Fairness in Contextual Resource Allocation Systems: Metrics and Incompatibility Results

Nathanael Jo*, Bill Tang*, Kathryn Dullerud, Sina Aghaei, Eric Rice, Phebe Vayanos
USC Center for AI in Society, Los Angeles, CA
nathanael.jo@gmail.com, {yongpeng, kdulleru, saghaei, ericr, phebe.vayanos}@usc.edu

Abstract

We study critical systems that allocate scarce resources to satisfy basic needs, such as homeless services that provide housing. These systems often support communities disproportionately affected by systemic racial, gender, or other injustices, so it is crucial to design these systems with fairness considerations in mind. To address this problem, we propose a framework for evaluating fairness in contextual resource allocation systems that is inspired by fairness metrics in machine learning. This framework can be applied to evaluate the fairness properties of a historical policy, as well as to impose constraints in the design of new (counterfactual) allocation policies. Our work culminates with a set of incompatibility results that investigate the interplay between the different fairness metrics we propose. Notably, we demonstrate that: *1)* fairness in allocation and fairness in outcomes are usually incompatible; *2)* policies that prioritize based on a vulnerability score will usually result in unequal outcomes across groups, even if the score is perfectly calibrated; *3)* policies using contextual information beyond what is needed to characterize baseline risk and treatment effects can be fairer in their outcomes than those using just baseline risk and treatment effects; and *4)* policies using group status in addition to baseline risk and treatment effects are as fair as possible given all available information. Our framework can help guide the discussion among stakeholders in deciding which fairness metrics to impose when allocating scarce resources.

1 Introduction

Many of our social and health service systems that operate in high-stakes settings are severely under-resourced. These systems allocate scarce resources that satisfy basic human needs, such as hospitals that provide treatments, social services that offer preventive care, or homeless services that provide shelter and housing. Inadequate allocation in these systems can have grave consequences; for instance, a social service program may exclude those most at risk of suicide from an appropriate preventive intervention.

Since these systems often support communities disproportionately affected by systemic racial, gender, or other injustices, they must be sensitive to the worsening of preexisting inequalities. As such, the fairness of a high-stakes resource allocation system must be easily scrutinized and understood by key stakeholders, which include policymakers and community members the system is designed to help. Indeed, even the perception of bias or lack of transparency can result in a loss of trust and participation. Systems that do not align with the values of their stakeholders may even get defunded, further exacerbating resource scarcity.

Our goals in this paper are a) to establish a simple framework for stakeholders to evaluate the fairness of such high-stakes resource allocation systems, and b) to guide policymakers in choosing which fairness requirements to impose in the design of new (counterfactual) allocation policies. In these systems, a central decision maker makes recommendations for how resources should be allocated. Such policies rely on *contextual* information, i.e., they are based on an individual's intrinsic characteristics (or covariates) which are validated by case managers, as opposed to an individual's preferences (as in the mechanism design literature). Therefore, our learning problem closely resembles that of the classical machine learning framework. Instead of predicting outcomes from features, however, we additionally consider the causal effect of resources (or treatments) on covariates.

We first present a menu of options for evaluating fairness at different stages of the allocation process that are inspired by popular fairness notions in the machine learning literature. We then present incompatibility results between these notions that illuminate the concrete decisions that policymakers need to make in deciding how to allocate resources. For instance, is it acceptable to use protected characteristics like race in deciding which resource an individual should receive? Is it sufficient to only use a vulnerability score to prioritize individuals for resources, or should one necessarily include more context (e.g., information on treatment effects)?

Our results in this paper apply to many high-stakes resource allocation systems. To streamline exposition, we focus on one concrete application, that of allocating scarce housing resources to people experiencing homelessness.

Housing Allocation for People Experiencing Homelessness in Los Angeles. Many cities face dramatic shortages of housing resources used to support those experiencing homelessness. In Los Angeles County, for example, there are over 69,000 homeless persons and only around 30,000 permanent housing units as of the latest counts in 2022 (Authority 2022a,b). Communities have attempted to address this problem by creating Coordinated Entry Systems (CES) where agencies pool their housing resources in a centralized system called a Continuum of Care (CoC). Agencies

*These authors contributed equally.

then distribute housing and services funding for people experiencing homelessness from this pooled resource. In Los Angeles, the CoC is directed by the Los Angeles Homeless Services Authority (LAHSA), who conducts the following process for allocating housing resources:

1. **Assessment** – Persons seeking housing are first assessed for eligibility and vulnerability by filling out a self-reported survey known as the Vulnerability Index - Service Prioritization Decision Assistance Tool (VI-SPDAT), which gives a score from 0 to 17. A higher score indicates higher vulnerability.
2. **Enrollment** – Case managers then use this score to "enroll" the individual into a housing resource, which may be one of the following interventions: Permanent Supportive Housing (PSH) or Rapid Re-Housing (RRH), which are intended for individuals with severe and moderate housing needs, respectively.
3. **Allocation** – Finally, individuals may receive a resource. Note the important distinction between what the individual has been *enrolled* in and what they have been *given*. Policymakers can only recommend connections to particular resources; the actual allocation depends on a variety of factors, including the availability of the resource, whether or not a landlord is willing to accept certain individual to their housing community, and many more.
4. **Outcome** – The impact of that resource is evaluated. While this step is not explicitly part of the matching process, we include it in order to be able to quantify the performance of a housing allocation policy. For example, the outcome being tracked may correspond to whether or not a homeless person returns to homelessness within 24 months of receiving a resource.

