Characterization of Overlap in Observational Studies

Fredrik D. Johansson, Dennis Wei, Michael Oberst
MIT-IBM Watson AI Lab, MIT CSAIL & IMES
fredrikj@mit.edu, dwei@us.ibm.com, moberst@mit.edu

Tian Gao, Gabriel Brat
MIT-IBM Watson AI Lab, Harvard Medical School
tgao@us.ibm.com, gbrat@bidmc.harvard.edu

David Sontag, Kush R. Varshney
MIT-IBM Watson AI Lab, MIT-IBM Watson AI Lab
dsontag@mit.edu, krvarshn@us.ibm.com

July 8th, 2019

Abstract
Overlap between treatment groups is required for nonparametric estimation of causal effects. If a subgroup of subjects always receives (or never receives) a given intervention, we cannot estimate the effect of intervention changes on that subgroup without further assumptions. When overlap does not hold globally, characterizing local regions of overlap can inform the relevance of any causal conclusions for new subjects, and can help guide additional data collection. To have impact, these descriptions must be interpretable for downstream users who are not machine learning experts, such as clinicians. We formalize overlap estimation as a problem of finding minimum volume sets and give a method to solve it by reduction to binary classification with Boolean rules. We also generalize our method to estimate overlap in off-policy policy evaluation. Using data from real-world applications, we demonstrate that these rules have comparable accuracy to black-box estimators while maintaining a simple description. In one case study, we perform a user study with clinicians to evaluate rules learned to describe treatment group overlap in post-surgical opioid prescriptions. In another, we estimate overlap in policy evaluation of antibiotic prescription for urinary tract infections.

1 Introduction
To estimate the causal effect of a binary intervention, we must use the outcomes of treated individuals to infer the outcomes of untreated individuals under treatment, and vice versa [21]. However, the set of subjects that were considered for either treatment group—the overlap set—is the only set for which we can learn provably optimal policies without assumptions on function class [27]. This motivates the construction of an explicit, interpretable characterization of the overlap set, to provide domain experts with insight into the relevance of learned policies for new subjects. For example, physicians...
who practice evidence-based medicine (EBM) need to know whether the results of a clinical trial or observational study apply to their specific patient \[29\]. Providing inclusion/exclusion criteria is insufficient in either setting, as these often apply to a larger cohort than the overlap set. For instance, if certain subgroups are included in the retrospective cohort, but never received the intervention, then the conclusions of the study should not be considered applicable for that subgroup.

Characterizing overlap between distributions is also relevant beyond causal inference. A closely related problem is unsupervised domain adaptation \[3\], in which predictive models trained on a source domain are applied to a target domain from which no labels are observed. By describing the overlap between domains, we gain insight into the inputs for which transfer is likely to succeed. Another use case is understanding areas of difficulty in classification tasks. Describing the overlap between inputs from different classes illustrates which inputs are “hard” to classify, and the complement of the overlap (limited to the observed distribution) describes which inputs are “easy”.

With this context in mind, our main contributions are as follows: (i) We propose desiderata in overlap estimation, and note how existing methods fail to satisfy them. (ii) We give a method for interpretable characterization of distributional overlap which satisfies these desiderata, by reducing the problem to two binary classification problems, and using a linear programming relaxation of learning optimal Boolean rules. (iii) We demonstrate that small rule sets often perform comparably to black-box estimators on a suite of real-world tasks. (iv) We show how a generalized definition and method applies to policy evaluation and apply it to describing overlap in policies for antibiotic prescription. (v) We evaluate the interpretability of rules for describing treatment group overlap in post-surgical opioid prescription in a user study with medical professionals.

2 Related work

Distributional (treatment group) overlap is a central assumption in the estimation of causal effects from observational data. A simple yet common method to estimate overlap is to compare group-specific covariate bounds and low-order moments \[26, 42, 9\]. A more flexible approach is to estimate the group propensity—the probability that a subject was prescribed treatment. Propensities bounded
away from 0 and 1 at a point indicates that groups overlap \([20, 27]\). This idea was used by \([9]\) to learn “interpretable study populations”, by identifying the largest axis-aligned box that contains only subjects with bounded propensity. Matching methods \([14, 25, 16]\), which compute an optimal cross-group matching of subjects with a limit on matching distance may also be used to estimate overlap but do not provide an explicit and generalizable description. Rule-based models have been widely considered also for classification tasks \([24, 11, 40, 7, 38, 10, 37]\) and subgroup discovery \([13]\), but have to the best of our knowledge not been applied to support or overlap estimation. Authors have proposed rule-based algorithms for density estimation, using decision trees \([23]\) and Bayesian histograms \([11]\), but small models of two densities are often not straightforward to combine into one small model of their overlap.

3 Problem Statement and Overlap Definition

We address rule-based overlap estimation—characterization of the intersection of two or more populations or densities. Our primary motivation is to aid policy making based on estimates of causal effects, the validity of which relies on knowing and communicating the set of subjects to which the policy applies. This often places restrictions on the overlap description \([2]\). We identify the following desiderata for estimates of overlap: (D.1) They include regions where all groups are well-represented; (D.2) They exclude all other regions, including those outside the support (see Figure 1a); (D.3) They can be expressed using a small set of simple rules. First, we discuss definitions of overlap satisfying \([D.1] and [D.2] \) and establish notation. We address \([D.3]\) in Sections 4 and 5.

Let subjects \(i = 1, \ldots, m\) be observed through samples \((x_i, t_i)\) of covariates \(X \in X \subseteq \mathbb{R}^d\) and a group indicator \(T \in T\). We assume that samples are independently and identically distributed according to a density \(p(X, T)\), that \(X\) is bounded. Let \(p_i(X) := p(X \mid T = t)\) denote the covariate density of group \(t \in T\). In causal effect estimation, it is most common to assess overlap between two groups with conditional densities \(p_0\) and \(p_1\) on \(X\), and \(T = \{0, 1\}\). In this case, overlap is often described as either a) the intersection of supports, \(\text{supp}(p_0) \cap \text{supp}(p_1)\), where \(\text{supp}(p) := \{x \in X : p(x) > 0\}\), or b) the set of covariates values for which the conditional probability of group membership (referred to as propensity) is bounded \([6, 20]\). We generalize this notion to an arbitrary set of groups \(T\):

\[
\mathcal{B}^e := \{x \in X \mid \forall t \in T : p(T = t \mid X = x) > \epsilon\} .
\]

Both of these definitions have shortcomings: the former is somewhat vacuous for variables with infinite support (e.g., a normal random variable), and even with finite support, we may wish to restrict it to “essential support” to avoid including distant outliers; The latter does not satisfy \([D.2]\) since a point may have bounded propensity but lie outside the support of the population \(p(X)\) (see Figure 1a).

Our preferred definition combines the propensity-based definition with a generalized notion of support in \(\alpha\)-minimum-volume sets \([30]\), based on the multi-dimensional quantile function \([8]\). Let \(\mathcal{C}\) be a set of measurable subsets of \(X\), let \(V(S)\) denote the volume (Lebesgue measure) of a set \(S \in \mathcal{C}\), and define \(P(S) := \int_{x \in S} dp\). An \(\alpha\)-minimum-volume set \(S^\alpha\) of \(p\) is then

\[
S^\alpha := \arg \min_C \{V(C) : P(C) \geq \alpha, C \in \mathcal{C}\} .
\]

When \(\alpha = 1\), \(S^\alpha = \text{supp}(p)\). \(S^\alpha\) may not always be unique, but the difference between any two \(\alpha\)-MV sets (for the same \(\alpha\)) is small for large \(\alpha\).\(^4\) In this work, we consider only \(\alpha < 1\) in order to

\(^4\)For two \(\alpha\)-MV sets \(C_1, C_2\) where \(\alpha = 1 - \delta\) for some \(\delta > 0\), their intersection is large in that \(P(C_1 \cap C_2) \geq 1 - 2\delta\).
handle distributions with infinite support and unwanted outliers, and refer to $S$ as the support of $p$. Defining overlap as the intersection of group-specific $\alpha$-MV sets is feasible but has the downside that empirical estimates scale poorly with $|T|$. Moreover, it does not facilitate the generalization to policy evaluation described at the end of this section, and the intersection of several descriptions is often less interpretable than a single description. Instead, we define the $\alpha,\epsilon$-overlap set, for $\alpha, \epsilon \in (0, 1)$, to be

$$O^{\alpha,\epsilon} := \{x \in S^\alpha : \forall t \in T : p(T = t \mid X = x) \geq \epsilon \} = S^\alpha \cap B^\epsilon. \quad (3)$$

We define the problem of overlap estimation under definition (3) as one of characterizing the set $O^{\alpha,\epsilon}$ given thresholds $\alpha$ and $\epsilon$. In line with (D.3), these characterizations should be useful in policy-making, and interpretable by domain experts, at small or no cost in accuracy. In the sequel, for notational convenience, we sometimes leave out superscripts from $S^\alpha$, $B^\epsilon$ and $O^{\alpha,\epsilon}$, assuming them fixed.

**Generalization to policy evaluation.** The definition of $B^\epsilon$ in (1) is motivated by causal effect estimation—comparison of outcomes under two or more alternative interventions. We may instead be interested in policy evaluation, which involves estimating the expected outcome under a conditional intervention $\pi$, which assigns (possibly stochastically) a treatment $t$ to each $x$ following a conditional distribution we write as $\pi(T \mid X)$ [22]. To perform this type of evaluation, we only require that the propensity $p(T \mid X)$ be bounded away from zero for treatments $T$ which have non-zero probability under the policy $\pi$. To describe the inputs for which this is satisfied, we may generalize $B^\epsilon$ to be a function of the target policy $\pi$ by defining $B^\epsilon(\pi) := \{x \in X : \forall t : \pi(t \mid x) > 0 : p(T = t \mid x) > \epsilon\}$. See the supplementary material for more details, and Section 6.4 for experimental results in this setting. The full application of this modified procedure is also given in the supplement for clarity.

