Fair and Actionable Causal Prescription Ruleset

ABSTRACT

Prescriptions, or actionable recommendations, are commonly generated across various fields to influence key outcomes such as improving public health, enhancing economic policies, or increasing business efficiency. While traditional association-based methods may identify correlations, they often fail to reveal the underlying causal factors needed for informed decision-making. On the other hand, in decision making for tasks with significant societal or economic impact, it is crucial to provide recommendations that are interpretable and justifiable, and equitable in terms of the outcome for both the protected and non-protected groups. Motivated by these two goals, this paper introduces a fairness-aware framework leveraging causal reasoning for generating a set of interpretable and actionable prescription rules (ruleset) toward betterment of an outcome while preventing exacerbating inequalities for protected groups. By considering group and individual fairness metrics from the literature, we ensure that both protected and non-protected groups benefit from these recommendations, providing a balanced and equitable approach to decision-making. We employ efficient optimizations to explore the vast and complex search space considering both fairness and coverage of the prescription ruleset. Empirical evaluation and case study on real-world datasets demonstrates the utility of our framework for different use cases.

1 INTRODUCTION

Prescriptions, or actionable recommendations, are commonly generated across various fields to influence key outcomes such as improving public health, enhancing economic policies, or increasing business efficiency. Policy makers in government, decision makers in businesses or health domains, and in general, data scientists in various fields, often rely on data-driven approaches to identify potential actions aimed at influencing an outcome of interest, such as increasing product satisfaction or income levels, or decreasing the likelihood of experiencing serious health conditions. While traditional association-or prediction-based methods are extensively used in practice to draw useful insights from data, they typically identify correlations among variables and may fail to reveal the underlying causal factors, i.e., which actions may result in an improved outcome, needed for an informed decision-making.

Causal analysis or *causal inference*, therefore, is considered one of the most important requirements to generate prescriptions that are *actionable* and aligned with human reasoning [33]. Causal inference, and in particular *observational studies* for causal inference on collected data (when controlled trials are impossible due to cost or ethical reasons), have been extensively studied in the statistics and artificial intelligence (AI) literature for several decades [60, 71]. Motivated by this foundational work on causal inference, the notion of causality has also influenced the field of database research. The causal models from AI have been extended to relational databases [74], and causality has been incorporated into various data management tasks such as finding responsibilities of inputs toward query answers [50, 51, 53], explanations for query answers [69, 97], data discovery [22, 96], data cleaning [62, 75], hypothetical reasoning [23], and large system diagnostics [5, 24, 29, 48]. We give a concrete example of the difference between association and causation in generating prescriptions or recommended actions below:

Example 1.1. **Importance of causal prescriptions:** *Consider the Stack Overflow (SO) annual developer survey [1], where respondents from around the world answer questions about their jobs and demographics. A sample of the dataset with a subset of the attributes is presented in Table 1. Alice, a researcher in the United Nations (UN) finance department, is interested in discovering ways to increase the salaries of high-tech employees worldwide. She is looking for a set of actionable recommendations to raise the overall average salary. Using association-based approaches [14, 40], she may discover that individuals residing in the US who identify as straight or heterosexual tend to earn higher salaries (see Section 7.2 for full details). However, this observation merely indicates a correlation: people living in the US, for example, generally earn more than those outside the country. Their comparatively higher salaries are primarily attributable to the country's economy and are unrelated to their sexual orientation. Thus, this observation cannot be used as a prescription rule to increase salary. Our causal analysis, on the other hand, reveals that individuals aged 25-34 with dependents would benefit from working as front-end developers. This results in a $44,009 annual salary increase on average. Another observation is that male students should exercise 3-4 times per week, and pursue an undergraduate major in CS. This can boost their salary by $25,578 per year (see details in Section 6).*

While generating prescriptions based on causal inference may help in robust decision making, causal prescriptions that solely consider betterment of an intended outcome (like salary) is not enough in practice. It is well-known that decision making in many high-stake applications (like hiring policy, policy for approving loans by banks, policy for improving revenue in businesses etc.) may lead to disparate societal or economic impact to different sub-populations. As a shocking example from a recent work called CauSumX [97] that generates a set of causal explanations for an aggregated view (generated by a SQL group by-average query), the explanations generated by CauSumX suggest that male individuals do a Bachelor's degree to increase salary while being an unmarried woman has the most adverse effect on the salary (borrowed directly from Figure 19 in [94]). We explored this further in the context of generating prescriptions, and observed that prescriptions that are not fairness-aware can indeed generate unfair outcomes to some sub-populations who we refer to as the *protected group* in this paper. Examples include women, Black, Latino, or Native Americans, individuals with disability, countries with a weaker economy or infrastructure, or other protected groups specific to an application.

Example 1.2. **Importance of fair prescriptions:** *Continuing Example 1.1, while those causal prescription rules are highly beneficial for the overall population, they are considerably less effective for individuals residing in countries with a low GDP (indicating a weaker economy). For this group, the average expected increase in*

Table 1: A subset of the Stack Overflow dataset.

ID	Gender	Ethnicity	Age	Role	Education	Country	Undergrad Major	Salary
1	Male	White	26	Data Scientist	PhD	US	Computer Science	180k
2	Non-binary	White	32	QA developer	Bachelor's degree	US	Mechanical Eng.	83k
3	Male	South Asian	29	C-suite executive	Bachelor's degree	India	Computer Science	24k
4	Female	South Asian	25	Back-end developer	Master's degree	India	Mathematics	7.5k
5	Male	East Asian	21	Back-end developer	Bachelor's degree	China	Computer Science	19k

salary is only approximately $13,000 per year (in contrast to $44,009 for the entire group). Consequently, implementing these rules would exacerbate the disparity between those living in countries with strong economies and those in countries with weaker economies.

The example above demonstrates that focusing solely on maximizing utility can result in a scenario where only some of the population receive significant improvement, while others experience no benefit. Additionally, even if a large portion of the population receives recommendations, a protected subpopulation might not share the benefits and, worse, their situation could deteriorate, exacerbating inequalities.

Examples 1.1 and 1.2 show that, while generating prescriptions toward betterment of an outcome, it is crucial to provide recommendations that are (1) *causal* for the outcome (beyond associations), and (2) also *fair or equitable* in terms of the outcome for both the protected and non-protected groups. While recent work in database research has focused on deriving *causal explanations* for individual data points, aggregated view, or entire datasets [47, 73, 95, 97], and in particular [97] has considered generating a set of causal explanations for an aggregated view that resemble a ruleset, the absence of fairness considerations in generating these causal explanations can lead to unfair outcomes to the protected group.

Our contributions. Motivated by the dual goals of generating causal and fair prescriptions for the betterment of an outcome, in this paper we introduce a *fairness-aware framework leveraging causal reasoning for generating a set of interpretable and actionable prescription rules (ruleset)* called FairCap (Fair CAusal Prescription). Following research on fairness in data management [23, 80], we assume the existence of a *protected subpopulation*, defined by an attribute such as gender or race for people, or GDP of a country. Motivated by the causal explanation rules for an aggregated view [97], each prescription rule in our ruleset applies to a sub-population defined by a *grouping attribute*, and prescribes a *treatment or intervention* to improve the *outcome* attribute for this sub-population. Fairness constraints ensure that the expected utility of the protected population is *comparable* to the utility of the unprotected individuals. We borrow the notions of *group and individual fairness* from the fairness literature, but tailor them for prescription rules. In addition to the fairness constraints, our coverage constraints for these rules ensure that a substantial fraction of the population and protected subpopulation receives at least one recommendation.

Example 1.3. *Continuing Examples 1.1 and 1.2, Alice uses our proposed system, called FairCap, to impose fairness and coverage constraints to discover useful and equitable recommendations for increasing salaries worldwide. In particular, Alice chooses to implement a coverage constraint to ensure that the selected rules apply to a significant portion of people worldwide, including a sufficiently large number of individuals from countries with low GDP (the protected group). She also imposes a fairness constraint to ensure that the expected gains for both protected and non-protected groups are comparable. She discovers, for example, that for individuals with 6-8 years of coding experience (a subpopulation comprising 21% of the entire dataset and 25% of the protected group), exercising 1-2 times per week, and pursuing a bachelor's degree will increase the expected salary by* $\$15.8k$ *for protected and by* $\$18.1k$ *for non-protected. (See Section 6 for more details.) This prescription rule applies to a large portion of the population and ensures fairness by providing a similar expected gain for both protected and non-protected groups, and the allowed difference of outcomes between these two populations may be adjusted by choosing appropriate thresholds in the fairness definitions.*

Our main contributions are as follows.

(1) We **develop a framework that generates a set of prescription rules to enhance an outcome of interest (Section 4)**. A prescription rule consists of a *grouping pattern* and an *intervention pattern*, representing the target subpopulation and the actionable recommendation for that group, respectively. The strength of the *conditional causal effect* (Section 3) of this intervention on the subgroup is used to measure the expected utility of a rule. Our objective is to identify the smallest set of rules that maximizes overall expected utility. We refer to this problem as the *Prescription Ruleset Selection* problem. We adopt several notions of fairness (individual vs. group, statistical parity vs. bounded group loss) from the literature to define the **fairness constraints** for our problem. In addition, **coverage constraints** (for individual rules or for a group) ensure that the solution for the Prescription Ruleset Selection problem is applied to a sufficient number of individuals and to minimize inequalities. We show NP-hardness for different variants of the problems and properties (matroid) useful in our algorithms.

(2) We **develop a general three-step algorithmic framework named FairCap to solve the optimization problem of selecting a fair prescription ruleset (Section 5)**. The first step involves mining frequent grouping patterns using the Apriori algorithm [4]. In the second step, we employ a lattice-based algorithm to find high utility and fair intervention patterns for each grouping pattern identified in the previous step. Finally, the third step applies a greedy approach to determine a solution. FairCap can be easily adapted to accommodate all variants of the Prescription Ruleset Selection problem.

(3) We provide a detailed case study (Section 6) and a thorough experimental analysis (Section 7) to evaluate our framework and algorithms. The case study shows the qualitative difference of different variants of our problem for different choices of the fairness and coverage constraints. The experimental analysis includes two datasets, three baselines, and 18 variations of our problem with different constraints. Our evaluations suggest that fairness may come at the cost of expected utility for everyone. However, without fairness constraints, we often observe a significant disparity between the protected and non-protected. We also observe that

Table 2: Positioning of our framework w.r.t. previous work.

Related Work		Causal	Fairness	Entire dataset
Aggregate Query Result Explanation	[47]	✓	✗	✗
	[51, 69, 73, 95, 97]	✓	✗	✓
	[43, 54]	✗	✗	✗
	[90]	✗	✗	✓
Interpretable Prediction Models	[14, 40]	✗	✗	✓
Multi dimensional data aggregation	[58, 59, 82]	✗	✓	✓
	[19, 37]	✗	✗	✓
FairCap		✓	✓	✓

achieving individual fairness is harder than group fairness, as most high-utility or high-coverage rules are unfair. Lastly, we show that FairCap can generate a set of prescription rules over large datasets in a reasonable time.

We discuss related work in Section 2, review background on causal inference in Section 3, and discuss the limitations of our framework and future work in Section 8.

2 RELATED WORK

Table 2 outlines the key distinctions between FairCap and prior work. The columns in bold emphasize our key contributions: we generate **causality-based** prescription rules aimed at improving outcomes for the **entire datasets**, while also **considering fairness**. In contrast, other approaches either produce non-causal rules (as shown in the Causal column), target only subsets of the data (as indicated in the Entire dataset column), or disregard fairness considerations (as highlighted in the Fairness column).

Rule mining. Association rule mining has been extensively studied [39] and is used to identify relationships between items that frequently co-occurring in datasets. These techniques are applied across various fields, such as data analysis and outcome improvement. Notable algorithms include STEM [26], FP-Growth [35], AIS [102], and the Apriori algorithm [4]. We leverage the Apriori algorithm to identify sufficiently large subpopulations for which we will generate causal interventions. Rule-based interpretable prediction models [14, 40, 92] often leverage association rule mining to generate predictive rules [40, 41], with the goal of balancing high predictive accuracy with interpretability [38, 41, 46, 72, 78].

Recent work has proposed generating rules based on causal relationships. In [63], a framework was introduced to understand and address biases in the data for fair causal analysis. The authors of [81] proposed a method to optimally allocate treatments with uncertain costs that vary based on confounders. Related research has also focused on estimating heterogeneous treatment effects, which refer to variations in treatment effects across subpopulations [88, 89, 91]. However, these approaches differ from ours, as they assume both treatment and outcome variables are known. In contrast, we assume only the outcome variable is provided and aim to identify treatments that influence the outcome for different subpopulations, potentially yielding different treatments for each group. Our approach ensures the rules apply broadly while maintaining fairness for minority groups.