In recent years, the fairness of CES has been put to question. In 2018, the Los Angeles Homeless Services Authority's Ad Hoc Committee on Black People Experiencing Homeless published their report mentioning that while CES seems to match Black individuals to housing resources at equal rates to their White counterparts, they still "experience a higher rate of returns to homelessness than all other race and ethnic groups" (Authority 2018). Our work aims to contextualize these claims by providing various metrics to evaluate the fairness of a resource allocation system that are more nuanced than the ones practitioners traditionally use.

1.1 Contributions

Our key contributions are:

- We propose a unifying framework for measuring fairness in high-stakes resource allocation systems which has several advantages: it is intuitive, in that any decision maker with little to no knowledge of algorithmic fairness can understand the fairness concepts; it is adaptable, in that our framework allows for the evaluation of arbitrary counterfactual policies; it is also comprehensive in considering the various sources of discrimination that could be present in a complex allocation system. Throughout this paper, we present our fairness framework in the general sense but will specifically focus on the motivating example of allocating resources for people experiencing homelessness in Los Angeles.
- We present incompatibility results showing that *1)* fairness in allocation and fairness in outcomes are usually incompatible; *2)* policies that prioritize based on a vulnerability score will usually result in unequal outcomes across groups, even if the score is perfectly calibrated; *3)* policies using contextual information beyond what is needed to characterize baseline risk and treatment effects can be fairer in their outcomes than those using just baseline risk and treatment effects; and *4)* policies using group status in addition to baseline risk and treatment effects are as fair as possible given all available information.

The remainder of this paper is organized as follows. The rest of Section 1 positions our paper within the related literature. In Section 2, we formalize our proposed notions of group fairness. Section 3 presents our incompatibility results when using a combination of these fairness notions together. Finally, we discuss the practical implications of our incompatibility results in Section 4. [1]

1.2 Literature Review

Our work is closely related to the fields of fairness in machine learning and resource allocation. We now position our work in these two literature streams.

Fairness in Machine Learning. The literature on algorithmic fairness is extensive. Of particular relevance to our work is enforcing "group fairness", i.e., *statistical fairness* over segments of the population. This is in contrast to "individual" fairness, which requires that individuals with similar covariates be classified in the same way (Dwork et al. 2012). We focus on group fairness notions because they are intuitive and are more aligned with how practitioners evaluate fairness. There are a number of fairness metrics that have been proposed in the literature. For instance, *statistical parity* enforces that the probability of receiving a positive class is equal across all sensitive groups (Dwork et al. 2012). As an extension to *statistical parity*, *conditional statistical parity* was introduced by Corbett-Davies et al. (2017), which stipulates that a classifier should assign a positive class at equal rates across all protected groups, conditional on some legitimate feature(s) that affect the outcome.

Other fairness notions incorporate the quality of a classifier's decisions, such as *equalized odds* where all protected groups should have the same true positive rates (TPR) and false positive rates (FPR) (Hardt, Price, and Srebro 2016). A simpler version of *equalized odds* is *equality of opportunity*, where only equal TPR is enforced. Approaches that include these two notions include Hardt, Price, and Srebro (2016); Zafar et al. (2017); Agarwal et al. (2018).

There are many other notions of fairness such as *error rate balance* (Feldman et al. 2015), *predictive parity* (Chouldechova 2017), *well-calibration* (Crowson, Atkinson, and Therneau 2016; Flores, Bechtel, and Lowenkamp 2016), and *balance of positive (negative) class* (Kleinberg,

[1] Full version at https://arxiv.org/abs/2212.01725

Mullainathan, and Raghavan 2016). We refer the interested reader to Barocas, Hardt, and Narayanan (2019); Mehrabi et al. (2021) for in-depth reviews of the literature.

An important line of work in fair machine learning investigates incompatibility/impossibility results among different notions of fairness. For instance, Kleinberg, Mullainathan, and Raghavan (2016) show the incompatibility of *well-calibration* and *balance of positive/negative class* and Chouldechova (2017) shows that *well-calibration* and *error rate balance* are incompatible except in trivial cases.

Another relevant branch of research in fair machine learning is the study of how "race-blind" practices might actually hinder fairness. Huq (2018) provides a perspective on how including group membership could work in a legal setting and Kleinberg et al. (2018) show that race-aware predictors outperform race-blind predictors in terms of fairness. Our work adds another dimension to this literature by expanding this stream of work into the resource allocation setting.

Fairness in Resource Allocation. Within the literature on fairness in resource allocation, works that are most closely related to ours are those that have attempted to bridge the gap between fair machine learning and fair resource allocation/mechanism design (Finocchiaro et al. 2021; Das 2021). However, these papers largely provide guiding frameworks for how problems connecting machine learning and resource allocation should be approached rather than defining fairness notions that bridge the two.