## 4 OverRule: Rule-Based Overlap Estimation

We propose OverRule, an algorithm for identifying the overlap set $O$ by separately estimating the $\alpha$-MV support set $S$ and the bounded-propensity set $B$, thereby satisfying desiderata (D.1)–(D.2). Our approach to estimating $S$ and $B$, described in Sections 4.1 and 4.2, aims to fulfill desideratum (D.3) by using Boolean rules—logical formulae in either disjunctive (DNF) or conjunctive (CNF) normal form approximators that have received renewed attention because of their interpretability [7, 35] (see Figure 3 for an example). It was observed in preliminary experiments that learning rules for $S$ and $B$ separately improved interpretability, as it makes clear which rules apply to which task and prevents the capacity of the function from being consumed by one task. The conjunction of the two rules yields a description of $O$. OverRule proceeds in the following steps: (i) Fit an estimate $\hat{S}$ of the marginal support using Boolean rules, (ii) Fit a group propensity-based estimator $\hat{b}$ indicating membership in $B$, (iii) Approximate $\hat{b}$ on $\hat{S}$ using Boolean rules.

Our main contribution in this section is to demonstrate how rule learning of $S$ and $B$, steps (i) & (iii), can be reduced to binary classification. This enables us to exploit the rich set of existing methods for rule-based classification [10] in our effort to improve the interpretability of the overlap estimate.

### 4.1 Estimation of $S^\alpha$ as Binary Classification

In the first step of OverRule, we learn a rule set to approximate the $\alpha$-MV set of the marginal distribution $p(X)$, $S^\alpha$, by reducing the $\alpha$-MV set problem to one of binary classification between
the observed samples and uniform background samples. With an abusive reuse of notation, let \( \mathcal{C} \subseteq \{ C \subseteq X \} \) denote a class of subsets of \( X \) where each subset \( C \) corresponds to a candidate \( \alpha \)-MV set. In Section [5] we parameterize \( \mathcal{C} \) using Boolean rules, where \( C \) is the set of inputs for which a rule holds (see Figure [1b]). Let \( \mathcal{D} = \{ x_i \}_{i=1}^m \) denote the observed set of covariates. Then, the empirical version of (2) specialized to rule sets is as follows:

\[
\arg \min_{C \in \mathcal{C}} V(C) + R(C) \quad \text{s.t.} \quad |\mathcal{D} \cap C| \geq \alpha m,
\]

(4)

with the addition of a regularization term \( R(C) \) to control the complexity of the rule set.

In practice, the volume \( V(C) \) may be difficult to compute during optimization for general classes \( \mathcal{C} \), and the size of \( \mathcal{C} \) is often too large to allow pre-computation of \( V(C) \) for all \( C \). In particular, in the case of DNF Boolean rules, each \( C \) is a union of several potentially overlapping rules (see Figure [1b]). Even if the volume spanned by each rule is known or quick to compute on the fly, \( V(C) \) may not be.

To estimate \( V(C) \), we use the fact that volume is a uniform measure on \( X \): The volume of \( C \) can be estimated as a fraction of the volume of \( X \) by means of uniform samples \( \{ x_{m+1}, \ldots, x_{m+n} \} \) over \( X \). Let \( U \) be the index set of these uniform samples. Then \( \frac{1}{n} \sum_{i \in U} 1[x_i \in C] \) is distributed as a scaled binomial random variable with mean \( \mu_C = \frac{\text{vol}(C)}{\text{vol}(X)} \) and variance \( \mu_C(1 - \mu_C)/n \).

If \( C \) places conditions on only \( s \) of the \( d \) dimensions, then \( \mu_C \) would be expected to scale exponentially with \( s \), not \( d \). Thus \( n \) does not need to be overly large to accurately estimate the volume of the rule set.

Given the above empirical estimator of volume, we reduce support estimation to a classification problem between the marginal density \( p(X) \) and a uniform distribution over \( X \). This is inspired by a similar reduction of support estimation (see Conclusion, p.695 in [31]).

\[
\hat{S} := \arg \min_C \frac{1}{|U|} \sum_{i \in U} 1[x_i \in C] + R(C) \quad \text{s.t.} \quad \sum_{i \in I} 1[x_i \in C] \geq \alpha m,
\]

(5)

where \( I = \{1, \ldots, m\} \). Problem (5) is a Neyman-Pearson-like classification problem with a false negative rate constraint of \( 1 - \alpha \) (instead of the usual false positive constraint).

### 4.2 Estimation of \( B^e \) as Binary Classification

Towards estimating \( B^e \), we follow in the tradition of using non-parametric (black-box) estimators of the group propensity \( p(T \mid X) \) to identify balanced cohorts in the study of causal effects [20, 9]. In particular, given an estimator of propensity \( \hat{p}(T \mid X) \), e.g. a random forest model, we assign labels to each data point indicating that propensity is bounded away from 0 and 1 in the following way:

\[
\forall i \in [m] : \hat{b}_i = \prod_{t \in T} 1[\hat{p}(T = t \mid X = x_i) \geq \epsilon] .
\]

(6)

Let \( \hat{B} = \{ x_i : \hat{b}_i = 1 \} \). Similar to the case of \( S^a \), we may now reduce rule set estimation of \( B^e \) to binary classification. Given \( \hat{S} \), the minimizer of (5), we restrict attention to samples in \( \hat{S} \) and again set up a Neyman-Pearson-like classification problem regarding the intersection \( \hat{S} \cap \hat{B} \) as the positive class:

\[
\hat{B} := \arg \min_C \frac{1}{|\hat{S} \setminus \hat{B}|} \sum_{i : x_i \in \hat{S} \setminus \hat{B}} 1[x_i \in C] + R(C) \quad \text{s.t.} \quad \sum_{i : x_i \in \hat{S} \cap \hat{B}} 1[x_i \in C] \geq \beta |\hat{S} \cap \hat{B}| .
\]

(7)
In this section, we derive an optimization procedure for (5) in the case where the hypothesis class $\mathcal{C}$ consists of Boolean DNF rules. The same procedure also solves (7). As the resulting DNF rule learning problem is an integer program (IP), we derive several heuristics for reducing computation.

DNF rules are also commonly known as rule sets, where the conjunctive clauses in the DNF correspond to individual rules in the set. As pointed out by [35], CNF rules can be learned by swapping class labels and fitting a DNF. Figure 1b exemplifies a DNF rule in $\mathbb{R}^2$. We assume that base features $X$ have been binarized to form literals such as (Age > 30) or (Sex = Female) indexed by a set $\mathcal{L}$, as is standard in e.g. decision tree learning. We let $\mathcal{K}$ index the set of all possible (exponentially many) conjunctions of literals in $\mathcal{L}$, e.g. (Age > 30) ∧ Female. Then, for $k \in \mathcal{K}$, let $a_{ik} \in \{0, 1\}$ denote the value taken by the $k$-th conjunction at sample $x_i$. Let the rule set be defined by $r \in \{0, 1\}^{|\mathcal{K}|}$ such that $r_k = 1$ indicates that the $k$-th conjunction is used in the rule set.

Recall that $\mathcal{U}$ indexes samples from a uniform reference measure and define an error variable $\xi_i$ for $i$ in $\mathcal{U}$ \cup $\mathcal{I}$ representing the penalty for covering or failing to cover point $i$, depending on its set membership.

Problem (5) may be reformulated accordingly as follows (similar to (7)),

$$\min_r \frac{1}{|\mathcal{U}|} \sum_{i \in \mathcal{U}} \xi_i + R(r) \quad \text{s.t.} \quad \left\{ r_k \in \{0, 1\}, \ k \in \mathcal{K}, \ \xi_i \geq 1 - \sum_{k \in \mathcal{K}} a_{ik} r_k, \ \xi_i \geq 0, \ i \in \mathcal{I}, \right.$$  

$$\left. \sum_{i \in \mathcal{I}} \xi_i \leq (1 - \alpha) m \quad \xi_i = \max_{k \in \mathcal{K}} (a_{ik} r_k), \ i \in \mathcal{U}. \right\} \quad \text{(8)}$$

Problem (8) is an IP with an exponential number of variables and is intractable as written. We follow the column generation approach of [7] to effectively manage the large number of variables and solve (8) approximately. As in that previous work, we bound from above the max operators in the constraints of (8) with sums (Hamming loss instead of zero-one loss) as it gives better numerical results. We also let $R(r) := \sum_{k \in \mathcal{K}} \lambda_k r_k$ with $\lambda_k = \lambda_0 + \lambda_1(\# \text{ conditions in conjunction } k)$ so that higher-degree conjunctions are more costly. These modifications yield an objective that is linear in $r$, $\sum_{k \in \mathcal{K}} (1/|\mathcal{U}|) \sum_{i \in \mathcal{U}} a_{ik} + \lambda_k) r_k$, with the same constraints as (8) except that $\xi_i, \ i \in \mathcal{U}$ has been absorbed into the objective. We then follow the overall procedure in [7] of solving the linear programming (LP) relaxation, using column generation to add variables only as needed.

We make the following departures from [7]. As noted, (8) has a constraint on false negative rate instead of a corresponding objective term and a penalty $R(r)$ on rule set complexity while [7] use a constraint. As a result, the LP reduced costs, used in column generation, are different. With $\mu_i \geq 0, \ i \in \mathcal{I}$ the dual variable for the top right constraint in (8), the reduced cost of conjunction $k$ is now $1/|\mathcal{U}| \sum_{i \in \mathcal{U}} a_{ik} + r_k - \sum_{i \in \mathcal{I}} \mu_i a_{ik}$, which remains a linear function of $a_{ik}$, allowing the same column generation method to be used. We also simplify the procedure of [7] to avoid the need for an IP solver by a) solving the column generation problem using a heuristic algorithm from [39], and b) once column generation terminates, we obtain an integer solution by taking the restriction of (8) to the final columns, converting to a weighted set cover problem, and applying a greedy set cover algorithm.
6 Experiments

We compare OverRule to black-box and rule-based estimators of overlap. The first baseline approximates the overlap region with the intersection of covariate (marginal) bounding boxes (CBB) to evoke classical balance checks in causal inference. The bounding boxes are selected to cover the $[(1-\alpha)/2, (1+\alpha)/2]$ percentiles of the data. Second, we use propensity score estimators as described in (6) with standard logistic regression (PS-LR) or $k$-nearest neighbors (PS-$k$NN) estimates of the propensity. Finally, we use One-Class Support Vector Machines (OSVM) to first estimate conditional supports and then overlap as their intersection.

