Efficient use of data is crucial for accurate estimation. Our experiments demonstrate that estimators based on the semi-offline RL view achieve greater data efficiency compared to both offline RL and missing data estimators.

While our semiparametric analysis shows that the efficiency of all estimators can still be improved, a closed-form efficient influence function does not exist. Some computationally intensive methods, though complex to implement, may still enhance efficiency if the respective strong positivity assumptions hold (Tsiatis, 2006; Liu et al., 2021).

Conclusion: Estimators derived from the semi-offline RL view demonstrate in experiments notably higher data efficiency compared to estimators from the offline RL and missing data (+ online RL) viewpoints.

When the answers to the above questions are uncertain, it is advisable to use multiple views and estimators in tandem as part of a broader sensitivity analysis. This approach enhances confidence in the reliability and safety of AFA agents before deployment.

Our study assumes that feature values change over time, making the timing of measurements critical. In our companion paper (von Kleist et al., 2023), we address how a static feature assumption can be incorporated into the AFAPE problem, and we explore how the semi-offline RL and missing data views can be combined when the NUC assumption is violated (i.e., in MNAR scenarios).

Looking forward, we aim to tackle the AFA optimization problem outlined in Section 3.4. Once the AFAPE problem is resolved and the estimation of the target parameter J is successful, optimization can commence. This includes training new AFA agents and classifiers. A natural next step is adapting established DTR (Murphy, 2003; Robins, 2004) or offline RL methods—such as off-policy policy gradient methods, actor-critic techniques, and model-based RL approaches (Levine et al., 2020)—to the semi-offline RL framework for further development. These adapted methods could also integrate insights from the online RL literature, such as employing an adaptive exploration policy to enhance the sampling process.

10. Conclusion

We study the problem of active feature acquisition performance evaluation (AFAPE), which involves estimating the acquisition and misclassification costs that an AFA agent would generate after being deployed, using retrospective data. We demonstrate that, depending on the assumptions, one can apply different existing viewpoints to solve AFAPE. Under the no unobserved confounding (NUC) assumption, one can apply identification and estimation methods from the offline RL literature. Under the no direct effect (NDE) assumption, which assumes the underlying feature values are not affected by their measurement, one can instead apply missing data methods. For settings where both the NUC and the NDE assumptions hold, we propose a novel semi-offline RL viewpoint, which requires weaker positivity assumptions for identification. Within the semi-offline RL viewpoint, we developed several novel estimators that correspond to semi-offline RL versions of the direct method (DM), inverse probability weighting (IPW), and double reinforcement learning (DRL). Finally, we conducted synthetic data experiments to highlight the significance of utilizing proper unbiased estimators for AFAPE to ensure the reliability and safety of AFA systems.

Acknowledgments

The present contribution is supported by the Helmholtz Association under the joint research school “HIDSS-006 - Munich School for Data Science @ Helmholtz, TUM & LMU”. Henrik von Kleist received a Carl-Duisberg Fellowship by the Bayer Foundation.

Appendix A. Literature Review for Active Feature Acquisition (AFA)

In this appendix, we explain in more detail the difference between AFA and related fields and introduce some common approaches to training AFA agents from the literature.

A.1 Distinction between AFA and Related Fields

AFA is different from active learning (Settles, 2009). In active learning, one assumes a classification task with a training data set that contains many unlabeled data points. The active learning task is then to decide which label acquisitions will improve the training performance the most. Similar research also exists for the acquisition of features for optimal improvement of training. This task has been referred to as "active selection of classification features" (Kok et al., 2021), and unfortunately also as "active feature acquisition" (Huang et al., 2018; Beyer et al., 2020), but its objective differs fundamentally from ours. Huang et al. (2018) attempt to find out which missing values within the retrospective data set would improve training the most when retroactively acquired. In this paper, we are, however, interested which features for a new data point would improve the individual prediction for that data point the most.

A.2 Approaches to Training AFA Agents

The AFA setting is most generally described as a sequential decision process, which motivates the use of RL-based solutions. One variant, model-based RL, focuses on learning a model for the state transitions. Under the NDE assumption, utilizing an imputation model to capture state transitions becomes feasible, exploiting the unique AFA structure for more straightforward learning (Yoon et al., 2018; Yin et al., 2020; Li and Oliva, 2021a,b; Ma et al., 2019). During deployment, this imputation model can simulate potential outcomes of feature acquisitions, facilitating the derivation of optimal acquisition strategies. Conversely, model-free RL methods do not require a state-transition function. One variant, Q-learning, involves estimating the expected cost of specific acquisition decisions (Chang et al., 2019; Janisch et al., 2020; Shim et al., 2018). For instance, Shim et al. (2018) illustrate the use of double Q-learning for the AFA agent, incorporating a deep neural network that shares network layers for the acquisition decision and classification tasks.

Appendix B. Review of Semiparametric Theory

We give here a short review of the basic concepts of semiparametric theory and some results for missing data problems. The review is based on work by Tsiatis (2006), which we recommend for more in-depth explanations.

B.1 General Semiparametric Theory

Semi-parametric theory aims at finding data-efficient estimators for a target parameter $J = J(p)$ without imposing unnecessarily strict assumptions on p. In this review, we restrict ourselves to only scalar parameters J. We let p denote the distribution $p(Z)$ over a set of random variables Z from which we have n independent and identically distributed samples $(Z^1,...,Z^n)$. It is possible in many cases to obtain estimators for J that are consistent at a

rate of $\sqrt{n}$ without imposing many assumptions. The derivation of such estimators relies on influence functions which are discussed next.

B.1.1 Influence functions and estimators

A central element of semi-parametric theory are influence functions as they characterize asymptotically linear estimators in the following sense. An estimator J_{est} is asymptotically linear and has an influence function $\varphi(Z) \equiv \varphi$ if it allows the following equality (Tsiatis, 2006):

$$J_{est}(n) - J = \frac{1}{n}\sum_{i=1}^{n}\varphi(Z^i) + o_p(\frac{1}{\sqrt{n}}) \tag{37}$$

where $\varphi \in \mathcal{H}$ and $\mathcal{H}$ represents the space of all random functions of zero mean and finite variance. The central limit theorem implies that J_{est} is asymptotically normally distributed (Tsiatis, 2006):

$$\sqrt{n}(J_{est}(n) - J) \rightsquigarrow \mathcal{N}\left(0, \mathbb{E}[\varphi^2]\right)$$

where $\rightsquigarrow$ denotes convergence in distribution. The estimation error is thus asymptotically bounded by the variance of the influence function. The efficient influence function φ_{eff} is the one with the smallest asymptotic variance.

Many influence functions, such as the ones of the DRL estimators in this work, depend linearly on the target parameter J, such that $\varphi = f(Z) + J$ for some function f. In these cases, one can very easily derive a corresponding, so-called "1-step" estimator by leveraging Eq. 37 to obtain:

$$J_{est} \equiv -J + \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i) = -J + \frac{1}{n}\sum_{i=1}^{n}f(Z_i) + J = \frac{1}{n}\sum_{i=1}^{n}f(Z_i).$$

B.1.2 Deriving influence functions

Deriving the space of influence functions or the efficient influence function for a new target parameter or new model restrictions can be complex. There are, however, some known properties that influence functions in general, or the efficient influence function in particular, have to fulfill and these can be used for their derivation. We examine these now in more detail:

1) An influence function must be in the orthocomp of the nuisance tangent space: $\varphi \in \Lambda^{\perp}_{\mathit{nuis}}$
To clarify this condition, we first separate the space of model parameters into J, the target parameter, and η, the nuisance parameters. We denote the nuisance tangent space as Λ_{nuis} and the space orthogonal to it, i.e., its orthocomplement (or orthocomp), as $\Lambda^{\perp}_{nuis}$. The nuisance tangent space can be seen as the collection of directions in which the nuisance parameters can vary without affecting the parameter of interest. It is defined as the mean

square closure of parametric submodel nuisance tangent spaces, where a parametric submodel nuisance tangent space is the linear subspace spanned by the nuisance scores, which are defined as:

$$S_\eta = \left. \frac{\partial \log p_Z(z, \eta, J)}{\partial \eta} \right|_{\eta=\eta_0}.$$

Here η_0 denotes the true value of η.

An influence function must be an element in $\Lambda_{nuis}^{\perp}$, but must fulfill also the following normalization (Theorem 3.2 from Tsiatis (2006)):

$$\mathbb{E}\left[\varphi(Z) S_J(Z)\right] = 1$$

where $S_J(Z)$ denotes the scores with respect to the target parameter.

A nuisance function can thus be obtained by taking any nonzero element $h(Z) \in \mathcal{H}$, projecting it onto the orthocomp of the nuisance tangent space $\Lambda_{nuis}^{\perp}$, and normalizing it. We denote the mentioned orthogonal projection by Π such that:

$$\varphi^*(Z) = \Pi(h(Z)|\Lambda_{nuis}^{\perp}) = h(Z) - \Pi(h(Z)|\Lambda_{nuis}).$$

where $\varphi^*(Z)$ represents an element of $\Lambda_{nuis}^{\perp}$ which if normalized would be an influence function.

2) The efficient influence function must be in the tangent space: $\varphi \in \Lambda$
The condition to obtain efficiency for an influence function is that we choose the influence function that is in the tangent space Λ. It can thus be obtained by projecting any influence function onto the tangent space:

$$\varphi_{\mathit{eff}}(Z) = \Pi(\varphi(Z)|\Lambda) = \varphi(Z) - \Pi(\varphi(Z)|\Lambda^{\perp}).$$

B.1.3 Constructing tangent spaces and projecting on them

Here, we provide some useful further details about the construction of tangent spaces with specific restrictions and the projection of random variables on them. We start by decomposing the space of all zero mean, finite variance functions $\mathcal{H}$ into orthogonal subspaces for the case of a multivarate Z. We then examine how tangent space restrictions given by conditional independence assumptions can be incorporated.

1) The decomposition of $\mathcal{H}$ for a multivariate variable Z
Firstly, we look at a useful decomposition of the space of a multivariate variable Z of dimensions d. The space $\mathcal{H}$ of all functions $h(Z_1, Z_2, ..., Z_d)$ separates into orthogonal subspaces (Theorem 4.5 from Tsiatis (2006)):

$$\mathcal{H} = \mathcal{H}_{Z_1} \oplus \mathcal{H}_{Z_2|Z_1} \oplus ... \oplus \mathcal{H}_{Z_d|Z_{d-1},...,Z_1} \tag{38}$$

where $\mathcal{H}_{Z_i|Z_{i-1},...,Z_1}$ denotes the space spanned by the conditional scores:

$$\mathcal{H}_{Z_i|Z_{i-1},...,Z_1} = \left\{ h(Z_i, Z_{i-1}, ..., Z_1) \in \mathcal{H} : \mathbb{E}[h(Z_i, Z_{i-1}, ..., Z_1)|Z_{i-1}, ..., Z_1] = 0 \right\}.$$

Fortunately, also the projection onto such a subspace is known and is given for an arbitrary element $h^*(Z) \in \mathcal{H}$ as:

$$\Pi\left(h^*(Z)|\mathcal{H}_{Z_i|Z_{i-1},...,Z_1}\right) = \mathbb{E}[h^*(Z)|Z_i, Z_{i-1}, ..., Z_1] - \mathbb{E}[h^*(Z)|Z_{i-1}, ..., Z_1].$$

2) Tangent space restrictions under a conditional independence assumption
It is also possible to define the tangent space under conditional independence restrictions and to project on it. In particular, let's revisit the decomposition of the tangent space from Eq. 38. Let's assume the following independence holds: $Z_i \perp\!\!\!\perp Z_j | Z_{i-1}, ..., Z_{j+1}, Z_{j-1}, ..., Z_1$ for $i > j$. We want to find Λ_r, the tangent space restricted by the conditional independence, its orthocomp $\Lambda_r^\perp$, and projections on it.

The independence restriction only affects the space of the respective conditional scores such that:

$$\Lambda_r = \mathcal{H}_{Z_1} \oplus \mathcal{H}_{Z_2|Z_1} \oplus ... \oplus \Lambda_{r,Z_i|Z_{i-1},...,Z_{j+1},Z_{j-1},...,Z_1} \oplus ... \oplus \mathcal{H}_{Z_d|Z_{d-1},...,Z_1}.$$

and $\Lambda_{r,Z_i|Z_{i-1},...,Z_{j+1},Z_{j-1},...,Z_1} = \mathcal{H}_{Z_i|Z_{i-1},...,Z_{j+1},Z_{j-1},...,Z_1}$. The projections onto Λ_r are thus straightforward:

$$\begin{aligned}\Pi\left(h^*(Z)|\Lambda_{r,Z_i|Z_{i-1},...,Z_{j+1},Z_{j-1},...,Z_1}\right) &= \Pi\left(h^*(Z)|\mathcal{H}_{Z_i|Z_{i-1},...,Z_{j+1},Z_{j-1},...,Z_1}\right) \\ &= \mathbb{E}[h^*(Z)|Z_i, Z_{i-1}, ..., Z_{j+1}, Z_{j-1}, ..., Z_1] - \mathbb{E}[h^*(Z)|Z_{i-1}, ..., Z_{j+1}, Z_{j-1}, ..., Z_1]\end{aligned}$$

Correspondingly, we can obtain the orthocomp as

$$\begin{aligned}\Lambda_r^\perp &= \mathcal{H}_{Z_i|Z_{i-1},...,Z_1} - \Pi(\mathcal{H}_{Z_i|Z_{i-1},...,Z_1}|\Lambda_{r,Z_i|Z_{i-1},...,Z_{j+1},Z_{j-1},...,Z_1}) \\ &= \Big\{ h(Z_i, Z_{i-1}, ..., Z_1) - \mathbb{E}[h(Z_i, Z_{i-1}, ..., Z_1)|Z_i, Z_{i-1}, ..., Z_{j+1}, Z_{j-1}, ..., Z_1] \\ &\quad + \underbrace{\mathbb{E}[h(Z_i, Z_{i-1}, ..., Z_1)|Z_{i-1}, ..., Z_{j+1}, Z_{j-1}, ..., Z_1]}_{=0} : \\ &\quad \mathbb{E}[h(Z_i, Z_{i-1}, ..., Z_1)|Z_{i-1}, ..., Z_1] = 0; h \in \mathcal{H} \Big\} \\ &= \Big\{ h(Z_i, Z_{i-1}, ..., Z_1) - \mathbb{E}[h(Z_i, Z_{i-1}, ..., Z_1)|Z_i, Z_{i-1}, ..., Z_{j+1}, Z_{j-1}, ..., Z_1] : \\ &\quad \mathbb{E}[h(Z_i, Z_{i-1}, ..., Z_1)|Z_{i-1}, ..., Z_1] = 0; h \in \mathcal{H} \Big\}.\end{aligned}$$

B.2 Semiparametric theory for missing data under the MAR assumption

Semiparametric methods have further been applied to missing data problems. As we take on a missing data view in this work, we now briefly introduce known results for such settings. We restrict our review to scenarios where the missingness process follows a missing-at-random (MAR) scenario. For more details, see also Tsiatis (2006).

We now distinguish between observed data (with missingness) and full data (without missingness). Let the full data be denoted by $X_{(1)} \in \mathbb{R}^d$, the missingness indicators by $A \in \{0,1\}^d$ and the observed data by $X \equiv G_A(X_{(1)})$ such that $X_i = X_{(1),i}$ if $A_i = 1$ and $X_i =$ "?", otherwise.

Semiparametric theory methods for missing data aim at finding observed data (efficient) influence functions from full data influence functions. Throughout this review and the whole paper, we assume no restrictions on the full data $X_{(1)}$, meaning that there is only one full data influence function. There are, however, in general multiple corresponding observed data influence functions. In order to find these, the observed data nuisance tangent space needs to be constructed. This construction is simplified in MAR scenarios, as the likelihood factorizes into two separate terms related to the acquisition process and the observed data part of the likelihood (Tsiatis, 2006):

$$p_{A,G_A(X_{(1)})}(A, G_A(X_{(1)}); \psi, J, \eta) = \underbrace{p_{A|G_A(X_{(1)})}(A|G_A(X_{(1)}); \psi)}_{\text{acquisition process}} \sum_{G_A(X_{(1)})} \underbrace{p_{X_{(1)}}(X_{(1)}; J, \eta)}_{\text{process of full data}} \tag{39}$$

where we let ψ denote the nuisance parameter of the acquisition process, and η denote the nuisance parameter of the full data generating process.

This factorization allows the following decomposition of the nuisance tangent space into orthogonal subspaces (Theorem 8.2 from Tsiatis (2006)):

$$\Lambda_{nuis} = \Lambda_{nuis,\psi} \oplus \Lambda_{nuis,\eta} = \Lambda_{\psi} \oplus \Lambda_{nuis,\eta}. \tag{40}$$

Here, we used $\Lambda_{\psi} = \Lambda_{nuis,\psi}$, which holds since the target parameter is not part of this acquisition process.

We continue with the discussion for the derivation of observed data influence functions in the main text in Section 7.