There are several papers that approach the problem of fair resource allocation through a central decision maker. Rodolfa et al. (2020) use the distribution of recall across protected groups to improve fairness, while also adjusting for how their measure of fairness might affect long-term outcomes. Rodolfa, Lamba, and Ghani (2021) build on the previous work by showing they do not sacrifice accuracy in applying their notion of fairness. Elzayn et al. (2019) present a notion of allocative fairness inspired by equality of opportunity in the classification setting, and Donahue and Kleinberg (2020) extend this work by providing an upper bound on the price of fairness. However, these papers focus on just one notion of fairness and thus none examine possible incompatibilities between different fairness notions.

Another line of work in this literature focuses on designing policies that satisfy fairness notions through linear constraints. Some applications include kidney allocation (Bertsimas, Farias, and Trichakis 2013) and allocation of housing resources (Azizi et al. 2018; Rahmattalabi et al. 2022). Some of the fairness notions imposed in these works overlap with the ones we propose in our framework. However, none of these papers study the interplay between different notions of fairness. Vayanos et al. (2021) propose a preference elicitation algorithm and show how it can be used to elicit the preferences of policymakers over various notions of fairness and efficiency. Yet, they do not present any incompatibility results between these fairness notions. There are also other works that compare fairness notions in non-contextual resource allocation problems such as facility location (Kumar and Kleinberg 2000), scheduling (Kumar and Kleinberg 2000; Baruah et al. 1993), and bandwidth assignments in networks (Kumar and Kleinberg 2000; Lan et al. 2010).

There are a number of papers that focus on fairness in mechanism design. However, they focus on the utility derived from allocating a resource to an individual. For example, Freeman, Shah, and Vaish (2020) develop algorithms for constructing allocation policies that are simultaneously exactly fair ex-ante and approximately fair ex-post and compare fairness notions such as envy-freeness, group fairness, and proportionality. Mashiat et al. (2022) outline several notions of fairness related to equalizing average regret and improvement in allocation across groups. Kash, Procaccia, and Shah (2014) and Benabbou, Chakraborty, and Zick (2019) focus on the incompatibility of fairness and efficiency in terms of envy-freeness and pareto-optimality. Other works model the allocation system using self-interested agents trying to obtain a good(s) through a market (Conitzer et al. 2019; Scarlett, Teh, and Zick 2021; Benabbou, Chakraborty, and Zick 2019; Benabbou et al. 2019; Barman et al. 2018; Fain, Munagala, and Shah 2018; Raman, Shah, and Dickerson 2021). Our work focuses specifically on the interplay of fairness notions at different phases of allocation and on contextual resource allocation systems.

2 Statistical Fairness

We consider a system that allocates scarce resources of different (finite) types, which we index in the set $\mathcal{T}$ (e.g., PSH or RRH). These resources must be allocated to individuals who are characterized by their feature vector $X \in \mathcal{X} \subseteq \mathbb{R}^n$, which includes their sensitive attribute(s) $G \in \mathcal{G}$, where $\mathcal{G}$ comprises the levels of (potentially intersection of multiple) sensitive attribute(s) (e.g., race or sex). Each individual is also characterized by their potential outcomes $Y^t \in \mathcal{Y} \subseteq \mathbb{R}$ under each resource $t \in \mathcal{T}$. These potential outcomes could be interpreted as one's utility given treatment (e.g., long-term health or return to homelessness), though they are not a function of an individual's preferences. The system uses a (possibly randomized) recommendation policy to assign each individual a treatment, $T = \pi(X) \in \mathcal{T}$, i.e., $\pi : \mathcal{X} \to \mathcal{T}$. This recommendation may not be the same as the treatment that the individual is given, denoted by $A \in \mathcal{T}$, which is determined by a (possibly randomized) allocation rule $\mu : \mathcal{X} \to \mathcal{T}$. While we have defined our policies π and μ in the general case for X, which includes G, in many sensitive settings using protected characteristics is prohibited to prevent discrimination. Therefore, we also define $X_{-G} \in \mathcal{X}_{-G}$ as the collection of features excluding group status G. Finally, for given π and μ, we assume an unknown joint distribution $\mathbb{P}$ over $(X, T, A, \{Y^t\}_{t\in\mathcal{T}})$. Note that the random quantity $\pi(x)$ is specified by the conditional probabilities $\mathbb{P}(\pi(x) = t|X = x)$ that each individual is assigned treatment t conditional on covariates x and $\mu(x)$ is specified by the conditional probabilities $\mathbb{P}(\mu(x) = t|X = x)$ that each individual receives treatment t conditional on covariates x. Furthermore, deterministic policies are a special case of the randomized policies we have defined where the conditional probabilities $\mathbb{P}(\pi(x) = t|X = x)$ and $\mathbb{P}(\mu(x) = t|X = x)$ equal 1 for a single treatment and 0 for all other treatments. Finally, we define uniformly randomized poli-

cies as the special case where $\mathbb{P}(\pi(x) = t|X = x) = p$ or $\mathbb{P}(\mu(x) = t|X = x) = p$ for some constant $p \in [0, 1]$ and for all $x \in \mathcal{X}$.