The PS estimators can be viewed as a binary version of overlap weights [20] and CBB as the standard practice of comparing basic covariate metrics. These baselines are also used as black-box estimators by OverRule. In addition, we compare our results to the MaxBox (MB) framework used by [9] to learn interpretable study populations.

When estimating support in OverRule, we use $m_R = m \cdot d$ uniform reference samples where $m$ is the number of data samples and $d$ their dimension. Parameters $\lambda_0, \lambda_1$ were selected from the range $[10^{-4}, 1]$ for estimation of both $S$ and $B$. The choice of $\lambda_1$ had small impact and was set to $\lambda_1 = 0$ unless otherwise specified. For propensity-based base estimators, we used a threshold $\epsilon = .1$. For $k$-NN we selected $k$ based on held-out accuracy in predicting group membership and used $1/k$ as threshold. For OSVM, we use a Gaussian RBF-kernel with bandwidth selected based on the held-out likelihood of kernel density estimator with the same bandwidth. To select hyperparameters for the rule-based models, and to assess quality when overlap is known, we use the balanced accuracy [5].

We evaluate OverRule in a series of real-world experiments. To give an understandable illustration of our definition and method, we estimate regions of classifier uncertainty in the famous Iris dataset. Estimation of treatment group overlap in the Jobs dataset [19, 32] provides an example with "ground truth". A study of overlap in opioid prescriptions gives us a large-scale real-world example with a user study, and policy evaluation for antibiotic prescriptions points to the versatility of the core methodology. A synthetic experiment showing the utility of characterizing overlap in observational studies of causal effects both before and after estimation can be found in the supplement.

6.1 Classifier Uncertainty: Iris

We use OverRule to identify the overlap between members of two species of Iris, as represented by their sepal and petal length and width, based on the famous Iris dataset. We fit OverRule using a $k$-NN base estimator ($k = 8$) and DNF Boolean rules with high regularization ($\lambda_0 = 0.7, \lambda_1 = 0.01$). In Figure 2 we present the rules learned to characterize $B$ and compare them with the rules learned for a binary classifier of group membership. In contrast, the coefficients of a logistic regression propensity score model, $\beta = [-1.7, -1.5, 2.5, 2.6]'$ reveal very little about which points lie in the overlap set.

6.2 Observational Study: Jobs

In a famous trial performed to study the effects of job training [19, 34], eligible US citizens were randomly selected into ($T = 1$), or left out of ($T = 0$) job training programs. The RCT ($E = 1$), which satisfies overlap by definition, has since been combined with non-experimental control samples ($E = 0, T = 0$), forming a larger observational set (Jobs), to serve as a benchmark for causal effect estimation [19]. Here, we aim to characterize the overlap between treated and control subjects.
Figure 2: Overlap in the Iris dataset identified by OverRule (left). Colored dots represent samples of the Versicolor (blue) and Virginica (red) species. Black circles indicate predicted overlap. Green rectangles indicate learned rules projected to 2D (hence there are points in green fields that are not in $O$). The learned rules are listed in the top right. In (b), we show the rules (red regions) learned by a group membership classifier illustrating that these rules are not sufficient for characterizing overlap. The code to reproduce the OverRule (left) figure is available at https://www.github.com/clinicalml/overrule-code

Table 1: Results for overlap estimation on the Jobs dataset.

|                | CBB         | OR          | OSVM        | MB          | OR          | PS-$k$NN      | MB          | OR          | PS-LR       | MB          | OR          |
|----------------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|
| Acc            | 0.75 ± 0.02 | 0.82 ± 0.02 | 0.81 ± 0.02 | 0.68 ± 0.01 | 0.85 ± 0.01 | 0.90 ± 0.02  | 0.84 ± 0.01 | 0.90 ± 0.01 | 0.97 ± 0.01  | 0.80 ± 0.02  | 0.88 ± 0.02  |
| FPR            | 0.12 ± 0.01 | 0.17 ± 0.02 | 0.23 ± 0.03 | 0.09 ± 0.02 | 0.24 ± 0.04 | 0.14 ± 0.03  | 0.03 ± 0.01 | 0.15 ± 0.01 | 0.07 ± 0.01  | 0.04 ± 0.01  | 0.13 ± 0.03  |
| FNR            | 0.38 ± 0.03 | 0.19 ± 0.03 | 0.15 ± 0.03 | 0.54 ± 0.01 | 0.06 ± 0.02 | 0.05 ± 0.02  | 0.29 ± 0.02 | 0.06 ± 0.02 | 0.11 ± 0.03  | 0.35 ± 0.04  | 0.10 ± 0.01  |
| #Literals      |             |             | 15          |             |             |              | 16          |             | 19           |             | 33          |

(a) Overlap estimation in Jobs. Balanced accuracy (Acc), false positive rate (FPR), false negative rate (FNR), with standard deviations over 5-fold CV. MB and OR indicate MaxBox and OverRule with base estimator from the line above. MB did not run with CBB.

(b) Example rule set describing $B$ for the Jobs dataset learned using the propensity base estimator. Used with support rules (see the supplement), it achieves 0.90 balanced accuracy in identifying overlap.

| Rule B.1         | Age ≤ 49                  | and Black | and RE74 ≤ $13700$ | and RE75 ≤ $13400$ |
|------------------|---------------------------|-----------|---------------------|---------------------|
|                  |                           |           |                     |                     |
| or Rule B.2      | Not married               | and       | Education ≤ 16 years | and RE74 ≤ $21390$  | and RE75 ≤ $17200$ |
|                  |                           |           |                     |                     |

As a consequence of the trial’s eligibility criteria, the experimental and non-experimental cohorts barely overlap; a standard logistic regression estimator separate the experimental and non-experimental groups with held-out balanced accuracy of 0.96. Since all treated subjects were part of the experiment, the experimental cohort perfectly represents the overlap region. For this reason, we use the experiment
indicator $E$ as ground truth for $O$. In studies of causal effects in this data, the following features were included to adjust for confounding: Age, #Years in education (Educ), Race (black/hispanic/other), Married, No degree (NoDegr), Real earnings in 1974 (RE74) and in 1975 (RE75). These are the features $X$ for which we estimate overlap between treated and controls.

The results for Jobs can be seen in Table 1. We present the results for the smallest rules that achieve balanced accuracy within 1% of that of the best performing model within each class. We see that for most base estimators, the OverRule approximations perform slightly worse than the base estimator, but with a simpler description. OverRule compares favorably to MaxBox in all cases. In the supplement, we give plots which show that the held-out balanced accuracy quickly converges with the number of literals in the rules and correlates strongly with the quality by which the rule set approximates the base estimator. The learned rules in Table 1b conform to our expectations as the eligibility criteria for the RCT allow only subjects who were currently unemployed and had been so for most of the time leading up to the trial—factors that correlate with education and marital status.

### 6.3 Observational Study: Opioid Misuse

Opioid misuse affects millions of Americans and understanding the factors that influence the risk of misuse is of great importance. To this end, [4] and [41] study the effect of choices in opioid prescriptions on the risk of future misuse. In this experiment, we study a group of post-surgical patients who were given opioid prescriptions within 7 days of surgery. We compare patients who were given doses, morphine milligram equivalent (MME), above and below the 85th percentile in the selected cohort, MME=450. We replicate the cohort eligibility criteria of [4], using a subset of the MarketScan insurance claims database. Subjects were represented by basic demographics (age, sex), diagnosis history and procedures billed as surgical on the index date. Note that surgical procedures are not mutually exclusive. We list first-order statistics of these features in the supplement.

We fit an OverRule model (OR) to a random forest base estimator with $\beta = 0.8$ for $B$ and $\alpha = 0.9$ for $S$ picked a priori. The hyperparameter $\lambda_0$ was set to $\lambda_0 = 1e^{-3}$ for $B$, chosen based on balanced accuracy w.r.t. the base estimator, and $\lambda_0 = 1e^{-5}$ for $S$ based on accuracy in classifying the reference measure. For comparison, we fit a MaxBox model (MB) [9] to the same base, and another OverRule model describing the complement of $O$ (OR-C). The balanced accuracy of these models w.r.t. the base were 0.90 (OR), 0.77 (MB) and 0.92 (OR-C). In Figure 3, we summarize the rules learned by OR which cover 27% of the overall population. MB learned the rule: Musculoskeletal surg. $\wedge$ Mediastinum surg. $\wedge$ Male genital surg. $\wedge$ Maternity surg. $\wedge$ Lumbosacral spondylosis without myelopathy which covers 17% of patients. The rules learned by OR-C are presented in the supplement. To evaluate the interpretability of learned overlap sets, we conducted a qualitative user study through a moderated discussion with three participants: two attending surgeons (P1 & P2) and a 4th year medical student (P3) at a large US teaching hospital. Before seeing the outputs of any method, the participants were asked to give their expectations for what to find in the overlap set. The full discussion was transcribed, anonymized, and included in the supplement.