We adapt the method proposed in [97] called CauSumX. CauSumX is designed to identify the treatment with the highest causal effect on the outcome for a given subpopulation, generating causal explanations for aggregate SQL queries. CauSumX focuses on generating explanations rather than interventions, meaning the treatments may be non-actionable (e.g., being male could be considered a treatment). Additionally, CauSumX does not consider fairness. We empirically show that using CauSumX to generate prescription rules can lead to significant disparities between protected and non-protected populations (See Section 7.2). Another main difference is the fact that CauSumX considers the *aggregate view* to generate explanations, whereas we consider the entire dataset, therefore, the search space is significantly larger. Our primary contribution lies in introducing fairness constraints on the generated rules, making the necessary adjustments to the algorithm to scale, and conducting an extensive experimental study to demonstrate the importance of fairness constraints.

Fairness in data management. Algorithmic fairness, especially in the context of predictions by ML algorithms for high-stake decision making, has been a prominent topic in ML and AI (e.g., [2, 12, 49, 61, 64, 65]). Popular notions of fairness include group and individual fairness [11, 25, 100]. Group fairness ensures that the decision-making process is fair to the group of protected individuals but may be unfair towards any specific individual. This notion is captured in the form of statistical parity or equalized odds. In contrast, individual notions of fairness enforce that the decisions are fair towards every individual. We refer the reader to Section 4.6 for more details. In recent years, fairness has emerged as a key consideration in data management research [23, 80, 83, 100]. This includes ensuring fairness during data acquisition [7, 55, 56], improving data cleaning processes to promote fairness [30, 75], and achieving fairness in ranking and in database queries [44, 99]. One of our contributions is the introduction of novel definitions of group and individual fairness for causal analysis.

Causal inference in data management. Causality, used as a generic term of cause-effect analysis, has been used in different contexts in data management research [51, 52, 67, 74]. This includes data discovery [22, 28, 32, 76, 96], data cleaning [62, 75], query result explanation [47, 69, 73, 95, 97], hypothetical reasoning [21], and large system diagnostics [5, 29, 48]. We use the interventional notion of causal inference on observational data (also called observational study) from AI and Statistics [60, 71] to define our prescription rules (more in Sections 3 and 4.3). to design interventions that improve an outcome of interest.

Aggregate query result explanation. A substantial line of research has focused on aggregate SQL query result explanations. Multiple works have utilized *data provenance* to provide explanations for query results [10, 13, 15, 42, 43, 51, 52, 85]. Other explanation methods include (non-causal) interventions [17, 68, 69, 84, 90], and counterbalancing patterns [54]. Recent work [47, 73, 95, 97] has proposed methods that use causal inference to explain query results.

Multi dimensional data aggregation. Previous work on multidimensional data aggregation developed methods that extend the traditional drill-down and roll-up operators to find the most interesting data parts for exploration [3, 34, 77, 93]. Other works have focused on assessing the similarity between data cubes [9]), or discovering intriguing data visualizations [87, 101].

Part of our goal is to identify subpopulations for which we can generate recommendations. We utilize existing solutions whenever

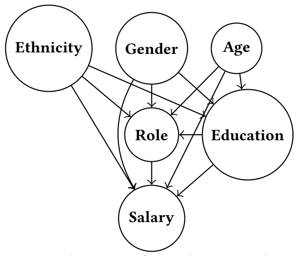

Figure 1: Partial causal DAG for the Stack Overflow dataset.

applicable (e.g., we use the Apriori algorithm [4] to find sufficiently large subpopulations) and develop novel methods when necessary.

3 BACKGROUND ON CAUSAL INFERENCE

In this section, we review the basic concepts and key assumptions for inferring the effects of an intervention on the outcome on collected datasets without performing randomized controlled experiments. We use *Pearl's graphical causal model* for *observational causal analysis* [60] to define these concepts.

Causal Inference and Causal DAGs. The primary goal of causal inference is to model causal dependencies between attributes and evaluate how changing one variable (referred to as intervention) would affect the other. Pearl's Probabilistic Graphical Causal Model [60] can be written as a tuple $(\mathbf{U}, \mathbf{V}, Pr_{\mathbf{U}}, \psi)$, where $\mathbf{U}$ is a set of *exogenous* variables, $\Pr_{\mathbf{U}}$ is the joint distribution of $\mathbf{U}$, and $\mathbf{V}$ is a set of observed *endogenous variables*. Here ψ is a set of structural equations that encode dependencies among variables. The equation for $A \in \mathbf{V}$ takes the following form:

$$\psi_A : \mathtt{dom}(Pa_{\mathbf{U}}(A)) \times \mathtt{dom}(Pa_{\mathbf{V}}(A)) \rightarrow \mathtt{dom}(A)$$

Here $Pa_{\mathbf{U}}(A) \subseteq \mathbf{U}$ and $Pa_{\mathbf{V}}(A) \subseteq \mathbf{V} \setminus \{A\}$ respectively denote the exogenous and endogenous parents of A. A causal relational model is associated with a directed acyclic graph (*causal DAG*) G, whose nodes are the endogenous variables $\mathbf{V}$ and there is a directed edge from X to O if $X \in Pa_{\mathbf{V}}(O)$. The causal DAG obfuscates exogenous variables as they are unobserved. The probability distribution $\Pr_{\mathbf{U}}$ on exogenous variables $\mathbf{U}$ induces a probability distribution on the endogenous variables $\mathbf{V}$ by the structural equations ψ. A causal DAG can be constructed by a domain expert as in the above example, or using existing *causal discovery* algorithms [27].

Example 3.1. *Figure 1 depicts a partial causal DAG for the SO dataset over the attributes in Table 1 as endogenous variables (we use a larger causal DAG with all 20 attributes in our experiments). Given this causal DAG, we can observe that the role that a coder has in their company depends on their education, age gender and ethnicity.*

Intervention. In Pearl's model, a treatment $T = t$ (on one or more variables) is considered as an *intervention* to a causal DAG by mechanically changing the DAG such that the values of node(s) of T in G are set to the value(s) in t, which is denoted by $\mathtt{do}(T = t)$. Following this operation, the probability distribution of the nodes in the graph changes as the treatment nodes no longer depend on the values of their parents. Pearl's model gives an approach to estimate the new probability distribution by identifying the confounding factors Z described earlier using conditions such as *d-separation* and *backdoor criteria* [60], which we do not discuss in this paper.

Average Treatment Effect. The effects of an intervention are often measured by evaluating *Conditional Average treatment effect (CATE)*, measuring the effect of an intervention on a subset of records [31, 70] by calculating the difference in average outcomes between the group that receives the treatment and the group that does not (called the *control* group), providing an estimate of how the intervention by T influences an outcome O for a given subpopulation. Given a subset of the records defined by (a vector of) attributes B and their values b, we can compute $CATE(T, O \mid B = b)$ as:

$$\mathbb{E}[O \mid \mathtt{do}(T = 1), B = b] - \mathbb{E}[O \mid \mathtt{do}(T = 0), B = b] \tag{1}$$

Setting $B = \phi$ is equivalent to the ATE estimate. The above definitions assumes that the treatment assigned to one unit does not affect the outcome of another unit (called the Stable Unit Treatment Value Assumption (SUTVA)) [71][1].

The ideal way of estimating the ATE and CATE is through *randomized controlled experiments*, where the population is randomly divided into two groups (treated and control, for binary treatments): denoted by $\mathtt{do}(T = 1)$ and $\mathtt{do}(T = 0)$ respectively) [60]. However, randomized experiments cannot always be performed due to ethical or feasibility issues (e.g., the effect of smoking on lung cancer). In these scenarios, observational data is used to estimate the treatment effect, which requires the following additional assumptions.

The first assumption is called *unconfoundedness* or *strong ignorability* [66] says that the independence of outcome O and treatment T conditioning on a set of confounder variables (covariates) Z, i.e., $O \perp\!\!\!\perp T | Z{=}z$. The second assumption called *overlap or positivity* says that there is a chance of observing individuals in both the treatment and control groups for every combination of covariate values, i.e., $0 < Pr(T{=}1 \mid Z{=}z) < 1$.

The unconfoundedness assumption requires that the treatment T and the outcome O be independent when conditioned on a set of variables Z. In SO, assuming that only Z ={`Gender`, `Age`, `Country`} affects T = `Education`, if we condition on a fixed set of values of Z, i.e., consider people of a given gender, from a given country, and at a given age, then T = `Education` and O = `Salary` are independent. For such confounding factors Z, Eq. (1) reduces to the following form (positivity gives the feasibility of the expectation difference):

$$CATE(T, O|B{=}b) {=} \mathbb{E}_Z \left[\mathbb{E}[O|T{=}1, B{=}b, Z{=}z] - \mathbb{E}[O|T{=}0, B{=}b, Z{=}z] \right]$$

This equation contains conditional probabilities and not $\mathtt{do}(T = b)$, which can be estimated from an observed data. Pearl's model gives a systematic way to find such a Z when a causal DAG is available.

4 PROBLEM FORMULATION

We consider a single-relation database over a schema $\mathbb{A}$. The schema is a vector of attribute names, i.e., $\mathbb{A}{=}(A_1, \ldots, A_s)$, where each A_i is associated with a domain $\mathtt{dom}(A_i)$, which can be categorical or continuous. A database instance D, populates the schema with a set of tuples $t{=}(a_1, \ldots, a_s)$ where $a_i \in \mathtt{dom}(A_i)$. We use $t[A_i]$ to denote the value of attribute A_i of tuple t.

Our high-level goal is to return a set of *prescription rules* (ruleset) with certain desired properties including fairness. In this section, first we define patterns on attributes, protected groups, prescription rules, and discuss the desired properties, finally defining our optimization problem in Section 4.7.

[1]This assumption does not hold for causal inference on multiple tables and even on a single table where tuples depend on each other.

4.1 Pattern and Protected Group

To define the notion of *prescription rules*, we build upon the commonly used concept of *patterns* [19, 45, 69, 90]. A pattern comprises *conjunctive predicates* on attribute values, and is defined as follows:

Definition 4.1 (Pattern). *Given a database instance D with schema $\mathbb{A}$, a predicate is an expression of the form $\varphi{=}A_i$ op a_i, where $A_i{\in}\mathbb{A}$, $a_i{\in}\mathsf{dom}(A_i)$, and* $\mathsf{op}{\in}\{=,\neq,<,>,\leq,\geq\}$. *A* pattern *is a conjunction of predicates $\mathcal{P}{=}\varphi_1 \wedge \ldots \wedge \varphi_k$.*

Example 4.1. *An example pattern over the Stack Overflow dataset (Table 1) is $\mathcal{P}$ ={*`Role = Designer` $\wedge$ `Country = US`*}. It defines a subset of the dataset comprised of designers from the US.*

Next we define *coverage* of a pattern $\mathcal{P}$ defined by the number of tuples from D that it captures.

Definition 4.2 (Coverage of a pattern). *Given a database instance D a pattern $\mathcal{P}$, and a tuple $t{\in}D$, $\mathcal{P}$ is said to* cover *t if t satisfies the predicates in $\mathcal{P}$. The subset of tuples in D covered by $\mathcal{P}$ is denoted by* Coverage$(\mathcal{P})$.

As is common in fairness research [12, 80, 98], we assume the presence of a protected group, defined by the pattern $\mathcal{P}_p$. The remainder of the population (i.e., $D \backslash \mathcal{P}_p(D)$) is referred to as the non-protected group. A protected group in the Stack Overflow dataset may be defined based on sensitive attributes such as age or ethnicity. For instance, it could be defined as $\mathcal{P}_p$ = { `Ethnicity≠ White`} to refer to non-white individuals.

4.2 Prescription Rules

A *prescription rule* outlines an *intervention* (*treatment*) designed to improve a target variable within a particular subpopulation. For example, a prescription rule might recommend that people under 25 pursue a Ph.D. to increase their salary. Before defining perception rules, we first discuss how attributes participate in such rules.

Mutable and immutable attributes. We assume the attributes $\mathbb{A}$ are partitioned into two disjoint sets. The first set contains the interventional attributes (e.g., programming language, education), which are the attributes that can be changed to improve the outcome. The second set contains immutable attributes (e.g., age, gender), which cannot be changed through prescription. Formally, let $\mathbf{I} \subseteq \mathbb{A}$, denote the set of immutable attributes and $\mathbf{M} \subseteq \mathbb{A}$ denote the set of mutable (interventional) attributes, where $\mathbf{M} \cap \mathbf{I} = \emptyset$ and the outcome $O \notin \mathbf{M} \cup \mathbf{I}$. This categorization intends to prohibit the infeasible or impractical recommendations to increase the outcome (e.g., changing one's age or ethnicity to improve one's income).