Appendix C. Glossary of Terms and Symbols

Term	Description
AFAPE	Active feature acquisition performance evaluation: The problem of estimating the counterfactual cost that would arise if an AFA agent was deployed.
NDE assumption	No direct effect assumption: States that the action of measuring a feature does not impact the values of any features or the label.
NUC assumption	No unobserved confounding assumption: States that acquisition decisions within the retrospective data set were only based on measured feature values.
Semi-offline RL	Novel framework that allows an agent to interact with the environment (the online part) but forbids the exploration of certain actions (the offline part).
DTR	Dynamic treatment regimes
G-formula	Identification formula from causal inference (Robins, 1986)

Plug-in of the G-formula	Estimation formula from causal inference that replaces unknown densities in the G-formula with estimated versions(Robins, 1986).
IPW	Inverse probability weighting: Estimator that is also known as importance sampling or the Horvitz-Thompson estimator.
DM	Direct method: Estimator based on a Q-function.
DRL	Double reinforcement learning: Double robust estimator that uses IPW weights and a Q-function.
m-graph	Missing data graph: Graph to visualize assumptions in missing data problems.
MI	Multiple imputation: Estimator for missing data problems that is a special case of the plug-in of the G-formula.
influence function	Function of mean zero and finite variance that is used to analyze the asymptotic properties of regular and asymptotically linear (RAL) estimators.
MCAR assumption	Missing-completely-at-random assumption: States that the reason for the missingness of certain features does not depend on any feature values.
MAR assumption	Missing-at-random assumption: States that the reason for missingness of certain features does only depend on observed feature values.
MNAR assumption	Missing-not-at-random assumption: States that the reason for the missingness of certain features may depend on feature values that are not observed.
nuisance function	A function that needs to be fitted from data in order to use a corresponding estimator but which is not of primary interest itself. Examples are the propensity score model and the Q-function.
local positivity assumption	Positivity assumption for semi-offline RL that ensures the simulation of a desired next action is possible from the retrospective data set.
regional positivity assumption	Positivity assumption for semi-offline RL that ensures the simulation of all future desired actions is possible from the retrospective data set.
global positivity assumption	Positivity assumption for semi-offline RL that ensures the simulation of all desired actions is possible from step 1 on.
maximal regional positivity assumption	A special, stronger version of the regional positivity assumption.

maximal global positivity assumption	A special, stronger version of the global positivity assumption.
tangent space	Space of scores (i.e., derivatives of the log-likelihood)
nuisance tangent space	Space of nuisance scores (i.e., the scores with respect to the nuisance parameters)

Symbol	**Description**
$t \in (0, ..., T)$	Time
U^t	Unobserved state variables at time t
d	Number of features, i.e., dimension of U^t
$X^t = G_{A^t}(U^t)$	Observed feature values at time t (retrospective data set)
A^t	Acquisition action at time t (retrospective data set)
$\mathcal{A}$	Space of A
Y	Label
$Y^* = f_{cl}(\underline{X}^T, \underline{A}^T)$	Predicted label based on classifier f_{cl}
C_a^t	Acquisition cost for action A^t
$C_{mc} = f_C(Y^*, Y)$	Misclassification cost (if Y and Y^* differ)
π_β	Retrospective acquisition policy
π_α	AFA policy
$C_{mc,(\pi_\alpha)}$	Counterfactual misclassification cost had π_α instead of π_β been applied
$g(.)$	known deterministic distribution
J / J_{mc}	Expected misclassification cost under the AFA policy and classifier
J_a	Expected acquisition cost under the AFA policy and classifier
ϕ_1^*, ϕ_2^*	Sets of parameters that parameterize the AFA policy and the classifier, respectively
$q(.)$	Counterfactual distribution
Q_{Off}^t	State-action value function from offline RL (at time t)
V_{Off}^t	State value function from offline RL (at time t)
π'	Blocked policy
π'_{sim}	(Blocked) simulation policy
$p'(.)$	Simulated distribution
C', Y'^*, X', A'	Simulated cost, predicted label, features, and actions

Symbol	Description
$\mathcal{D}$	Retrospective data set
$\mathcal{D}'$	Simulated data set
$\mathcal{A}_{adm}$	Local admissible set
$\tilde{\mathcal{A}}_{adm}$	Regional admissible set
π_{id}	Distribution for A that allows identification of J under the semi-offline RL view (subject to support restrictions)
Q^t_{Semi}	State-action value function from semi-offline RL (at time t)
$\Xi \subseteq \{O, Y\}$	Arbitrary subset of the always observed features $O \subset X_{(1)}$ and the label Y
$q'(.)$	Counterfactual simulated distribution
φ	Influence function
Λ	(Observed data) tangent space
$\Lambda^{\perp}$	Orthocomplement of the (observed data) tangent space
$\Pi([.]\|\Lambda)$	(Orthogonal) projection onto the tangent space
Λ_{nuis}	(Observed data) nuisance tangent space
Λ^F	Full data tangent space
Λ^F_{nuis}	Full data nuisance tangent space
$\Lambda_{nuis,\psi} = \Lambda_{\psi}$	(Nuisance) tangent space of the acquisition process
$\Lambda_{nuis,\eta}$	Nuisance tangent space of the observed part of the full data process
Λ_{IPW}	IPW space
Λ_2	Augmentation space
$\Lambda_{2,Semi}(\Xi)$	Subspace of Λ_2 proposed for projection onto under the semi-offline RL view

Appendix D. Identification of the Block-conditional Model

In this appendix, we demonstrate how identification of $p(X_{(1)}, Y)$ can be achieved when the NUC assumption (Assumption 5) is violated. This corresponds to the block-conditional model (Zhou et al., 2010). We show that the propensity score model $p(A = \vec{1}|X_{(1)}, Y)$ is identified, which in turn results in identification of $p(X_{(1)}, Y)$, as $p(X_{(1)}, Y) = \frac{p(X_{(1)}, Y, A=\vec{1})}{p(A=\vec{1}|X_{(1)}, Y)}$.

Identification of $p(A = \vec{1}|X_{(1)}, Y)$*:*
The propensity score is identified by

$$\begin{aligned} p(A = \vec{1}|X_{(1)}, Y) &= \prod_{t=1}^{T} p(A^t = \vec{1}|X_{(1)}, Y, \underline{A}^{t-1} = \vec{1}) \\ &\overset{*_1}{=} \prod_{t=1}^{T} \pi_\beta^t(A^t = \vec{1}|\underline{X}_{(1)}^{t-1}, \underline{A}^{t-1} = \vec{1}) \\ &\overset{*_2}{=} \prod_{t=1}^{T} \pi_\beta^t(A^t = \vec{1}|\underline{X}^{t-1}, \underline{A}^{t-1} = \vec{1}) \end{aligned}$$

where we used in $*1$) the fact that future feature values do not affect current acquisition decisions: $A^t \perp\!\!\!\perp \overline{X}_{(1)}^t, Y|\underline{X}_{(1)}^{t-1}, \underline{A}^{t-1}$. We further use in $*2$) that counterfactual feature values $\underline{X}_{(1)}^{t-1}$ are equal to $\underline{X}^{t-1}$ if $\underline{A}^{t-1} = \vec{1}$. The last expression is a function of only observed variables and is thus identified.

Appendix E. Multiple Imputation (MI) for the AFAPE Problem

In this appendix, we aim to delve deeper into the multiple imputation (MI) estimator in the AFAPE context and highlight advantages as well as some common pitfalls associated with using MI approaches in AFA.

Let us begin by emphasizing a significant advantage of the MI estimator compared to other estimators discussed in this paper. It offers an elegant solution to the temporal coarsening problem. In time-series settings, where fixed time intervals are assumed ($t \in \{0, 1, ..., T\}$), employing a very fine resolution of time steps would inevitably result in a considerable increase in missingness, thereby making the AFAPE problem more challenging. The MI estimator can typically overcome this issue by assuming an often justifiable temporal smoothness of the feature distributions.

However, there are drawbacks to MI. MI requires modeling joint distributions, which is a complex task in practice, particularly in high-dimensional settings and when dealing with complex missingness patterns. For instance, the multiple imputation by chained equations (MICE) method (van Buuren, 2007) necessitates fitting d conditional densities for d partially observed features in static settings. In comparison, IPW only requires the specification of the propensity score, which is often more feasible. This effect is especially drastic for high-dimensional features such as images, which necessitate modeling for each pixel when using multiple imputation, but only the modeling of one joint missingness indicator when using IPW.

Furthermore, the MI estimator implies imputation of the missing features X_m by conditioning on the observed features X_o and the label Y (i.e., estimating $\hat{p}(X_m|X_o, Y)$). This introduces the risk of data leakage, as the imputed features may carry predictive information not because of the true data generation mechanism but due to the imputation itself, resulting in a potentially over-optimistic estimation of prediction performance. A common alternative, frequently employed in machine learning, is to impute the data without conditioning on Y. However, this assumption implies that a missing feature $X_{(1),i} \in X_m$ is

conditionally independent of the label given the observed features ($X_{(1),i} \perp\!\!\!\perp Y|X_o$). Determining the marginal predictive value of a feature for predicting Y is, however, the whole task of AFA, which renders this approach impractical.

Conditional mean imputation represents a simplified imputation approach that reduces the complexity of modeling. It has been applied in AFA settings (An et al., 2022; Erion et al., 2021; Janisch et al., 2020). In this approach, missing values are imputed using a conditional mean model for $\hat{\mathbb{E}}[X_m|X_o]$ (or $\hat{\mathbb{E}}[X_m|X_o, Y]$). Therefore, conditional mean imputation assumes:

$$\begin{aligned}\hat{J}_{\textit{MI-Miss}} &= \sum_{X_m,X_o,Y} \mathbb{E}[C_{(\pi_\alpha)}|X_m, X_o, Y]\hat{p}(X_m|X_o, Y)p(X_o, Y) \\ &\approx \sum_{X_o,Y} \mathbb{E}[C_{(\pi_\alpha)}|\hat{\mathbb{E}}[X_m|X_o], X_o, Y]p(X_o, Y)\end{aligned}$$

which does not hold in general and can lead to strongly biased results when $\mathbb{E}[C_{(\pi_\alpha)}|X_m, X_o, Y]$ is nonlinear as is the case generally in AFA settings.

Appendix F. Proof of Lemma 6

In this appendix, we prove Lemma 6, which we repeat here for readability:

Lemma 6. *(Sufficiency conditions for global positivity). The global positivity assumption for semi-offline RL (Assumption 6.3) holds if the positivity assumption from offline RL (Assumption 6.1) or from missing data (Assumption 6.2) holds.*

We split the proof into the following propositions:

Proposition 18. *If the positivity assumption for offline RL (Assumption 6.1) holds, then the global positivity assumption for semi-offline RL also holds.*

Proposition 19. *If the positivity assumption for missing data (Assumption 6.2) holds, then the global positivity assumption for semi-offline RL also holds.*

We begin by proving Proposition 18:

Proof Consider the initial time step $t = 1$, focusing on data points (a'^1, x^0) such that $p(x^0)\pi_\alpha(a'^1|x^0) > 0$. To establish the global positivity assumption, we need to ensure that the regional positivity assumption is satisfied.

We will show that $a^1 = a'^1$ belongs to the regional admissible set $\tilde{\mathcal{A}}^1_{adm}(x^0, a'^1)$, which implies that the regional positivity condition is met. Two conditions must be fulfilled for $a^1 = a'^1$ to be included in $\tilde{\mathcal{A}}^1_{adm}(x^0, a'^1)$. First, $a^1 = a'^1$ must belong to the local admissible set $\mathcal{A}^1_{adm}(x^0, a'^1)$, which directly follows from the offline RL positivity assumption. Second, we must ensure that for $a^1 = a'^1$, the regional admissible set exists at the subsequent time step. Specifically, since $a^1 = a'^1$ (and therefore $x^1 = x'^1$), the regional admissible set at time step 2, $\tilde{\mathcal{A}}^2_{adm}(\underline{a}'^2, \underline{x}^1, \underline{a}^1)$, must exist for all x^1 and a'^2 such that:

$$p(x^1|x^0, a^1)\pi_\alpha(a'^2|\underline{x}^1, \underline{a}^1) > 0.$$

For these conditions, the positivity assumption under the offline RL view again implies that $a^2 = a'^2$ is included in the local admissible set $\mathcal{A}^2_{adm}(a'^2, \underline{x}^1, \underline{a}^1)$. This reasoning can be extended iteratively through all time steps up to T, thus proving that regional positivity holds at every prior time point, which in turn establishes global positivity. ■

Next, we prove Proposition 19, following a similar strategy:

Proof Again, consider the initial time step $t = 1$, focusing on data points (a'^1, x^0) such that $p(x^0)\pi_\alpha(a'^1|x^0) > 0$. We show that $a^1 = \vec{1}$ is included in the regional admissible set $\tilde{\mathcal{A}}^1_{adm}(x^0, a'^1)$, irrespective of the value of a'^1. First, $a^1 = \vec{1}$ must be included in the local admissible set $\mathcal{A}^1_{adm}(x^0, a'^1)$, a condition that directly follows from the missing data positivity assumption.

Second, $a^1 = \vec{1}$ must permit the existence of a regional admissible set at the next time step. Specifically, given $a^1 = \vec{1}$, the regional admissible set at time step 2, $\tilde{\mathcal{A}}^2_{adm}(\underline{a}'^2, \underline{x}^1, \underline{a}^1) = \tilde{\mathcal{A}}^2_{adm}(\underline{a}'^2, \underline{x}^1, \underline{a}^1 = \vec{1})$, must exist for all x^1 and a'^2 such that:

$$p(x^1|x^0, A^1 = \vec{1})\pi_\alpha(a'^2|\underline{x}'^1, \underline{a}'^1) > 0.$$

Under these conditions, the missing data positivity assumption again directly implies that $a^2 = \vec{1}$ belongs to the local admissible set at time step 2, $\tilde{\mathcal{A}}^2_{adm}(\underline{a}'^2, \underline{x}^1, \underline{a}^1 = \vec{1})$, regardless of the values of a'^2 and x'^1. This reasoning can be extended step-by-step until T, thereby ensuring that regional positivity holds at each prior time point and, consequently, that global positivity is satisfied as well. ■

Appendix G. Proof of Theorems 7 and 8

In this Appendix, we prove Theorems 7 and 8. We also demonstrate how the positivity assumption arises. We restate the theorems here for clarity and ease of reference.

Theorem 7. *(Identification of J for the semi-offline RL view). The reformulated AFAPE problem of estimating J under the semi-offline RL view (Eq. 17) is under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and Assumption 6.3 (global positivity) identified by*

$$J = \mathbb{E}_{p'}[C'_{(\pi_\alpha)}] = \sum_{A', A, G_A(X_{(1)}), Y} f_C(A', X', Y) q'(A', A, X, Y) \tag{19}$$

with the distribution

$$q'(A', A, X, Y) = \prod_{t=1}^{T} \underbrace{\pi^t_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})}_{\text{distr. subject to constraints}} \underbrace{\pi^t_\alpha(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1})}_{\text{target policy}} \prod_{t=0}^{T} p(X^t|\underline{X}^{t-1}, \underline{A}^t, Y) p(Y) \tag{20}$$

where

$$\pi^t_{id}(A^t|\underline{A}'^t,\underline{X}^{t-1},\underline{A}^{t-1}) = \underbrace{\mathbb{I}(A^t \in \tilde{\mathcal{A}}^t_{adm}(\underline{X}^{t-1},\underline{A}^{t-1},\underline{A}'^t))}_{\textit{support restriction}} f^t_{id}(\underline{A}'^t,\underline{X}^{t-1},\underline{A}^{t-1}) \tag{21}$$

for any function f^t_{id} s.t. π^t_{id} is a valid density.

Theorem 8. *(Bellman equation for semi-offline RL). The semi-offline RL view admits under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and the local positivity assumption at datapoint $\underline{x}^{t-1},\underline{a}^{t-1},a'^t$ (from Definition 4), the following semi-offline RL version of the Bellman equation:*

$$Q_{Semi}(\underline{A}'^t,\underline{X}^{t-1},\underline{A}^{t-1},\Xi) = \sum_{X^t} V_{Semi}(\underline{A}'^t,\underline{X}^t,\underline{A}^{t-1},A^t=a^t,\Xi)p(X^t|\underline{X}^{t-1},\underline{A}^{t-1},A^t=a^t,\Xi) \tag{22}$$

for any $a^t \in \mathcal{A}^t_{adm}(\underline{X}^{t-1},\underline{A}^{t-1},A'^t)$

$$V_{Semi}(\underline{A}'^t,\underline{X}^t,\underline{A}^t,\Xi) = \sum_{A'^{t+1}} Q_{Semi}(\underline{A}'^{t+1},\underline{X}^t,\underline{A}^t,\Xi)\pi^{t+1}_\alpha(A'^{t+1}|\underline{X}'^t,\underline{A}'^t) \tag{23}$$

with semi-offline RL versions of the state-action value function Q_{Semi} and state value function V_{Semi}:

$$Q^t_{Semi} \equiv Q_{Semi}(\underline{A}'^t,\underline{X}^{t-1},\underline{A}^{t-1},\Xi) \equiv \mathbb{E}_{p'}[C'_{(\overline{\pi}^{t+1}_\alpha)}|\underline{A}'^t,\underline{X}^{t-1},\underline{A}^{t-1},\Xi]$$
$$V^t_{Semi} \equiv V_{Semi}(\underline{A}'^t,\underline{X}^t,\underline{A}^t,\Xi) \equiv \mathbb{E}_{p'}[C'_{(\overline{\pi}^{t+1}_\alpha)}|\underline{A}'^t,\underline{X}^t,\underline{A}^t,\Xi]$$

where $C'_{(\overline{\pi}^{t+1}_\alpha)}$ denotes the potential outcome of C' under interventions from time step $t+1$ onwards. $\Xi \subseteq \{Y,O\}$, with O denoting all features that are always available, denotes an optional subset of additional variables that can be conditioned on. Furthermore, Q^t_{Semi} and V^t_{Semi} are identified if the regional positivity assumption (from Definition 5) holds at $\underline{X}^{t-1},\underline{A}^{t-1},\underline{A}'^t$ and $a^t \in \tilde{\mathcal{A}}^t_{adm}(\underline{X}^{t-1},\underline{A}^{t-1},\underline{A}'^t)$.

Proof Firstly, we factorize the counterfactual distribution, denoted by q', expressing it as a function of the observed (simulated) data. We factorize the graph step-by-step to show how the semi-offline RL version of the Bellman equation arises. We split identification in each step into two parts to emphasize the two parts of the Bellman equation. To help guide the identification, we duplicate Figure 6 of the causal graph describing the simulation process in Figure 11A). Alongside it, we show the counterfactual graph (for identification step $t=1$) in Figure 11B).