The participants expected that the overlap set would mostly correspond to patients in the higher dose range, as these patients are often considered also for smaller doses, and that overlap would be driven largely by surgery type. All participants expected Musculoskeletal and Cardiovascular surgery patients to be predominantly in the higher dose group, and sometimes in the lower, and one suggested

---

2This may introduce a small number false negatives in the label used as ground truth.
Support rules $\hat{S}$

Rule S.1:
- History:
  - $\neg$ Injury of face and neck
  - $\neg$ Unspecified septicemia
  - $\neg$ Other injury of chest wall
  - $\neg$ Acute respiratory failure
  - $\neg$ Altered mental status
  - Surgical procedure:
    - $\neg$ Endocrine system
    - $\neg$ Mediastinum (thoracic cavity)
    - $\neg$ Auditory system

Propensity overlap rules $\hat{B}$

Rule B.1:
- Surgical procedure:
  - Musculoskeletal

or Rule B.2:
- Age $> 44$
- Male
- Surgical procedure:
  - Cardiovascular
  - $\neg$ Urinary system (e.g., bladder)
  - $\neg$ Male genital system

Rule B.3:
- Surgical procedure:
  - Nervous (e.g., epidural)
  - $\neg$ Maternity (e.g., C-section)
  - $\neg$ Female genital system

or Rule B.4:
- Age $> 23$
- Surgical procedure:
  - Thoracic or lumbosacral neuritis or radiculitis

$\hat{D} = S.1 \land (B.1 \lor B.2 \lor B.3 \lor B.4)$

Figure 3: OverRule description of the overlap between post-surgical patients with higher and lower opioid prescriptions. If the support rule (left) applies and either propensity overlap rule (right) applies, a patient is considered to be in the overlap set. $\neg$ indicates a negation. The rules cover 27% of patients with balanced accuracy of 0.90 w.r.t. the base estimator. Procedures are not mutually exclusive.

that Maternity surgeries (e.g., C-sections) would be only in the lower range. These comments are all consistent with the findings of OverRule, which identified all of these surgery types as important. MaxBox identified only Musculoskeletal surgery patients as overlapping. One participant expected history of psychiatric disease and Tobacco use disorder to be predictive of higher prescription doses for some patients, and thus overlap. Neither method identified psychiatric disease, but Tobacco use disorder was identified by OR-C as anti-correlated with exclusion from overlap (see the supplement).

The participants found the support rules ($\hat{S}$) output by OR (Figure 3 left) intuitive and P1 stated that Endocrine surgeries are not typically followed by opioid prescriptions. They found the MaxBox and OR rule descriptions easy to interpret, and discussion focused on their clinical meaning. The first three propensity overlap rules B.1-B.3 were all consistent with expectation as described above, with the caveat that Cardiovascular patients are not typically stratified by Urinary and Genital surgeries. This was later partially explained by catheters being billed as Urinary and P3 interpreted it as a proxy for more severe Cardiovascular surgeries. P1 pointed out the value in discovering such surprising patterns that may be hidden in black-box analyses. The OR-C rules were found hard to interpret due to many double negatives (“excluded from exclusion”), but were ultimately deemed clinically sound.

6.4 Observational Study: Policy Evaluation of Antibiotic Prescription Guidelines

Using the policy evaluation formulation of $B^*(\pi)$ (see Section 3), we apply OverRule to assess the overlap set for a policy that follows clinical guidelines published by the Infectious Disease Society of America (IDSA) for treatment of uncomplicated urinary tract infections (UTIs) in female patients [12]. We use data derived from the electronic medical record of two academic medical centers.

We apply the OverRule algorithm to a broad cohort (e.g., including men) of 65,000 UTI patients to test whether or not it can recover a clinically meaningful overlap set. From a qualitative perspective, we discussed the results with an infection disease specialist, who verified that the resulting rules have a clear clinical interpretation which aligns with how the guidelines are applied in practice, identifying primarily female outpatient cases and uncomplicated female inpatient cases. From a quantitative
perspective, we compared the learned region (covering 42k patients, 64% of total) with a subset of patients selected a-priori to be eligible for the guidelines (14k patients, 21% of total). We found that the former covers 96% of the latter while also including a much broader, but clinically intuitive, cohort. The experimental setup and results are described in more detail in the supplement.

7 Conclusion

We have presented OverRule—an algorithm for learning rule-based characterizations of overlap between distributions, or the inputs for which policy evaluation is feasible. The algorithm learns to exclude points marginally out-of-distribution, as well as points where either distribution/policy has low density. We evaluated the algorithm in characterizing overlap in causal effect estimation, and demonstrated that our rule descriptions often have similar accuracy to black-box estimators and outperform a competitive baseline. In an application to study treatment group overlap in post-surgical opioid prescription, a qualitative user study found the results interpretable and clinically meaningful. Similar observations were made in an application to evaluation of antibiotic prescription policies.

Acknowledgments

We thank Chloe O’Connell and Charles S. Parsons for providing clinical feedback on the opioid misuse experiment, and Sanjat Kanjilal for providing clinical feedback on the antibiotic prescription experiment. We also thank Bhanukiran Vinzamuri for assistance with the opioids data, David Amirault for insightful suggestions and feedback, and members of the Clinical Machine Learning group for feedback on earlier drafts.

References

[1] Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, and Cynthia Rudin. Learning certifiably optimal rule lists. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pages 35–44, 2017.

[2] Susan Athey and Stefan Wager. Efficient policy learning. arXiv preprint arXiv:1702.02896, 2017.

[3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. Machine learning, 79(1-2):151–175, 2010.

[4] Gabriel A Brat, Denis Agniel, Andrew Beam, Brian Yorkgitis, Mark Bicket, Mark Homer, Kathe P Fox, Daniel B Knecht, Cheryl N McMahill-Walraven, Nathan Palmer, et al. Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: retrospective cohort study. Bmj, 360:j5790, 2018.

[5] Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. The balanced accuracy and its posterior distribution. In 2010 20th International Conference on Pattern Recognition, pages 3121–3124. IEEE, 2010.
[6] Alexander D’Amour, Peng Ding, Avi Feller, Lihua Lei, and Jasjeet Sekhon. Overlap in observational studies with high-dimensional covariates. *arXiv preprint arXiv:1711.02582*, 2017.

[7] Sanjeeb Dash, Oktay Gunluk, and Dennis Wei. Boolean decision rules via column generation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 4660–4670. Curran Associates, Inc., 2018.

[8] John HJ Einmahl and David M Mason. Generalized quantile processes. *The Annals of Statistics*, pages 1062–1078, 1992.

[9] Colin B Fogarty, Mark E Mikkelsen, David F Gaieski, and Dylan S Small. Discrete optimization for interpretable study populations and randomization inference in an observational study of severe sepsis mortality. *Journal of the American Statistical Association*, 111(514):447–458, 2016.

[10] Alex A Freitas. Comprehensible classification models: a position paper. *ACM SIGKDD explorations newsletter*, 15(1):1–10, 2014.

[11] Siong Thye Goh and Cynthia Rudin. Cascaded high dimensional histograms: A generative approach to density estimation. *arXiv preprint arXiv:1510.06779*, 2015.

[12] Kalpana Gupta, Thomas M Hooton, Kurt G Naber, Bjorn Wullt, Richard Colgan, Loren G Miller, Gregory J Moran, Lindsay E Nicolle, Raul Raz, Anthony J Schaeffer, and David E Soper. International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: A 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. *Clinical Infectious Diseases*, 52(5):e103–20, mar 2011.

[13] Francisco Herrera, Cristóbal José Carmona, Pedro González, and María José Del Jesus. An overview on subgroup discovery: foundations and applications. *Knowledge and information systems*, 29(3):495–525, 2011.

[14] Stefano M Iacus, Gary King, and Giuseppe Porro. Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24, 2012.

[15] Herman Kahn. Use of Different Monte Carlo Sampling Techniques. Technical report, RAND Corporation, Santa Monica, California, 1955.

[16] Nathan Kallus. Generalized optimal matching methods for causal inference. *arXiv preprint arXiv:1612.08321*, 2016.

[17] Nathan Kallus and Angela Zhou. Policy evaluation and optimization with continuous treatments. *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, 84:1243–1251, 09–11 Apr 2018.

[18] Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. 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*, pages 1675–1684. ACM, 2016.
[19] Robert J LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, pages 604–620, 1986.

[20] Fan Li, Kari Lock Morgan, and Alan M Zaslavsky. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400, 2018.

[21] Judea Pearl. *Causality*. Cambridge university press, 2009.

[22] Doina Precup, Richard S Sutton, and Satinder P Singh. Eligibility Traces for Off-Policy Policy Evaluation. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML)*, pages 759–766, 2000.

[23] Parikshit Ram and Alexander G Gray. Density estimation trees. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 627–635. ACM, 2011.

[24] Ronald L Rivest. Learning decision lists. *Machine learning*, 2(3):229–246, 1987.

[25] Paul R Rosenbaum. Optimal matching for observational studies. *Journal of the American Statistical Association*, 84(408):1024–1032, 1989.

[26] Paul R Rosenbaum. *Design of observational studies*, volume 10. Springer, 2010.

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

[28] Guillermo V Sanchez, Ahmed Babiker, Ronald N Master, Tony Luu, Anisha Mathur, and Jose Bordon. Antibiotic Resistance among Urinary Isolates from Female Outpatients in the United States in 2003 and 2012. *Antimicrobial Agents and Chemotherapy*, 60(5):2680–2683, 2016.

[29] Connie Schardt, Martha B Adams, Thomas Owens, Sheri Keitz, and Paul Fontelo. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC medical informatics and decision making*, 7:16, jun 2007.

[30] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[31] Clayton D Scott and Robert D Nowak. Learning minimum volume sets. *Journal of Machine Learning Research*, 7(Apr):665–704, 2006.

[32] Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3076–3085. JMLR. org, 2017.

[33] Daniel J Shapiro, Lauri A Hicks, Andrew T Pavia, and Adam L Hersh. Antibiotic prescribing for adults in ambulatory care in the USA, 2007–09. *Journal of Antimicrobial Chemotherapy*, 69(1):234–240, 2013.

[34] Jeffrey A Smith and Petra E Todd. Does matching overcome lalonde’s critique of nonexperimental estimators? *Journal of econometrics*, 125(1-2):305–353, 2005.
[35] Guolong Su, Dennis Wei, Kush R. Varshney, and Dmitry M. Malioutov. Learning sparse two-level Boolean rules. In Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP), pages 1–6, September 2016.

[36] Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. MIT Press, 2nd edition, 2017.

[37] Fulton Wang and Cynthia Rudin. Falling rule lists. In Artificial Intelligence and Statistics, pages 1013–1022, 2015.

[38] Tong Wang, Cynthia Rudin, Finale Doshi-Velez, Yimin Liu, Erica Klampfl, and Perry MacNeille. A Bayesian framework for learning rule sets for interpretable classification. Journal of Machine Learning Research, 18(70):1–37, 2017.

[39] Dennis Wei, Sanjeeb Dash, Tian Gao, and Oktay Gunluk. Generalized linear rule models. In Proceedings of the 36th International Conference on Machine Learning (ICML), 2019.