2.1 Enrollment Fairness

We now outline several fairness considerations during the enrollment stage, i.e., the results following a recommendation rule π.

Definition 2.1 (Statistical Parity in Enrollments)**.** *Statistical parity in enrollments is satisfied if the probabilities of enrolling in any given treatment are equal across sensitive groups, i.e.,*

$$\mathbb{P}(\pi(X) = t|G = g) = \mathbb{P}(\pi(X) = t|G = g') \quad \forall g, g' \in \mathcal{G}, t \in \mathcal{T}.$$

As mentioned in the introduction, statistical parity is perhaps one of the most well-known fairness notions in the machine learning literature and used in evaluating many real-world systems, see e.g., LAHSA's Ad Hoc Committee on Black People Experiencing Homelessness report (Authority 2018). While it is simple and intuitive, this notion may not sufficiently protect groups that are more vulnerable at baseline. For instance, recall that people experiencing homelessness in Los Angeles are assigned a risk score from 0 to 17 that characterizes how vulnerable they are. Suppose now that we design a policy satisfying statistical parity on race as the sensitive attribute. Since we do not condition on risk score, it is possible that such an allocation policy disproportionately assigns treatments to those with lower vulnerability within each racial group, which is undesirable. This observation motivates us to introduce the following notion of fairness.

Definition 2.2 (Conditional Statistical Parity in Enrollments)**.** *Conditional statistical parity in enrollments is satisfied if the probabilities of enrolling in any given treatment are equal across sensitive groups, conditional on some "legitimate feature(s)", i.e.,*

$$\mathbb{P}(\pi(X) = t|G = g, L = \ell) = \mathbb{P}(\pi(X) = t|G = g', L = \ell) \quad \forall g, g' \in \mathcal{G}, t \in \mathcal{T}, \ell \in \mathcal{L},$$

where $L = f(X)$, $f : \mathcal{X} \to \mathcal{L}$, collects all legitimate features and takes values in the set $\mathcal{L}$.

In our motivating example, L, which is a function of covariates collected from a survey assessment, is the risk score from 0 to 17. Policies that assign or recommend based on a risk score are used in many applications (see Ustun and Rudin (2019) for examples within medicine, criminal justice, and finance). Beyond conditioning on vulnerability, we also propose considering features measuring treatment effects as legitimate features since policymakers may be interested in targeting treatment effects to achieve outcome related fairness, which is detailed in Section 2.3.

2.2 Allocative Fairness

After being enrolled in a treatment, an individual may not actually receive their recommended treatment. We therefore define separate fairness notions for the allocation stage. While most of the definitions are similar to that of the enrollment stage, we include one more definition that is inspired by equalized odds. This is motivated by the need to ensure that the discrepancy between enrollment and allocation is roughly equal among all groups.

Definition 2.3 (Equalized Faithfulness in Allocation)**.** *Equalized faithfulness in allocation is satisfied if the probabilities of receiving treatments other than the one initially recommended are equal across groups, i.e.,*

$$\mathbb{P}(\mu(X) = t'|G = g, \pi(X) = t) = \mathbb{P}(\mu(X) = t'|G = g', \pi(X) = t) \quad \forall g, g' \in \mathcal{G}, t, t' \in \mathcal{T}.$$

Equalized faithfulness in allocation may be a useful mechanism when there are unaccounted biases towards certain groups between the enrollment and allocation stages. In our motivating problem of homelessness in Los Angeles, for example, some resources require that a housing complex's landlord approves of the person experiencing homelessness. Unfortunately, caseworkers have shared numerous cases of Black individuals being recommended a resource but not receiving it because no landlord is willing to accept them (Milburn et al. 2021).

Simply requiring equalized faithfulness between groups may lead to undesirable results when there are distributional differences in vulnerability between groups. For example, suppose group g is generally more vulnerable than group g'. Imposing equalized faithfulness may result in an allocation where individuals from g receive their recommended resource at much lower rates than their counterparts in g' with the same vulnerability. This is possible so long as faithfulness is *overall* the same across the two groups. A more reasonable policy might instead be more likely to allocate the recommended resource to more vulnerable individuals. We therefore introduce the notion of *conditional* equalized faithfulness, which accounts for vulnerability.

Definition 2.4 (Conditional Equalized Faithfulness in Allocation)**.** *Conditional equalized faithfulness in allocation is satisfied if the probabilities of receiving treatments other than the one initially recommended are equal across groups, conditional on some "legitimate feature(s)", i.e.,*

$$\mathbb{P}(\mu(X) = t'|G = g, \pi(X) = t, L = \ell) = \mathbb{P}(\mu(X) = t'|G = g', \pi(X) = t, L = \ell) \quad \forall g, g' \in \mathcal{G}, t, t' \in \mathcal{T}, \ell \in \mathcal{L}.$$

We may also alter the definitions in Section 2.1 to get similar notions in the allocation setting.