Step 0

Counterfactual factorization (step $t=0$, part 1):

$$p'\left(C'_{(\pi_\alpha)}\right) \overset{*_1}{\equiv} p'\left(C'_{(\overline{\pi}^1_\alpha)}\right) = \sum_{X^0,\Xi} p'\left(C'_{(\overline{\pi}^1_\alpha)}\middle|X^0,\Xi\right) p(X^0,\Xi)$$

where we denote in $*1)$ $C'_{(\overline{\pi}^1_\alpha)}$ as the counterfactual C' under an intervention of π_α from step $t = 1$ onwards. The extension by X^0 is needed for adjustment. The inclusion of $\Xi \subseteq \{Y, O\}$, where O denotes the subset of always observed features amongst $X_{(1)}$, is optional.

Counterfactual factorization (step $t = 0$, part 2):

$$
\begin{aligned}
p'\left(C'_{(\overline{\pi}^1_\alpha)}\middle|X^0,\Xi\right) &= \sum_{\boldsymbol{a'^1}} p'\left(C'_{(\overline{\boldsymbol{\pi}}^{\mathbf{2}}_{\boldsymbol{\alpha}},\boldsymbol{a'^1})}\middle|X^0,\Xi\right)\boldsymbol{\pi}^{\mathbf{1}}_{\boldsymbol{\alpha}}(\boldsymbol{a'^1}|\boldsymbol{X^0}) \\
&\overset{*_1}{=} \sum_{a'^1} p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1,\boldsymbol{\pi}^{\mathbf{1}}_{\boldsymbol{id}})}\middle|X^0,\Xi\right)\pi^1_\alpha(a'^1|X^0) \\
&= \sum_{a'^1,\boldsymbol{a^1}} p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1,\boldsymbol{a^1})}\middle|X^0,\Xi\right)\boldsymbol{\pi}^{\mathbf{1}}_{\boldsymbol{id}}(\boldsymbol{a^1}|\boldsymbol{X^0},\boldsymbol{a'^1})\pi^1_\alpha(a'^1|X^0) \\
&\overset{*_2}{=} \sum_{a'^1,a^1} p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1,a^1)}\middle|X^0,\boldsymbol{a^1},\Xi\right)\pi^1_{id}(a^1|X^0,a'^1)\pi^1_\alpha(a'^1|X^0) \\
&\overset{*_3}{=} \sum_{a'^1,a^1} p'\left(C'_{(\overline{\boldsymbol{\pi}}^{\mathbf{2}}_{\boldsymbol{\alpha}},\boldsymbol{a'^1})}\middle|X^0,a^1,\Xi\right)\pi^1_{id}(a^1|X^0,a'^1)\pi^1_\alpha(a'^1|X^0) \\
&\overset{*_4}{=} \sum_{a'^1,a^1} p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1)}\middle|X^0,a'^1,\boldsymbol{a^1},\Xi\right)\pi^1_{id}(a^1|X^0,a'^1)\pi^1_\alpha(a'^1|X^0) \\
&\overset{*_5}{=} \sum_{a'^1,a^1} p'\left(C'_{(\overline{\boldsymbol{\pi}}^{\mathbf{2}}_{\boldsymbol{\alpha}})}\middle|X^0,a'^1,a^1,\Xi\right)\pi^1_{id}(a^1|X^0,a'^1)\pi^1_\alpha(a'^1|X^0) \\
&\overset{*_6}{=} \sum_{a'^1,a^1} p'\left(C'_{(\overline{\pi}^2_\alpha)}\middle|\boldsymbol{X^0},\boldsymbol{a'^1},\boldsymbol{\Xi}\right)\pi^1_{id}(a^1|X^0,a'^1)\pi^1_\alpha(a'^1|X^0)
\end{aligned}
$$

with the following explanations:

- $*1)$: We notice that $C'_{(\overline{\pi}^2_\alpha,a'^1)}$ is independent of any interventions π^1_{id} on A^1. This step prevents positivity problems in subsequent steps.
- $*2)$: We use the exchangeability $C'_{(\overline{\pi}^2_\alpha,a'^1,a^1)} \perp\!\!\!\perp A^1|X^0,\Xi$, which follows from the NUC assumption.
- $*3)$: We use the consistency assumption: $p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1,a^1)}\middle|X^0,a^1,\Xi\right) = p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1)}\middle|X^0,a^1,\Xi\right)$
- $*4)$: We use the exchangeability: $C'_{(\overline{\pi}^2_\alpha,a'^1)} \perp\!\!\!\perp A'^1|X^0,A^1,\Xi$
- $*5)$: We use the consistency assumption: $p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1)}\middle|X^0,a'^1,a^1,\Xi\right) = p'\left(C'_{(\overline{\pi}^2_\alpha)}\middle|X^0,a'^1,a^1,\Xi\right)$
- $*6)$: We use the conditional independence $C'_{(\overline{\pi}^2_\alpha)} \perp\!\!\!\perp A^1|X^0,A'^1,\Xi$

We must also ensure that $p'\left(C'_{(\overline{\pi}^2_\alpha,a'^1)}\middle|X^0,a'^1,a^1,\Xi\right)$, i.e. conditioning on X^0,A'^1,A^1,Ξ, is well specified in $*4)$. To understand what positivity requirements are necessary, we first factorize the "observational" (i.e., simulated) distribution for step $t = 0$. By observational distribution for step $t = 0$, we refer to a distribution which only contains interventions from step $t = 2$ onwards:

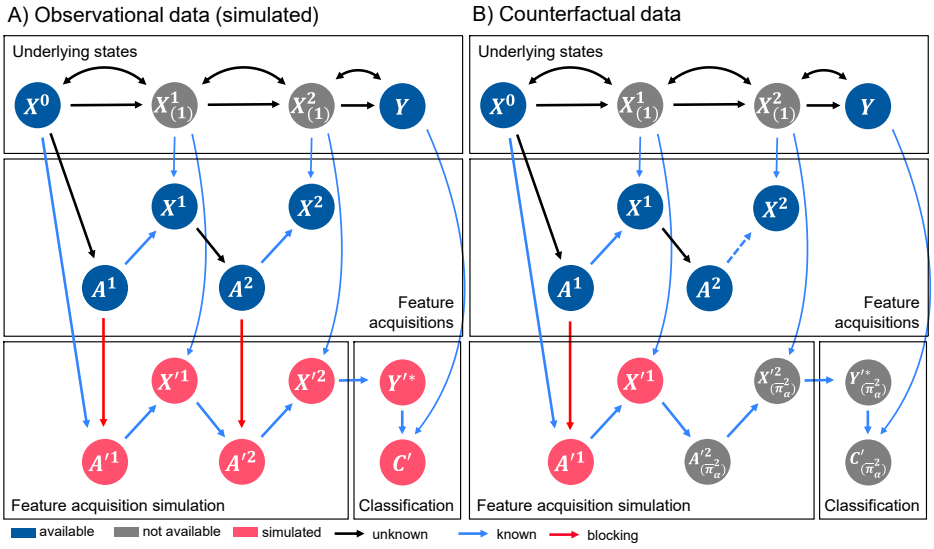


Figure 11: Causal graph for the distribution p'. A) Simulated ("observational") distribution. B) Counterfactual distribution under the intervention $\overline{\pi}^2_\alpha$. Edges showing long-term dependencies are omitted from the graphs for visual clarity. These include: $\underline{X}^{t-1}_{(1)} \leftrightarrow X^t_{(1)}$; $\underline{X}^T_{(1)} \leftrightarrow Y$; $\underline{X}^{t-1}, \underline{A}^{t-1} \rightarrow A^t$; $\underline{X}'^{t-1}/\underline{X}'^{t-1}_{(\overline{\pi}^2_\alpha)}, \underline{A}'^{t-1}/\underline{A}'^{t-1}_{(\overline{\pi}^2_\alpha)} \rightarrow A'^t/A'^t_{(\overline{\pi}^t_\alpha)}$; and $\underline{X}'^T/\underline{X}'^T_{(\overline{\pi}^2_\alpha)}, \underline{A}'^T/\underline{A}'^T_{(\overline{\pi}^2_\alpha)} \rightarrow Y'^*/Y'^*_{(\overline{\pi}^2_\alpha)}$.

Observational factorization (step $t = 0$):

$$p'\left(C'_{(\overline{\pi}^2_\alpha)}\right) = \sum_{X^0, A^1, A'^1, \Xi} p'\left(C'_{(\overline{\pi}^2_\alpha)} \middle| X^0, A'^1, \Xi\right) \underbrace{\pi'^1_{sim}(A'^1|X^0, A^1)}_{\text{simulation policy}} \underbrace{\pi^1_\beta(A^1|X^0)}_{\text{retro. acq. policy}} p(X^0, \Xi)$$

By comparing the observational and counterfactual factorizations, we see that the following positivity assumption is required:

$$\begin{aligned} &\text{if} && p(x^0)q'(a'^1, a^1|x^0) = p(x^0)\pi^1_\alpha(a'^1|x^0)\pi^1_{id}(a^1|x^0, a'^1) > 0 \\ &\text{then} && p(x^0)p'(a'^1, a^1|x^0) = p(x^0)\pi'^1_{sim}(a'^1|x^0, a^1)\pi^1_\beta(a^1|x^0) \geq \mathcal{O} \\ & && \forall x^0, a'^1, a^1, \text{and some constant } \mathcal{O} > 0 \end{aligned} \tag{41}$$

Since none of the distributions π_α, π_{id}, π'_{sim}, and π_β depend on Ξ, the choice for Ξ will not influence the positivity requirements. We can further simplify the positivity assumption by using knowledge about the known simulation policy π'_{sim}. By the construction of the blocking operation of the simulation policy π'_{sim} (Definition 2), one observes that

$$\text{if } \pi^1_\alpha(a'^1|x^0) > 0, \qquad \text{then } \pi'^1_{sim}(a'^1|x^0, a^1) \geq \mathcal{O}_1, \qquad \text{if and only if } a'^1 \leq a^1$$

where, $\mathcal{O}_1$ is some constant > 0, and, as before, we let $a'^1 \leq a^1$ denote the element-wise comparison. The resulting positivity violation for the case $a'^1 \not\leq a^1$ can be avoided by restricting π_{id} in the following way:

Restriction 1 for π_{id} (step $t = 0$):

$$\text{if } a'^1 \not\leq a^1, \qquad \text{then } \pi^1_{id}(a^1|x^0, a'^1) = 0 \qquad \forall x^0, a'^1, a^1.$$

A second possible positivity violation arises if $\pi^1_\beta(a^1|x^0) = 0$ for some values of a^1. This poses a second requirement for π_{id}:

Restriction 2 for π_{id} (step $t = 0$):

$$\text{if } \pi^1_\beta(a^1|x^0) = 0, \qquad \text{then } \pi^1_{id}(a^1|x^0, a'^1) = 0 \qquad \forall x^0, a'^1, a^1.$$

Since π_{id} is required to be a valid probability distribution (it cannot be 0 for all a^1), this imposes the following requirement for π_β:

$$\text{if } p(x^0)\pi^1_\alpha(a'^1|x^0) > 0, \qquad \text{then } \pi^1_\beta(A^1 \geq a'^1|x^0) \geq \mathcal{O} \quad \forall x^0, a'^1, \text{ and some constant } \mathcal{O} > 0.$$

The positivity assumption implies that for any desired action a'^1 by the target policy π_α, there exists at least positive support for one set of acquisitions a^1 that include equal or more acquisitions than what is contained in a'^1. This is equivalent to the local positivity assumption at x^0, a'^0 (i.e. the existence of $\mathcal{A}^1_{adm}$ from Definition 4). In the next steps, we show that these are only minimal requirements for $\pi^1_{id}(A^1|X^0, A'^1)$. To avoid running into positivity violations in later time steps, a further restriction can be necessary.

Step 1

In the following, we continue the identification for step $t = 1$.

Counterfactual factorization (step $t = 1$, part 1):

$$\begin{aligned} p'\left(C'_{(\overline{\pi}^2_\alpha)}\middle|X^0, A'^1, \Xi\right) &= p'\left(C'_{(\overline{\pi}^2_\alpha)}\middle|X^0, A'^1, a^1, \Xi\right) = \\ &= \sum_{X^1} p'\left(C'_{(\overline{\pi}^2_\alpha)}\middle|\underline{A}'^1, \underline{X}^1, a^1, \Xi\right) p(X^1|X^0, a^1, \Xi) \end{aligned}$$

which holds for any $a^1 \in \mathcal{A}^1_{adm}(X^0, A'^1)$ (because local positivity must hold). Therefore, the term $p'\left(C'_{(\overline{\pi}^2_\alpha)}\middle|\underline{A}'^1, \underline{X}^1, a^1, \Xi\right)$ needs to be only identified for *at least one value* $a^1 \in \mathcal{A}^1_{adm}(X^0, A'^1)$.

Counterfactual factorization (step $t = 1$, part 2):

$$
\begin{aligned}
&p'\left(C'_{(\overline{\pi}^2_\alpha)}\middle|\underline{A}'^1, \underline{X}^1, \underline{A}^1, \Xi\right) = \\
&\quad= \sum_{\boldsymbol{a'^2}} p'\left(C'_{(\boldsymbol{\overline{\pi}^3_\alpha}, \boldsymbol{a'^2})}\middle|\underline{A}'^1, \underline{X}^1, \underline{A}^1, \Xi\right) \boldsymbol{\pi^2_\alpha(a'^2|\underline{X}'^1, \underline{A}'^1)} \\
&\quad\overset{*_1}{=} \sum_{a'^2} p'\left(C'_{(\overline{\pi}^3_\alpha, a'^2, \boldsymbol{\pi^2_{id}})}\middle|\underline{A}'^1, \underline{X}^1, \underline{A}^1, \Xi\right) \pi^2_\alpha(a'^2|\underline{X}'^1, \underline{A}'^1) \\
&\quad= \sum_{a'^2, \boldsymbol{a^2}} p'\left(C'_{(\overline{\pi}^3_\alpha, a'^2, \boldsymbol{a^2})}\middle|\underline{A}'^1, \underline{X}^1, \underline{A}^1, \Xi\right) \boldsymbol{\pi^2_{id}(a^2|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1)} \pi^2_\alpha(a'^2|\underline{X}'^1, \underline{A}'^1) \\
&\quad\overset{*_2}{=} \sum_{a'^2, a^2} p'\left(C'_{(\boldsymbol{\overline{\pi}^3_\alpha}, \boldsymbol{a'^2})}\middle|\underline{A}'^1, \underline{X}^1, \underline{A}^1, \boldsymbol{a^2}, \Xi\right) \pi^2_{id}(a^2|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1) \pi^2_\alpha(a'^2|\underline{X}'^1, \underline{A}'^1) \\
&\quad\overset{*_3}{=} \sum_{a'^2, a^2} p'\left(C'_{(\boldsymbol{\overline{\pi}^3_\alpha})}\middle|\underline{A}'^1, \boldsymbol{a'^2}, \underline{X}^1, \underline{A}^1, a^2, \Xi\right) \pi^2_{id}(a^2|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1) \pi^2_\alpha(a'^2|\underline{X}'^1, \underline{A}'^1) \\
&\quad\overset{*_4}{=} \sum_{a'^2, a^2} p'\left(C'_{(\overline{\pi}^3_\alpha)}\middle|\boldsymbol{\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1}, \Xi\right) \pi^2_{id}(a^2|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1) \pi^2_\alpha(a'^2|\underline{X}'^1, \underline{A}'^1)
\end{aligned}
$$

where we denote $\underline{X}'^1 \equiv G_{\underline{A}'^1}(X_{(1)})$ which contains a subset of the features in $\underline{X}^1 = G_{\underline{A}^1}(X_{(1)})$. The derivation was based on the following arguments:

- $*1$): We use that $C'_{(\overline{\pi}^3_\alpha, a'^2)}$ is independent of any interventions π^2_{id} on A^2.
- $*2$): We use exchangeability : $C'_{(\overline{\pi}^3_\alpha, a'^2, a^2)} \perp\!\!\!\perp A^2 | \underline{A}'^1, \underline{X}^1, \underline{A}^1, \Xi$ and consistency: $p'\left(C'_{(\overline{\pi}^3_\alpha, a'^2, a^2)}\middle|\underline{A}'^1, \underline{X}^1, \underline{A}^1, a^2, \Xi\right) = p'\left(C'_{(\overline{\pi}^3_\alpha, a'^2)}\middle|\underline{A}'^1, \underline{X}^1, \underline{A}^1, a^2, \Xi\right)$ for A^2.
- $*3$): We use the exchangeability $C'_{(\overline{\pi}^3_\alpha, a'^2)} \perp\!\!\!\perp A'^2 | \underline{A}'^1, \underline{X}^1, \underline{A}^1, a^2, \Xi$ and consistency: $p'\left(C'_{(\overline{\pi}^3_\alpha, a'^2)}\middle|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1, a^2, \Xi\right) = p'\left(C'_{(\overline{\pi}^3_\alpha)}\middle|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1, a^2, \Xi\right)$ for A'^2.
- $*4$): We use the conditional independence $C'_{(\overline{\pi}^3_\alpha)} \perp\!\!\!\perp A^2 | \underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1, \Xi$

We must also ensure in $*3$) that $p'\left(C'_{(\overline{\pi}^3_\alpha, a'^2)}\middle|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1, a^2, \Xi\right)$, i.e., conditioning on $\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1, a^2, \Xi$, is well specified. To understand what positivity requirements are necessary, we factorize the "observational" (i.e., simulated) distribution for step $t = 1$.