Our prescription rules defined below as a combination of grouping and intervention patterns are motivated by the causal explanations defined in [97]. However, the focus of this paper is to study the interplay between utility and fairness of a ruleset that was not considered in [97]. As a result, the specific objectives and optimization problem are defined differently as discussed next.

Definition 4.3 (Prescription Rule and Ruleset, Grouping and intervention patterns, and Coverage). *Given a database D with mutable attributes $\mathbf{M}$ and immutable attribute $\mathbf{I}$, a* prescription rule *r is a pair of patterns $r = (\mathcal{P}_{\mathtt{grp}}, \mathcal{P}_{\mathtt{int}})$, where (1) $\mathcal{P}_{\mathtt{grp}}$ is called the* grouping pattern *and consists exclusively of the immutable attributes in $\mathbf{I}$, and (2) $\mathcal{P}_{\mathtt{int}}$ is called the* intervention pattern *and consists exclusively of the mutable attributes in $\mathbf{M}$.*

By overloading notations, Coverage(r) = Coverage$(\mathcal{P}_{\mathtt{grp}})$ *is called the* coverage of rule *r since it captures the subset of tuples in D on which the rule r applies, i.e., each tuple $t \in D$ is either* covered *or not by $\mathcal{P}_{\mathtt{grp}}$ of rule r. $\mathcal{P}_{\mathtt{int}}$ defines the recommended intervention in the prescription rule aimed at betterment of the outcome O for the subgroup that* Coverage(r) *defines.*

Given a set of prescription rules R (called a ruleset), Coverage$(R) = \cup_{r \in R}$Coverage(r), *i.e., coverage of a ruleset corresponds to the subset of tuples in D that are covered by at least one of the rules in R.*

For a prescription rule $r = (\mathcal{P}_{\mathtt{grp}}, \mathcal{P}_{\mathtt{int}})$, the intervention pattern $\mathcal{P}_{\mathtt{int}}$ partitions the tuples defined by $\mathcal{P}_{\mathtt{grp}}$ into treated ($T = 1$ if $\mathcal{P}_{\mathtt{int}}$ evaluates to true for a tuple) and control groups ($T = 0$ if $\mathcal{P}_{\mathtt{int}}$ evaluates to false). This partition is then used to assess the causal effects of the intervention $\mathcal{P}_{\mathtt{int}}$ on the outcome O within the subpopulation Coverage(r) which the rule r applies to.

Example 4.2. *An example prescription rule suggests that individuals aged 25-34 with dependents, should work as front-end developers. ($\mathcal{P}_{\mathtt{grp}}$:* `age = 25-34` $\wedge$ `dependents = yes`*), the intervention is working as front-end developers ($\mathcal{P}_{\mathtt{int}}$:* `role = frontend developer`*). The expected CATE value is $44,009, namely, the expected salary increase for a 25-34-year-old individual with dependents working as a frontend developer is $44,009 per year (compared to a 25-34-year-old individual with dependents working in a different role).*

4.3 Utility of a Prescription Ruleset

Utility of a single rule: To evaluate the effectiveness of a prescription rule $r = (\mathcal{P}_{\mathtt{grp}}, \mathcal{P}_{\mathtt{int}})$ toward improving the outcome O, we define its utility. The utility measures the expected impact (as CATE) of the recommended intervention on the outcome O within the subpopulation Coverage(r) where r applies to. We define the overall utility, and utility for the protected and non-protected groups.

Definition 4.4 (Utility of a prescription rule - overall, protected, non-protected). *Given a database instance D with schema $\mathbb{A}$, an outcome variable O, a causal model $\mathcal{G}_D$ on $\mathbb{A}$, a protected group p defined by a pattern $\mathcal{P}_p$, and a prescription rule $r = (\mathcal{P}_{\mathtt{grp}}, \mathcal{P}_{\mathtt{int}})$,*

(1) the overall utility *of r is defined as:*

$$utility(r) {:=} CATE_{\mathcal{G}_D}(\mathcal{P}_{\mathtt{int}}, O \mid \mathcal{P}_{\mathtt{grp}}) \tag{2}$$

(2) the utility of r for the protected group p *is defined as:*

$$utility_p(r) {:=} CATE_{\mathcal{G}_D}(\mathcal{P}_{\mathtt{int}}, O \mid \mathcal{P}_{\mathtt{grp}} \wedge \mathcal{P}_p) \tag{3}$$

(3) the utility of r for the non-protected group p *is defined as:*

$$utility_{\bar{p}}(r) {:=} CATE_{\mathcal{G}_D}(\mathcal{P}_{\mathtt{int}}, O \mid \mathcal{P}_{\mathtt{grp}} \wedge \mathcal{P}_{\neg p}) \tag{4}$$

The subscript $\mathcal{G}_D$ denotes that the CATE is estimated using the causal model, and we drop the subscript when it is clear from context.

If Coverage(r) = Coverage$(\mathcal{P}_{\mathtt{grp}}) = \emptyset$, *i.e., if the rule does not apply to any tuple in D, then we assume that $utility(r) = 0$; similarly $utility_p(r) = 0$ if* Coverage$(\mathcal{P}_{\mathtt{grp}} \wedge \mathcal{P}_p) = \emptyset$, *and $utility_{\bar{p}}(r) = 0$ if* Coverage$(\mathcal{P}_{\mathtt{grp}} \wedge \mathcal{P}_{\neg p}) = \emptyset$.

The goal of prescription rules is to improve the outcome O as desired. If the goal is to increase the outcome O (e.g., increase salary), we discard rules with negative utility, as they do not help achieve this objective. Similarly, if the aim is to decrease the outcome, we ignore rules with negative utility. Throughout the paper, without loss of generality, we assume that the goal is to increase the outcome, thereby focusing on maximizing utility.

Prescription to individuals when multiple rules apply: When dealing with a ruleset R, it is possible for multiple rules to apply to the same subpopulation. Specifically, if two rules r_i=$(\mathcal{P}^i_{\texttt{grp}}, \mathcal{P}^i_{\texttt{int}})$ and r_j=$(\mathcal{P}^j_{\texttt{grp}}, \mathcal{P}^j_{\texttt{int}})$∈$R$ share a non-empty intersection between their coverage, namely $\textsc{Coverage}(\mathcal{P}^i_{\texttt{grp}})\cap\textsc{Coverage}(\mathcal{P}^j_{\texttt{grp}})\neq\emptyset$, then the subpopulation defined by the pattern $\mathcal{P}^i_{\texttt{grp}}\wedge\mathcal{P}^j_{\texttt{grp}}$ will have more than one relevant rule. In our definition below for utility of a ruleset, we refrain from applying more than one prescription rule to a subpopulation for two reasons. First, two rules may conflict with each other. For instance, if one rule suggests individuals above 25 to earn a Ph.D., while another recommends women over 20 pursue an MBA, women above 25 would receive conflicting recommendations (whether to do a Ph.D. or an MBA). Second, CATE is known to be non-monotonic [97], implying that appending a predicate to an intervention pattern can either increase or decrease the CATE value. Therefore, employing multiple rules simultaneously for a subpopulation might yield a utility gain smaller than the individual rules. Hence when multiple rules apply to a tuple in D, we assume that only one is chosen by the decision-maker, and we leave it to the application scenario which rule is chosen.

Utility of a ruleset: For a prescription ruleset R, we use its *expected utility* on D as the utility of R.

Definition 4.5 (Expected Utility of a Ruleset). *The* expected utility *of a prescription ruleset R is defined as the average maximum utility of an individual from $\textsc{Coverage}(R)$ from the rules in R that applies to the individual, i.e.,*

$$ExpUtility(R) = \frac{1}{n}\sum_{t\in\textsc{Coverage}(R)}\max_{r\in R_t}(utility(r)) \quad (5)$$

where $R_t \subseteq R$ denotes the set of rules covering the tuple t, and $n = |D|$. Note that if a rule does not apply to a tuple, its utility is zero, so the sum above is also over all $t \in D$.

Given a protected pattern $\mathcal{P}_p$, the expected utility for the protected and non-protected groups are defined as follows:

$$ExpUtility_p(R) = \frac{1}{n_p}\sum_{t\in\textsc{Coverage}_p(R)}\min_{r\in R_t}(utility(r)) \quad (6)$$

$$ExpUtility_{\bar{p}}(R) = \frac{1}{n_{\bar{p}}}\sum_{t\in\textsc{Coverage}_{\bar{p}}(R)}\max_{r\in R_t}(utility(r)) \quad (7)$$

where $\textsc{Coverage}_p(R)$ denotes the set of protected individuals covered by R and $n_p = |\textsc{Coverage}_p(R)|$ (similarly $\textsc{Coverage}_{\bar{p}}()$ and $n_{\bar{p}}$).

Note the difference between formulas (6) and (5, 7). Since we do not assume any restriction on which rule is chosen for a tuple when multiple rules apply, we do a conservative worst-case analysis on fairness. We assume that protected individuals choose the worst possible rule, while the rest choose the best possible one. This ensures that the expected utility for the protected group in reality (irrespective of the rule chosen for each protected tuple) will be at least as high as the expected utility from the least beneficial relevant prescription rule for this group.

4.4 Size of a Prescription Ruleset

The size of a prescription ruleset is the number of rules in that set, denoted by $size(R)$. Ideally, we want to find a small-size ruleset. The intuition is that, the fewer the rules in a set, the easier it is to understand the suggested interventions. Suppose we want to find a ruleset R with high utility without specifying a constraint on the size. The following lemma shows that the best strategy is to return the *optimal rule* that applies to each individual. That is, to maximize utility, prescribing a personalized rule for each individual may lead to the best utility. Specifically, we can show that for every rule $r = (\mathcal{P}_{\texttt{grp}}, \mathcal{P}_{\texttt{int}})$, there exists a subgroup $g' \subseteq \textsc{Coverage}(\mathcal{P}_{\texttt{grp}})$ and a intervention $\mathcal{P}_{t'}$ (not necessarily $\mathcal{P}_{\texttt{int}}$) s.t the utility of the rule $r' = (g', \mathcal{P}_{t'})$ is greater than that of the original rule r (proved in the full version [6]).

Lemma 4.1. *Given a rule r=$(\mathcal{P}_{\texttt{grp}}, \mathcal{P}_{\texttt{int}})$, there exists a rule r'=$(\mathcal{P}_{g'}, \mathcal{P}_{t'})$ s.t $\mathcal{P}_{g'} \subset \textsc{Coverage}(\mathcal{P}_{\texttt{grp}})$ and $utility(r') \geq utility(r)$.*

This property implies that the number of prescription rules in the optimal solution is $O(|D|)$, making it impractical to implement in real-world scenarios. For instance, consider a policy enacted by a government official to allocate healthcare resources based on patient data. If the number of rules scales linearly with the size of the dataset, it would become infeasible to apply the policy effectively across a large population. Therefore, we limit the number of recommended rules. One approach is to impose a strict limit on the number of rules selected. However, pre-setting this constraint often requires tuning to balance utility and comprehensibility. Therefore, we incorporate the number of rules as an objective, considering rulesets with fewer rules to be more desirable, as was done in [40].

4.5 Coverage Constraints

We consider two types of coverage constraints: *group coverage*, where the goal is to find a solution that covers a predefined fraction of protected individuals and a certain fraction of the entire population, and *rule coverage*, where every selected rule must cover a certain fraction of the population and protected individuals.

Group Coverage Given two thresholds $\theta, \theta_p \in [0, 1]$, we say that a ruleset R satisfies the group coverage constraint if R covers at least a θ fraction of the population, and a θ_p fraction of the protected subpopulation. Formally, both conditions are satisfied: (i) $Coverage(R) \geq \theta \cdot |D|$, (ii) $Coverage_p(R) \geq \theta_p \cdot |\mathcal{P}_p(D)|$, where $Coverage_p(R)$ denotes the number of covered protected individuals by R.

Rule Coverage Given two thresholds $\theta, \theta_p \in [0, 1]$, we say that a ruleset R satisfies the rule coverage constraint if every rule $r \in R$ covers at least a θ fraction of the population, and at least a θ_p fraction of the protected subpopulation. Formally, both of the following conditions hold: (i) For every $r \in R$: $coverage(r) \geq \theta \cdot |D|$, (ii) For every $r \in R$: $coverage_p(r) \geq \theta_p \cdot |\mathcal{P}_p(D)|$, where $coverage_p(r)$ denotes the number of covered protected individuals by r.

4.6 Fairness Constraints

We study two definitions of fairness: statistical parity (SP) [49], and bounded group loss (BGL) [2]. Those definitions are based on equivalent notions of fairness for regression tasks [2]. We next provide an extension for these definitions to causal estimates.

Group and individual fairness are two key concepts in algorithmic fairness [11, 25, 80]. Group fairness aims to ensure that different groups receive similar outcomes. Individual fairness focuses on treating similar individuals similarly, meaning that if two individuals are alike in relevant aspects, they should receive similar outcomes. Both approaches aim to reduce bias, with the choice of which approach to adopt depending on the specific context. Next, we present four types of fairness constraints: SP and BGL, each of which can be applied to ensure group or individual fairness.