Definition 2.5 (Statistical Parity in Allocation)**.** *Statistical parity in allocation is satisfied if the probabilities of allocating any given treatment are equal across groups, i.e.,*

$$\mathbb{P}(\mu(X) = t|G = g) = \mathbb{P}(\mu(X) = t|G = g') \quad \forall g, g' \in \mathcal{G}, t \in \mathcal{T}.$$

Definition 2.6 (Conditional Statistical Parity in Allocation)**.** *Conditional statistical parity in allocation is satisfied if the probabilities of allocating any given treatment are equal*

across groups, conditional on some legitimate feature(s), i.e.,

$$\mathbb{P}(\mu(X) = t|G = g, L = \ell) = \mathbb{P}(\mu(X) = t|G = g', L = \ell) \quad \forall g, g' \in \mathcal{G}, t \in \mathcal{T}, \ell \in \mathcal{L}.$$

While fairness in enrollment is something that policymakers control (because they assign the interventions), allocative fairness may not be possible to control since the actual allocation might depend on external factors like the availability of resources or discrimination (e.g., landlords might reject certain individuals from their housing communities). However, one may be able to affect allocative fairness by changing enrollment fairness assuming that faithfulness in allocation remains unchanged. Despite these challenges, allocative fairness is nonetheless important to measure because it reflects which resources individuals truly receive. We thus introduce these definitions separately to highlight the distinct challenges faced in the allocation stage versus the enrollment stage.

2.3 Outcome Fairness

Lastly, we may impose similar notions of fairness from the outcome perspective, where outcomes could be some personal utility upon receiving treatment (e.g., a return to homelessness or health outcome). These definitions are useful in situations where imposing fairness in allocation is inappropriate because they exacerbate existing inequalities. For instance, Black individuals experiencing homelessness in Los Angeles may be assigned housing resources at equal rates to their White counterparts, but they still "experience a higher rate of returns to homelessness than all other race and ethnic groups" (Authority 2018).

For simplicity, we introduce these fairness notions assuming that potential outcome Y is binary, where $Y = 1$ (resp. 0) corresponds to a positive (resp. negative) outcome. This assumption will carry over to our incompatibility results in Section 3. Note, however, that the proposed definitions can be amended to reflect (conditional) expectations or higher order moments when $Y \in \mathbb{R}$.

Definition 2.7 (Statistical Parity in Outcomes)**.** *Statistical parity in outcomes is satisfied if the probabilities of experiencing a positive outcome are equal across groups, i.e.,*

$$\mathbb{P}(Y(\mu(X)) = 1|G = g) = \mathbb{P}(Y(\mu(X)) = 1|G = g') \quad \forall g, g' \in \mathcal{G}.$$

Definition 2.8 (Conditional Statistical Parity in Outcomes)**.** *Conditional statistical parity in outcomes is satisfied if the probabilities of experiencing a positive outcome are equal across sensitive groups, conditional on some legitimate feature(s), i.e.,*

$$\mathbb{P}(Y(\mu(X)) = 1|G = g, L = \ell) = \mathbb{P}(Y(\mu(X)) = 1|G = g', L = \ell) \quad \forall g, g' \in \mathcal{G}, \ell \in \mathcal{L}.$$

When the distribution of legitimate features vary widely between protected groups, we have discussed that conditional statistical parity is often preferable over statistical parity in the enrollment and allocation stages. This is not the case, however, when we want to ensure fairness in outcomes. We argue that statistical parity in outcomes is in fact the stronger condition for an allocation policy to satisfy. In the housing allocation problem, Black people are more often classified as being at higher risk to homelessness than White people. Enforcing conditional statistical parity merely protects the subpopulation in both groups who happen to have the same risk scores. However, we are more interested in seeing that both races *overall* have an equal probability of experiencing a positive outcome, which may require allocating better/more resources to Black people because they are more at risk of homelessness in the first place. In general, our argument holds when there is existing and apparent discrimination between two or more groups, and ultimately our goal in enforcing statistical parity is to repair said discrimination.

The aforementioned definitions can be amended in multiple ways depending on the allocation setting. For instance, they can be altered so that one (historically marginalized) group has a higher probability of success than others. This change may be particularly helpful in settings involving affirmative action or reparations.

3 Incompatibility Results

3.1 Incompatibility of Allocation and Outcome Fairness

In this section, we present incompatibility results between fairness in allocation and fairness in outcomes by defining necessary conditions for different metrics to jointly hold and then providing intuitions as to why these conditions would not hold in general. Such results are important since policymakers may often want to see if they can achieve both allocation and outcome fairness between groups. Throughout this section, our results focus on the binary treatment case, i.e., $\mathcal{T} = \{0, 1\}$, where $t = 0$ corresponds to the no-treatment case (control). Each condition for the binary case can be extended to multiple treatments at the cost of complicating notation. While we do not explicitly consider budget constraints in deriving incompatibility results and in studying the fairness behavior of policies under various metrics, budget constraints are captured by the assignment probabilities. For example, $\mathbb{P}(\mu(X) = 1)$, the population treatment probability, is bounded by the budget. For simplicity, we assume non-negative conditional average treatment effects, i.e., $\mathbb{E}[Y^1 - Y^0 \mid X = x] \geq 0 \ \forall x \in \mathcal{X}$, which in our motivating example means that giving someone (anybody) a housing resource is expected to help them. We discuss the incompatibility of allocation and outcomes in terms of the allocation rule μ and not the recommendation policy π since the actual allocation is what affects the outcomes. Finally, our focus is on the case where policymakers can only influence outcomes through the allocation rule μ of limited resources and cannot improve the actual intervention effects. All proofs can be found in the Supplementary Materials.