[40] Hongyu Yang, Cynthia Rudin, and Margo Seltzer. Scalable Bayesian rule lists. In Proc. Int. Conf. Mach. Learn. (ICML), pages 1013–1022, 2017.

[41] Jinghe Zhang, Vijay Iyengar, Dennis Wei, Bhanukiran Vinzamuri, Hamsa S. Bastani, Alexander R. Macalalad, Anne E. Fischer, Gigi Yuen-Reed, Aleksandra Mojsilovic, and Kush R. Varshney. Exploring the causal relationships between initial opioid prescriptions and outcomes. In AMIA Workshop on Data Mining for Medical Informatics, Washington, DC, November 2017.

[42] José R Zubizarreta. Using mixed integer programming for matching in an observational study of kidney failure after surgery. Journal of the American Statistical Association, 107(500):1360–1371, 2012.
Appendix

A Generalization to Policy Evaluation

In this section we give the detailed algorithm for applying OverRule to policy evaluation, as described in the main paper. In this context, we wish to evaluate not a specific treatment decision (e.g., the average treatment effect of giving a drug vs. withholding it), but rather a conditional policy representing a personalized treatment regime, which we will refer to as the target policy. This problem falls under the setting of off-policy policy evaluation when this target policy differs from the policy which generated the data, which we observe in the observational data as \( p(T = t | x) \).

**Rationale for \( \mathcal{B}^\epsilon(\pi) \)** In the main paper, we drew a connection between the set \( \mathcal{B}^\epsilon \) and the following set, which is a function of the target policy \( \pi \), \( \mathcal{B}^\epsilon(\pi) := \{ x \in X; \forall t : \pi(t | x) > 0 : p(T = t | x) > \epsilon \} \). In this section, we will provide theoretical justification for why we are restricted to this set, if we wish to evaluate the policy \( \pi \) given samples generated according to \( p(T = t | x) \).

Following similar notation to [17], we will let \( X \in \mathcal{X} \) correspond to covariates, \( Y \in \mathcal{Y} \) to an outcome of interest, \( T \in \mathcal{T} \) to a treatment decision. We write \( \pi(t | x) \) as the probability of each treatment under the policy, which may be stochastic. We write \( Y(t) \) to represent the potential outcome under treatment \( t \). In this setting, we wish to evaluate the expected value of \( Y \) under the target policy, which we denote as \( \mathbb{E}[Y(\pi)] \). For our purposes, we note the following as motivation for our definition of \( \mathcal{B}^\epsilon(\pi) \),

**Proposition 1 (Informal).** The expectation \( \mathbb{E}[Y(\pi)] \) is only defined w.r.t. the observed distribution \( p(X, T, Y) \) for the subset \( B \in \mathcal{X} \) such that \( \forall x \in B \), \( \pi(T = t | X = x) > 0 \implies p(T = t | X = x) > 0 \).

**Proof.** Under the assumption that ignorability [21] holds, we can write out our desired quantity as follows in terms of observed distribution \( p(X, T, Y) \)

\[
\mathbb{E}[Y(\pi)] = \int_{X,T,Y} y \cdot p(X = x) \pi(T = t | X = x)p(Y(t) = y | X = x, T = t) dx dtdy
\]

\[= \int_{X,T,Y} y \cdot p(X = x) \frac{\pi(T = t | X = x)}{p(T = t | X = x)} p(Y(t) = y | X = x, T = t)p(T = t | X = x) dx dtdy \tag{10} \]

\[= \int_{X,T,Y} y \cdot p(X = x)p(T = t | X = x)p(Y = y | X = x, T = t) \frac{\pi(T = t | X = x)}{p(T = t | X = x)} dx dtdy \tag{11} \]

\[= \int_{X,T,Y} y \cdot p(X = x, T = t, Y = y) \frac{\pi(T = t | X = x)}{p(T = t | X = x)} dx dtdy \tag{12} \]

Where in Equation (10) we multiply by one, in Equation (11) we use the assumption of ignorability to write \( p(Y(t) = y | X = x, T = t) = p(Y = y | X = x, T = t) \) and rearrange terms, and in Equation (12) we collect the terms which represent the observed distribution. For our purposes, it is sufficient to look at the integral in Equation (12) to see that it requires the condition that for all \( (x, t) \in X \times \mathcal{T} \), the relationship \( \pi(T = t | X = x) > 0 \implies p(T = t | X = x) > 0 \) must hold. □

The condition given in Proposition 1 is sometimes referred to as the condition of coverage [see 36, Section 5.5] in off-policy evaluation. Rewriting Equation (12) as an expectation over the observed
distribution, we can see that this leads naturally to the importance sampling estimator
\[
\mathbb{E} \left[ \frac{\pi(T = t \mid X = x)}{p(T = t \mid X = x)} \right] \approx \frac{1}{n} \sum_{i=1}^{n} y_i \frac{\pi(t_i \mid x_i)}{p(t_i \mid x_i)},
\]
which approximates our desired quantity. If \( \epsilon > p(t \mid x) > 0 \) for some small value of \( \epsilon \), then the variance of the importance sampling estimator increases dramatically. This motivates our notion of “strict” coverage, that for each value of \( x \in B^\epsilon(\pi) \), we require that for all actions \( t \) such that \( \pi(t \mid x) > 0 \), the condition \( p(t \mid x) > \epsilon \) must hold.

Note that this differs conceptually from the binary treatment case in an important respect: Since we are not seeking to contrast all treatments, we do not require that \( \mu(t \mid x) > \epsilon \), \( \forall t \in \mathcal{T} \), but rather just for those treatments which have positive probability of being taken under the target policy.

**Algorithmic Details**  As described in the main paper, applying OverRule to the policy evaluation setting only requires a single change to the procedure, which is that the set \( \hat{B}^\epsilon(\pi) \) is used in place of the set \( \hat{B}^\epsilon \) in Equation 7 in Section 4.2. Nonetheless, we provide an explicit self-contained sketch of the procedure here to avoid any confusion:

1. Given a dataset, find an \( \alpha \)-MV set \( S^\alpha \) using the approach given in the main paper.

2. Using this set, learn the conditional probabilities of each possible treatment \( t \in \mathcal{T} \), resulting in estimated propensities \( \hat{p}(T = t \mid X = x) \).

3. For each data point in the support set \( S^\alpha \), assign the label
\[
\hat{b}_i(\pi) = \prod_{t \in \pi(x_i)} \mathbb{1}[\hat{p}(T = t \mid X = x) \geq \epsilon],
\]
where \( \pi(x_i) := \{ t : \pi(t \mid x_i) > 0 \} \). The set \( \hat{B}^\epsilon(\pi) \) is the collection of data points such that \( \hat{b}_i(\pi) = 1 \). Note that we know the target policy \( \pi \) that we are evaluating, so we can evaluate \( \pi(t \mid x_i) \) for each data point.

4. Solve the following Neyman-Pearson-like classification problem, using the techniques discussed in the main paper. Note that this is identical to solving Equation 7 in Section 4.2, with the substitution of \( \hat{B}^\epsilon(\pi) \) for \( \hat{B}^\epsilon \):
\[
\hat{B}(\pi) := \arg \min_C \frac{1}{|\hat{S} \setminus \hat{B}|} \sum_{i \in \hat{S} \setminus \hat{B}^\epsilon(\pi)} \mathbb{1}[x_i \in C] + R(C) \quad \text{s.t.} \quad \sum_{i \in \hat{S} \cap \hat{B}^\epsilon(\pi)} \mathbb{1}[x_i \in C] \geq \beta |\hat{S} \cap \hat{B}^\epsilon(\pi)|.
\]

**B Additional experimental results**

**B.1 Synthetic task: Estimating causal effects**

To illustrate the utility of characterizing overlap before estimating causal effects, we perform a synthetic observational study of the average treatment effect (ATE) of a treatment \( T \) on an outcome \( Y \) under confounding by a variable \( X \). Under the distribution \( p(X, T, Y) \), the true ATE on the overlap set \( O \) is \( \text{ATE}_O = \mathbb{E}_X [\mathbb{E}_Y (Y \mid T = 1, X) - \mathbb{E}_Y (Y \mid T = 0, X) \mid X \in O] \). We compare estimates of \( \text{ATE}_O \)
Table 2: Synthetic task. Mean absolute error (MAE) in estimates of the ATE on $O$. (Lower is better)

|               | DiM ($D$) | OLS ($D$) | DiM ($\hat{O}_0$) | OLS ($\hat{O}_0$) | DiM ($\hat{O}_r$) | OLS ($\hat{O}_r$) |
|---------------|-----------|-----------|--------------------|--------------------|--------------------|--------------------|
| MAE           | 0.39 ± 0.01 | 0.10 ± 0.02 | 0.04 ± 0.01 | 0.02 ± 0.01 | 0.03 ± 0.01 | 0.01 ± 0.01 |

Table 3: Example rule set describing the support $S$ for the Jobs dataset learned using OverRule. The support rule set had parameters $\alpha = 0.95, \lambda_0 = 10^{-2}$.

| Rule S.1          | or Rule S.2 | or Rule S.3  |
|-------------------|-------------|--------------|
| Age > 20          | Not Hispanic | Not black    |
| and educ > 8      | and RE74 ≤ $33220$ | and Hispanic |
| and Not Black     | and RE75 ≤ $322000$ | and Education ≤ 12 years |
| and Not Hispanic  | and RE74 ≤ $322000$ | and RE74 ≤ $21900$ |
| and Married       | and RE75 ≤ $322000$ | and RE75 ≤ $9850$ |
| and RE74 > $6270$ |             |              |
| and RE75 > $1162$ |             |              |

based on a) all available data $D$ and b) an estimate of the overlap region $\hat{O}$. For global estimators that are sensitive to misspecification, using samples outside of $O$ often worsens estimates of $\text{ATE}_O$.