4.6.1 Statistical parity. In SP, the goal is to ensure that the gain in the utility of a protected individual is similar to that of any individual from the non-protected group.
Group Fairness: Intuitively, if we randomly sample a protected individual, the expected gain should be almost the same as that of an individual from the non-protected group. Formally: $|\text{ExpUtility}_p(R) - \text{ExpUtility}_{\bar{p}}(R)| \leq \epsilon$, where $\epsilon > 0$ is a threshold.

Individual Fairness: Individual fairness says that the expected gain of every protected individual is similar to that of an individual from the non-protected group. That means that the expected utility of each rule $r \in R$ on a protected individual should be similar to that of an individual from the non-protected group. Formally, for every $r \in R$, $|utility_p(r) - utility_{\bar{p}}| \leq \epsilon$, where $\epsilon > 0$ is a threshold.

4.6.2 Bounded group loss (BGL). : Fair regression with BGL minimizes the overall loss while controlling the worst loss in the protected group [2]. In our context, this translates to the following constraint: When selecting an individual from the protected group, the utility increase should exceed a specified threshold $\tau \geq 0$.
Group Fairness: We aim to ensure that the expected utility of a randomly sampled protected individual within $Coverage(R)$ is above a given threshold τ. Formally, $\text{ExpUtility}_p(R) \geq \tau$.

Individual Fairness: We aim to ensure that the gain of every protected individual from $Coverage(R)$ exceeds a threshold τ. Therefore, a ruleset R satisfies the individual loss constraint if the utility of every rule r on protected individuals is at least τ. Formally, for every rule $r \in R$, $utility_p(r) \geq \tau$.

4.7 The Prescription Ruleset Selection Problem

We are finally ready to present the problem we study in this paper. If we did not have *fairness or coverage constraints*, then our goal is to select a small-size perception ruleset with high expected utility. However, as demonstrated in the introduction, not considering fairness constraints may result in a ruleset that are only highly beneficial to a small, non-protected subset of the population. Therefore, we extend our problem definition to include coverage and fairness constraints. We can apply any of SP or BGL group or individual fairness constraints (Section 4.6), as well as rule or group coverage constraints Section 4.5), along with no fairness or coverage constraints, resulting in 18 distinct problem variants. The choice of which constraints to apply is left to the user as it may be application-dependent, and is discussed below. To define the generic problem, we use $R \models \mathcal{F}$ and $R \models C$ to denote that a ruleset R satisfies a given fairness constraint $\mathcal{F}$ and a given coverage constraint C respectively (if no constraints are given, these conditions are trivially satisfied). We assume that a set of candidate rules $\{r_i\}_{i=1}^{l}$ has been already mined and is available as an input to the problem.

Definition 4.6 (Prescription Ruleset Selection under fairness and coverage constraints). *Given a database D, a causal model $\mathcal{G}_D$, an outcome attribute O, a fairness constraint $\mathcal{F}$, a coverage constraint C, and a collection of prescription rules $\{r_i\}_{i=1}^{l}$, a subset $R \subseteq \{r_i\}_{i=1}^{l}$ of prescription rules is called* valid *if (1) $R \models \mathcal{F}$ and (2) $R \models C$. The goal is to find a valid subset of rules $R^* \subseteq \{r_i\}_{i=1}^{l}$ s.t*

$$R^* = argmax_{R \subseteq \{r_i\}_{i=1}^{l}} \left[\lambda_1 \cdot (l - size(R)) + \lambda_2 \cdot ExpUtility(R)\right] \quad (8)$$

where λ_1, λ_2 are non-negative weights that scale the relative influence of each objective.

Here λ_1 and λ_2 may be application-dependent or they can be tuned by the user. The above optimization problem, as expected, is NP-hard even for simple variants, although some constraints are matroid constraints and therefore are amenable to greedy approaches (discussions and proofs in the full version [**?**]). Therefore, we obtain efficient algorithms that work well in practice in Section 5 and experimentally demonstrate the effect of different constraints on the results in Section 6.

Remark. We did not incorporate additional objectives to the Prescription Ruleset Selection problem related to the *overlap between rules* and their generality in terms of the *number of predicates defining the patterns*. Our initial empirical study revealed that these objectives are effectively addressed by the coverage constraints. To simplify the problem, we decided to omit them.

Selecting the appropriate problem variant. Since we have several options for fairness and coverage constraints, one natural question is which version to use. We observe that there is no one-size-fits-all solution and the best choice depends on the specific application and the constraints at hand. For instance, going for individual fairness gives a stronger fairness guarantee at the expense of possible lower utility to everyone. In addition, the complexity of different versions can vary. To assist in making this decision, we summarize the process through a decision tree that guides users in selecting the most suitable variant for their needs, presented in Figure 2. The decision to choose between SP or BGL fairness is left to the user. In Section 6, we present a case study that empirically compares the obtained rulesets under different coverage and fairness constraints, demonstrating the trade-off between coverage, fairness and utility.

5 THE FAIRCAP ALGORITHM

A brute-force approach, which considers all grouping and intervention patterns to form prescription rules, results in long runtimes (as we show in Section 7.3). Instead, we propose a more efficient algorithmic framework, called FairCap (Fair CAusal Prescription), which avoids generating every possible prescription rule. FairCap can be adapted for any variant of the Prescription Ruleset Selection problem, including group or individual fairness and rule or group coverage constraints. To simplify the presentation, we first describe FairCap for the Prescription Ruleset Selection problem with SP group fairness and group coverage constraints. We then explain how it can be modified to accommodate other variants.

Our algorithmic framework is outlined in Algorithm 1. FairCap consists of three parts: (1) generating grouping patterns by using the Apriori algorithm [4] (line 2), (2) identifying promising intervention

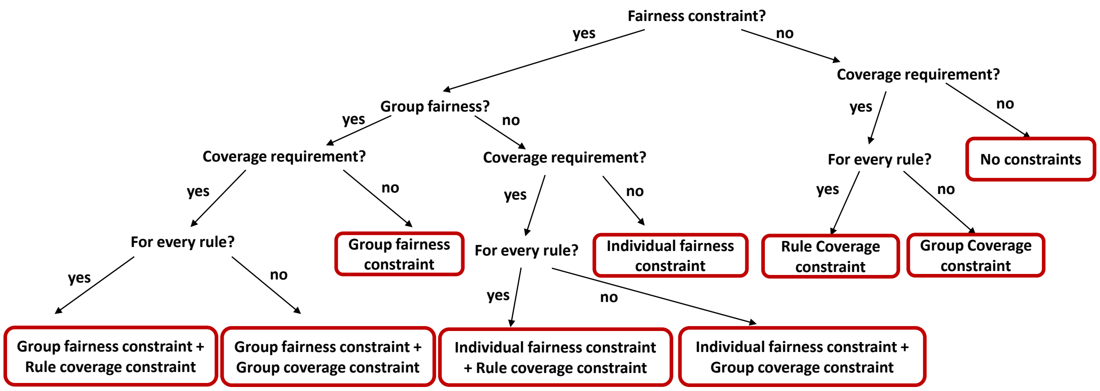

Figure 2: A decision tree for selecting the appropriate problem variant.

Algorithm 1: The FairCap algorithmic framework.

input : A database relation D, a protected group defined by the pattern $\mathcal{P}_p$ and an outcome variable O
output : A set Φ of prescription rules.

$\Phi \leftarrow \emptyset$;
$\mathcal{G} \leftarrow$ `GetGroupingPatterns`(D, O); `// Section 5.1`
for $\mathcal{P}_g \in \mathcal{G}$ **do**
  $\mathcal{P}_t \leftarrow$ `GetIntervention`$(\mathcal{P}_g, O, \mathcal{P}_p, D)$; `// Section 5.2`
  $\Phi \leftarrow \Phi \cup (\mathcal{P}_g, \mathcal{P}_p)$
$\Phi \leftarrow$ `ApplyGreedy`$(\Phi, O, \mathcal{P}_p)$; `// Section 5.3`
return Φ

patterns for each grouping pattern by using a lattice traversal approach [7], and (3) finding a set of prescription rules using a greedy approach. We leverage existing solutions (e.g., [4, 7, 59, 97]) where applicable, and develop novel techniques where necessary. Specifically, the first step follows the same approach as CauSumX [97], while the second and third steps introduce novel methods.

Note that FairCap lacks theoretical guarantees due to its design, which avoids generating all possible rules (as their number grows exponentially with the database size). If steps 1 and 2 were replaced by a brute-force approach that generates all rules, then a greedy approach for selecting a ruleset could approximate the optimal solution for certain problem variants, as the objective is a non-negative, monotone submodular function (even with a rule coverage or individual fairness constraints which are matroid constraints). However, other constraints are harder to satisfy. Future work will explore the complexity of various problem variants and establish theoretical bounds for finding approximate solutions.

5.1 Step 1: Mining Grouping Patterns

Considering every possible grouping pattern is infeasible as their number is exponential ($O(agrmax_{A_i \in \mathbb{A}} |\mathsf{dom}(A_i)|^{|\mathbb{A}|}$). Instead, as done in previous work [59, 97], we utilize the Apriori algorithm [4] to generate candidate patterns. The Apriori algorithm gets a threshold τ, and ensures that the mined grouping patterns are present in at least τ tuples of D. The algorithm guarantees that each mined pattern covers at least τ tuples from D, making them promising candidates for covering many tuples from D.

5.2 Step 2: Mining Intervention Patterns

Our next goal is to identify an intervention pattern $\mathcal{P}_{\mathtt{int}}$ for each mined grouping pattern $\mathcal{P}_{\mathtt{grp}}$ that maximizes utility (i.e., treatments with the highest CATE for $\mathcal{P}_{\mathtt{grp}}$) while ensuring fairness to the protected group. Unlike step 1, this step requires novel techniques for finding treatments that are both fair and have high utility.

Since the number of potential intervention patterns for $\mathcal{P}_{\mathtt{grp}}$ can be large (exponential in $|\mathbb{A}|$), we employ a greedy lattice-traversal [7, 16] approach, inspired by [59, 97]. This allows us to materialize and assess the CATE only for promising patterns.

Concretely, the space of all intervention patterns can be represented as a lattice where nodes correspond to patterns and there is an edge between $\mathcal{P}^1_{\mathtt{int}}$ and $\mathcal{P}^2_{\mathtt{int}}$ if $\mathcal{P}^2_{\mathtt{int}}$ can be obtained from $\mathcal{P}^1_{\mathtt{int}}$ by adding a single predicate. This lattice can be traversed in a top-down fashion. Since not all nodes correspond to treatments with a positive CATE, we only materialize nodes if all their parents have a positive CATE. We note that this might lead the algorithm to overlook certain relevant intervention patterns. However, as shown in [97], combining patterns that exhibit a positive CATE is highly likely to result in an intervention with a positive CATE as well.

When a group fairness constraint is imposed, instead of searching for the treatment with the highest CATE, we search for the treatment that is "fair" by that it maximizes CATE for both protected and non-protected groups, while minimizing disparities.

To identify the most fair treatment, we define the *benefit* of an intervention pattern as follows. Intuitively, we penalize the treatment based on the difference between the utility for the non-protected group and the utility provided to the protected group. The larger the difference, the lower the benefit of the treatment. Formally, the benefit of a rule $r = (\mathcal{P}_{\mathtt{grp}}, \mathcal{P}_{\mathtt{int}})$ is defined as:

$$benefit(r) = \begin{cases} \frac{utility(r)}{1+utility_{\bar{p}}(r)-utility_{p}(r)}, & \text{if } utility_{\bar{p}}(r) \geq utility_{p}(r) \\ utility(r), & \text{otherwise} \end{cases}$$

We implement two optimizations to improve efficiency: **(i)** we discard attributes that do not have a causal relationship with the outcome, since such attributes have no impact on CATE values. We can detect such attributes by utilizing the input causal DAG. **(ii)** The process of extracting intervention patterns for each grouping pattern can be performed in parallel since this procedure is dependent only on the grouping pattern.

5.3 Step 3: A Greedy Approach

The final step involves finding a solution from the rules mined in Steps 1 and 2. We propose a greedy algorithm that optimizes the problem's objectives. Intuitively, the algorithm operates as follows:

at each iteration, it selects the next best rule that maximizes expected utility, benefit (as defined in Section 5.2), and coverage. Once the coverage constraints are met, the focus shifts to maximizing benefit and utility. The algorithm stops when the additional gain becomes negligible, as the number of rules is not predetermined.