First, we look at when we can expect statistical parity in allocation and statistical parity in outcomes to jointly hold. For a policy μ to satisfy statistical parity in alloca-

tion (Definition 2.5), it must be that group level assignment probabilities are equal across groups, $\mathbb{P}(\mu(X) = 1|G = g) = \mathbb{P}(\mu(X) = 1|G = g')\ \forall g, g' \in \mathcal{G}$, or equivalently, $\mathbb{P}(\mu(X) = 1) = \mathbb{P}(\mu(X) = 1|G = g)\ \forall g \in \mathcal{G}$. Note that such a μ always exists since we can take μ to be a uniformly randomized policy, which will satisfy statistical parity in allocation. The following result shows that statistical parity in allocation and statistical parity in outcome will hold jointly only in exceptional circumstances.

Proposition 1. *A policy μ satisfies both statistical parity in allocation (Definition 2.5) and statistical parity in outcomes (Definition 2.7) if and only if*

$$\begin{aligned}&\Big[\big\{\mathbb{P}(Y^1 = 1|G = g, \mu(X) = 1) - \\ &\quad \mathbb{P}(Y^0 = 1|G = g, \mu(X) = 0)\big\} - \\ &\quad \big\{\mathbb{P}(Y^1 = 1|G = g', \mu(X) = 1) - \\ &\quad \mathbb{P}(Y^0 = 1|G = g', \mu(X) = 0)\big\}\Big]\mathbb{P}(\mu(X) = 1) = \\ &\mathbb{E}[Y^0|G = g', \mu(X) = 0] - \mathbb{E}[Y^0|G = g, \mu(X) = 0] \\ &\quad \forall g, g' \in \mathcal{G}.\end{aligned} \tag{1}$$

In the special case where μ is a uniformly randomized policy, condition (1) *simplifies to*

$$\begin{aligned}&(\mathbb{E}[Y^1 - Y^0|G = g'] - \mathbb{E}[Y^1 - Y^0|G = g])\mathbb{P}(\mu(X) = 1) \\ &= \mathbb{E}[Y^0|G = g] - \mathbb{E}[Y^0|G = g'] \qquad \forall g, g' \in \mathcal{G}.\end{aligned} \tag{2}$$

We provide some intuition in the case where μ is uniformly randomized since such a policy always exists and satisfies statistical parity in allocation. Condition (2) requires that the probability of treatment assignment of the uniformly randomized policy μ, when scaled by the treatment effect differences between g and g', is equal to the baseline outcome differences under no treatment between g and g'. Since the probability of assignment is between 0 and 1, in order for (2) to hold, it must be that group g' has greater treatment effects than g in order to make up for group g having higher outcomes under no treatment. However, policymakers have no control over baseline outcomes and average treatment effects, and g could have both better outcomes under no treatment and better treatment effects. Therefore, in general there may not exist a policy μ that satisfies condition (1).

Since we may not be able to have statistical parity in both allocation and outcomes, perhaps we can have both conditional statistical parity (CSP) in allocation, where we condition on a set of legitimate features or risk scores, and statistical parity in outcomes. This is of practical interest since in our motivating example, under the U.S. Department of Housing and Urban Development's (HUD) guidance of prioritizing the most vulnerable populations (U.S. Department of Housing and Urban Development 2014), scarce housing allocations are made based on risk scores measuring vulnerability under no treatment like the VI-SPDAT (Greater Los Angeles Coordinated Entry System 2018).

Let $\mathcal{L}_0$ denote the set of risk scores/levels that capture an individual's risk and L_0, which is just a function of the covariates as defined in Section 2, denote a random variable that can take on values $\ell_0 \in \mathcal{L}_0$. We assume L_0 is "well chosen" such that baseline outcomes under no treatment are independent of group membership for each score, i.e., $\mathbb{P}(Y^0 = 1|L_0 = \ell) = \mathbb{P}(Y^0 = 1|G = g, L_0 = \ell)\ \forall g \in \mathcal{G}, \ell_0 \in \mathcal{L}_0$. While this "well chosen" property may not hold for existing deployed risk scores, it is possible to construct risk scores that satisfy this property with well-calibrated probabilistic estimates. As in our motivating example where the risk scores are discrete, we also assume that $\mathcal{L}_0$ is a discrete set. We can then decompose group outcomes under no treatment for each group g as $\mathbb{P}(Y^0 = 1|G = g) = \sum_{\ell_0 \in \mathcal{L}_0} \mathbb{P}(Y^0 = 1|L_0 = \ell)\mathbb{P}(L_0 = \ell|G = g)$, which suggests group outcome differences are a result of group distribution differences across the risk scores in $\mathcal{L}_0$. It is natural then to enforce CSP in allocation conditioned on L_0 while trying to achieve outcome parity because we are accounting for distributional differences in vulnerability between groups. Let $\mu_{\mathcal{L}_0}$ be a policy that satisfies conditional statistical parity in allocation conditional on L_0. In particular, assume that $\mu_{\mathcal{L}_0} : \mathcal{L}_0 \to \mathcal{T}$ since in practice many policies are often simply a function of L_0 as described after Definition 2.2. Also for ease of notation, let $\tau_{\ell,g} := \mathbb{E}[Y^1 - Y^0|G = g, L_0 = \ell]$ be the average treatment effect for individuals in group g with $L_0 = \ell$. We have the following incompatibility result.