Observational factorization (step $t = 1$):

$$
\begin{aligned}
p'\left(C'_{(\overline{\pi}^3_\alpha)}\middle|A'^1, X^0, A^1, \Xi\right) = \sum_{X^1, a^2, a'^2} & p'\left(C'_{(\overline{\pi}^3_\alpha)}\middle|\underline{A}'^1, a'^2, \underline{X}^1, \underline{A}^1, \Xi\right) \\
& \cdot \underbrace{\pi^{\prime 2}_{sim}(a'^2|\underline{X}'^1, \underline{A}'^1, a^2)}_{\text{known simulation policy}} \underbrace{\pi^2_\beta(a^2|\underline{X}^1, \underline{A}^1)}_{\text{retro. acquisition policy}} \quad p(X^1|X^0, A^1, \Xi)
\end{aligned}
$$

By comparing the observational and counterfactual factorizations, we see that the following positivity assumption is required:

$$\begin{aligned}
\text{if} \quad & q'(\underline{a}'^1, a'^2, \underline{x}^1, \underline{a}^1, a^2) = \\
& \quad = q'(\underline{a}'^1, \underline{x}^1, \underline{a}^1)\pi^2_\alpha(a'^2|\underline{x}'^1, \underline{a}'^1)\pi^2_{id}(a^2|\underline{a}'^1, a'^2, \underline{x}^1, \underline{a}^1) > 0 \\
\text{then} \quad & p'(\underline{x}'^1, a'^2, \underline{x}^1, \underline{a}^1, a^2) = \\
& \quad = p'(\underline{a}'^1, \underline{x}^1, \underline{a}^1)\pi'^2_{sim}(a'^2|\underline{x}'^1, \underline{a}'^1, a^2)\pi^2_\beta(a^2|\underline{x}^1, \underline{a}^1) \geq \mathcal{O} \\
& \forall \underline{x}^1, \underline{a}'^1, \underline{a}^1, a'^2, a^2, \text{ and some constant } \mathcal{O} > 0
\end{aligned}$$

with the following factorizations:

$$\begin{aligned}
q'(\underline{A}'^1, \underline{X}^1, \underline{A}^1) &= q'(\underline{A}'^1, \underline{X}^0, \underline{A}^1)p(X^1|\underline{X}^0, \underline{A}^1) \\
p'(\underline{A}'^1, \underline{X}^1, \underline{A}^1) &= p'(\underline{A}'^1, \underline{X}^0, \underline{A}^1)p(X^1|\underline{X}^0, \underline{A}^1)
\end{aligned}$$

Note that we can again ignore Ξ since it doesn't affect positivity requirements. The positivity condition can again be simplified through the two restrictions on π^2_{id}:

Restrictions 1 and 2 for π_{id} (step $t = 2$):

$$\text{if } a'^2 \not\leq a^2 \text{ or } \pi^2_\beta(a^2|\underline{x}^1, \underline{a}^1) = 0, \quad \text{then } \pi^2_{id}(a^2|\underline{a}'^1, a'^2, \underline{x}^1, \underline{a}^1) = 0 \qquad \forall a^2, a'^2$$

This imposes the requirement for π^2_β that there exists at least one value a^2 such that $a'^2 \leq a^2$ and $\pi^2_\beta(a^2|\underline{x}^1, \underline{a}^1) \geq \mathcal{O}$ (i.e. local positivity at $\underline{x}^1, \underline{a}^1, a'^1$). Notice, however, that this has to hold for *all values* a^1 that were "allowed" in step $t = 1$ (i.e. where $\pi^1_{id}(a^1|x^0, a'^1) \geq \mathcal{O}$). As π^1_{id} only needs to have support for at least one $a^1 \in \mathcal{A}^1_{adm}$, we can restrict π^1_{id} at step $t = 0$ further to reduce the positivity assumption for step $t = 2$. We do, however, only want to restrict π^1_{id} as much as necessary because if π^1_{id} has wider support, this means that more data points are used in the analysis. Therefore, we introduce the notion of regional positivity and the regional admissible set $\tilde{\mathcal{A}}_{adm}$ (from Definition 5). In particular, Definition 5 defines $\tilde{\mathcal{A}}^1_{adm}(x^0, a'^1)$ as the subset of $\mathcal{A}^1_{adm}(x^0, a'^1)$ such that local positivity holds at step $t = 1$ for all possible values of x^1, and a'^2. As this has to hold for future time-steps as well (as will be shown next), the definition for $\tilde{\mathcal{A}}_{adm}$ even states regional positivity has to hold recursively, i.e., also at $t = 1$.

In summary, local positivity at step $t = 2$ ensures that the available data allows the simulation of the currently desired action a'^2. Regional positivity at step $t = 0$ ensures that only those simulations are used at step $t = 0$ such that simulations of desired actions in the future (at step $t = 1$) are possible with the data.

Step t

Now, we generalize the factorization to step t.

Counterfactual factorization (step t, part 1):

$$\begin{aligned}
p'\left(C'_{(\overline{\pi}^{t+1}_\alpha)}\middle|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi\right) &= p'\left(C'_{(\overline{\pi}^{t+1}_\alpha)}\middle|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, a^t, \Xi\right) = \\
&= \sum_{X'^t, X^t} p'\left(C'_{(\overline{\pi}^{t+1}_\alpha)}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^{t-1}, a^t, \Xi\right) p(X^t|\underline{X}^{t-1}, \underline{A}^{t-1}, a^t, \Xi)
\end{aligned} \tag{42}$$

which holds for any $a^t \in \mathcal{A}^t_{adm}(\underline{X}^{t-1}, \underline{A}^{t-1}, A'^t)$ (because local positivity must hold). Therefore, the term $p'\left(C'_{(\overline{\pi}^{t+1}_\alpha)}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^{t-1}, a^t, \Xi\right)$ needs to be only identified for *at least one value* $a^t \in \mathcal{A}^t_{adm}(\underline{X}^{t-1}, \underline{A}^{t-1}, A'^t)$.

Counterfactual factorization (step t, part 2):

$$
\begin{aligned}
&p'\left(C'_{(\overline{\pi}^{t+1}_\alpha)}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi\right) = \qquad (43)\\
&= \sum_{\boldsymbol{a'^{t+1}}} p'\left(C'_{(\boldsymbol{\overline{\pi}^{t+2}_\alpha, a'^{t+1}})}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi\right) \boldsymbol{\pi^{t+1}_\alpha(a'^{t+1}|\underline{X}'^t, \underline{A}'^t)}\\
&\overset{*1}{=} \sum_{a'^{t+1}} p'\left(C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1}, \boldsymbol{\pi^{t+1}_{id}})}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi\right) \pi^{t+1}_\alpha(a'^{t+1}|\underline{X}'^t, \underline{A}'^t)\\
&= \sum_{a'^{t+1}, \boldsymbol{a^{t+1}}} p'\left(C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1}, \boldsymbol{a^{t+1}})}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi\right) \boldsymbol{\pi^{t+1}_{id}(a^{t+1}|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t)} \pi^{t+1}_\alpha(a'^{t+1}|\underline{X}'^t, \underline{A}'^t)\\
&\overset{*2}{=} \sum_{a'^{t+1}, a^{t+1}} p'\left(C'_{\boldsymbol{(\overline{\pi}^{t+2}_\alpha, a'^{t+1})}}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, \boldsymbol{a^{t+1}}, \Xi\right) \pi^{t+1}_{id}(a^{t+1}|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t) \pi^{t+1}_\alpha(a'^{t+1}|\underline{X}'^t, \underline{A}'^t)\\
&\overset{*3}{=} \sum_{a'^{t+1}, a^{t+1}} p'\left(C'_{\boldsymbol{(\overline{\pi}^{t+2}_\alpha)}}\middle|\underline{A}'^t, \boldsymbol{a'^{t+1}}, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi\right) \pi^{t+1}_{id}(a^{t+1}|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t) \pi^{t+1}_\alpha(a'^{t+1}|\underline{X}'^t, \underline{A}'^t)\\
&\overset{*4}{=} \sum_{a'^{t+1}, a^{t+1}} p'\left(C'_{(\overline{\pi}^{t+2}_\alpha)}\middle|\boldsymbol{\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t}, \Xi\right) \underbrace{\pi^{t+1}_{id}(a^{t+1}|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t)}_{\text{arbitrary dist. subject to constraints}} \underbrace{\pi^{t+1}_\alpha(a'^{t+1}|\underline{X}'^t, \underline{A}'^t)}_{\text{target policy}}
\end{aligned}
$$

with the following explanations:

- $*1$): We use that $C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1})}$ is independent of any interventions π^{t+1}_{id} on A^{t+1}.
- $*2$): We use exchangeability : $C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1}, a^{t+1})} \perp\!\!\!\perp A^{t+1}|\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi$ and consistency: $p'\left(C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1}, a^{t+1})}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi\right) = p'\left(C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1})}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi\right)$ for A^{t+1}.
- $*3$): We use the exchangeability $C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1})} \perp\!\!\!\perp A'^{t+1}|\underline{A}'^t, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi$ and consistency: $p'\left(C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1})}\middle|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi\right) = p'\left(C'_{(\overline{\pi}^{t+2}_\alpha)}\middle|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi\right)$ for A'^{t+1}.
- $*4$): We use the conditional independence $C'_{(\overline{\pi}^{t+2}_\alpha)} \perp\!\!\!\perp A^{t+1}|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t, \Xi$

As before, we have to make sure in $*3$) that $p'\left(C'_{(\overline{\pi}^{t+2}_\alpha, a'^{t+1})}\middle|\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi\right)$, i.e. conditioning on $\underline{A}'^t, a'^{t+1}, \underline{X}^t, \underline{A}^t, a^{t+1}, \Xi$, is well specified. To understand what positivity requirements are necessary, we factorize the "observational" (i.e., simulated) distribution for step t:

Observational factorization (step t):

$$p'\left(C'_{(\overline{\pi}_\alpha^{t+2})}\middle|\underline{A}'^t,\underline{X}^{t-1},\underline{A}^t,\Xi\right)=\sum_{X^t,a^{t+1},a'^{t+1}} p'\left(C'_{(\overline{\pi}_\alpha^{t+2})}\middle|\underline{A}'^t,a'^{t+1},\underline{X}^t,\underline{A}^t,\Xi\right)$$
$$\cdot\underbrace{\pi_{sim}'^{t+1}(a'^{t+1}|\underline{X}'^t,\underline{A}'^t,a^{t+1})}_{\text{known simulation policy}}\ \underbrace{\pi_\beta^{t+1}(a^{t+1}|\underline{X}^t,\underline{A}^t)}_{\text{retro. acquisition policy}}\ p(X^t|\underline{X}^{t-1},\underline{A}^t,\Xi)$$

By comparing the observational and counterfactual factorizations, we see that the following positivity assumption is required:

$$\begin{aligned}
\text{if}\quad & q'(\underline{a}'^t,a'^{t+1},\underline{x}^t,\underline{a}^t,a^{t+1})=\\
&\quad=q'(\underline{a}'^t,\underline{x}^t,\underline{a}^t)\pi_\alpha^{t+1}(a'^{t+1}|\underline{x}'^t,\underline{a}'^t)\pi_{id}^{t+1}(a^{t+1}|\underline{a}'^t,a'^{t+1},\underline{x}^t,\underline{a}^t)>0\\
\text{then}\quad & p'(a'^{t+1},\underline{x}^t,\underline{a}^t,a^{t+1})=\\
&\quad=p'(\underline{a}'^t,\underline{x}^t,\underline{a}^t)\pi_{sim}'^{t+1}(a'^{t+1}|\underline{x}'^t,\underline{a}'^t,a^{t+1})\pi_\beta^{t+1}(a^{t+1}|\underline{x}^t,\underline{a}^t)\geq\mathcal{O}\\
&\forall\underline{x}^t,\underline{a}'^t,\underline{a}^t,a'^{t+1},a^{t+1},\ \text{and some constant}\ \mathcal{O}>0
\end{aligned}$$

with the following factorizations:

$$\begin{aligned}
q'(\underline{A}'^t,\underline{X}^t,\underline{A}^t)&=q'(\underline{A}'^t,\underline{X}^{t-1},\underline{A}^t)p(X^t|\underline{X}^{t-1},\underline{A}^t)\\
p'(\underline{A}'^t,\underline{X}^t,\underline{A}^t)&=p'(\underline{A}'^t,\underline{X}^{t-1},\underline{A}^t)p(X^t|\underline{X}^{t-1},\underline{A}^t)
\end{aligned}$$

The positivity condition can again be simplified through the two restrictions on π_{id}:

Restrictions 1 and 2 for π_{id} (step t):

$$\begin{aligned}
\text{if}\quad & a'^{t+1}\not\leq a^{t+1}\ \text{or}\ \pi_\beta^{t+1}(a^{t+1}|\underline{x}^t,\underline{a}^t)=0,\\
\text{then}\quad & \pi_{id}^{t+1}(a^{t+1}|\underline{a}'^t,a'^{t+1},\underline{x}^t,\underline{a}^t)=0\\
& \forall a^{t+1},a'^{t+1}
\end{aligned}$$

This imposes the requirement for π_β that there exists at least one value a^{t+1} such that $a'^{t+1}\leq a^{t+1}$ and $\pi_\beta^{t+1}(a^{t+1}|\underline{x}^t,\underline{a}^t)\geq\mathcal{O}$ (i.e. local positivity at $\underline{x}^t,\underline{a}^t,a'^t$). This has to hold for *all values* $\underline{a}^t$ that were "allowed" in all previous steps (i.e. all a^t, for all $\tau\leq t$, s.t. $\pi_{id}^t(a^\tau|\underline{a}'^{\tau-1},a'^\tau,\underline{x}^{\tau-1},\underline{a}^{\tau-1})>0$ and which could later on have led to the current state). As π_{id} only needs to have support for at least one $a^t\in\mathcal{A}_{adm}^t$ per step, we can restrict π_{id} at *all* previous steps to reduce the positivity assumption for step t. Again, we do not want to restrict π_{id} too much because if π_{id} has wider support, this means that more data points are used in the analysis. The regional positivity assumption (from Definition 5) ensures in this case that only those simulations are used (and exist) at all previous steps such that simulations of the desired actions can be performed at step t (and for future steps).

Full factorization

Bringing all time-steps $t = 0, ..., T$ together and including Y (as a part of Ξ) one obtains the full factorization of the identifying distribution q':

$$q'(A', A, X, Y) = \prod_{t=1}^{T} \underbrace{\pi_{id}^t(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})}_{\text{arb. distr. subject to constraints}} \underbrace{\pi_{\alpha}^t(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1})}_{\text{target policy}} \prod_{t=0}^{T} p(X^t|\underline{X}^{t-1}, \underline{A}^t, Y)p(Y)$$

$$= \prod_{t=1}^{T} \pi_{id}^t(A^t|\underline{A}'^t, G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^{t-1})\pi_{\alpha}^t(A'^t|G_{\underline{A}'^{t-1}}(X_{(1)}), \underline{A}'^{t-1})$$

$$\cdot \prod_{t=0}^{T} p(G_{A^t}(X_{(1)})|G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^t, Y)p(Y)$$

$$= q'(A', A, G_A(X_{(1)}), Y)$$

In order for this expression to hold, π_{id}^t must be restricted to have support only on $\tilde{\mathcal{A}}_{adm}$. This leads to the following restriction:

$$\pi_{id}^t(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \underbrace{\mathbb{I}(A^t \in \tilde{\mathcal{A}}_{adm}^t(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}))}_{\text{support restriction}} f_{id}^t(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}).$$

where f_{id}^t is an arbitrary function that ensures that π_{id}^t is a valid density. This concludes the proof of Theorem 7.

Bellman equation

Equations 42 and 43 correspond to the two parts of the semi-offline RL version of the Bellman equation:

$$\mathbb{E}\left[C'_{(\overline{\pi}_{\alpha}^{t+1})}\middle|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi\right] = \sum_{X^t} \mathbb{E}\left[C'_{(\overline{\pi}_{\alpha}^{t+1})}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^{t-1}, a^t, \Xi\right] p(X^t|\underline{X}^{t-1}, \underline{A}^{t-1}, a^t, \Xi)$$

$$= \sum_{G_{A^t}(X_{(1)})} \mathbb{E}\left[C'_{(\overline{\pi}_{\alpha}^{t+1})}\middle|\underline{A}'^t, G_{\underline{A}^t}(X_{(1)}), \underline{A}^{t-1}, a^t, \Xi\right] p(G_{A^t}(X_{(1)})|G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^{t-1}, a^t, \Xi)$$

for any $a^t \in \mathcal{A}_{adm}^t(\underline{X}^{t-1}, \underline{A}^{t-1}, A'^t)$

$$\mathbb{E}\left[C'_{(\overline{\pi}_{\alpha}^{t+1})}\middle|\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi\right] = \sum_{A'^{t+1}} \mathbb{E}\left[C'_{(\overline{\pi}_{\alpha}^{t+2})}\middle|\underline{A}'^{t+1}, \underline{X}^t, \underline{A}^t, \Xi\right] \pi_{\alpha}^{t+1}(A'^{t+1}|\underline{X}'^t, \underline{A}'^t)$$

$$= \sum_{A'^{t+1}} \mathbb{E}\left[C'_{(\overline{\pi}_{\alpha}^{t+2})}\middle|\underline{A}'^{t+1}, G_{\underline{A}^t}(X_{(1)}), \underline{A}^t, \Xi\right] \pi_{\alpha}^{t+1}(A'^{t+1}|G_{\underline{A}'^{t-1}}(X_{(1)}), \underline{A}'^t)$$

The factorization holds under local positivity (if $\mathcal{A}_{adm}^t \neq \emptyset$ exists). Furthermore, the individual terms are identified if regional positivity holds, which concludes the proof of Theorem 8. ■

Appendix H. Proof of Corollary 10

In this appendix, we prove Corollary 10, stating identification under the maximal global positivity assumption. We repeat it here for ease of reference:

Corollary 10. *(Identification of J for the semi-offline RL view under maximal global positivity). The reformulated AFAPE problem of estimating J under the semi-offline RL view (Eq. 17) is under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and Assumption 6.4 (maximum global positivity) identified by Eqs. 19 and 20 where*

$$\pi_{id}^t(A^t|\underline{A}'^{t-1}, A'^t = a'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \\ = \mathbb{I}(A^t \geq a'^t)\pi_\beta^t(A^t|\underline{X}^{t-1}, \underline{A}^{t-1}) f_{id}^t(\underline{A}'^{t-1}, A'^t = a'^t, \underline{X}^{t-1}, \underline{A}^{t-1})$$

for any function f_{id}^t s.t. π_{id}^t is a valid density. This holds in particular for the choice of a truncated π_β:

$$\pi_{id}^t(A^t|\underline{A}'^{t-1}, A'^t = a'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \pi_\beta^t(A^t|A^t \geq a'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) = \\ = \frac{\mathbb{I}(A^t \geq a'^t)\pi_\beta^t(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})}{\pi_\beta^t(A^t \geq a'^t|\underline{X}^{t-1}, \underline{A}^{t-1})}.$$

Proof Under the maximal global positivity assumption, we have $\tilde{\mathcal{A}}_{adm}^t(\underline{a}'^t, \underline{x}^{t-1}, \underline{a}^{t-1}) = \mathcal{A}_{adm}^t(\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t)$. We can now insert this assumption into Eq. 21 which states the identification of J under global positivity:

$$\begin{aligned}\pi_{id}^t(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) &= \mathbb{I}(A^t \in \tilde{\mathcal{A}}_{adm}^t(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})) f_{id}^t(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) \\ &= \mathbb{I}(A^t \in \mathcal{A}_{adm}^t(\underline{X}^{t-1}, \underline{A}^{t-1}, A'^t)) f_{id}^t(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}) \\ &= \mathbb{I}(A^t \geq a'^t)\pi_\beta^t(A^t|\underline{X}^{t-1}, \underline{A}^{t-1}) f_{id}^{*t}(\underline{X}'^{t-1}, \underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})\end{aligned}$$

where we let f_{id}^{*t} denote another arbitrary function that ensures that π_{id}^t is a valid density. This concludes the proof for Corollary 10.