Formally, the next best rule is determined as follows. Let $\{r_j\}_{j=1}^{l}$ denote the candidate rules and R_i is the ruleset selected in the first i iterations. The score of a rule r w.r.t R_i is defined as:

$$score(r) :=$$

$$Coverage(R_i \cup \{r_{i+1}\}) + benefit(R_i \cup \{r_{i+1}\}) + ExpUtility(R_i \cup \{r_{i+1}\})$$

The next best rule r_{i+1} to add in case the coverage constraints are not met yet is defined as:

$$r^*_{i+1} = argmax_{r_{i+1} \in \{r_j\}_{j=1}^{l} \setminus R_i} score(r_{i+1})$$

In case the coverage constraints are met, ignore the coverage term. The algorithm stops at the first iteration i where the score of the selected rule r_i falls below a predefined threshold, indicating that the marginal gain from r_i is negligible.

5.4 Adjustments to Other Variants

We explain how FairCap can be adjusted to solve other problem variants. We set the Apriori's threshold to ensure that each mined grouping pattern covers a sufficient number of individuals when a rule coverage constraint is imposed (step 1). Without coverage constraints, the Apriori threshold can be set to any value.

Without fairness constraints, in Step 2, the goal is to identify the intervention with the highest CATE value (as was done in [97]). Whit an individual fairness constraint, each rule must satisfy this constraint, so we only select interventions that are guaranteed to meet the constraint while maximizing CATE (step 2).

In case a group BGL fairness constraint is imposed, we define the benefit of a rule $r = (\mathcal{P}_{\texttt{grp}}, \mathcal{P}_{\texttt{int}})$ as follows. Intuitively, we penalize the treatment based on the difference between the minimum required utility for the protected group and the utility provided to the protected group by this treatment. The larger the difference, the lower the benefit of the treatment. Formally:

$$benefit(r) = \begin{cases} \frac{utility(r)}{1+\tau-utility_p(r)}, & \text{if } \tau \geq utility_p(r) \\ utility(r), & \text{otherwise} \end{cases}$$

where τ is the threshold for the BGL fairness constraint. This benefit definition is applied in Step 2 of the algorithm to identify fair and effective treatments for the mined grouping patterns.

Runtime complexity analysis: The maximum number of rules in a database D with attributes $\mathbb{A}$ is bounded by $|D|^{|\mathbb{A}|}$ (considering both grouping and intervention patterns and the active domain of attributes), which is polynomial in terms of *data complexity*, assuming a fixed schema [86]. The final greedy step is also polynomial in the number of rules considered. Additional operations, such as calculating CATE values, are polynomial in D, leading to a worst-case polynomial data complexity. As we demonstrate in Section 7.3, our algorithm is capable of efficiently handling large datasets..

6 CASE STUDY

The objective of this case study is to empirically evaluate the impact of various constraints on the solution for the Prescription Ruleset Selection problem. We analyze two datasets, (1) German Credit (German in short) and (2) Stack Overflow (SO in short), each with a corresponding protected group, and assess the rules chosen by FairCap under different constraints. We aim to understand how these constraints influence coverage, utility, and disparities (for fairness) between protected and non-protected groups.

Table 3: Examined datasets.

Dataset	Tuples	Atts	Mutable Atts	Protected Group
SO [1]	38K	20	10	People from countries with a low GDP
German Credit [8]	1000	20	15	Single Females

Datasets. We examine two commonly used datasets: (1) Stack Oveflow (SO) [1], as described in Example 1.1. Here, the goal is to increase salary. (2) German Credit [8], which contains details of bank account holders, including demographic and financial information. Here, the goal is to increase the credit score (binary). The corresponding causal DAG was constructed using [96]. The datasets' statistics are presented in Table 3. The protected groups were selected to represent subgroups where the desired outcome was relatively low and sufficiently large (approximately 30% of the population) to ensure the discovery of statistically significant rules.

Default parameters. Unless otherwise specified, the threshold of the Apriori algorithm is set to 0.1. For the SO dataset, the coverage thresholds are set to 0.5. The threshold for the SP and BFL fairness constraint is set at \$10k. For the German dataset, the coverage thresholds are set at 30% and the fairness thresholds are set at 0.1. This configuration allows for the generation of multiple rules.

Result Summary:

⋆ Fairness may come at the cost of expected utility for everyone.
⋆ Achieving individual fairness is harder than group fairness, as most rules are unfair.
⋆ Achieving rule coverage is harder than group coverage, as many useful rules apply only to a small fraction of the population.
⋆ Without fairness constraints, we observe a significant disparity in the expected utility between the protected and non-protected.
⋆ With SP fairness constraints, the difference in expected utility between protected and non-protected is bounded.
⋆ With BGL fairness constraints, which consider only the minimal gain for the protected without regard for non-protected, we may still observe a disparity between the two groups.

The results are shown in Table 4, illustrating the trade-off between utility, coverage, and fairness. Without constraints, the expected utility is substantially higher, but this comes at the expense of greater disparities between protected and non-protected groups (as indicated by the unfairness score — the difference between the expected utility of protected and non-protected). In the examined scenarios, coverage for both groups was achieved without constraints, but other protected group definitions may require them.

Stack Overflow. Observe that while the expected utility for both protected and non-protected groups reaches its highest value in the no-constraints variant, the unfairness score is very high. This indicates that achieving SP fairness requires compromising on the expected utility for both protected and non-protected groups. Interestingly, rule coverage and individual fairness are difficult to

Table 4: Comparison of Solutions in Terms of Size, Coverage, Expected Utility and Unfairness

Stack Overflow (SP fairness)	# rules	coverage	coverage pro	exp utility	exp utility non-pro	exp utility pro	unfairness
No constraints	20	99.91%	99.98%	**32634.2**	**32626.98**	**18432.66**	14194.32
Group coverage	20	99.84%	99.88%	32597.02	32595.1	18340.29	14254.81
Rule coverage	10	**99.99**%	**99.99**%	22301.77	22292.02	16604.92	**5687.1**
Group fairness	8	97.52%	97.81%	27870.77	27998.47	17998.66	9999.81
Individual fairness	20	**99.99**%	**99.99**%	28014.58	28256.35	14241.07	14015.28
Group coverage, Group fairness	11	97.95%	98.85%	27934.76	28144.58	18145.23	9999.35
Rule coverage, Group fairness	12	99.96%	99.89%	22284.1	22279.93	16594.77	5685.16
Group coverage, Individual fairness	20	99.74%	99.88%	28057.78	28284.25	15128.91	13155.34
Rule coverage, Individual fairness	13	**99.99**%	**99.99%**	18591.41	18606.68	12797.15	5809.53
German Credit (BGL fairness)	**# rules**	**coverage**	**coverage pro**	**exp utility**	**exp utility non-pro**	**exp utility pro**	**unfariness**
No constraints	17	**100.0%**	**100.0%**	**0.39**	**0.39**	0.27	0.12
Group coverage	18	**100.0%**	**100.0%**	**0.39**	**0.39**	0.3	0.09
Rule coverage	6	96.0%	**100.0%**	0.31	0.31	0.3	0.01
Group fairness	18	**100.0%**	**100.0%**	**0.39**	**0.39**	0.3	0.09
Individual fairness	20	**100.0%**	**100.0%**	0.37	0.37	0.23	0.14
Group coverage, Group fairness	6	**100.0%**	**100.0%**	0.36	0.37	**0.31**	0.06
Rule coverage, Group fairness	3	90.0%	**100.0%**	0.29	0.29	**0.31**	**-0.02**
Group coverage, Individual fairness	20	**100.0%**	**100.0%**	0.37	0.37	0.23	0.14
Rule coverage, Individual fairness	8	96.8%	**100.0%**	0.29	0.29	0.23	0.06

achieve, as most rules fail to meet these criteria. This leads to lower expected utility for all groups. On the other hand, group coverage and fairness constraints are easier to satisfy, as they offer more flexibility by allowing the selection of some unfair rules alongside those specifically designed for the protected group.

3 Selected Rules out of 11 for SO (SP group fairness):

▹($S1_a$) For individuals aged 24-34, pursue an undergraduate major in CS (expected utility protected: 10,292, expected utility non-protected: 22,586).
▹($S1_b$) For individuals with 6-8 years of coding experience, exercise 1-2 times per week, and pursue a bachelor's degree.(expected utility protected: 15,864, expected utility non-protected: 18,157).
▹($S1_c$) For males whose parents have a secondary school education, exercise 3-4 times per week, and work with a computer 9-12 hours a day (expected utility protected: 58,548, expected utility non-protected: 41,733).

3 Selected Rules out of 20 for SO (SP individual fairness):

▹($S2_a$) For males aged 18-24 with no dependents, exercise 3-4 times per week and pursue a bachelor's degree (expected utility protected: 17,957, expected utility non-protected: 20,021).
▹($S2_b$) For individuals aged 25-34, exercise 1-2 times per week, and work as back-end developers.(expected utility protected: 13,714, expected utility non-protected: 15,703).
▹($S2_c$) For individuals with dependents, exercise 3-4 times per week, and pursue an undergraduate major in CS (expected utility protected: 20,113, expected utility non-protected: 21,533).

We show above the three example rules selected under group fairness constraint. The first rule $S1_a$ is more advantageous for the non-protected group, the second ($S1_b$) benefits both protected and non-protected groups similarly, while the third rule ($S1_c$) is more beneficial for the protected group. Overall, all these rules together satisfy the group fairness requirement. We also present three example rules selected under individual fairness constraint. In this case, all rules ($S2_a, S2_b, S2_c$) are nearly equally beneficial for both groups, but the overall expected utility is lower. Finally, consider the three example rules selected with no constraints. Here, all rules ($S3_a, S3_b, S3_c$ in the figure below) favor the non-protected group, highlighting the importance of including fairness constraints.

3 Selected Rules out of 20 for SO (no fairness constraints):

▹($S3_a$) For White aged 25-34 with dependents, exercise 3-4 times per week and work as back-end developers (expected utility protected: 13,182, expected utility non-protected: 42,146).
▹($S3_b$) For males aged 25-34 with dependents, exercise 1-2 times per week, and work as front-end developers.(expected utility protected: 12,180, expected utility non-protected: 44,009).
▹($S3_c$) For male students, exercise 3-4 times per week, and pursue an undergraduate major in CS (expected utility for protected: 13,329, expected utility for non-protected: 25,578).

German. While the expected utility for both protected and non-protected peaks in the no-constraints variant, the unfairness score is relatively high. This suggests that achieving BGL fairness necessitates compromising utility for both groups. Notably, to reduce the unfairness, it is feasible to impose either a rule coverage constraint or a rule coverage constraint combined with group fairness. We show three rules selected under BGL group fairness constraints below. Since we are focusing on BGL fairness, which considers only the minimal gain for the protected group without regard for the gains of the non-protected group, we still observe a disparity between the two, even with a fairness constraint in place.

3 Selected Rules out of 20 for German (group BGL fairness):

▹($G1_a$) For people aged 24-30 with 0-2 dependents, maintain a minimum balance of 200 DM in the checking account and pursue skilled employment (expected utility protected: 0.26, expected utility non-protected: 0.35).
▹($G1_b$) For people seeking a loan to purchase furniture or equipment, maintain a minimum balance of 200 DM in the checking account (expected utility protected: 0.38, expected utility non-protected: 0.29).
▹($G1_c$) For people seeking a loan for an unspecified purpose, maintain a minimum balance of 200 DM in the checking account and own a house. (expected utility protected: 0.54, expected utility non-protected: 0.41).

Table 5: Comparison of Solutions in Terms of Fairness

Stack Overflow (SP fairness)	# rules	coverage	coverage pro	exp utility	exp utility non-pro	exp utility pro	unfairness
Group SP (2.5K)	4	97.82%	99.0%	20973.55	20772.77	18275.44	**2497.33**
Group SP (5K)	7	97.31%	98.24%	22805.52	23069.98	18071.12	4998.86
Group fairness (10K)	8	97.52%	97.81%	27870.77	27998.47	17998.66	9999.81
Group SP (20K)	20	**99.88%**	**99.94%**	**32671.11**	**32664.45**	**18423.64**	14240.81
Individual SP (2.5K)	20	99.95%	99.98%	24070.94	24433.55	12784.62	11648.93
Individual SP (5K)	20	**99.99%**	**99.99%**	25526.1	25911.22	15327.21	**10584.01**
Individual SP(10K)	20	**99.99%**	**99.99%**	28014.58	28256.35	14241.07	14015.28
Individual SP (20K)	20	99.51%	99.63%	**29984.0**	**29966.29**	14929.7	15036.59

7 EXPERIMENTAL EVALUATION

We present an experimental evaluation that evaluates FairCap effectiveness and efficiency. We aim to address the following questions: **Q1**: How does the quality of our generated rulesets compare to that of existing methods? **Q2** : What is the efficiency of FairCap and how is it affected by various data and system parameters?