Proposition 2. *A policy $\mu_{\mathcal{L}_0}$ satisfies conditional statistical parity in allocation (Definition 2.6) and statistical parity in outcome (Definition 2.7) if and only if the following holds*

$$\begin{aligned}&\textstyle\sum_{\ell_0 \in \mathcal{L}_0} \mathbb{P}(\mu_{\mathcal{L}_0}(\ell) = 1)\,(\tau_{\ell,g}\mathbb{P}(L_0 = \ell|G = g) - \\ &\qquad \tau_{\ell,g'}\mathbb{P}(L_0 = \ell|G = g')) = \\ &\mathbb{P}(Y^0 = 1|G = g') - \mathbb{P}(Y^0 = 1|G = g) \\ &\qquad \forall g, g' \in \mathcal{G}.\end{aligned} \tag{3}$$

We emphasize that Proposition 2 holds even if L_0 is not "well-chosen" since the proof does not depend on this property. What is interesting is that the result holds even if L_0 satisfies the desirable "well-chosen" property. Similar to Proposition 1, there is no guarantee in general that there exists $\mu_{\mathcal{L}_0}$ satisfying (3). While policymakers can target specific risk groups through a $\mu_{\mathcal{L}_0}$ policy, they cannot control the treatment effects within each level, $\tau_{\ell,g}$. Therefore it might make sense to explicitly account for treatment effects as well.

We propose another type of conditional statistical parity policy that uses an additional set of legitimate features L_1 in addition to L_0, $\mu_{\mathcal{L}_0,\mathcal{L}_1} : \mathcal{L}_0 \times \mathcal{L}_1 \to \mathcal{T}$, where L_1 is "well chosen" to capture individuals' treatment effects such that treatment effects are independent of G and X conditioned on L_1, i.e., $\mathbb{E}[Y^1 - Y^0|L_1 = \ell_1] = \mathbb{E}[Y^1 - Y^0|G = g, X = x, L_1 = \ell_1]\ \forall g \in \mathcal{G}, x \in \mathcal{X}, \ell_1 \in \mathcal{L}_1$. Since the treatment effect can be captured by just L_1, for ease of notation we let $\tau_l := \mathbb{E}[Y^1 - Y^0|L_1 = \ell_1]$ denote the treatment effect for individuals with treatment effects defined by $L_1 = \ell_1$. In practice, L_1 could be estimated via a heterogeneous treatment effects model, see e.g., Wager and Athey (2018). While $\mu_{\mathcal{L}_0}$ is commonly used in practice for allocation, we show that $\mu_{\mathcal{L}_0,\mathcal{L}_1}$ can mitigate disparities as well as, if not better than, $\mu_{\mathcal{L}_0}$. Intuitively this is because $\mu_{\mathcal{L}_0}$, as can be seen

in (3), does not account for treatment effects. By incorporating information about treatment effects, we can hope to further mitigate outcome disparities.

Proposition 3. *For any given conditional statistical parity policy $\mu_{\mathcal{L}_0}$, and under any distribution $\mathbb{P}$ over $(X, T, A, \{Y^t\}_{t\in\mathcal{T}})$, there exists $\mu_{\mathcal{L}_0,\mathcal{L}_1}$ such that*

$$\begin{aligned}&\mathbb{P}(Y(\mu_{\mathcal{L}_0,\mathcal{L}_1}(L_0, L_1)) = 1|G = g)-\\&\mathbb{P}(Y(\mu_{\mathcal{L}_0,\mathcal{L}_1}(L_0, L_1)) = 1|G = g')\\&\leq\ \mathbb{P}(Y(\mu_{\mathcal{L}_0}(L_0)) = 1|G = g)-\\&\mathbb{P}(Y(\mu_{\mathcal{L}_0}(L_0)) = 1|G = g')\ \ \forall g, g' \in \mathcal{G}.\end{aligned} \tag{4}$$

Moreover, there exists a joint distribution $\mathbb{P}$ such that the inequality (4) *is strict for at least one pair g, g'.*

Proposition 3 suggests that using more contextual information about group distributions or treatment effects may allow us to define conditional statistical parity policies that are more fair in terms of outcomes. While we have explicitly defined L_0 for baseline outcomes and L_1 for treatment effects, we can simply define $L := \{(\ell_0, \ell_1) \mid \ell_0 \in \mathcal{L}_0, \ell_1 \in \mathcal{L}_1\}$ and relabel each (ℓ_0, ℓ_1) combination as some ℓ instead with the understanding that we are really partitioning our population by L_0 and L_1. In the remaining results, we use the combined L and define $\tau_\ell := \mathbb{E}[Y^1 - Y^0|L = \ell]$. We also use $\mu_{\mathcal{L}}$ to denote a CSP in allocation policy conditioned on our new L. While using both L_0 and L_1 gives us more information and allocation flexibility, we show in the proposition below that incorporating both baseline outcome information and treatment effects may still not be enough to achieve statistical parity in outcomes across all groups.