■

Appendix I. Comparison of the Semi-offline RL IPW Estimator with Related Methods

In this appendix, we demonstrate that our proposed IPW estimator $J_{IPW\text{-}Semi}$ is a more general version of an adapted version of the IPW estimator introduced by Caniglia et al. (2019). We refer to this estimator as $\hat{J}_{IPW\text{-}Cen}$ since it is derived from a censoring viewpoint. While $\hat{J}_{IPW\text{-}Cen}$ was developed for a scenario where both feature acquisition decisions and treatment decisions are made by the agent, it can be adapted to the AFA setting. However, $\hat{J}_{IPW\text{-}Cen}$ is only applicable to simpler settings with one acquisition option per time-point

($A^t \in \{0,1\}$). $\hat{J}_{IPW\text{-}Cen}$ is also only consistent if the maximal global positivity assumption holds, as will be shown.

Caniglia et al. (2019) derived $J_{IPW\text{-}Cen}$ under the NDE and NUC assumptions. Instead of using a semi-offline sampling policy that avoids the acquisition of non-available features as proposed in this paper, Caniglia et al. (2019) simply sample from π_α, even without knowledge about $X_{(1)}$. As the feature revelation is not possible if a non-available feature is acquired, they treat the resulting trajectory as censored. Known missing data methods are then applied to adjust for this censoring. Hence, in the wording of this paper, we would describe this viewpoint as an *online RL + censoring viewpoint*.

Adapted to the AFA setting under the consideration of deterministic AFA policies π_α, $\hat{J}_{IPW\text{-}Cen}$ becomes:

$$\hat{J}_{IPW\text{-}Cen} = \hat{\mathbb{E}}_{n,uncen}[\hat{\rho}^T_{Cen} C_{(\pi_\alpha)}] \text{ where } \hat{\rho}^T_{Cen} = \prod_{t=1}^{T} \left(\frac{\mathbb{I}(A^t = 1)}{\hat{\pi}^t_\beta(A^t = 1 | \underline{X}^{t-1}, \underline{A}^{t-1})} \right)^{A^t_{(\pi_\alpha)}} \tag{44}$$

where $\hat{\mathbb{E}}_{n',uncen}[.]$ denotes the empirical average over the uncensored data points which have the known deterministic counterfactuals $A^t_{(\pi_\alpha)}$,$X^t_{(\pi_\alpha)}$ and $C_{(\pi_\alpha)}$.

Since $A^t \in \{0,1\}$, it can be observed that the propensity score for a specific time-point t only appears in the factorization if the corresponding action is $A^t_{(\pi_\alpha)} = 1$. An "acquire nothing" AFA policy (where $\pi^t_\alpha(A^t = 0 | \underline{X}^{t-1}, \underline{A}^{t-1}) = 1 \; \forall t$) would thus require no adjustment ($\rho^T_{Cen} = 1$). In their example, this estimator achieved a 50-fold increase in data efficiency compared to the standard offline RL IPW estimator (Caniglia et al., 2019).

We establish the equivalence of our estimator $\hat{J}_{IPW\text{-}Semi}$ and $\hat{J}_{IPW\text{-}Cen}$ in the following proposition:

Proposition 20. *(Equivalance of $\hat{J}_{IPW\text{-}Cen}$ and $\hat{J}_{IPW\text{-}Semi}$). The estimators $\hat{J}_{IPW\text{-}Cen}$ and $\hat{J}_{IPW\text{-}Semi}$ are equivalent for AFA settings with one action option per time-point, deterministic AFA policies π_α, the maximal global positivity assumption (Assumption 6.4), and a simulation policy $\pi_{sim} = \pi_\alpha$.*

Proof Firstly, we clarify the blocking operation (from Definition 2) for this setting:

$$\pi'_{sim}(A'^t = a'^t | \underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t = a^t) = \begin{cases} 1, & \text{if } a'^t = 0 \;\&\; a^t = 0 \\ 0, & \text{if } a'^t = 1 \;\&\; a^t = 0 \\ \pi^t_\alpha(A'^t = a'^t | \underline{X}'^{t-1}, \underline{A}'^{t-1}), & \text{if } a^t = 1. \end{cases}$$

The inverse probability weights of $\hat{J}_{IPW\text{-}Semi}$ become:

$$\begin{aligned} \rho^T_{Semi} &= \prod_{t=1}^{T} \frac{\pi^t_\alpha(A'^t = a'^t | \underline{X}'^{t-1}, \underline{A}'^{t-1})}{\pi'^t_{sim}(A'^t = a'^t | \underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t = a^t)} \frac{\mathbb{I}(A^t \geq a'^t)}{\pi^t_\beta(A^t \geq a'^t | \underline{X}^{t-1}, \underline{A}^{t-1})} \\ &\overset{*_1}{=} \prod_{t=1}^{T} \left(\pi^t_\alpha(A'^t = a'^t | \underline{X}'^{t-1}, \underline{A}'^{t-1})\right)^{1-a^t} \left(\frac{\mathbb{I}(A^t = 1)}{\pi^t_\beta(A^t = 1 | \underline{X}^{t-1}, \underline{A}^{t-1})} \right)^{a'^t} \\ &\overset{*_2}{=} \prod_{t=1}^{T} \mathbb{I}(A'^t = A^t_{(\pi_\alpha)}) \left(\frac{\mathbb{I}(A^t = 1)}{\pi^t_\beta(A^t = 1 | \underline{X}^{t-1}, \underline{A}^{t-1})} \right)^{A^t_{(\pi_\alpha)}} \end{aligned}$$

where we used in $*1$) the above definition of π'_{sim} and that $\mathbb{I}(A^t \geq 0) = 1 = \pi_\beta(A^t \geq 0|\underline{X}^{t-1}, \underline{A}^{t-1})$. In $*2$), we see that the first term corresponds to whether π_α could be applied without running into censoring. It thus gives 0 weights to all datapoints where blocking occurred (i.e., which are censored under the online RL + censoring viewpoint). The second term then corresponds to the same weights as ρ^T_{Cen}, which concludes the proof for Proposition 20. ■

We have demonstrated that, although the two estimators are derived from different concepts (online RL with censoring vs semi-offline RL), they are equal in this specific AFA setting of one action option per time-step, deterministic policies and under the maximal global positivity assumption. However, the key distinction lies in the generality of our estimator. Unlike $\hat{J}_{IPW\text{-}Cen}$, which is limited to the described setting, we developed an IPW estimator that can be applied for multiple acquisition options (i.e., higher dimensional A^t) under the weaker global positivity assumption and in a modified version for static features settings as we show in our companion paper (von Kleist et al., 2023). It can further be combined with a Q-model to build the DRL estimator.

Appendix J. Proof of Theorem 13

In this appendix, we proof Theorem 13, which we repeat here:

Theorem 13. *(Double robustness of $\hat{J}_{DRL\text{-}Semi}$). The estimator $\hat{J}_{DRL\text{-}Semi}$ is doubly robust, in the sense that it is consistent if either the Q-function $\hat{Q}_{Semi}$ or the propensity score model $\hat{\pi}_\beta$ is correctly specified.*

Proof To prove the double robustness property of the semi-offline RL version of the DRL estimator (i.e., Theorem 13), we decompose $\hat{J}_{DRL\text{-}Semi}$ in two different ways:

Scenario 1: If $\hat{\pi}_\beta$ is correctly specified, we find

$$\begin{aligned}\hat{J}_{DRL\text{-}Semi} &= \hat{\mathbb{E}}_n\left[\hat{\mathbb{E}}_{n'}\left[\rho^T_{Semi}C' + \sum_{t=1}^{T}\left(-\hat{\rho}^t_{Semi}\hat{Q}^t_{Semi} + \rho^{t-1}_{Semi}\hat{V}^{t-1}_{Semi}\right)\Big|A, X, Y\right]\right] \\ &= \underbrace{\mathbb{E}[\rho^T_{Semi}C']}_{=J} + \sum_{t=1}^{T}\underbrace{\mathbb{E}\left[-\rho^t_{Semi}\hat{Q}^t_{Semi} + \rho^{t-1}_{Semi}\hat{V}^{t-1}_{Semi}\right]}_{=0},\end{aligned}$$

where the first term is just the IPW estimator. As $\hat{\pi}_\beta = \pi_\beta$ is correctly specified, the IPW estimator consistently estimates J. The fact that the second term equals 0 is shown in the

following:

$$
\begin{aligned}
&\mathbb{E}\left[-\rho^t_{Semi}\hat{Q}^t_{Semi} + \rho^{t-1}_{Semi}\hat{V}^{t-1}_{Semi}\right] = \\
&\quad \overset{*1}{=} \mathbb{E}\Bigg[\rho^{\boldsymbol{t-1}}_{Semi}\Bigg(-\frac{\boldsymbol{\pi^t_\alpha}(\boldsymbol{A't}|\underline{\boldsymbol{X}}'^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}'^{\boldsymbol{t-1}})}{\boldsymbol{\pi'^t_{sim}}(\boldsymbol{A't}|\underline{\boldsymbol{X}}'^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}'^{\boldsymbol{t-1}}, \boldsymbol{A^t})}\frac{\boldsymbol{\pi^t_{id}}(\boldsymbol{A^t}|\underline{\boldsymbol{A}}'^{\boldsymbol{t}}, \underline{\boldsymbol{X}}^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}^{\boldsymbol{t-1}})}{\boldsymbol{\pi^t_\beta}(\boldsymbol{A^t}|\underline{\boldsymbol{X}}^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}^{\boldsymbol{t-1}})}\hat{Q}^t_{Semi} \\
&\qquad\qquad + \sum_{\boldsymbol{A'^t}} \boldsymbol{\pi^t_\alpha}(\boldsymbol{A't}|\underline{\boldsymbol{X}}'^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}'^{\boldsymbol{t-1}})\hat{\boldsymbol{Q}}^{\boldsymbol{t}}_{\boldsymbol{Semi}}\Bigg)\Bigg] \\
&\quad \overset{*2}{=} \mathbb{E}\Bigg[\rho^{t-1}_{Semi}\Bigg(-\sum_{\boldsymbol{A'^t}, \boldsymbol{A^t}} \boldsymbol{\pi'^t_{sim}}(\boldsymbol{A't}|\underline{\boldsymbol{X}}'^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}'^{\boldsymbol{t-1}}, \boldsymbol{A^t})\boldsymbol{\pi^t_\beta}(\boldsymbol{A^t}|\underline{\boldsymbol{X}}^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}^{\boldsymbol{t-1}})\frac{\pi^t_\alpha(A't|\underline{X}'^{t-1}, \underline{A}'^{t-1})}{\pi'^t_{sim}(A't|\underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t)}\cdot \\
&\qquad\qquad \cdot \frac{\pi^t_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})}{\pi^t_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})}\hat{Q}^t_{Semi} + \sum_{A'^t}\pi^t_\alpha(A't|\underline{X}'^{t-1}, \underline{A}'^{t-1})\hat{Q}^t_{Semi}\Bigg)\Bigg] \\
&\quad \overset{*3}{=} \mathbb{E}\Bigg[\rho^{t-1}_{Semi}\Bigg(-\sum_{\boldsymbol{A'^t}} \boldsymbol{\pi^t_\alpha}(\boldsymbol{A't}|\underline{\boldsymbol{X}}'^{\boldsymbol{t-1}}, \underline{\boldsymbol{A}}'^{\boldsymbol{t-1}})\hat{Q}^t_{Semi} + \sum_{A'^t}\pi^t_\alpha(A't|\underline{X}'^{t-1}, \underline{A}'^{t-1})\hat{Q}^t_{Semi}\Bigg)\Bigg] = 0
\end{aligned}
$$

with the following explanations:

- $*1$): We use the relationship $\hat{V}^{t-1}_{Semi} = \mathbb{E}_{\pi_\alpha}[\hat{Q}^t_{Semi}]$ and the decomposition of ρ_{Semi}.
- $*2$): We use the fact that one can pull the expected value with respect to $\pi'^t_{sim}(A't|\underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t)\pi^t_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})$ inside.
- $*3$): We use the fact that $\hat{Q}^t_{Semi}$ is independent of $\pi^t_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})$ as long as it fulfills the positivity assumption.

Scenario 2: If $\hat{Q}_{Semi}$ is correctly specified, we find

$$
\begin{aligned}
\hat{J}_{DRL\text{-}Semi} &= \mathbb{E}[V^0_{Semi}] + \mathbb{E}\left[\hat{\rho}^T_{Semi}\left(C' - Q^T_{Semi}\right)\right] + \mathbb{E}\left[\sum_{t=1}^{T-1}\hat{\rho}^t_{Semi}\left(-Q^t_{Semi} + V^t_{Semi}\right)\right] \\
&= \underbrace{\mathbb{E}\left[V^0_{Semi}\right]}_{=J} + \mathbb{E}\left[\hat{\rho}^T_{Semi}\underbrace{\left(f_C(A', X', Y) - Q^T_{Semi}\right)}_{=0}\right] \\
&\quad + \mathbb{E}\Bigg[\sum_{t=1}^{T-1}\hat{\rho}^t_{Semi}\underbrace{\left(-Q^t_{Semi} + \sum_{X^t} V^t_{Semi}\, p(X^t|\underline{X}^{t-1}, \underline{A}^t, \Xi)\right)}_{=0}\Bigg]
\end{aligned}
$$

where $\mathbb{E}\left[V^0_{Semi}\right]$ corresponds to the DM estimator which is consistent if $\hat{Q}_{Semi} = Q_{Semi}$ is correctly specified. For the last term, we used that we can pull in the expected value with respect to the conditional distributions of X^t. The resulting term equals the first part of the semi-offline RL version of Bellman's equation. This concludes the proof of Theorem 13. ■

Appendix K. Estimation of Other Target Parameters from the Semi-offline RL View

In this appendix, we extend the target parameter to include time-dependent costs C^t (s.t. $C = \overline{C}^1$). The newly defined target parameter becomes $J = \mathbb{E}\left[\sum_{t=1}^{T} C^t_{(\pi_\alpha)}\right]$. In particular, these costs may include acquisition costs or misclassification costs for predictions at each time step. The acquisition costs are given by the known deterministic $f^t_{C_a}(A^t)$. When considering misclassification costs, we assume a per-step label Y^t to be available at each time step (s.t. $Y = \overline{Y}^1$). The per-step misclassification costs can be computed by:

$$f^t_{C_{mc}}(Y^{*t}, Y^t) = f^t_{C_{mc}}(f_{cl}(\underline{A}^t, \underline{X}^t), Y^t).$$

We combine both costs such that the target parameter is redefined to be:

$$\begin{aligned} J &= \mathbb{E}\left[\sum_{t=1}^{T}\left(C^t_{a,(\pi_\alpha)} + C^t_{mc,(\pi_\alpha)}\right)\right] = \mathbb{E}\left[\sum_{t=1}^{T}\left(f^t_{C_a}(A^t_{(\pi_\alpha)}) + f^t_{C_{mc}}(\underline{A}^t_{(\pi_\alpha)}, \underline{X}^t_{(\pi_\alpha)}, Y^t)\right)\right] \\ &\equiv \mathbb{E}\left[\sum_{t=1}^{T} f^t_C(\underline{A}^t_{(\pi_\alpha)}, \underline{X}^t_{(\pi_\alpha)}, Y^t)\right] = \mathbb{E}\left[\sum_{t=1}^{T} C^t_{(\pi_\alpha)}\right]. \end{aligned}$$

The reformulation, identification, and estimation steps from the semi-offline RL view can be extended to per-step costs. For this setting, we provide corollaries of the identification and estimation theorems from the main body. We do not provide additional proofs, as the extensions are straightforward.

K.1 Identification

We start with a corollary that extends Theorem 7 for the per-step costs setting.