7.1 Experimental Setup

FairCap was implemented in Python, and is publicly available in [6]. CATE values computation was performed using the DoWhy library [79]. The generated rules were translated into natural language using ChatGPT [57]. We perform experiments on CloudLab [18] xl170 machines (10-core 2.4 GHz CPU, 64 GB RAM).

The datasets, protected groups, and default parameters considered are the same as those described in Section 6.

Baselines. We compare FairCap with the following baselines:
CauSumX: CauSumX [97] is designed to find a summarized causal explanation for a group-by-average SQL query. When applied directly to the datasets (without involving a SQL query), it can be viewed as a solution to our problem, but without fairness constraints and only with an overall coverage constraint.
IDS [40] is a framework for generating Interpretable Decision Sets for prediction tasks. IDS incorporates parameters restricting the percentage of uncovered tuples and the number of rules. These parameters were assigned the same values in our system.
FRL: The authors of [14] introduced a framework for producing Falling Rule Lists as a probabilistic classification model. FRLs consist of a sequence of if-then rules, with the if-clauses containing antecedents and the then-clauses containing probabilities of the desired outcome. The order of rules in a Falling Rule List reflects the probabilities associated with each outcome.

Since IDS and FRL assume a binary outcome, we binned the salary variable in SO using the average value. To address fairness considerations, we run the baseline algorithms twice (excluding Brute-Force). First, we apply them to the entire dataset to obtain a set of rules applicable to the entire population. Then, we run them again solely on the tuples belonging to the protected population to generate rules specifically tailored for them. The final solution for each baseline is considered the union of these two sets of rules.

7.2 Quality Evaluation (Q1)

We compare the set of rules chosen by each baseline and FairCap.
Stack Overflow. As discussed in Section 6, prescription rules selected without fairness constraints, similar to the behavior of

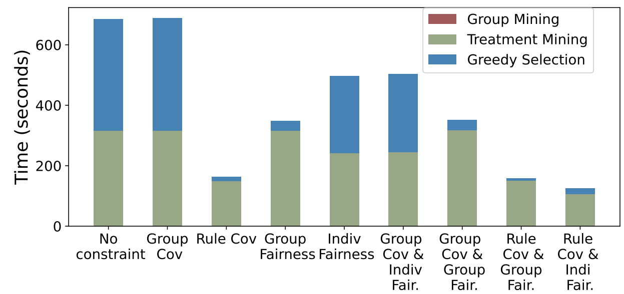

Figure 3: Runtime by-step of the FairCap algorithm (SO)

CauSumX, were significantly more advantageous for non-protected. This highlights the importance of imposing fairness constraints. IDS generated 16 rules for the overall population and 21 rules for the protected group. Notably, these rules do not suggest interventions to improve outcomes. For example, one rule states that if Country = Turkey and Age = 18-24 years, then the expected salary is low (with the outcome binned). Another key distinction is that these rules are not causal, as they are based solely on correlations in the data. For example, one rule indicates that if the years coding = 0-2 and Sexual Orientation = Gay or Lesbian, then the expected salary is low. The FRL baseline generated 9 rules for the overall population and 7 for the protected group. Similar to the IDS baseline, these rules do not propose interventions to improve outcomes and are not causal. For example, one rule states that if Country = United States and Sexual Orientation = Straight or Heterosexual, then the expected salary is high. In contrast, FairCap generates interventions aimed at improving the outcome variable of interest by leveraging causal relationships. It also allows users to impose fairness constraints, ensuring that the protected group benefits from these interventions.
German. Here again, with no fairness constraint (akin to CauSumX), the selected rules were mostly beneficial for the non-protected. IDS generated 12 rules for the overall population and 20 for the protected group. Here again, the rules are not causal and do not offer an intervention. For example, one of the rules suggested that single females at the age of 35-41 are unlikely to get the loan. FRL generated 13 rules for the overall population and 11 for the protected group. As before, the rules are not causal and do not propose ways to improve the credit risk score. For example, one rule suggests that if a person has lived in a house for 4-7 years, their credit risk score is likely to be high. Another rule states that if the purpose of the loan is to buy a used car, the credit risk score is also likely to be high. Clearly, these rules rely on correlations in the data rather than causal relationships. In contrast, FairCap generated a ruleset that offers interventions to improve the credit risk score based on causal relationships. Example selected rules are shown in Section 6.

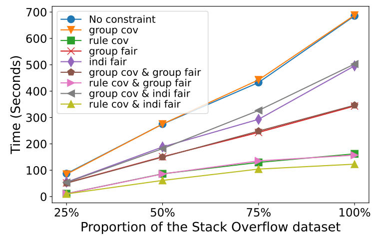

Figure 4: Runtime as a function of the dataset size (SO)

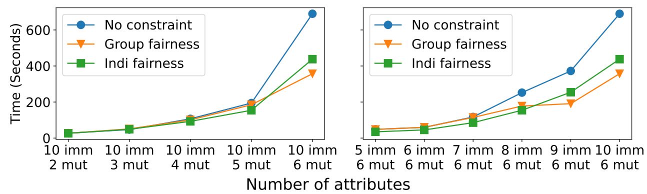

Figure 5: Runtime as a function of number of mutable and immutable attributes for SO with statistical parity

7.3 Scalability Evaluation (Q2)

Breakdown analysis by step. Figure 3 shows the runtime comparison of FairCap for different problem settings. Observe that using rule coverage constraint has the lowest runtime because it helps to prune rules which do not satisfy the coverage constraint. Employing rule coverage with individual fairness is the fastest among all settings, while no constraint setting takes the longest time. The time taken by the group mining phase is less than 2 seconds across all setups, and is therefore not visible in the plot. The intervention mining phase (Step 2) is the most inefficient phase, which takes around 6 mins for the unconstrained setting. The running time of these components aligns with our time complexity analysis (Section 5). Due to space restrictions, we do not present the corresponding plot for German dataset. All conclusions remain the same but the overall running time is $\approx 10\times$ faster due to its smaller size. We now analyze the impact of system parameters and data size on performance.

Data Size. Figure 4 compares the running time of FairCap for varying dataset sizes. We observe that the time taken by FairCap increases linearly for most of the settings. We also observed that the quality of rules returned by sampling 25% of the data points is comparable with the rules returned by using the whole dataset. Therefore, sampling based optimizations can help to reduce the running time from 11 min to less than 2min for the unconstrained setting and less than a minute with fairness constraints.

Number of Attributes. Figure 5 shows the runtime of FairCap while increasing the number of mutable and immutable attributes. On increasing the number of mutable attributes, the number of intervention patterns increases exponentially while on increasing immutable attributes, the number of grouping patterns increases exponentially. Therefore, both have a similar impact on runtime.

Fairness Threshold. Table 5 presents the results for varying ϵ for group and individual fairness. We observe that the unfairness of the returned solution increases with the increase in ϵ. Additionally, the overall expected utility increases but the expected utility of the protected individuals decreases. This result matches our intuition as highly unfair rules are selected for higher values of ϵ. We also notice that the greedy algorithm satisfies the group fairness constraint in all scenarios (unfairness is always less than the desired threshold).

For individual fairness, the overall utility increases monotonically with ϵ. However, the rate of growth for individual fairness is slower than that of group fairness. One interesting observation about individual fairness is that when all rules have statistical parity difference less than 2500, the overall unfairness is still around $11K$. This sudden increase in unfairness when considering multiple fair rules together is because we evaluate the upper bound of unfairness by taking the difference between max utility of unprotected and min utility of protected individuals. On manual inspection, we observed that all rules are indeed individually fair.

Coverage Threshold. With the change in coverage thresholds, we do not observe major difference in the overall results because the majority of the rules exhibit very high coverage (Table 4).

Apriori Threshold. We observe that increasing the Apriori threshold τ leads to a reduction in the number of grouping patterns considered, and thus to a decrease in runtime. However, our findings indicate that higher τ values lead to a decrease in both utility and fairness. Based on our findings, we recommend using a default value of 0.1, which provides satisfactory results in terms of coverage, utility, fairness and runtime.

8 LIMITATIONS AND FUTURE WORK

FairCap generates actionable, causal-based recommendations to improve a target outcome while incorporating coverage and fairness constraints. FairCap currently supports a single-relation database without dependencies among tuples to ensure compliance with the SUTVA assumption [71] (discussed in Section 3). However, this assumption breaks down even in single-table databases with tuple dependencies. In single-table settings, intervention and grouping patterns are straightforward to define. Extending these definitions to multi-table databases, where grouping attributes and interventions may originate from different tables, introduces a significant challenge. This complexity arises due to many-to-many relationships and cross-table patterns. Previous work, such as [21, 74], has extended causal models to handle multi-table data, but they have not explicitly targeted recommendations for subgroups. Expanding our framework to support multi-relational databases with complex dependencies remains an important direction for future research. Notably, prior work leveraging causal inference [47, 73, 95, 97] has also primarily focused on single-table settings.

We acknowledge that the generated rules may lack robustness, meaning they might not impact all individuals receiving the recommendation in the same way. An exciting direction for future research would be to focus on ensuring the robustness of prescription rules. Another direction for future work is to account for the cost of interventions. Some interventions may be impractical (e.g., pursuing a bachelor's degree in CS for someone who already holds a PhD in CS) or vary significantly in cost (e.g., moving to the US versus learning Python). Future research will incorporate intervention costs to generate budget-constrained prescription rules.

Finally, it is essential to recognize that the generated rules may be influenced by several factors, including the method used to evaluate causal effects, the quality of the causal DAG, the thresholds set for the constraints, and the overall quality of the data.

REFERENCES

[1] 2021. 2021 Stackoverflow Developer Survey. https://insights.stackoverflow.com/survey/2021.

[2] Alekh Agarwal, Miroslav Dudík, and Zhiwei Steven Wu. 2019. Fair regression: Quantitative definitions and reduction-based algorithms. In *International Conference on Machine Learning*. PMLR, 120–129.

[3] Sameet Agarwal, Rakesh Agrawal, Prasad Deshpande, Ashish Gupta, Jeffrey F. Naughton, Raghu Ramakrishnan, and Sunita Sarawagi. 1996. On the Computation of Multidimensional Aggregates. In *VLDB*. Morgan Kaufmann, 506–521.

[4] Rakesh Agrawal, Ramakrishnan Srikant, et al. 1994. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, Vol. 1215. Santiago, Chile, 487–499.

[5] Abdullah Alomar, Pouya Hamadanian, Arash Nasr-Esfahany, Anish Agarwal, Mohammad Alizadeh, and Devavrat Shah. 2023. CausalSim: A Causal Framework for Unbiased Trace-Driven Simulation. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 1115–1147.

[6] Anonymous Author(s). 2024. Git Repository. https://anonymous.4open.science/r/FairCap-D4AD/README.md. Accessed: Oct 17, 2024.

[7] Abolfazl Asudeh, Zhongjun Jin, and HV Jagadish. 2019. Assessing and remedying coverage for a given dataset. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 554–565.

[8] Arthur Asuncion and David Newman. 2007. UCI machine learning repository.

[9] Eftychia Baikousi, Georgios Rogkakos, and Panos Vassiliadis. 2011. Similarity measures for multidimensional data. In *2011 IEEE 27th International Conference on Data Engineering*. IEEE, IEEE, 171–182.

[10] Nicole Bidoit, Melanie Herschel, and Katerina Tzompanaki. 2014. Query-based why-not provenance with nedexplain. In *Extending database technology (EDBT)*.

[11] Reuben Binns. 2020. On the apparent conflict between individual and group fairness. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 514–524.

[12] Simon Caton and Christian Haas. 2024. Fairness in machine learning: A survey. *Comput. Surveys* 56, 7 (2024), 1–38.

[13] Adriane Chapman and HV Jagadish. 2009. Why not?. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. 523–534.

[14] Chaofan Chen and Cynthia Rudin. 2018. An optimization approach to learning falling rule lists. In *International conference on artificial intelligence and statistics*. PMLR, 604–612.

[15] Daniel Deutch, Nave Frost, and Amir Gilad. 2020. Explaining Natural Language query results. *VLDB J.* 29, 1 (2020), 485–508.