We generate $X, T$ according to a 2D Gaussian mixture model with $T \sim U\{0, 1\}$ the mixture component and $X \mid T \sim N(2T, 1)$. We let $Y = \sigma(W_1X + W_2XT + b + \epsilon_n)$, with $\sigma$ the logistic function, $W_1, W_2$ linear weights, $b$ a constant offset, and $\epsilon_n \sim N(0, 1)$ noise. We compare two estimators of ATE: difference in means (DiM) ($\text{ATE}_{\text{DiM}}(D) = 1/n_1 \sum_{i \in D, t_i = 1} y_i - 1/n_0 \sum_{i \in D, t_i = 0} y_i$) and ordinary least-squares regression ($\text{ATE}_{\text{OLS}}(D) = 1/N \sum_{i \in D} (f_1(x_i) - f_0(x_i))$) where $f_1$ and $f_0$ are OLS models fit to $X, Y$ for treatment groups 0 and 1, respectively. We also fit two OverRule models ($\hat{O}_r$ and $\hat{O}_u$) with $\lambda_0 = 0.01$ and $\lambda_0 = 0$, respectively, corresponding to weak and no regularization.

We report the mean absolute $O$ over repeated experiments to measure performance. As seen in Table 2, models based on $\hat{O}_u$ or $\hat{O}_r$, using only samples in the estimated overlap regions, give better mean absolute error (MAE) than using the whole data, DiM($D$) and OLS($D$). $\text{ATE}_{\text{OLS}}$ performs better than $\text{ATE}_{\text{DiM}}$ for all three cases as it adjusts for confounding and regularized $\hat{O}_r$ outperforms unregularized $\hat{O}_u$. The results show that, under model mis-specification, restricting the study to comply with the overlap assumption can be beneficial also prior to estimation.

B.2 Jobs

In figure 5 we see the correlation between held-out AUC for the rule set w.r.t. the experimental label, and the AUC for the rule set in approximating the base estimator. AUC is equal to balanced accuracy for binary predictions.

B.3 Opioids

For a full table of covariate statistics for the Opioids dataset, see Table 4. For a illustration of the rules learned by OverRule to describe the complement of the overlap set, see Figure 4.
Support rules $\hat{\delta}$

Rule S.1:

History:
- $\neg$ Injury of face and neck
- $\neg$ Unspecified septicemia
- $\neg$ Other injury of chest wall
- $\neg$ Acute respiratory failure
- $\neg$ Altered mental status
- Surgical procedure:
  - $\neg$ Endocrine system
  - $\neg$ Mediastinum (thoracic cavity)
  - $\neg$ Auditory system

Propensity overlap complement rules $\overline{\delta}^c$

Rule B.1:

Surgical procedure:
- $\neg$ Respiratory
- $\neg$ Nervous
- $\neg$ Musculoskeletal
- $\neg$ Cardiovascular
- History:
  - $\neg$ Tobacco use disorder
  - $\neg$ Thoracic or lumbosacral neuritis or radiculitis: unspecified
  - $\neg$ Lumbosacral spondylosis without myelopathy
  - $\neg$ Degeneration of cervical intervertebral disc
  - $\neg$ Degeneration of lumbar or lumbosacral intervertebral disc

Rule B.2:

Surgical procedure:
- Maternity
- History:
  - $\neg$ Degeneration of lumbar or lumbosacral intervertebral disc

$\hat{\delta} = \text{S.1} \land \neg (\text{B.1} \lor \text{B.2})$

Figure 4: OverRule description of the complement of the overlap between post-surgical patients with higher and lower opioid prescriptions. If the support rule (left) applies and neither propensity overlap rule (right) applies, a patient is considered to be in the overlap set. $\neg$ indicates a negation. The rules cover 36% of patients with balanced accuracy 0.92 w.r.t. the base estimator (random forest). Procedures are not mutually exclusive.

Table 4: Population averages for covariates in Opioids in order of difference between the overlapping and non-overlapping set. DMME, MME and Duration are the medians of daily MME, total MME and prescription duration days in each group.

|                    | Total    | DMME  | MME  | Duration |
|--------------------|----------|-------|------|----------|
| Total sample       | 35106    | 46    | 225  | 5        |
| Male               | 9301     | 50    | 300  | 5        |
| Female             | 25805    | 45    | 225  | 5        |
| Age groups         |          |       |      |          |
| <15                | 847      | 20    | 100  | 5        |
| 15-24              | 3334     | 45    | 200  | 5        |
| 25-34              | 9994     | 45    | 210  | 4        |
| 35-44              | 6820     | 46    | 225  | 5        |
| 45-54              | 6196     | 50    | 250  | 5        |
| 55-64              | 7915     | 50    | 300  | 5        |
| >=65               | 0        | 0     | 0    | 0        |
| Surgery type       |          |       |      |          |
| Auditory           | 29       | 18    | 135  | 6        |
| Cardiovascular     | 3633     | 45    | 270  | 5        |
| Integumentary      | 1507     | 48    | 225  | 5        |
| Mediastinum        | 54       | 47    | 300  | 5        |
| Female genital     | 3913     | 48    | 225  | 5        |
| Hemic              | 885      | 50    | 225  | 5        |
| Diagnosis category                                                                 | Count | Code | Value | Year |
|---------------------------------------------------------------------------------|-------|------|-------|------|
| Respiratory                                                                     | 665   | 45   | 250   | 5    |
| Endocrine                                                                       | 214   | 45   | 200   | 5    |
| Nervous                                                                         | 4350  | 60   | 375   | 6    |
| Urinary                                                                         | 1476  | 45   | 225   | 5    |
| Musculoskeletal                                                                  | 6678  | 60   | 450   | 7    |
| Maternity                                                                        | 13553 | 45   | 200   | 4    |
| Male genital                                                                     | 585   | 45   | 225   | 5    |
| Year                                                                             |       |      |       |      |
| 2011                                                                             | 7547  | 45   | 225   | 5    |
| 2012                                                                             | 10743 | 46   | 225   | 5    |
| 2013                                                                             | 9651  | 50   | 225   | 5    |
| 2014                                                                             | 7165  | 45   | 225   | 5    |
| Diagnosis history (until day before surgery)                                    |       |      |       |      |
| Other specified gastritis: without mention of hemorrhage                          | 491   | 42   | 225   | 5    |
| Other ascites                                                                    | 233   | 45   | 225   | 5    |
| Lumbosacral spondylosis without myelopathy                                       | 1135  | 60   | 400   | 6    |
| Nausea with vomiting                                                             | 1914  | 45   | 225   | 5    |
| Other respiratory abnormalities                                                  | 1935  | 45   | 225   | 5    |
| Vomiting alone                                                                   | 765   | 45   | 200   | 5    |
| Myalgia and myositis: unspecified                                                | 1522  | 50   | 250   | 5    |
| Attention deficit disorder with hyperactivity                                    | 370   | 45   | 225   | 5    |
| Attention deficit disorder without mention of hyperactivity                      | 444   | 45   | 225   | 5    |
| Depressive disorder: not elsewhere classified                                    | 2221  | 50   | 225   | 5    |
| Dysphoric disorder                                                               | 752   | 50   | 225   | 5    |
| Tachycardia: unspecified                                                         | 631   | 45   | 225   | 5    |
| Degeneration of cervical intervertebral disc                                     | 904   | 56   | 337   | 6    |
| Flatulence: eructation: and gas pain                                            | 427   | 45   | 225   | 5    |
| Generalized anxiety disorder                                                     | 833   | 45   | 225   | 5    |
| Other symptoms referable to back                                                | 368   | 50   | 300   | 5    |
| Cellulitis and abscess of leg: except foot                                       | 450   | 45   | 225   | 5    |
| Constipation: unspecified                                                       | 1136  | 45   | 225   | 5    |
| Thoracic or lumbosacral neuritis or radiculitis: unspecified                    | 1676  | 60   | 326   | 6    |
| Anxiety state: unspecified                                                      | 2205  | 50   | 225   | 5    |
| Lumbago                                                                          | 4559  | 50   | 250   | 5    |
| Abdominal pain: generalized                                                     | 1607  | 45   | 225   | 5    |
| Degeneration of lumbar or lumbosacral intervertebral disc                       | 1542  | 60   | 388   | 6    |
| Other and unspecified noninfectious gastroenteritis and colitis                 | 1254  | 45   | 225   | 5    |
| Major depressive affective disorder: recurrent episode: moderate                 | 507   | 45   | 225   | 5    |
| Asthma: unspecified type: unspecified                                           | 2044  | 45   | 225   | 5    |
| Arthrodesis status                                                              | 178   | 60   | 450   | 7    |
| Chest pain: unspecified                                                         | 4701  | 45   | 225   | 5    |
| Routine general medical examination at a health care facility                    | 9529  | 50   | 225   | 5    |
| Diarrhea                                                                        | 1714  | 50   | 225   | 5    |
| Fitting and adjustment of vascular catheter                                     | 318   | 45   | 225   | 5    |
| Hypopotassemia                                                                  | 721   | 45   | 225   | 5    |
| Bariatric surgery status                                                        | 302   | 40   | 200   | 5    |
| Sprain of neck                                                                  | 816   | 50   | 225   | 5    |
| Unspecified gastritis and gastroduodenitis: without mention of hemorrhage       | 960   | 45   | 225   | 5    |
| Injury of face and neck                                                         | 271   | 46   | 300   | 5    |
Backache: unspecified 2471 50 225 5
Unspecified septicemia 222 45 225 5
Acute pharyngitis 4219 45 225 5
Acute bronchitis 3311 46 225 5
Abdominal pain: other specified site 2890 45 225 5
Atrophic gastritis: without mention of hemorrhage 537 45 225 5
Cough 3946 45 225 5
Altered mental status 202 45 225 5
Cervicalgia 2758 50 250 5
Abdominal pain: unspecified site 6339 45 225 5
Other chronic pain 346 56 300 6
Headache 3514 45 225 5
Tobacco use disorder 1834 50 225 5
Other screening mammogram 5722 50 240 5
Observation and evaluation for other specified suspected conditions 337 45 225 5
Unspecified sinusitis (chronic) 1624 46 225 5
Rheumatoid arthritis 353 50 300 5
Brachial neuritis or radiculitis NOS 1147 50 300 5
Loss of weight 455 46 225 5
Hypersonomnia with sleep apnea: unspecified 424 42 225 5
Insomnia: unspecified 968 50 225 5
Other malaise and fatigue 5178 46 225 5
Other injury of chest wall 210 50 300 5
Dehydration 841 45 225 5
Acute respiratory failure 120 40 225 5

B.4 Observational Study: Policy Evaluation of Antibiotic Prescription Guidelines

Antibiotic resistance is a growing problem in the treatment of urinary tract infections (UTI) [28], a common infection for which more than 1.6 million prescriptions are given annually in the United States [33]. With this in mind, we are interested in the following clinical problem: When a patient presents with a UTI, the physician needs to choose between a range of antibiotics, with the dual goals of (a) treating the infection, and (b) minimizing the use of broad-spectrum antibiotics, which are more likely to select for drug-resistant strains of bacteria.