Corollary 21. *(Identification of J (for per-step costs) for the semi-offline RL view). The reformulated AFAPE problem of estimating J (for per-step costs) under the semi-offline RL view (Eq. 17) is under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and Assumption 6.3 (global positivity) identified by*

$$J = \mathbb{E}_{p'}\left[\sum_{t=1}^{T} C'_{(\pi_\alpha)}\right] = \sum_{A', A, G_A(X_{(1)}), Y} \sum_{t=1}^{T} f^t_C(\underline{A}'^t, \underline{X}'^t, Y^t) q'(A', A, X, Y) \tag{45}$$

where q' is given by Eq. 20.

Next, we continue with a corollary that extends Theorem 8 for the per-step costs setting.

Corollary 22. *(Bellman equation for semi-offline RL (for per-step costs)). The semi-offline RL view admits under Assumption 1 (no measurement noise), Assumption 2 (consistency), Assumption 3 (no interference), Assumption 4 (NDE), Assumption 5 (NUC) and*

the local positivity assumption at datapoint $\underline{x}^{t-1}, \underline{a}^{t-1}, a'^t$ *(from Definition 4), the following semi-offline RL version of the Bellman equation for per-step costs:*

$$Q_{Semi}(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi) = \sum_{X^t} V_{Semi}(\underline{A}'^t, \underline{X}^t, \underline{A}^{t-1}, A^t = a^t, \Xi) p(X^t | \underline{X}^{t-1}, \underline{A}^{t-1}, A^t = a^t, \Xi) \tag{46}$$

for any $a^t \in \mathcal{A}^t_{adm}(\underline{X}^{t-1}, \underline{A}^{t-1}, A'^t)$

$$\begin{aligned} V_{Semi}(\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi) = &\sum_{Y^t} f^t_C(\underline{X}'^t, \underline{A}'^t, Y^t) p(Y^t | \underline{X}^t, \underline{A}^t, \Xi) \\ &+ \sum_{A'^{t+1}} Q_{Semi}(\underline{A}'^{t+1}, \underline{X}^t, \underline{A}^t, \Xi) \pi^{t+1}_\alpha(A'^{t+1} | \underline{X}'^t, \underline{A}'^t) \end{aligned} \tag{47}$$

with semi-offline RL versions of the state-action value function Q_{Semi} *and state value function* V_{Semi}*:*

$$Q^t_{Semi} \equiv Q_{Semi}(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi) \equiv \mathbb{E}_{p'} \left[\sum_{\tau=t}^{T} C'^\tau_{(\overline{\pi}^{t+1}_\alpha)} \middle| \underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}, \Xi \right]$$

$$V^t_{Semi} \equiv V_{Semi}(\underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi) \equiv \mathbb{E}_{p'} \left[\sum_{\tau=t}^{T} C'^\tau_{(\overline{\pi}^{t+1}_\alpha)} \middle| \underline{A}'^t, \underline{X}^t, \underline{A}^t, \Xi \right].$$

Furthermore, Q^t_{Semi} *and* V^t_{Semi} *are identified if the regional positivity assumption holds at* $\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1}$ *and* $a^t \in \tilde{\mathcal{A}}^t_{adm}(\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})$*.*

K.2 Estimation

The estimation formulas can be extended to the per-step setting as follows:

1) Inverse probability weighting (IPW):
The target cost (for per-step costs) that is estimated by the semi-offline IPW estimator is

$$\hat{J}_{IPW\text{-}Semi} = \hat{\mathbb{E}}_n \left[\hat{\mathbb{E}}_{n'} \left[\sum_{t=1}^{T} \hat{\rho}^t_{Semi} \, C'^t \middle| A, X, Y \right] \right],$$

with the same options for ρ^t_{Semi} as in the setting described in the main body.

2) Direct method (DM):
The target cost (for per-step costs) that is estimated by the semi-offline DM estimator is

$$\hat{J}_{DM\text{-}Semi} = \hat{\mathbb{E}}_{n'}[\hat{V}^0_{Semi}]$$

with the adapted per-step cost version of V_{Semi} from Corollary 22.

3) Double reinforcement learning (DRL):
The target cost (for per-step costs) that is estimated by the semi-offline DRL estimator is

$$\hat{J}_{DRL\text{-}Semi} = \hat{\mathbb{E}}_n \left[\hat{\mathbb{E}}_{n'} \left[\sum_{t=1}^{T} \left(\hat{\rho}^t_{Semi} C'^t - \hat{\rho}^t_{Semi} \hat{Q}^t_{Semi} + \hat{\rho}^{t-1}_{Semi} \hat{V}^{t-1}_{Semi} \right) \middle| A, X, Y \right] \right].$$

with the adapted per-step cost version of V_{Semi} and Q_{Semi}.

Appendix L. Derivation of the Missing Data Semiparametric Theory Approach

In this appendix, we show why Eq. 35 and Eq. 36 constitute valid choices for an element of the IPW space Λ_{IPW} and the augmentation space Λ_2 but rely on the positivity assumption for missing data (Assumption 6.2). This is a simplified derivation from Tsiatis (2006).

IPW space:
Clearly, $h_{Miss} = \rho_{Miss}\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)}, Y] - J$ is a function of the observed data, since ρ_{Miss} is non-zero only for complete cases. It thus remains to be shown that

$$\mathbb{E}\left[h_{Semi}(A, G_A(X_{(1)}), Y)|X_{(1)}, Y\right] = \varphi^F(X_{(1)}, Y).$$

This is shown in the following:

$$\begin{aligned}
\mathbb{E}[h_{Semi}(A, G_A(X_{(1)}), Y)|X_{(1)}, Y] &= \mathbb{E}\left[\rho_{Miss}\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)}, Y]\Big|X_{(1)}, Y\right] - J \\
&= \mathbb{E}\left[\prod_{t=1}^{T} \frac{\mathbb{I}(A^t = \vec{1})}{\pi_\beta^t(A^t = \vec{1}|\underline{X}_{(1)}^{t-1}, \underline{A}^{t-1} = \vec{1})}\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)}, Y]\Big|X_{(1)}, Y\right] - J \\
&= \mathbb{E}\left[\prod_{t=1}^{T} \frac{\mathbb{E}[\mathbb{I}(A^t = \vec{1})|\underline{X}_{(1)}^{t-1}, \underline{A}^{t-1} = \vec{1})]}{\pi_\beta^t(A^t = \vec{1}|\underline{X}_{(1)}^{t-1}, \underline{A}^{t-1} = \vec{1})}\mathbb{E}[C_{(\pi_\alpha)}|X_{(1)}, Y]\Big|X_{(1)}, Y\right] - J \\
&= \mathbb{E}\left[1 \cdot \mathbb{E}\left[C_{(\pi_\alpha)}\Big|X_{(1)}, Y\right]\Big|X_{(1)}, Y\right] - J \\
&= \mathbb{E}\left[C_{(\pi_\alpha)}\Big|X_{(1)}, Y\right] - J = \varphi^F(X_{(1)}, Y).
\end{aligned}$$

Augmentation space Λ_2:
To derive Λ_2 under the missing data view, one redefines any function $b(A, G_A(X_{(1)}), Y)$ using the fact that A is a categorical variable:

$$b(A, G_A(X_{(1)}), Y) = \sum_{a \in \mathcal{A}} \mathbb{I}(A = a) b_a(G_a(X_{(1)}), Y) \tag{48}$$

where $b_a(G_a(X_{(1)}), Y)$ is any mean zero, finite variance function of $G_a(X_{(1)}), Y$. This allows the enforcement of the zero conditional mean condition that defines Λ_2:

$$\mathbb{E}[b(A, G_A(X_{(1)}), Y)|X_{(1)}, Y] = \sum_{a \in \mathcal{A}} p(A = a|G_a(X_{(1)}), Y) b_a(G_a(X_{(1)}), Y) = 0.$$

Under the missing data positivity assumption, one can now solve for $b_{\vec{1}}$:

$$b_{\vec{1}}(G_{\vec{1}}(X_{(1)}), Y) = -\frac{1}{p(A = \vec{1}|G_{\vec{1}}(X_{(1)}), Y)} \sum_{a \in \mathcal{A} \setminus \vec{1}} p(A = a|G_a(X_{(1)}), Y) b_a(G_a(X_{(1)}), Y)$$

Substituting $b_{\vec{1}}(G_{\vec{1}}(X_{(1)}), Y)$ into Eq. 48 and applying the known factorization of the propensity score model for our AFA setting gives the desired space Λ_2 consisting of all

$$b(A, G_A(X_{(1)}), Y) = \sum_{a \in \mathcal{A} \setminus \vec{1}} \left[\mathbb{I}(A = a) - \frac{\mathbb{I}(A = \vec{1}) p(A = a | G_a(X_{(1)}), Y)}{p(A = \vec{1} | G_{\vec{1}}(X_{(1)}), Y)} \right] b_a(G_a(X_{(1)}), Y)$$

$$= \sum_{a \in \mathcal{A} \setminus \vec{1}} \left[\mathbb{I}(A = a) - \prod_{t=1}^{T} \frac{\mathbb{I}(A^t = \vec{1}) \pi_\beta^t(A^t = a^t | G_{\underline{a}^{t-1}}(X_{(1)}), \underline{a}^{t-1})}{\pi_\beta^t(A^t = \vec{1} | G_{\underline{a}^{t-1}}(X_{(1)}), \underline{a}^{t-1})} \right] b_a(G_a(X_{(1)}), Y).$$

Appendix M. Proof of Lemma 15 and Lemma 16

In this appendix, we prove Lemma 15 and Lemma 16 which establish the equivalence of the missing data and offline RL semiparametric theory approaches to AFAPE under the NUC and NDE assumptions. We start with Lemma 15.

For ease of reference, we repeat it here:

Lemma 15. *(Relating the offline RL IPW estimator to the IPW space). The functional $h_{Off} \equiv \rho_{Off} C - J$, based on the IPW estimator from the offline RL view, is a valid element of the IPW space: $h_{Off} \in \Lambda_{IPW}$.*

Proof We need to show that $h_{Off} = h_{Off}(A, G_A(X_{(1)}), Y) = \rho_{Off}^T C - J \in \Lambda_{IPW}$. Clearly, h_{Off} is a function of the observed data. It thus remains to be shown that

$$\mathbb{E}\left[h_{Off}(A, G_A(X_{(1)}), Y) | X_{(1)}, Y\right] = \varphi^F(X_{(1)}, Y).$$

This is shown in the following:

$$\mathbb{E}\left[h_{Off}(A, G_A(X_{(1)}), Y) | X_{(1)}, Y\right] = \mathbb{E}\left[\rho_{Off}^T C - J \middle| X_{(1)}, Y\right]$$

$$= \mathbb{E}\left[\prod_{t=1}^{T} \frac{\pi_\alpha^t(A^t | G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^{t-1})}{\pi_\beta^t(A^t | G_{\underline{A}^{t-1}}(X_{(1)}), \underline{A}^{t-1})} f_C(Y, G_A(X_{(1)}), A) \middle| X_{(1)}, Y\right] - J$$

$$= \sum_{a \in \mathcal{A}} \prod_{t=1}^{T} \frac{\pi_\alpha^t(a^t | G_{\underline{a}^{t-1}}(X_{(1)}), \underline{a}^{t-1})}{\pi_\beta^t(a^t | G_{\underline{a}^{t-1}}(X_{(1)}), \underline{a}^{t-1})} f_C(Y, G_a(X_{(1)}), a) \pi_\beta^t(a^t | G_{\underline{a}^{t-1}}(X_{(1)}), \underline{a}^{t-1}) - J$$

$$= \sum_{a \in \mathcal{A}} \prod_{t=1}^{T} \pi_\alpha^t(a^t | G_{\underline{a}^{t-1}}(X_{(1)}), \underline{a}^{t-1}) f_C(Y, G_a(X_{(1)}), a) - J = \varphi^F(X_{(1)}, Y)$$

which completes the proof. ■

Next, we prove Lemma 16. We also repeat it here:

Lemma 16. *(Λ_* is equal to the augmentation space). The augmentation space Λ_2 is equal to Λ_*.*

We reuse properties about Λ_* shown by Liu et al. (2021) and do not repeat the corresponding proofs for these properties as they involve cumbersome notation.

Proof To demonstrate that $\Lambda_* = \Lambda_2$, we first introduce a new space, denoted as Λ_*^{AF}, containing all functions of $b(A, X_{(1)}, Y)$ for which $\mathbb{E}[b(A, X_{(1)}, Y)|X_{(1)}, Y] = 0$ holds:

$$\Lambda_*^{AF} \equiv \left\{ b^t(\underline{A}^{t-1}, X_{(1)}, Y) \left(\frac{A^t}{\pi_\beta^t} - 1 \right) : b^t(\underline{A}^{t-1}, X_{(1)}, Y) \in \mathcal{H}; t \in \{1, ..., T\} \right\}.$$

The space Λ_*^{AF} indeed includes all functions $b(A, X_{(1)}, Y)$ with mean zero and finite variance for which $\mathbb{E}[b(A, X_{(1)}, Y)|X_{(1)}, Y] = 0$ holds. Specifically, the elements $b^t(\underline{A}^{t-1}, X_{(1)}, Y)$ can represent any function of $\underline{A}^{t-1}, X_{(1)}, Y$. Since A^t is binary (taking values 0 or 1), adding a term $\left(\frac{A^t}{c(\underline{A}^{t-1}, X_{(1)}, Y)} - 1 \right)$ (for some $c(A, X_{(1)}, Y)$) generalizes the space to contain all functions of $A^t, X_{(1)}, Y$. The specific choice $c(\underline{A}^{t-1}, X_{(1)}, Y) = \pi_\beta^t$ is enforced by the condition $\mathbb{E}[b(A, X_{(1)}, Y)|X_{(1)}, Y] = 0$:

$$\begin{aligned}
\mathbb{E}\left[b^t(\underline{A}^{t-1}, X_{(1)}, Y) \left(\frac{A^t}{\pi_\beta^t} - 1 \right) |X_{(1)}, Y \right] &= \\
&= \mathbb{E}\left[b^t(\underline{A}^{t-1}, X_{(1)}, Y) \mathbb{E}\left[\left(\frac{A^t}{\pi_\beta^t} - 1 \right) |X_{(1)}, Y, \underline{A}^{t-1} \right] |X_{(1)}, Y \right] \\
&= \mathbb{E}\left[b^t(\underline{A}^{t-1}, X_{(1)}, Y) \underbrace{\left(\frac{\pi_\beta^t}{\pi_\beta^t} - 1 \right)}_{0} |X_{(1)}, Y \right] = 0.
\end{aligned}$$

Now, in order to find Λ_2, we must find the subspace of Λ_*^{AF}, which contains only the functions of the observed data. This is exactly Λ_* as was shown by Liu et al. (2021) (in the proof of Remark 13). In fact, they showed that there aren't observed data elements that are in Λ_*^{AF}, but not in Λ_*. They also showed that Λ_* only contains observed data elements. This concludes the proof. ■

Appendix N. Proof of Lemma 17

In this appendix, we prove Lemma 17. The proof follows a similar approach as for Lemma 15 for the offline RL IPW estimator. We repeat the lemma here for ease of reference.

Lemma 17. *(Relating the semi-offline RL IPW estimator to the IPW space).* $h_{Semi} \equiv h_{Semi}(A, G_A(X_{(1)}), Y) = \hat{\mathbb{E}}_{n'}[\rho_{Semi}^T C' | A, G_A(X_{(1)}), Y] - J$ *is an element of the IPW space* Λ_{IPW}.

Proof We need to show that $h_{Semi} = h_{Semi}(A, G_A(X_{(1)}), Y) = \hat{\mathbb{E}}_{n'}[\rho_{Semi}^T C' | A, G_A(X_{(1)}), Y] - J \in \Lambda_{IPW}$. Clearly, h_{Semi} is a function of the observed data. It thus remains to be shown that

$$\mathbb{E}\left[h_{Semi}(A, G_A(X_{(1)}), Y) | X_{(1)}, Y \right] = \varphi^F(X_{(1)}, Y).$$

This is shown in the following:

$$
\begin{aligned}
&\mathbb{E}\left[h_{Semi}(A, G_A(X_{(1)}), Y)\middle| X_{(1)}, Y\right] = \mathbb{E}\left[\mathbb{E}\left[\rho^T_{Semi} f_C(A', G_{A'}(X_{(1)}), Y)|A, G_A(X_{(1)}), Y\right] - J\middle| X_{(1)}, Y\right] \\
&= \mathbb{E}\left[\prod_{t=1}^{T} \frac{\pi^t_\alpha(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1})}{\pi'^t_{sim}(A'^t|\underline{X}'^{t-1}, \underline{A}'^{t-1}, A^t)} \frac{\pi^t_{id}(A^t|\underline{A}'^t, \underline{X}^{t-1}, \underline{A}^{t-1})}{\pi^t_\beta(A^t|\underline{X}^{t-1}, \underline{A}^{t-1})} f_C(A', G_{A'}(X_{(1)}), Y)\middle| X_{(1)}, Y\right] - J \\
&= \sum_{a\in\mathcal{A}} \sum_{a'\in\mathcal{A}} \prod_{t=1}^{T} \frac{\pi^t_\alpha(a'^t|\underline{X}'^{t-1}, \underline{a}'^{t-1})}{\pi'^t_{sim}(a'^t|\underline{X}'^{t-1}, \underline{a}'^{t-1}, a^t)} \frac{\pi^t_{id}(a^t|\underline{a}'^t, \underline{X}^{t-1}, \underline{a}^{t-1})}{\pi^t_\beta(a^t|\underline{X}^{t-1}, \underline{a}^{t-1})} f_C(a', G_{a'}(X_{(1)}), Y) \\
&\cdot \pi'^t_{sim}(a'^t|\underline{X}'^{t-1}, \underline{a}'^{t-1}, a^t)\pi^t_\beta(a^t|\underline{X}^{t-1}, \underline{a}^{t-1}) - J \\
&= \sum_{a\in\mathcal{A}} \sum_{a'\in\mathcal{A}} \prod_{t=1}^{T} \pi^t_\alpha(a'^t|\underline{X}'^{t-1}, \underline{a}'^{t-1})\pi^t_{id}(a^t|\underline{a}'^t, \underline{X}^{t-1}, \underline{a}^{t-1}) f_C(a', G_{a'}(X_{(1)}), Y) - J \\
&= \sum_{a'\in\mathcal{A}} \prod_{t=1}^{T} \pi^t_\alpha(a'^t|\underline{X}'^{t-1}, \underline{a}'^{t-1}) f_C(a', G_{a'}(X_{(1)}), Y) - J \\
&= \varphi^F(X_{(1)}, Y)
\end{aligned}
$$

which completes the proof. ■

Appendix O. Derivation of the Influence Function for Semi-offline RL

In this Appendix, we provide the complete derivation of the projection of h_{Semi} onto $\Lambda_{2,Semi}(\Xi)$, thereby completing the proof for the class of influence functions under the semi-offline RL view, as proposed in Theorem 14.