[16] Daniel Deutch and Amir Gilad. 2019. Reverse-Engineering Conjunctive Queries from Provenance Examples. In *Advances in Database Technology - 22nd International Conference on Extending Database Technology, EDBT 2019, Lisbon, Portugal, March 26-29, 2019*, Melanie Herschel, Helena Galhardas, Berthold Reinwald, Irini Fundulaki, Carsten Binnig, and Zoi Kaoudi (Eds.). OpenProceedings.org, 277–288. https://doi.org/10.5441/002/edbt.2019.25

[17] Daniel Deutch, Amir Gilad, Tova Milo, Amit Mualem, and Amit Somech. 2022. FEDEX: An Explainability Framework for Data Exploration Steps. *Proc. VLDB Endow.* 15, 13 (2022), 3854–3868. https://www.vldb.org/pvldb/vol15/p3854-gilad.pdf

[18] Dmitry Duplyakin, Robert Ricci, Aleksander Maricq, Gary Wong, Jonathon Duerig, Eric Eide, Leigh Stoller, Mike Hibler, David Johnson, Kirk Webb, Aditya Akella, Kuangching Wang, Glenn Ricart, Larry Landweber, Chip Elliott, Michael Zink, Emmanuel Cecchet, Snigdhaswin Kar, and Prabodh Mishra. 2019. The Design and Operation of CloudLab. In *Proceedings of the USENIX Annual Technical Conference (ATC)*. 1–14. https://www.flux.utah.edu/paper/duplyakin-atc19

[19] Kareem El Gebaly, Parag Agrawal, Lukasz Golab, Flip Korn, and Divesh Srivastava. 2014. Interpretable and informative explanations of outcomes. *Proceedings of the VLDB Endowment* 8, 1 (2014), 61–72.

[20] Uriel Feige. 1998. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)* 45, 4 (1998), 634–652.

[21] Sainyam Galhotra, Amir Gilad, Sudeepa Roy, and Babak Salimi. 2022. Hyper: Hypothetical reasoning with what-if and how-to queries using a probabilistic causal approach. In *Proceedings of the 2022 International Conference on Management of Data*. 1598–1611.

[22] Sainyam Galhotra, Yue Gong, and Raul Castro Fernandez. 2023. Metam: Goal-oriented data discovery. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 2780–2793.

[23] Sainyam Galhotra, Karthikeyan Shanmugam, Prasanna Sattigeri, and Kush R Varshney. 2022. Causal feature selection for algorithmic fairness. In *Proceedings of the 2022 International Conference on Management of Data*. 276–285.

[24] Yu Gan, Mingyu Liang, Sundar Dev, David Lo, and Christina Delimitrou. 2021. Sage: Practical and Scalable ML-Driven Performance Debugging in Microservices. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*. 135–151.

[25] David García-Soriano and Francesco Bonchi. 2021. Maxmin-fair ranking: individual fairness under group-fairness constraints. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 436–446.

[26] Manisha Girotra, Kanika Nagpal, Saloni Minocha, and Neha Sharma. 2013. Comparative survey on association rule mining algorithms. *International Journal of Computer Applications* 84, 10 (2013).

[27] Clark Glymour, Kun Zhang, and Peter Spirtes. 2019. Review of causal discovery methods based on graphical models. *Frontiers in genetics* 10 (2019), 524.

[28] Yue Gong, Sainyam Galhotra, and Raul Castro Fernandez. 2024. Nexus: Correlation Discovery over Collections of Spatio-Temporal Tabular Data. *Proceedings of the ACM on Management of Data* 2, 3 (2024), 1–28.

[29] Helga Gudmundsdottir, Babak Salimi, Magdalena Balazinska, Dan RK Ports, and Dan Suciu. 2017. A demonstration of interactive analysis of performance measurements with viska. In *Proceedings of the 2017 ACM International Conference on Management of Data*. 1707–1710.

[30] Shubha Guha, Falaah Arif Khan, Julia Stoyanovich, and Sebastian Schelter. 2024. Automated data cleaning can hurt fairness in machine learning-based decision making. *IEEE Transactions on Knowledge and Data Engineering* (2024).

[31] Paul W Holland. 1986. Statistics and causal inference. *Journal of the American statistical Association* 81, 396 (1986), 945–960.

[32] Zezhou Huang, Jiaxiang Liu, Haonan Wang, and Eugene Wu. 2023. The Fast and the Private: Task-based Dataset Search. *arXiv preprint arXiv:2308.05637* (2023).

[33] Guido W Imbens. 2024. Causal inference in the social sciences. *Annual Review of Statistics and Its Application* 11 (2024).

[34] Manas Joglekar, Hector Garcia-Molina, and Aditya Parameswaran. 2017. Interactive data exploration with smart drill-down. *IEEE Transactions on Knowledge and Data Engineering* 31, 1 (2017), 46–60.

[35] Gagandeep Kaur and Shruti Aggarwal. 2013. Performance analysis of association rule mining algorithms. *International Journal of Advanced Research in Computer Science and Software Engineering* 3, 8 (2013), 856–58.

[36] Samir Khuller, Anna Moss, and Joseph Seffi Naor. 1999. The budgeted maximum coverage problem. *Information processing letters* 70, 1 (1999), 39–45.

[37] Alexandra Kim, Laks VS Lakshmanan, and Divesh Srivastava. 2020. Summarizing hierarchical multidimensional data. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 877–888.

[38] Been Kim, Cynthia Rudin, and Julie A Shah. 2014. The bayesian case model: A generative approach for case-based reasoning and prototype classification. *Advances in neural information processing systems* 27 (2014).

[39] Trupti A Kumbhare and Santosh V Chobe. 2014. An overview of association rule mining algorithms. *International Journal of Computer Science and Information Technologies* 5, 1 (2014), 927–930.

[40] Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. 2016. Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1675–1684.

[41] Connor Lawless, Sanjeeb Dash, Oktay Gunluk, and Dennis Wei. 2023. Interpretable and fair boolean rule sets via column generation. *Journal of Machine Learning Research* 24, 229 (2023), 1–50.

[42] Seokki Lee, Bertram Ludäscher, and Boris Glavic. 2020. Approximate summaries for why and why-not provenance (extended version). *arXiv preprint arXiv:2002.00084* (2020).

[43] Chenjie Li, Zhengjie Miao, Qitian Zeng, Boris Glavic, and Sudeepa Roy. 2021. Putting Things into Context: Rich Explanations for Query Answers using Join Graphs. In *Proceedings of the 2021 International Conference on Management of Data*. 1051–1063.

[44] Jinyang Li, Yuval Moskovitch, Julia Stoyanovich, and HV Jagadish. 2023. Query Refinement for Diversity Constraint Satisfaction. *Proceedings of the VLDB Endowment* 17, 2 (2023), 106–118.

[45] Yin Lin, Brit Youngmann, Yuval Moskovitch, HV Jagadish, and Tova Milo. 2021. On detecting cherry-picked generalizations. *Proceedings of the VLDB Endowment* 15, 1 (2021), 59–71.

[46] Yin Lou, Rich Caruana, Johannes Gehrke, and Giles Hooker. 2013. Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 623–631.

[47] Pingchuan Ma, Rui Ding, Shuai Wang, Shi Han, and Dongmei Zhang. 2023. Xinsight: explainable data analysis through the lens of causality. *Proceedings of the ACM on Management of Data* 1, 2 (2023), 1–27.

[48] Markos Markakis, An Bo Chen, Brit Youngmann, Trinity Gao, Ziyu Zhang, Rana Shahout, Peter Baile Chen, Chunwei Liu, Ibrahim Sabek, and Michael Cafarella. 2024. Sawmill: From Logs to Causal Diagnosis of Large Systems. In *SIGMOD*. 444–447.

[49] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–35.

[50] Alexandra Meliou, Wolfgang Gatterbauer, Joseph Y Halpern, Christoph Koch, Katherine F Moore, and Dan Suciu. 2010. Causality in databases. *IEEE Data Engineering Bulletin* 33, 3 (2010), 59–67.

[51] Alexandra Meliou, Wolfgang Gatterbauer, Katherine F Moore, and Dan Suciu. 2009. Why so? or why no? functional causality for explaining query answers. *arXiv preprint arXiv:0912.5340* (2009).

[52] Alexandra Meliou, Wolfgang Gatterbauer, Katherine F Moore, and Dan Suciu. 2010. The complexity of causality and responsibility for query answers and non-answers. *arXiv preprint arXiv:1009.2021* (2010).

[53] Alexandra Meliou, Sudeepa Roy, and Dan Suciu. 2014. Causality and explanations in databases. *Proceedings of the VLDB Endowment* 7, 13 (2014), 1715–1716.

[54] Zhengjie Miao, Qitian Zeng, Boris Glavic, and Sudeepa Roy. 2019. Going beyond provenance: Explaining query answers with pattern-based counterbalances. In *Proceedings of the 2019 International Conference on Management of Data*. 485–502.

[55] Fatemeh Nargesian, Abolfazl Asudeh, and HV Jagadish. 2021. Tailoring data source distributions for fairness-aware data integration. *Proceedings of the VLDB Endowment* 14, 11 (2021), 2519–2532.

[56] Fatemeh Nargesian, Abolfazl Asudeh, and H. V. Jagadish. 2022. Responsible Data Integration: Next-generation Challenges. In *Proceedings of the 2022 ACM SIGMOD International Conference on Management of data*.

[57] OpenAI. 2023. ChatGPT (October 2023 version). https://chat.openai.com/. Large language model.

[58] Eliana Pastor, Elena Baralis, and Luca de Alfaro. 2023. A hierarchical approach to anomalous subgroup discovery. In *2023 IEEE 39th international conference on data engineering (ICDE)*. IEEE, 2647–2659.

[59] Eliana Pastor, Luca De Alfaro, and Elena Baralis. 2021. Looking for trouble: Analyzing classifier behavior via pattern divergence. In *Proceedings of the 2021 International Conference on Management of Data*. 1400–1412.

[60] Judea Pearl. 2009. Causal inference in statistics: An overview. (2009).

[61] Dana Pessach and Erez Shmueli. 2022. A review on fairness in machine learning. *ACM Computing Surveys (CSUR)* 55, 3 (2022), 1–44.

[62] Alireza Pirhadi, Mohammad Hossein Moslemi, Alexander Cloninger, Mostafa Milani, and Babak Salimi. 2024. Otclean: Data cleaning for conditional independence violations using optimal transport. *Proceedings of the ACM on Management of Data* 2, 3 (2024), 1–26.

[63] Drago Plecko and Elias Bareinboim. 2022. Causal fairness analysis. *arXiv preprint arXiv:2207.11385* (2022).

[64] Yuji Roh, Kangwook Lee, Steven Euijong Whang, and Changho Suh. 2020. Fairbatch: Batch selection for model fairness. *arXiv preprint arXiv:2012.01696* (2020).

[65] Yuji Roh, Kangwook Lee, Steven Euijong Whang, and Changho Suh. 2023. Improving Fair Training under Correlation Shifts. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 29179–29209.

[66] Paul R Rosenbaum and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55.

[67] Sudeepa Roy. 2022. Toward interpretable and actionable data analysis with explanations and causality. *Proc. VLDB Endow.* 15, 12 (2022), 3812–3820.

[68] Sudeepa Roy, Laurel Orr, and Dan Suciu. 2015. Explaining query answers with explanation-ready databases. *Proceedings of the VLDB Endowment* 9, 4 (2015), 348–359.

[69] Sudeepa Roy and Dan Suciu. 2014. A formal approach to finding explanations for database queries. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 1579–1590.

[70] Donald Bruce Rubin. 1971. *The use of matched sampling and regression adjustment in observational studies.* Ph. D. Dissertation. Harvard University.

[71] Donald B Rubin. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* 100, 469 (2005), 322–331.

[72] Omer Sagi and Lior Rokach. 2021. Approximating XGBoost with an interpretable decision tree. *Information Sciences* 572 (2021), 522–542.

[73] Babak Salimi, Johannes Gehrke, and Dan Suciu. 2018. Bias in olap queries: Detection, explanation, and removal. In *Proceedings of the 2018 International Conference on Management of Data*. 1021–1035.

[74] Babak Salimi, Harsh Parikh, Moe Kayali, Lise Getoor, Sudeepa Roy, and Dan Suciu. 2020. Causal relational learning. In *Proceedings of the 2020 ACM SIGMOD international conference on management of data*. 241–256.

[75] Babak Salimi, Luke Rodriguez, Bill Howe, and Dan Suciu. 2019. Interventional fairness: Causal database repair for algorithmic fairness. In *Proceedings of the 2019 International Conference on Management of Data*. 793–810.

[76] Aécio Santos, Aline Bessa, Fernando Chirigati, Christopher Musco, and Juliana Freire. 2021. Correlation sketches for approximate join-correlation queries. In *Proceedings of the 2021 International Conference on Management of Data*. 1531–1544.

[77] Gayatri Sathe and Sunita Sarawagi. 2001. Intelligent rollups in multidimensional OLAP data. In *VLDB*. 307–316.

[78] Holger Schielzeth. 2010. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* 1, 2 (2010), 103–113.

[79] Amit Sharma and Emre Kiciman. 2020. DoWhy: An End-to-End Library for Causal Inference. *arXiv preprint arXiv:2011.04216* (2020).

[80] Julia Stoyanovich, Bill Howe, and Hosagrahar Visvesvaraya Jagadish. 2020. Responsible data management. *Proceedings of the VLDB Endowment* 13, 12 (2020).