In this context, we might be interested in evaluating a range of potential treatment policies. For our purposes, we will use a pre-defined policy: The clinical guidelines published by the Infectious Disease Society of America (IDSA) for treatment of uncomplicated UTIs in female patients [12]. Using the policy evaluation formulation of $\mathcal{B}^\pi(\pi)$, we will apply OverRule to a conservative interpretation of the IDSA guidelines, using data curated from the Electronic Medical Record (EMR) of two academic medical centers.

The official guidelines discuss the importance of patient and population level risk factors in predicting resistance, and include some factors that we do not observe in our data (such as drug allergies). In order to characterize the guideline explicitly as a policy that we can evaluate in our dataset, we used the following interpretation:

- Choose the first-line agent, either Nitrofurantoin (NIT) or Trimethoprim/Sulfamethoxazole (SXT), to which the patient did not have previous antibiotic exposure or resistance in the prior 90 days. Additionally, if local rates of resistance to SXT are $\geq 20\%$ in the prior 30-90 days,
Figure 5: Results from the Jobs datasets for OverRule approximations of different base estimators, sweeping $\lambda_0, \lambda_1$. AUC is measured with respect to the experimental indicator.

then avoid prescription of SXT.

- If neither of the first-line agents are indicated, then prescribe Ciprofloxacin (CIP), a second-line agent.

**Experimental details** From our data set, we selected all patients from 2007–2017 which had a UTI, and were prescribed one of the four most common antibiotics: NIT, SXT, CIP, or Levofloxacin (LVX). Features include demographics (race, gender, age, and veteran status), comorbidities observed in the past 90 days, information about previous infections (organism, antibiotics given, and resistance profile), hospital ward (inpatient, outpatient, ER, and ICU), and indicators for pregnancy and nursing home residence in the past 90 days. The local rates of resistance (for each hospital ward) are given over the past 30–90 days, and used at the patient level as a feature, as well as an input to the decision of the guidelines.

We preprocess our data first, removing any binary feature with a prevalence of less than 0.1%, and any associated subject: This results in the removal of 48 binary features with less than 0.1% prevalence and 888 corresponding subjects. This leaves a total of 156 (150 binary, 6 continuous) features and 64593 subjects. Detail on all remaining features are given in Table 5. For the purposes of running our algorithm, we convert all continuous variables into binary variables by using indicator functions for deciles.

We then characterize the support set $S^*$ as described in the main paper, using $\alpha = 0.95$, $\lambda_0 = 0.01$, $\lambda_1 = 0$. Using the data points which fall into the support set, we then estimate the propensity $p(t|x)$ of prescribing each of the four drugs using a random forest classifier, with hyperparameter selection done using 5-fold cross-validation on 80% of the remaining cohort used as a training set, over the following parameter grid: Number of estimators $\in [100, 500]$, Minimum samples per leaf (as fraction of total) $\in [0.005, 0.01, 0.02]$. The resulting calibration curves for each antibiotic are given in Figure 6 using the remaining held-out 20% of the data. Using these propensity scores, we apply the procedure described in Section A to estimate the region of strict coverage, $\hat{B}^*(\pi)$ using Boolean rules, and the resulting rules are given in Figure 7. For this stage, we used hyperparameters of $\beta = 0.9, \lambda_0 = 0.03, \lambda_1 = 0$.

**Clinical Validity / Interpretation** Towards understanding the clinical validity of these rules, we interviewed a clinician who specialises in infectious diseases. First, we asked them, based on the
available features, which they would expect to differentiate between subjects for whom the policy is or is not followed. They noted that the guidelines are designed for uncomplicated cases: In particular, patients who have a Foley catheter (a catheter used to drain urine from the bladder) are not covered under these guidelines, because infections in these patients tend to be more complex (e.g., the infection could have been introduced by the catheter itself). The use of the Foley catheter is common during intensive care (e.g., in the ICU), so complex hospitalized patients are less likely to...
Support rules $\mathcal{S}$

Rule S.1 (99.0%):

Previous Resistance:
- $\neg$ Amikacin
- $\neg$ Ertapenem
- $\neg$ Linezolid
- $\neg$ Meropenem
- $\neg$ Nalidixic Acid

and

Previous Prescription:
- $\neg$ Amikacin
- $\neg$ Daptomycin
- $\neg$ Tetracycline Metronidazole
- $\neg$ Trimethoprim

and

Previous Infections:
- $\neg$ Morganella

Propensity overlap rules $\mathcal{B}$

Rule B.1 (27.3%):

Age $< 41$ years
and
Female
and
Location of care
- $\neg$ Intensive Care Unit (ICU)
and
Secondary infection sites
- $\neg$ Bloodstream
and
Medical History:
- $\neg$ Congestive Heart Failure
- $\neg$ Fluid/Electrolyte Disorders
- $\neg$ Metastatic Cancer
- $\neg$ Pulmonary Circ. Disorders
and
Previous Prescription:
- $\neg$ Imipenem
- $\neg$ Posaconazole
and
Previous Resistance:
- $\neg$ Streptomycin (synergistic)
and
Previous Medical Care:
- $\neg$ Mechanical Ventilation
and
$\neg$ Nursing Home

or Rule B.2 (58.4%):

Female
and
Location of care:
- Outpatient
and
$\neg$ Inpatient

or Rule B.3 (3.6%):

Previous Resistance:
- Nitrofurantoin

$\hat{\mathcal{S}} = S.1 \land (B.1 \lor B.2 \lor B.3)$

Figure 7: OverRule description of the coverage region for policy evaluation of the clinical guidelines. Beside each rule we give the percentage of subjects that are covered by the rule in the test set. Overall, the rules for $\hat{\mathcal{B}}$ cover 65.4% of the data points in the support region (compared to the 71% of points labelled by our base estimator), and they have an balanced accuracy of 0.96 versus the base estimator.

With that in mind, they reviewed the available features and noted the following: (i) While UTIs are common for women, they are rare for men; Men with UTIs tend to be more complicated cases, because it is indicative of deeper abnormalities. Similarly, pregnant women are excluded from the guidelines. (ii) Of the comorbidities given, none of them should directly disqualify patients from the guidelines, except potentially for complicated diabetes. (iii) Prior organisms / resistance / prescriptions should not directly disqualify patients from the guidelines, though they will influence the type of antibiotic given. In particular, if a patient has had previous resistance to an antibiotic, they are unlikely to be prescribed it again. (iv) The previous procedures given (with the exception of surgery) are associated with ICU patients. For instance, mechanical ventilation and parenteral nutrition are exclusive to the ICU, and those patients likely have a Foley catheter as well. Surgery is too broad of a category to draw any conclusions. (v) In terms of locations besides the ICU, patients who are admitted to the hospital and who are on intravenous (IV) antibiotics already will be treated differently. The guidelines are focused on oral antibiotics, whereas if an IV already exists, additional IV antibiotics are likely to be given instead.

Having discussed these points first, we then showed them the rules learned by the OverRule algorithm, and asked for their interpretation, as well as for any critiques of the rules based on their clinical knowledge. Their reaction to each of the rules was as follows:
• **Rule B.1**: This appears to correspond to a relatively straightforward young inpatient female (given that Rule B.2 covers all outpatient females). In particular, it rules out ICU patients directly, as well as those with recent mechanical ventilation, which would indicate a recent ICU stay. It also rules out patients with current bloodstream infections, and those who had previously been tested for (and found to be) resistance to Streptomycin (synergistic): This is only tested for in the context of bloodstream infections by enterococcus, and would be an indicator of previous bloodstream infections. Imipenem is an IV antibiotic only given in inpatient settings, and posaconazole is an antifungal used in bone marrow transplant patients. Patients who are both young and in a nursing home tend to be more complex, e.g., they may be paralysed or otherwise unable to perform activities independently. Finally, the excluded comorbidities are less intuitive, because some of them (e.g., congestive heart failure) manifest with a range of severity: For patients with controlled congestive heart failure, this is not a contraindication for following the guidelines, but if they are fully decompensated, then they would likely be on a Foley catheter.

• **Rule B.2**: This concisely describes the most common manifestation of UTI and the set of patients who are most likely to be treated according to the guidelines.

• **Rule B.3**: The conjecture is that this represents patients who have had an uncomplicated UTI in the past, since patients are usually tested for the antibiotics under consideration by a physician, and since nitrofurantoin is one of the first-line treatments for uncomplicated UTIs.

From a quantitative perspective, we compared the learned region with an explicitly constructed cohort of patients whose inclusion criteria were explicitly designed to make them eligible for application of the IDSA guidelines. In particular, we defined this cohort as including non-pregnant women between the ages of 18 to 55 years of age with no record of genitourinary surgery or instrumentation, immunosuppression, indwelling catheters, or neurologic dysfunction in the preceding 90 days.

In relationship to this conservative subset, the learned region (covering 42k patients, 64% of total) covers 96% of the explicitly constructed cohort, while also demonstrating that a broader set of patients are treated according to these guidelines in practice.