We now show all in-between steps of the following equalities shown in Section 7:

$$
\begin{aligned}
&\varphi_{Semi}(A, G_A(X_{(1)}), Y; \Xi) = h_{Semi} - \Pi(h_{Semi}|\Lambda_{2,Semi}(\Xi)) \\
&\quad = h_{Semi} - \sum_{t=1}^{T} \mathbb{E}\left[h_{Semi}\middle| A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] + \sum_{t=1}^{T} \mathbb{E}\left[h_{Semi}\middle| \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] \\
&\quad \overset{*_1}{=} \mathbb{E}\left[\rho^T_{Semi} f_C(A', G_{A'}(X_{(1)}), Y)|Y, G_A(X_{(1)}), A\right] \\
&\quad - \sum_{t=1}^{T} \mathbb{E}\left[\rho^t_{Semi} Q^t_{Semi}\middle| A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] \\
&\quad + \sum_{t=1}^{T} \mathbb{E}\left[\rho^{t-1}_{Semi} V^{t-1}_{Semi}\middle| \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] - J \\
&\quad = \mathbb{E}\left[\rho^T_{Semi} f_C(A', G_{A'}(X_{(1)}), Y) - \sum_{t=1}^{T} \rho^t_{Semi} Q^t_{Semi} + \sum_{t=1}^{T} \rho^{t-1}_{Semi} V^{t-1}_{Semi}\middle| A, G_A(X_{(1)}), Y\right] - J.
\end{aligned}
$$

We now go into more detail about why $*1$) holds. We begin with the term including Q_{Semi}:

$$
\begin{aligned}
&\mathbb{E}\left[h_{Semi}\middle|A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] = \\
&= \mathbb{E}\left[\mathbb{E}\left[\rho_{Semi}^T f_C(A', G_{A'}(X_{(1)}), Y)|Y, G_A(X_{(1)}), A\right]|A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] - J \\
&= \mathbb{E}\left[\rho_{Semi}^T f_C(A', G_{A'}(X_{(1)}), Y)|A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] - J \\
&= \mathbb{E}\Bigg[\prod_{\tau=1}^{T} \frac{\pi_\alpha^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1})}{\pi_{sim}'^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1}, A^\tau)} \frac{\pi_{id}^\tau(A^\tau|\underline{A}'^\tau, \underline{X}^{\tau-1}, \underline{A}^{\tau-1})}{\pi_\beta^\tau(A^\tau|\underline{X}^{\tau-1}, \underline{A}^{\tau-1})} \\
&\quad \cdot f_C(A', G_{A'}(X_{(1)}), Y)\Big|A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Bigg] - J \\
&= \mathbb{E}\Big[\prod_{\tau=1}^{t} \frac{\pi_\alpha^\tau(A'^1|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1})}{\pi_{sim}'^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1}, A^\tau)} \frac{\pi_{id}^\tau(A^\tau|\underline{A}'^\tau, \underline{X}^{\tau-1}, \underline{A}^{\tau-1})}{\pi_\beta^\tau(A^\tau|\underline{X}^{\tau-1}, \underline{A}^{\tau-1})} \\
&\cdot \mathbb{E}\Big[\prod_{\tau=t+1}^{T} \frac{\pi_\alpha^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1})}{\pi_{sim}'^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1}, A^\tau)} \frac{\pi_{id}^\tau(A^\tau|\underline{A}'^\tau, \underline{X}^{\tau-1}, \underline{A}^{\tau-1})}{\pi_\beta^\tau(A^\tau|\underline{X}^{\tau-1}, \underline{A}^{\tau-1})} \\
&\cdot f_C(A', G_{A'}(X_{(1)}), Y)\Big|A'^t, \underline{A}'^{t-1}, A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Big]\Big|A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Big] - J \\
&= \mathbb{E}\Big[\rho_{Semi}^t Q_{Semi}(\underline{A}'^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi)\Big|A^t, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Big] - J \\
&= \mathbb{E}\Big[\rho_{Semi}^t Q_{Semi}^t\Big|A, G_A(X_{(1)}), Y\Big] - J
\end{aligned}
$$

Similarly, we can show for the term including V_{Semi}:

$$
\begin{aligned}
&\mathbb{E}\left[h_{Semi}\middle|\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] = \\
&= \mathbb{E}\left[\mathbb{E}\left[\rho_{Semi}^T f_C(A', G_{A'}(X_{(1)}), Y)|Y, G_A(X_{(1)}), A\right]|\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] - J \\
&= \mathbb{E}\left[\rho_{Semi}^T f_C(A', G_{A'}(X_{(1)}), Y)|\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\right] - J \\
&= \mathbb{E}\Bigg[\prod_{\tau=1}^{T} \frac{\pi_\alpha^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1})}{\pi_{sim}'^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1}, A^\tau)} \frac{\pi_{id}^\tau(A^\tau|\underline{A}'^\tau, \underline{X}^{\tau-1}, \underline{A}^{\tau-1})}{\pi_\beta^\tau(A^\tau|\underline{X}^{\tau-1}, \underline{A}^{\tau-1})} \\
&\quad \cdot f_C(A', G_{A'}(X_{(1)}), Y)\Big|\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Bigg] - J \\
&= \mathbb{E}\Big[\prod_{\tau=1}^{t-1} \frac{\pi_\alpha^\tau(A'^1|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1})}{\pi_{sim}'^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1}, A^\tau)} \frac{\pi_{id}^\tau(A^\tau|\underline{A}'^\tau, \underline{X}^{\tau-1}, \underline{A}^{\tau-1})}{\pi_\beta^\tau(A^\tau|\underline{X}^{\tau-1}, \underline{A}^{\tau-1})} \\
&\cdot \mathbb{E}\Big[\prod_{\tau=t}^{T} \frac{\pi_\alpha^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1})}{\pi_{sim}'^\tau(A'^\tau|\underline{X}'^{\tau-1}, \underline{A}'^{\tau-1}, A^\tau)} \frac{\pi_{id}^\tau(A^\tau|\underline{A}'^\tau, \underline{X}^{\tau-1}, \underline{A}^{\tau-1})}{\pi_\beta^\tau(A^\tau|\underline{X}^{\tau-1}, \underline{A}^{\tau-1})} \\
&\cdot f_C(A', G_{A'}(X_{(1)}), Y)\Big|\underline{A}'^{t-1}, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Big]\Big|\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Big] - J \\
&= \mathbb{E}\Big[\rho_{Semi}^{t-1} V_{Semi}(\underline{A}'^{t-1}, \underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi)\Big|\underline{A}^{t-1}, G_{\underline{A}^{t-1}}(X_{(1)}), \Xi\Big] - J \\
&= \mathbb{E}\Big[\rho_{Semi}^{t-1} V_{Semi}^{t-1}\Big|A, G_A(X_{(1)}), Y\Big] - J.
\end{aligned}
$$

Appendix P. Experiment Details

In this section, we describe the experiment setup in more detail. We also provide a detailed list of the parameters and configurations for each experiment in Tables 3 and 4.

P.1 Data, Costs and Missingness Mechanisms

For the experiments, we defined a "superfeature" as a feature that comprises multiple subfeatures, which are acquired jointly and which have a single cost. Furthermore, we assumed a subset of features is available at no cost (free features) and set fixed acquisition costs c_{acq} for the remaining features. A prediction was to be performed at each time step, which corresponds to the setting described in Appendix K. We chose misclassification costs such that good policies must find a balance between the feature acquisition cost and the predictive value of the features.

We evaluated and compared the described methods on synthetic data sets with and without violation of either the NDE or NUC assumption. In experiments where the NDE assumption holds, the features are distributed according to:

$$X^t_{(1),i} = \begin{cases} \gamma_i X^{t-1}_{(1),i} + (1-\gamma_i)\epsilon_i, & \text{if } t > 0 \\ \epsilon_i, & \text{if } t = 0. \end{cases}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma)$. In experiments with a violation of the NDE assumption, the unobserved variables U were distributed according to:

$$U^t_i = \begin{cases} \gamma_i U^{t-1}_i + (1-\gamma_i)\epsilon_i + 0.5 \sum_i A^{t-1}_i & \text{if } t > 1 \\ \gamma_i U^{t-1}_i + (1-\gamma_i)\epsilon_i, & \text{if } t = 1 \\ \epsilon_i, & \text{if } t = 0. \end{cases}$$

The labels are distributed according to

$$p(Y^t = 1) = \begin{cases} 1, & \text{if } \zeta_1 \sum_i W_i X^t_{(1),i} + \zeta_2 \sum_i W_i X^{t-1}_{(1),i} > 0 \\ 0.3, & \text{otherwise.} \end{cases}$$

This choice for Y simulates a scenario where not all data points are equally easy to classify.

The retrospective policy π_β follows different logistic models depending on whether a MAR assumption (NUC holds) or MNAR assumption (NUC is violated) is assumed, as specified in Table 3. To evaluate the convergence of different estimators when the NDE assumption holds, we consider the average cost of running the AFA agent on the data set over all data points in the ground truth test set (without missingness) as the true expected cost J. When NDE is violated, we sample the ground truth data generating process while running the agent and do so the same number of times as there are data points in the test set.

We performed five different experiments:

- **Experiment 1:** A standard experiment where NUC, NDE, and all three positivity assumptions (Assumptions 6.1, 6.2, and 6.4) hold.

- **Experiment 2:** The NUC and NDE assumptions hold, but the missing data positivity assumption (Assumption 6.2) is violated. This is achieved by reducing the number of complete cases to 0.007%.
- **Experiment 3:** The NUC and NDE assumptions hold, but the offline RL (Assumption 6.1) positivity assumption is violated. This is achieved by letting A_2 be always 1 under π_β.
- **Experiment 4:** The NDE assumption holds, but the NUC assumption is violated. This corresponds to an MNAR missing data scenario.
- **Experiment 5:** The NUC assumption holds, but the NDE assumption is violated.

For full experiment configurations for the acquisition processes, please see Table 4.

P.2 Training

We used an impute-then-regress classifier (Le Morvan et al., 2021) with unconditional mean imputation and a logistic regression classifier for the classification task and trained it on the available and further randomly subsampled data (where $p(A_i^t = 1) = 0.5$). We tested random and fixed acquisition policies that acquire each costly feature with a 50% or 100% probability. Furthermore, we evaluated a proximal policy optimization (PPO) RL agent (Schulman et al., 2017), which was trained on the semi-offline sampling distribution p' using π_α as the semi-offline sampling policy, but without adjustment for the blocking of actions. The data sets were split into training set (for the training of the agent and the classifier), nuisance function training set, and test set, where the estimators were evaluated. The splitting of the data set in a nuisance function training set and a test set is necessary due to the complexity of the used nuisance model functions classes (Kennedy, 2022). The resulting loss of efficiency may, however, be avoided using a cross-fitting approach (Kennedy, 2022).

Data and environment	
Sample size n_D	$100'000$ divided into 30% training set (for agent and classifier), 30% nuisance function training set, and 40% test set.
Superfeatures	superX_0: $[X_0]$, superX_1: $[X_1]$, superX_2: $[X_2, X_3]$
Label	$Y^t \in \{0, 1\}$ and for $t \leq T = 3$.
Data generation parameters	$\gamma_i = 0.2\ \forall i$, $\sigma = 1$, $\zeta_1 = 1$, $\zeta_2 = 0.3$, $W = [1, 1, 2, 2]/6$
Feature acquisition cost	$c_{acq} = [0, 1, 1]$
Misclassification cost	$c_{mc} = 12$
Models	
Classifier	Logistic regression
Agents	Random 50%, Fixed 100%, PPO (learning rate: 0.0001, number of layers: 2, hidden layer neurons per layer: 64, hidden layer activation function: tanh)
Nuisance functions	$\hat{\pi}_\beta$ (logistic regression), $\hat{Q}_{Semi}$ ($\Xi = \emptyset$, learning rate: 0.001, number of layers: 2, hidden layer neurons per layer: 16, hidden layer activation function: ReLU)

Table 3: Full experiment details except for the acquisition process

Missingness mechanisms	
Exp 1	$p(A_0^t = 1) = 1.0$, $p(A_1^t = 1) = \sigma(0.8 - 3.0X_0^{t-1} + 0.02X_1^{t-1} - 0.02X_2^{t-1})$, $p(A_2^t = 1) = \sigma(0.8 - 3.0X_0^{t-1} + 0.02X_1^{t-1} - 0.02X_2^{t-1})$ Complete cases ratio: $p(A = \vec{1}) = 11.71\%$
Exp 2	$p(A_0^t = 1) = 1.0$, $p(A_1^t = 1) = 0.2$, $p(A_2^t = 1) = 0.2$ Complete cases ratio: $p(A = \vec{1}) = 0.007\%$
Exp 3	$p(A_0^t = 1) = 1.0$, $p(A_1^t = 1) = 1.0$, $p(A_2^t = 1) = \sigma(-0.5 - 2.0X_0^{t-1} - 0.1X_1^{t-1} - 0.1X_2^{t-1})$ Complete cases ratio: $p(A = \vec{1}) = 7.09\%$
Exp 4	$p(A_0^t = 1) = 1.0$, $p(A_1^t = 1) = 1.0$, $p(A_2^t = 1) = \sigma(-0.6 - 1.5X_{(1),2}^{t-1} - 1.5X_{(1),3}^{t-1})$ Complete cases ratio: $p(A = \vec{1}) = 9.63\%$
Exp 5	$p(A_0^t = 1) = 1.0$, $p(A_1^t = 1) = \sigma(0.8 - 0.2X_0^{t-1} - 0.1X_1^{t-1} + 0.5X_2^{t-1})$, $p(A_2^t = 1) = \sigma(0.8 - 0.2X_0^{t-1} - 0.1X_1^{t-1} + 0.5X_2^{t-1})$ Complete cases ratio: $p(A = \vec{1}) = 11.71\%$

Table 4: Acquisition process details for all five experiments

References

Chaojie An, Qifeng Zhou, and Shen Yang. A reinforcement learning guided adaptive cost-sensitive feature acquisition method. *Applied Soft Computing*, 117:108437, March 2022. ISSN 1568-4946. doi: 10.1016/j.asoc.2022.108437. URL https://www.sciencedirect.com/science/article/pii/S1568494622000163.

Qi An, Hui Li, Xuejun Liao, and Lawrence Carin. Active feature acquisition with POMDP models. *Submitted to Pattern Recognition Letters*, 2006.

Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 1. Athena scientific, 2012.

Christian Beyer, Maik Büttner, Vishnu Unnikrishnan, Miro Schleicher, Eirini Ntoutsi, and Myra Spiliopoulou. Active feature acquisition on data streams under feature drift. *Annals of Telecommunications*, 75(9-10):597–611, October 2020. ISSN 0003-4347, 1958-9395. doi: 10.1007/s12243-020-00775-2. URL https://link.springer.com/10.1007/s12243-020-00775-2.

Rohit Bhattacharya, Razieh Nabi, Ilya Shpitser, and James M. Robins. Identification In Missing Data Models Represented By Directed Acyclic Graphs. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, pages 1149–1158. PMLR, August 2020. URL https://proceedings.mlr.press/v115/bhattacharya20b.html.

Marc F. Botteman, Chris L. Pashos, Alberto Redaelli, Benjamin Laskin, and Robert Hauser. The health economics of bladder cancer. *PharmacoEconomics*, 21(18):1315–1330, December 2003. ISSN 1179-2027. doi: 10.1007/BF03262330. URL https://doi.org/10.1007/BF03262330.

Ellen C. Caniglia, James M. Robins, Lauren E. Cain, Caroline Sabin, Roger Logan, Sophie Abgrall, Michael J. Mugavero, Sonia Hernández-Díaz, Laurence Meyer, Remonie Seng, Daniel R. Drozd, George R. Seage III, Fabrice Bonnet, Fabien Le Marec, Richard D. Moore, Peter Reiss, Ard van Sighem, William C. Mathews, Inma Jarrín, Belén Alejos, Steven G. Deeks, Roberto Muga, Stephen L. Boswell, Elena Ferrer, Joseph J. Eron, John Gill, Antonio Pacheco, Beatriz Grinsztejn, Sonia Napravnik, Sophie Jose, Andrew Phillips, Amy Justice, Janet Tate, Heiner C. Bucher, Matthias Egger, Hansjakob Furrer, Jose M. Miro, Jordi Casabona, Kholoud Porter, Giota Touloumi, Heidi Crane, Dominique Costagliola, Michael Saag, and Miguel A. Hernán. Emulating a trial of joint dynamic strategies: An application to monitoring and treatment of HIV-positive individuals. *Statistics in Medicine*, 38(13):2428–2446, 2019. ISSN 1097-0258. doi: 10.1002/sim.8120. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8120. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8120.

Chun-Hao Chang, Mingjie Mai, and Anna Goldenberg. Dynamic Measurement Scheduling for Event Forecasting using Deep RL. In *Proceedings of the 36th International Conference on Machine Learning*, pages 951–960. PMLR, May 2019. URL https://proceedings.mlr.press/v97/chang19a.html.