[81] Hao Sun, Evan Munro, Georgy Kalashnov, Shuyang Du, and Stefan Wager. 2021. Treatment allocation under uncertain costs. *arXiv preprint arXiv:2103.11066* (2021).

[82] Tanmay Surve and Romila Pradhan. 2024. Example-based Explanations for Random Forests using Machine Unlearning. *arXiv preprint arXiv:2402.05007* (2024).

[83] Ki Hyun Tae, Hantian Zhang, Jaeyoung Park, Kexin Rong, and Steven Euijong Whang. 2024. Falcon: Fair Active Learning using Multi-armed Bandits. *Proc. VLDB Endow.* 17, 5 (2024), 952–965.

[84] Yuchao Tao, Amir Gilad, Ashwin Machanavajjhala, and Sudeepa Roy. 2022. DPXPlain: Privately Explaining Aggregate Query Answers. *Proc. VLDB Endow.* 16, 1 (2022), 113–126. https://www.vldb.org/pvldb/vol16/p113-tao.pdf

[85] Balder ten Cate, Cristina Civili, Evgeny Sherkhonov, and Wang-Chiew Tan. 2015. High-level why-not explanations using ontologies. In *Proceedings of the 34th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. 31–43.

[86] Moshe Y. Vardi. 1982. The Complexity of Relational Query Languages (Extended Abstract). In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing* (San Francisco, California, USA) *(STOC '82)*. ACM, New York, NY, USA, 137–146. https://doi.org/10.1145/800070.802186

[87] Manasi Vartak, Sajjadur Rahman, Samuel Madden, Aditya Parameswaran, and Neoklis Polyzotis. 2015. Seedb: Efficient data-driven visualization recommendations to support visual analytics. In *VLDB*, Vol. 8. NIH Public Access, 2182.

[88] Stefan Wager and Susan Athey. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113, 523 (2018), 1228–1242.

[89] Tong Wang and Cynthia Rudin. 2022. Causal rule sets for identifying subgroups with enhanced treatment effects. *INFORMS Journal on Computing* 34, 3 (2022), 1626–1643.

[90] Eugene Wu and Samuel Madden. 2013. Scorpion: Explaining away outliers in aggregate queries. (2013).

[91] Yu Xie, Jennie E Brand, and Ben Jann. 2012. Estimating heterogeneous treatment effects with observational data. *Sociological methodology* 42, 1 (2012), 314–347.

[92] Hongyu Yang, Cynthia Rudin, and Margo Seltzer. 2017. Scalable Bayesian rule lists. In *International conference on machine learning*. PMLR, 3921–3930.

[93] Brit Youngmann, Sihem Amer-Yahia, and Aurélien Personnaz. 2022. Guided Exploration of Data Summaries. *Proc. VLDB Endow.* 15, 9 (2022).

[94] Brit Youngmann, Michael Cafarella, Amir Gilad, and Sudeepa Roy. 2024. Summarized Causal Explanations For Aggregate Views (Full version). arXiv:2410.11435 [cs.DB] https://arxiv.org/abs/2410.11435

[95] Brit Youngmann, Michael Cafarella, Yuval Moskovitch, and Babak Salimi. 2023. On Explaining Confounding Bias. *2023 IEEE 39th International Conference on Data Engineering (ICDE)* (2023).

[96] Brit Youngmann, Michael Cafarella, Babak Salimi, and Anna Zeng. 2023. Causal Data Integration. *arXiv preprint arXiv:2305.08741* (2023).

[97] Brit Youngmann, Michael J. Cafarella, Amir Gilad, and Sudeepa Roy. 2024. Summarized Causal Explanations For Aggregate Views. *Proc. ACM Manag. Data* 2, 1 (2024), 71:1–71:27.

[98] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. Fa* ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1569–1578.

[99] Meike Zehlike, Ke Yang, and Julia Stoyanovich. 2021. Fairness in ranking: A survey. *arXiv preprint arXiv:2103.14000* (2021).

[100] Hantian Zhang, Ki Hyun Tae, Jaeyoung Park, Xu Chu, and Steven Euijong Whang. 2023. iFlipper: Label Flipping for Individual Fairness. *Proc. ACM Manag. Data* 1, 1 (2023), 8:1–8:26.

[101] Xiaozhong Zhang, Xiaoyu Ge, Panos K Chrysanthis, and Mohamed A Sharaf. 2021. Viewseeker: An interactive view recommendation framework. *Big Data Research* 25 (2021), 100238.

[102] Qiankun Zhao and Sourav S Bhowmick. 2003. Association rule mining: A survey. *Nanyang Technological University, Singapore* 135 (2003), 18.

9 MISSING PROOFS

Proof of Lemma 4.1. The utility of a rule r denotes the expected increase in outcome O when all individuals within the subgroup $\mathcal{P}_g$ are treated with $\mathcal{P}_t$.

$$utility(r) \quad = \quad \frac{1}{|\mathcal{P}_g|} \sum_{i \in coverage(\mathcal{P}_g)} utility_i(\mathcal{P}_t)$$

where $utility_i(\mathcal{P}_t)$ denotes the utility for tuple i with respect to treatment $\mathcal{P}_t$. Since utility(r) is an average over multiple different utilities, the utility will be higher than the expected value for certain tuples in $coverage(\mathcal{P}_g)$. Let $i^* = \arg\max utility_i(\mathcal{P}_t)$.

Consider a new prescription rule $r'(i^*, \mathcal{P}_t)$ which considers the same treatment $\mathcal{P}_t$ for the tuple i^*. Therefore, $utility(r') = utility_i(\mathcal{P}_t) > utility(r)$. □

9.1 Hardness Results

We next study the complexity of the Prescription Ruleset Selection problem under different constraint combinations. We show that Prescription Ruleset Selection is equivalent to optimizing a non-negative and monotone submodular function. Furthermore, the individual fairness constraint and rule coverage constraints are matroid constraints. Therefore, a greedy algorithm is appropriate approach to solve the problem.

Proposition 9.1. *The optimization objective of Prescription Ruleset Selection problem is a non-negative submodular function.*

Proof of Proposition 9.1. According to [40], the size objective is a non-negative and submodular function. Similarly, the expected utility—assuming each individual receives a single rule and selects the best option—is also a non-negative submodular function. As a result, the linear combination of these functions remains a non-negative submodular function, and maximizing it is known to be NP-hard [36]. □

Proposition 9.2. *Individual fairness and rule coverage constraints are matroid constraint*

Proof of Proposition 9.2. We will show these constraints satisfy the following properties:

(1) **Hereditary Property**: If S is an independent set, then every subset of S is also an independent set.
(2) **Exchange Property**: If S and T are independent sets and $|S| < |T|$, then there exists an element $e \in T \setminus S$ such that $S \cup \{e\}$ is also an independent set.

These two properties ensure that the set system behaves like a matroid.

In our setting Start by specifying the ground set is all possible rules and what qualifies as an "independent set" is a subset of rules satisfying a constraint.

Individual Fairness. If a set of rules R satisfies the individual fairness constraint, this means each rule within R individually satisfies the constraint. Consequently, any subset $R' \subseteq R$ also upholds individual fairness. This further implies the exchange property, as any rule that satisfies individual fairness can be added to an individually fair set of rules while preserving individual fairness.

Rule coverage. If a set of rules R satisfies the rule coverage constraint, this means each rule within R individually satisfies the rule coverage constraint. Consequently, any subset $R' \subseteq R$ also satisfies the rule coverage constraint. This also implies the exchange property, as any rule that satisfies rule coverage can be added to a set of rules satisfying rule coverage while preserving the rule coverage constraints. □

For the group coverage, we can show that merely finding a solution that satisfies the constraints, even without maximizing expected utility, is NP-hard via a reduction from the Set Cover problem [20].

Proposition 9.3. *Prescription Ruleset Selection with a group-coverage constraint is NP-hard*

Proof of 9.3. In the decision version of the Set Cover problem, we are given a universe of elements $U = \{x_1, \ldots, x_{n'}\}$, a collection of m subsets $S_1, \ldots, S_{m'} \subseteq U$ and a number k. The question is whether there exists a cover of U of at most k' subsets.

In the decision version of Prescription Ruleset Selection we are searching for a set of rules R such that: $f(R) \geq \tau$, where:

$$f(R) = \lambda_1 \cdot (l - size(R)) + \lambda_2 \cdot ExpUtility(R)$$

such that R satisfies the group-coverage constraints, defined by the parameter θ. In this proof, we assume no protected group is given, namely the constraint requires that the selected ruleset R would cover at least a θ fraction of the population (i.e., the protected group to be the empty group).

Given an instance of the set cover problem, we build an instance of the Prescription Ruleset Selection problem as follows. We build a relation R with $m'+1$ attributes, $\mathbb{A} = (A_1, \ldots, A_{m'}, O)$, and containing $n'+m'$ tuples. For each element $x_i \in U$, we create a tuple t_i, such that $t_i[A_j] = 1$ iff $x_i \in S_j$. We further add m' tuples t_{S_j} such that $t_{S_j}[A_j] = 1$, $t_{S_j}[O] = 0$, and $t_{S_j}[A_p] = l \neq 0$ for all $p \neq j$ where l is a unique number not used anywhere else in an attribute of R. We set the outcome variable to be O.

Here, $\mathcal{P}_g$ can be any predicate. Note that each set of tuples defined by a pattern can only have an outcome of 0, as the outcome of all tuples is 0. Therefore, the utility of all intervention patterns is 0. For Prescription Ruleset Selection, we further define the threshold for the group coverage constraint $\theta = \frac{n'+k'}{n'+m'}$. The underlying causal DAG, G, only contains the edges of the form $A_j \to O$ for all $1 \leq j \leq m'$. We claim that there exists a cover of U with at most k sets iff there exists a solution R to Prescription Ruleset Selection such that $f(R) \geq (l - k)$.

($\Rightarrow$) Assume we have a collection $S_{j_1}, \ldots, S_{j_k}$ such that $\cup_{j=j_1}^{j_k} S_j = U$. We show that there is a solution for Prescription Ruleset Selection as follows. For each S_{j_1}, we choose for the solution the pattern $\mathcal{P}_g^{j_i} : A_{j_i} = 1$. We show that $R = \{(\mathcal{P}_g^{j_1}, \emptyset), \ldots, (\mathcal{P}_g^{j_k}, \emptyset)\}$ is a solution to Prescription Ruleset Selection. The intervention pattern can be any pattern, as the utility of every intervention pattern is 0. First, we note that all tuples of the form t_i are covered by at least one grouping pattern by their definition. For the m remaining tuples, we have coverage of at most k tuples. These are the tuples $t_{S_{j_i}}$ that have $A_{j_i} = 1$. Thus, the number of covered tuples is exactly $n' + k'$ out of $n' + m'$ tuples in R. If there are fewer than k tuples we can augment the original cover with arbitrary sets to obtain a cover of size k.

($\Leftarrow$) Assume we have a solution R to Prescription Ruleset Selection with the aforementioned parameters. We show that we can find a solution to the set cover problem. First, note that since $f(R) \geq (l - k)$ and the expected utility is always 0, that means we have selected no more than k rules.

Suppose $R = \{(\mathcal{P}_g^{j_1}, \emptyset), \ldots, (\mathcal{P}_g^{j_k}, \emptyset)\}$. We first claim that no grouping pattern that includes $A_i = 0$ in a conjunction can be included in R as such a pattern will not cover any tuple t_{S_j} since these tuples do not have an attribute with value 0 by definition (and any other number other than 1 will only cover a single tuple). Thus, the number of covered tuples will be $< \frac{n'+k'}{n'+m'} = \theta$, which would contradict the assumption that this is a valid solution to Prescription Ruleset Selection that satisfies the coverage constraint. Hence, all patterns are of conjunctions of $A_i = 1$. For each intervention pattern of the form $\mathcal{P}_g = \wedge_{j=i_a}^{i_b} (A_j = 1) \wedge (A_p = l)$, we choose an arbitrary attribute in the conjunction A_j if $(A_j = 1) \in \mathcal{P}_g$ and choose S_j for the cover. Finally, if there is an uncovered element x in U and R includes a pattern in of the form $\mathcal{P}_g = (A_j = l)$ where $l \neq 1$, we choose for the cover a set S that covers x arbitrarily. We claim that the chosen collection of sets is a cover of U. To see this, recall that we claimed that the coverage of R is at least $n + k$. If the coverage includes tuples of the form t_{S_j}, then each pattern covers a single tuple. Suppose these patterns are

$\mathcal{P}_a, \ldots, \mathcal{P}_b$. When building the coverage, instead of these patterns, we add a set that covers elements not yet covered by existing patterns. Thus, there are at least $b - a$ covered elements from U in addition to the $n + k - (b - a)$ tuples covered by the patterns. Thus, the set cover we have assembled contains $n - (b - a) + (b - a) = n$ elements and covers all elements in U. □