---

2Note that outpatient and “not inpatient” can appear in the same rule without being redundant, because multiple specimens collected on the same day for the same patient are collapsed into a single subject.
Table 5: Population averages for the 156 features in the UTI cohort. Mean values and total (for binary features) are given, and there are 64593 subjects in total.

| Demographics          | Mean | Total  |
|-----------------------|------|--------|
| Age                   | 55.1 |        |
| Male                  | 16.53% | 10685 |
| White                 | 72.17% | 46662 |
| Veteran               | 4.61% | 2981   |

| Current Location      | Mean  | Total  |
|-----------------------|-------|--------|
| Outpatient            | 64.89% | 41957 |
| Emergency Room        | 15.69% | 10142 |
| Inpatient             | 17.26% | 11159 |
| Intensive Care Unit (ICU) | 2.69% | 1736   |

| Local Resistance Rates (Past 30-90 days, at this location) | Mean |
|------------------------------------------------------------|------|
| Trimethoprim/Sulfamethoxazole                               | 18.61% |
| Nitrofurantoin                                             | 19.85% |
| Ciprofloxacin                                              | 22.70% |
| Levofloxacin                                               | 24.19% |

| Secondary Site of Infection                                  | Mean  | Total  |
|-------------------------------------------------------------|-------|--------|
| Skin / Soft Tissue                                          | 0.20% | 132    |
| Blood                                                       | 1.59% | 1031   |
| Respiratory Tract                                           | 0.53% | 341    |
| Nasal or Rectal Swab                                        | 0.19% | 124    |

| Medical History (Past 90 Days)                               | Mean  | Total  |
|-------------------------------------------------------------|-------|--------|
| Alcohol abuse                                               | 1.66% | 1074   |
| Deficiency anemia                                           | 2.84% | 1837   |
| Cardiac arrhythmias                                         | 17.08%| 11041  |
| Blood loss anemia                                           | 0.49% | 315    |
| Congestive heart failure                                    | 10.16%| 6571   |
| Coagulopathy                                                | 3.81% | 2466   |
| Diabetes, uncomplicated                                     | 14.13%| 9135   |
| Diabetes, complicated                                       | 5.00% | 3232   |
| Depression                                                  | 11.80%| 7627   |
| Drug abuse                                                  | 1.72% | 1114   |
| Fluid and electrolyte disorders                             | 13.84%| 8946   |
| AIDS/HIV                                                    | 0.43% | 281    |
| Hypertension, uncomplicated                                 | 32.51%| 21017  |
| Hypertension, complicated                                  | 5.43% | 3513   |
| Hypothyroidism                                             | 7.86% | 5085   |
| Liver disease                                               | 4.36% | 2822   |
| Lymphoma                                                    | 1.63% | 1051   |
| Metastatic cancer                                           | 5.50% | 3559   |
| Other neurological disorders                                | 6.68% | 4319   |
| Obesity                                                    | 6.70% | 4332   |
| Pulmonary circulation disorders                             | 3.13% | 2025   |
| Peptic ulcer disease, excluding bleeding                    | 0.61% | 393    |
| Peripheral vascular disorders                               | 5.68% | 3672   |
| Paralysis                                                   | 3.08% | 1992   |
| Condition                                                                 | Percentage | Count   |
|---------------------------------------------------------------------------|------------|---------|
| Psychoses                                                                 | 2.42%      | 1563    |
| Chronic pulmonary disease                                                 | 11.29%     | 7299    |
| Renal                                                                     | 8.87%      | 5735    |
| Rheumatoid arthritis / collagen vascular diseases                         | 3.76%      | 2428    |
| Solid tumor without metastasis                                            | 12.00%     | 7760    |
| Valvular disease                                                          | 7.79%      | 5034    |
| Weight loss                                                               | 3.59%      | 2319    |
| Pregnant                                                                  | 3.08%      | 1989    |

**Previous Care (Past 90 days)**

| Care Type                        | Percentage | Count   |
|----------------------------------|------------|---------|
| Inpatient Stay                   | 18.38%     | 11882   |
| Nursing Home Stay                | 1.20%      | 779     |

**Previous Procedures (Past 90 days)**

| Procedure                        | Percentage | Count   |
|----------------------------------|------------|---------|
| Central Venous Catheder          | 5.27%      | 3410    |
| Hemodialysis                     | 0.66%      | 427     |
| Mechanical Ventilation           | 5.74%      | 3714    |
| Parenteral Nutrition             | 0.67%      | 434     |
| Surgery                          | 59.84%     | 38689   |

**Previous Organisms (Past 90 days)**

| Organism                        | Percentage | Count   |
|---------------------------------|------------|---------|
| Citrobacter species             | 0.42%      | 270     |
| Coagulate negative Staphylococcus species | 1.15% | 741     |
| Enterobacter aerogenes          | 0.15%      | 95      |
| Escherichia coli                | 7.82%      | 5057    |
| Enterococcus species            | 2.66%      | 1718    |
| Enterobacter cloacae            | 0.29%      | 186     |
| Group B Streptococcus           | 0.17%      | 109     |
| Klebsiella pneumoniae           | 2.02%      | 1307    |
| Morganella species              | 0.11%      | 73      |
| Pseudomonas aeruginosa          | 0.92%      | 594     |
| Proteus species                 | 0.69%      | 445     |
| Staph aureus                    | 1.55%      | 1003    |
| Serratia species                | 0.22%      | 145     |

**Previous Resistance, measured by culture (Last 90 Days)**

| Antibiotic                      | Percentage | Count   |
|---------------------------------|------------|---------|
| Amoxicillin Clavulanate         | 2.34%      | 1511    |
| Amikacin                        | 0.10%      | 67      |
| Ampicillin                      | 7.44%      | 4808    |
| Aztreonam                       | 0.95%      | 616     |
| Ceftazidime                     | 0.30%      | 197     |
| Cefazolin                       | 9.22%      | 5962    |
| Chloramphenicol                 | 0.17%      | 111     |
| Ciprofloxacin                   | 4.62%      | 2984    |
| Clindamycin                     | 0.97%      | 624     |
| Ceftriaxone                     | 1.24%      | 804     |
| Doxycycline                     | 0.39%      | 249     |
| Ertapenem                       | 0.14%      | 88      |
| Erythromycin                    | 3.71%      | 2399    |
| Cefepime                        | 0.54%      | 351     |
| Cefoxitin                       | 0.49%      | 319     |
| Gentamicin                      | 1.65%      | 1066    |
| Gentamicin (Synergistic)        | 0.47%      | 307     |
| Antibiotic                | Frequency | Quantity |
|--------------------------|-----------|----------|
| Imipenem                 | 0.47%     | 303      |
| Levofloxacin             | 5.32%     | 3439     |
| Linezolid                | 0.09%     | 58       |
| Meropenem                | 0.13%     | 85       |
| Moxifloxacin             | 0.86%     | 556      |
| Nalidixic Acid           | 0.09%     | 60       |
| Nitrofurantoin           | 4.06%     | 2628     |
| Oxacillin                | 1.79%     | 1158     |
| Penicillin               | 2.41%     | 1559     |
| Piperacillin             | 0.62%     | 402      |
| Polymyxin B              | 1.22%     | 790      |
| Rifampin                 | 0.80%     | 518      |
| Ampicillin Sulbactam     | 1.63%     | 1056     |
| Streptomycin (Synergistic)| 0.23%    | 150      |
| Trimethoprim Sulfamethoxazole | 3.10% | 2006     |
| Tetracycline             | 5.33%     | 3443     |
| Ticarcillin              | 0.24%     | 153      |
| Tobramycin               | 0.31%     | 203      |
| Piperacillin Tazobactam  | 0.53%     | 341      |
| Vancomycin               | 0.92%     | 598      |

**Previous Antibiotic Prescription (Last 90 Days)**

| Antibiotic                | Frequency | Quantity |
|--------------------------|-----------|----------|
| Amikacin                 | 0.09%     | 60       |
| Amoxicillin              | 2.47%     | 1596     |
| Amoxicillin/Clavulanate  | 2.15%     | 1388     |
| Amphotericin B           | 0.16%     | 102      |
| Ampicillin/Sulbactam     | 0.34%     | 217      |
| Azithromycin             | 2.86%     | 1847     |
| Aztreonam                | 0.25%     | 159      |
| Cefadroxil               | 0.15%     | 96       |
| Cefazolin                | 4.87%     | 3150     |
| Cefepime                 | 2.30%     | 1489     |
| Cefixime                 | 0.26%     | 166      |
| Cefotetan                | 0.18%     | 114      |
| Cefoxitin                | 0.25%     | 161      |
| Cefpodoxime              | 0.88%     | 570      |
| Ceftazidime              | 0.73%     | 475      |
| Ceftriaxone              | 2.75%     | 1775     |
| Cefuroxime               | 0.24%     | 156      |
| Cephalexin               | 2.31%     | 1496     |
| Ciprofloxacin            | 11.09%    | 7170     |
| Clarithromycin           | 0.35%     | 226      |
| Clindamycin              | 1.84%     | 1187     |
| Daptomycin               | 0.10%     | 63       |
| Dicloxacillin            | 0.19%     | 126      |
| Doxycycline              | 1.73%     | 1119     |
| Ertapenem                | 0.22%     | 140      |
| Erythromycin             | 0.39%     | 249      |
| Fluconazole              | 3.56%     | 2301     |
| Fosfomycin               | 0.36%     | 232      |
| Gentamicin               | 0.94%     | 607      |
| Drug                        | Frequency | Count |
|-----------------------------|-----------|-------|
| Imipenem                   | 0.33%     | 216   |
| Levofloxacin               | 5.94%     | 3838  |
| Linezolid                  | 0.73%     | 470   |
| Meropenem                  | 0.40%     | 256   |
| Metronidazole              | 4.49%     | 2906  |
| Micafungin                 | 0.24%     | 154   |
| Minocycline                | 0.20%     | 129   |
| Moxifloxacin               | 0.27%     | 174   |
| Nafcillin                  | 0.24%     | 157   |
| Nitrofurantoin             | 2.73%     | 1767  |
| Norfloxacin                | 4.25%     | 2749  |
| Penicillin                 | 0.31%     | 199   |
| Piperacillin               | 0.41%     | 268   |
| Piperacillin/Tazobactam    | 0.23%     | 148   |
| Polymyxin B                | 0.52%     | 333   |
| Posaconazole               | 0.18%     | 118   |
| Tetracycline Metronidazole | 0.09%     | 59    |
| Trimethoprim               | 0.12%     | 79    |
| Trimethoprim/Sulfamethoxazole | 3.96% | 2558  |
| Vancomycin                 | 8.80%     | 5690  |
| Vancomycin Gentamicin      | 3.35%     | 2165  |