Li-Fang Cheng, Niranjani Prasad, and Barbara E. Engelhardt. An Optimal Policy for Patient Laboratory Tests in Intensive Care Units. In *Biocomputing 2019*, pages 320–331. WORLD SCIENTIFIC, October 2018. ISBN 978-981-327-981-0. doi: 10.1142/9789813279827_0029. URL https://www.worldscientific.com/doi/abs/10.1142/9789813279827_0029.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, February 2018. ISSN 1368-4221. doi: 10.1111/ectj.12097. URL https://doi.org/10.1111/ectj.12097.

Srijita Das, Rishabh Iyer, and Sriraam Natarajan. A Clustering based Selection Framework for Cost Aware and Test-time Feature Elicitation. In *8th ACM IKDD CODS and 26th COMAD*, pages 20–28, Bangalore India, January 2021. ACM. ISBN 978-1-4503-8817-7. doi: 10.1145/3430984.3431008. URL https://dl.acm.org/doi/10.1145/3430984.3431008.

Miroslav Dudik, John Langford, and Lihong Li. Doubly Robust Policy Evaluation and Learning, May 2011. URL http://arxiv.org/abs/1103.4601. arXiv:1103.4601 [cs, stat].

Gabriel Erion, Joseph D. Janizek, Carly Hudelson, Richard B. Utarnachitt, Andrew M. McCoy, Michael R. Sayre, Nathan J. White, and Su-In Lee. CoAI: Cost-Aware Artificial Intelligence for Health Care. Technical report, medRxiv, January 2021. URL https://www.medrxiv.org/content/10.1101/2021.01.19.21249356v1.

US Preventive Services Task Force*. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Annals of internal medicine*, 151(10):716–726, 2009. Publisher: American College of Physicians.

Wenbo Gong, Sebastian Tschiatschek, Sebastian Nowozin, Richard E Turner, José Miguel Hernández-Lobato, and Cheng Zhang. Icebreaker: Element-wise Efficient Information Acquisition with a Bayesian Deep Latent Gaussian Model. In *Advances in Neural Information Processing Systems*, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/c055dcc749c2632fd4dd806301f05ba6-Abstract.html.

John P Gould. Risk, stochastic preference, and the value of information. *Journal of Economic Theory*, 8(1):64–84, May 1974. ISSN 0022-0531. doi: 10.1016/0022-0531(74)90006-4. URL https://www.sciencedirect.com/science/article/pii/0022053174900064.

James Hess. Risk and the Gain from Information. *Journal of Economic Theory*, 27(1): 231–238, 1982.

Ronald W. Hilton. The Determinants of Cost Information Value: An Illustrative Analysis. *Journal of Accounting Research*, 17(2):411–435, 1979. ISSN 0021-8456. doi: 10.2307/2490511. URL https://www.jstor.org/stable/2490511. Publisher: [Accounting Research Center, Booth School of Business, University of Chicago, Wiley].

Daniel G. Horvitz and Donovan J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260): 663–685, 1952. Publisher: Taylor & Francis.

Sheng-Jun Huang, Miao Xu, Ming-Kun Xie, Masashi Sugiyama, Gang Niu, and Songcan Chen. Active Feature Acquisition with Supervised Matrix Completion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1571–1579, July 2018. ISBN 978-1-4503-5552-0. doi: 10.1145/3219819.3220084. URL https://dl.acm.org/doi/10.1145/3219819.3220084.

Jaromír Janisch, Tomáš Pevný, and Viliam Lisý. Classification with costly features as a sequential decision-making problem. *Machine Learning*, 109(8):1587–1615, August 2020. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-020-05874-8. URL http://link.springer.com/10.1007/s10994-020-05874-8.

Daniel Jarrett and Mihaela van der Schaar. Inverse Active Sensing: Modeling and Understanding Timely Decision-Making. *arXiv:2006.14141 [cs, stat]*, June 2020. URL http://arxiv.org/abs/2006.14141.

Nathan Kallus and Masatoshi Uehara. Double Reinforcement Learning for Efficient Off-Policy Evaluation in Markov Decision Processes. *Journal of Machine Learning Research*, 21(167):1–63, 2020. ISSN 1533-7928. URL http://jmlr.org/papers/v21/19-827.html.

Jeffrey M. Keisler, Zachary A. Collier, Eric Chu, Nina Sinatra, and Igor Linkov. Value of information analysis: the state of application. *Environment Systems and Decisions*, 34 (1):3–23, March 2014. ISSN 2194-5411. doi: 10.1007/s10669-013-9439-4. URL https://doi.org/10.1007/s10669-013-9439-4.

Edward H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *arXiv:2203.06469 [stat]*, March 2022. URL http://arxiv.org/abs/2203.06469. arXiv: 2203.06469.

Thomas T. Kok, Rachel M. Brouwer, Rene M. Mandl, Hugo G. Schnack, and Georg Krempl. Active Selection of Classification Features. *arXiv:2102.13636 [cs]*, February 2021. URL http://arxiv.org/abs/2102.13636. arXiv: 2102.13636.

Murray D. Krahn, John E. Mahoney, Mark H. Eckman, John Trachtenberg, Stephen G. Pauker, and Allan S. Detsky. Screening for Prostate Cancer: A Decision Analytic View. *JAMA*, 272(10):773–780, September 1994. ISSN 0098-7484. doi: 10.1001/jama.1994.03520100035030. URL https://doi.org/10.1001/jama.1994.03520100035030.

Noémi Kreif, Oleg Sofrygin, Julie A. Schmittdiel, Alyce S. Adams, Richard W. Grant, Zheng Zhu, Mark J. van der Laan, and Romain Neugebauer. Exploiting nonsystematic covariate monitoring to broaden the scope of evidence about the causal effects of adaptive treatment strategies. *Biometrics*, 77(1):329–342, 2021. ISSN 1541-0420. doi: 10.1111/biom.13271. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13271. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/biom.13271.

Irving H. LaValle. On cash equivalents and information evaluation in decisions under uncertainty part I: Basic theory. *Journal of the American Statistical Association*, 63(321): 252–276, 1968a. Publisher: Taylor & Francis.

Irving H. LaValle. On cash equivalents and information evaluation in decisions under uncertainty Part II: Incremental information decisions. *Journal of the American Statistical Association*, 63(321):277–284, 1968b. Publisher: Taylor & Francis.

Marine Le Morvan, Julie Josse, Erwan Scornet, and Gael Varoquaux. What's a good imputation to predict with missing values? In *Advances in Neural Information Processing Systems*, volume 34, pages 11530–11540, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/5fe8fdc79ce292c39c5f209d734b7206-Abstract.html.

Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems. *arXiv:2005.01643 [cs, stat]*, November 2020. URL http://arxiv.org/abs/2005.01643. arXiv: 2005.01643.

Yang Li and Junier Oliva. Active Feature Acquisition with Generative Surrogate Models. In *Proceedings of the 38th International Conference on Machine Learning*, pages 6450–6459. PMLR, July 2021a. URL https://proceedings.mlr.press/v139/li21p.html. ISSN: 2640-3498.

Yang Li and Junier B. Oliva. Dynamic Feature Acquisition with Arbitrary Conditional Flows. *arXiv:2006.07701 [cs, stat]*, March 2021b. URL http://arxiv.org/abs/2006.07701.

Yang Li, Siyuan Shan, Qin Liu, and Junier B. Oliva. Towards Robust Active Feature Acquisition. *arXiv:2107.04163 [cs]*, July 2021. URL http://arxiv.org/abs/2107.04163.

Lin Liu, Zach Shahn, James M. Robins, and Andrea Rotnitzky. Efficient Estimation of Optimal Regimes Under a No Direct Effect Assumption. *Journal of the American Statistical Association*, 116(533):224–239, January 2021. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2020.1856117. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2020.1856117.

Chao Ma, Sebastian Tschiatschek, Konstantina Palla, Jose Miguel Hernandez-Lobato, Sebastian Nowozin, and Cheng Zhang. EDDI: Efficient Dynamic Discovery of High-Value Information with Partial VAE. In *Proceedings of the 36th International Conference on Machine Learning*, pages 4234–4243. PMLR, May 2019. URL https://proceedings.mlr.press/v97/ma19c.html.

Karthika Mohan and Judea Pearl. Graphical Models for Processing Missing Data. *Journal of the American Statistical Association*, 116(534):1023–1037, April 2021. ISSN 0162-1459. doi: 10.1080/01621459.2021.1874961. URL https://doi.org/10.1080/01621459.2021.1874961. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2021.1874961.

Karthika Mohan, Judea Pearl, and Jin Tian. Graphical Models for Inference with Missing Data. In *Advances in Neural Information Processing Systems*, volume 26, 2013. URL https://proceedings.neurips.cc/paper/2013/hash/0ff8033cf9437c213ee13937b1c4c455-Abstract.html.

Susan A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):331–355, 2003. URL https://academic.oup.com/jrsssb/article-abstract/65/2/331/7092852. Publisher: Oxford University Press.

Alvin I. Mushlin and Lou Fintor. Is screening for breast cancer cost-effective? *Cancer*, 69(S7):1957–1962, 1992. ISSN 1097-0142. doi: 10.1002/1097-0142(19920401)69:7+⟨1957::AID-CNCR2820691716⟩3.0.CO;2-T. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%2819920401%2969%3A7%20%3C1957%3A%3AAID-CNCR2820691716%3E3.0.CO%3B2-T. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0142%2819920401%2969%3A7%2B%3C1957%3A%3AAID-CNCR2820691716%3E3.0.CO%3B2-T.

Razieh Nabi, Rohit Bhattacharya, and Ilya Shpitser. Full law identification in graphical models of missing data: Completeness results. In *International Conference on Machine Learning*, pages 7153–7163. PMLR, 2020. URL http://proceedings.mlr.press/v119/nabi20a.html.

Sriraam Natarajan, Srijita Das, Nandini Ramanan, Gautam Kunapuli, and Predrag Radivojac. On Whom Should I Perform this Lab Test Next? An Active Feature Elicitation Approach. In *IJCAI*, pages 3498–3505, 2018.

Romain Neugebauer, Julie A. Schmittdiel, Alyce S. Adams, Richard W. Grant, and Mark J. van der Laan. Identification of the Joint Effect of a Dynamic Treatment Intervention and a Stochastic Monitoring Intervention Under the No Direct Effect Assumption. *Journal of Causal Inference*, 5(1):20160015, September 2017. ISSN 2193-3685, 2193-3677. doi: 10.1515/jci-2016-0015. URL https://www.degruyter.com/document/doi/10.1515/jci-2016-0015/html.

Maya L Petersen, Kristin E Porter, Susan Gruber, Yue Wang, and Mark J van der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, 21(1):31–54, February 2012. ISSN 0962-2802. doi: 10.1177/0962280210386207. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4107929/.

Sophia M. Rein, Jing Li, Miguel Hernan, and Andrew Beam. Deep Learning Methods for the Noniterative Conditional Expectation G-Formula for Causal Inference from Complex Observational Data, October 2024. URL http://arxiv.org/abs/2410.21531. arXiv:2410.21531.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, January 1986. ISSN 0270-0255.

doi: 10.1016/0270-0255(86)90088-6. URL https://www.sciencedirect.com/science/article/pii/0270025586900886.

James Robins, Liliana Orellana, and Andrea Rotnitzky. Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine*, 27(23):4678–4721, 2008. ISSN 1097-0258. doi: 10.1002/sim.3301. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3301. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.3301.

James M. Robins. Optimal Structural Nested Models for Optimal Sequential Decisions. In P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, S. Zeger, D. Y. Lin, and P. J. Heagerty, editors, *Proceedings of the Second Seattle Symposium in Biostatistics*, volume 179, pages 189–326. Springer New York, New York, NY, 2004. ISBN 978-0-387-20862-6 978-1-4419-9076-1. doi: 10.1007/978-1-4419-9076-1_11. URL http://link.springer.com/10.1007/978-1-4419-9076-1_11. Series Title: Lecture Notes in Statistics.

Patrick Rockenschaub, Ela Marie Akay, Benjamin Gregory Carlisle, Adam Hilbert, Falk Meyer-Eschenbach, Anatol-Fiete Näher, Dietmar Frey, and Vince Istvan Madai. Generalisability of AI-based scoring systems in the ICU: a systematic review and meta-analysis, October 2023a. URL https://www.medrxiv.org/content/10.1101/2023.10.11.23296733v1. Pages: 2023.10.11.23296733.

Patrick Rockenschaub, Adam Hilbert, Tabea Kossen, Falk von Dincklage, Vince Istvan Madai, and Dietmar Frey. From Single-Hospital to Multi-Centre Applications: Enhancing the Generalisability of Deep Learning Models for Adverse Event Prediction in the ICU, April 2023b. URL http://arxiv.org/abs/2303.15354. arXiv:2303.15354 [cs].

Andrea Rotnitzky, James Robins, and Lucia Babino. On the multiply robust estimation of the mean of the g-functional, May 2017. URL http://arxiv.org/abs/1705.08582. arXiv:1705.08582 [stat].

Daniel O. Scharfstein, Andrea Rotnitzky, and James M. Robins. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999. ISSN 0162-1459. doi: 10.2307/2669923. URL https://www.jstor.org/stable/2669923. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms, August 2017. URL http://arxiv.org/abs/1707.06347. arXiv:1707.06347 [cs].

Shaun R. Seaman and Ian R. White. Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, 22(3):278–295, 2013. Publisher: Sage Publications Sage UK: London, England.

Burr Settles. Active learning literature survey. 2009.

Hajin Shim, Sung Ju Hwang, and Eunho Yang. Joint Active Feature Acquisition and Classification with Variable-Size Set Encoding. *Advances in Neural Information Processing Systems*, 31, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/e5841df2166dd424a57127423d276bbe-Abstract.html.

Ilya Shpitser, Karthika Mohan, and Judea Pearl. Missing data as a causal and probabilistic problem. Technical report, CALIFORNIA UNIV LOS ANGELES DEPT OF COMPUTER SCIENCE, 2015.

Jonathan A. C. Sterne, Ian R. White, John B. Carlin, Michael Spratt, Patrick Royston, Michael G. Kenward, Angela M. Wood, and James R. Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338: b2393, June 2009. ISSN 0959-8138, 1468-5833. doi: 10.1136/bmj.b2393. URL https://www.bmj.com/content/338/bmj.b2393. Publisher: British Medical Journal Publishing Group Section: Research Methods & Reporting.

Fengyi Tang, Lifan Zeng, Fei Wang, and Jiayu Zhou. Adversarial Precision Sensing with Healthcare Applications. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 521–530, November 2020. doi: 10.1109/ICDM50108.2020.00061.

Morteza Tavakol, Salman Ashraf, and Sorin J. Brener. Risks and Complications of Coronary Angiography: A Comprehensive Review. *Global Journal of Health Science*, 4(1):65–93, January 2012. ISSN 1916-9736. doi: 10.5539/gjhs.v4n1p65. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4777042/.

Philip Thomas and Emma Brunskill. Data-Efficient Off-Policy Policy Evaluation for Reinforcement Learning. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2139–2148. PMLR, June 2016. URL https://proceedings.mlr.press/v48/thomasa16.html. ISSN: 1938-7228.

Anastasios A. Tsiatis. *Semiparametric theory and missing data*. Springer series in statistics. Springer, New York, 2006. ISBN 978-0-387-32448-7.

Stef van Buuren. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 16(3):219–242, June 2007. ISSN 0962-2802. doi: 10.1177/0962280206074463. URL https://doi.org/10.1177/0962280206074463.

Thomas S. Verma and Judea Pearl. Equivalence and Synthesis of Causal Models. Technical Report R-150, Department of Computer Science, University of California, Los Angeles, 1990.

Henrik von Kleist, Alireza Zamanian, Ilya Shpitser, and Narges Ahmidi. Evaluation of Active Feature Acquisition Methods for Time-varying Feature Settings, December 2023. URL http://arxiv.org/abs/2312.01530. arXiv:2312.01530 [cs, stat].

Lan Wen, Jessica G. Young, James M. Robins, and Miguel A. Hernán. Parametric G-Formula Implementations for Causal Survival Analyses. *Biometrics*, 77(2):740–753, June

2021. ISSN 0006-341X. doi: 10.1111/biom.13321. URL https://doi.org/10.1111/biom.13321.

Xiaoyong Chai, Lin Deng, Qiang Yang, and C. X. Ling. Test-cost sensitive naive Bayes classification. In *Fourth IEEE International Conference on Data Mining (ICDM'04)*, pages 51–58, November 2004. doi: 10.1109/ICDM.2004.10092.

Haiyan Yin, Yingzhen Li, Sinno Jialin Pan, Cheng Zhang, and Sebastian Tschiatschek. Reinforcement Learning with Efficient Active Feature Acquisition. *arXiv:2011.00825 [cs]*, November 2020. URL http://arxiv.org/abs/2011.00825.

Jinsung Yoon, William R. Zame, and Mihaela van der Schaar. Deep sensing: Active sensing using multi-directional recurrent neural networks. In *International Conference on Learning Representations*, 2018.

Jinsung Yoon, James Jordon, and Mihaela van der Schaar. ASAC: Active Sensing using Actor-Critic models. In *Machine Learning for Healthcare Conference*, pages 451–473. PMLR, October 2019. URL http://proceedings.mlr.press/v106/yoon19a.html. ISSN: 2640-3498.

Pin Zhang. A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model. *Applied Soft Computing*, 85: 105859, December 2019. ISSN 1568-4946. doi: 10.1016/j.asoc.2019.105859. URL http://www.sciencedirect.com/science/article/pii/S1568494619306404.

Yan Zhou, Roderick J. A. Little, and John D. Kalbfleisch. Block-Conditional Missing at Random Models for Missing Data. *Statistical Science*, 25(4), November 2010. ISSN 0883-4237. doi: 10.1214/10-STS344. URL https://projecteuclid.org/journals/statistical-science/volume-25/issue-4/Block-Conditional-Missing-at-Random-Models-for-Missing-Data/10.1214/10-STS344.full.