Proposition 4. *An allocation policy $\mu_{\mathcal{L}}$ satisfies conditional statistical parity in allocation (Definition 2.6) and statistical parity in outcome (Definition 2.7) if and only if the following holds:*

$$\begin{aligned}&\textstyle\sum_{\ell\in\mathcal{L}}\mathbb{P}(\mu_{\mathcal{L}}(\ell) = 1)\tau_\ell\times\\&\quad[\mathbb{P}(L = \ell|G = g) - \mathbb{P}(L = \ell|G = g')] =\\&\quad\mathbb{P}(Y^0 = 1|G = g') - \mathbb{P}(Y^0 = 1|G = g)\\&\qquad\forall g, g' \in \mathcal{G}.\end{aligned} \tag{5}$$

To gain some intuition on the implications of Proposition 4 and when there exists $\mu_{\mathcal{L}}$ such that (5) holds, let us look at a simple example. Suppose $|\mathcal{G}| = 2, |\mathcal{L}| = 2$, that is, there are only two groups and risk levels, $\mathbb{P}(Y^0 = 1|G = g) = 0.6$, and $\mathbb{P}(Y^0 = 1|G = g') = 0.4$. Let us also define $a_\ell := \tau_\ell[\mathbb{P}(L = \ell|G = g) - \mathbb{P}(L = \ell|G = g')], \ell \in \{1, 2\}$. Then in order for $\mu_{\mathcal{L}}$ to satisfy CSP in allocation conditional on L and outcome parity, it follows from Proposition 4 that

$$\begin{aligned}&\mathbb{P}(\mu_{\mathcal{L}}(1) = 1)a_1 + \mathbb{P}(\mu_{\mathcal{L}}(2) = 1)a_2 =\\&\mathbb{P}(Y^0 = 1|G = g') - \mathbb{P}(Y^0 = 1|G = g) = -0.2\end{aligned}$$

if our allocation satisfies both CSP in allocation and outcome parity. Without loss of generality, let us assume $\mathbb{P}(L = 1|G = g) > \mathbb{P}(L = 1|G = g')$ and thus $\mathbb{P}(L = 2|G = g) < \mathbb{P}(L = 2|G = g')$, so that $a_1 > 0$ and $a_2 < 0$. We then have two possibilities: *(i)* $a_2 \leq -0.2 \leq 0$ or *(ii)* $-0.2 < a_2$. In case (i), since a_1 and a_2 have opposite signs, we can easily find values $0 < \mathbb{P}(\mu_{\mathcal{L}}(1) = 1), \mathbb{P}(\mu_{\mathcal{L}}(2) = 1) < 1$ such that CSP in allocation and outcome parity are satisfied. However, in case (ii), there does not exist such CSP allocation. The closest we could come to satisfying our conditions would be to set $\mathbb{P}(\mu_{\mathcal{L}}(1) = 1) = 0$ and $\mathbb{P}(\mu_{\mathcal{L}}(2) = 1) = 1$. From this example, we see that if the baseline difference in outcome between our two groups $|\mathbb{P}(Y^0 = 1|G = g) - \mathbb{P}(Y^0 = 1|G = g')|$ is very large, then it becomes harder to find an allocation that satisfies both CSP in allocation and outcome parity. This makes sense – as the disparity in outcome under no treatment becomes very large, it becomes harder to make up this difference by allocation.

3.2 Policies with Outcome Fairness

In Section 3.1, we established that in general there may not exist allocation policies that satisfy fairness in allocation (as measured by statistical and conditional statistical parity) and parity in outcomes. However, if a policymaker wants parity in outcomes, it is not immediately clear that a policy can even achieve this since outcomes depend on treatment effects and distribution of baseline risk within each group. In this section, we show that using feature information without group status, X_{-G}, in a policy can lead to policies with less outcome disparity than just using risk scores alone, but in general outcome parity may not be possible unless a policy explicitly uses group status. Let $\mu_{\mathcal{L},\mathcal{X}_{-G}}$ be a (possibly) randomized policy that uses X_{-G} and L for assignment. In the propositions that follow, for each inequality of outcome disparity between g, g' under different policies, we use g to denote the group with better outcomes under no treatment compared to those of g'. Therefore we do not have absolute values below since we only consider pairs of g, g' where the outcome difference between g and g' is positive. Proposition 5 shows that using extra information in the form of X_{-G} leads to same, if not lower, disparities than policies relying only on L across each comparison $g, g' \in \mathcal{G}, g \neq g'$.

Proposition 5. *For any given policy $\mu_{\mathcal{L}}$, and under any distribution $\mathbb{P}$ over $(X, T, A, \{Y^t\}_{t\in\mathcal{T}})$, there exists $\mu_{\mathcal{L},\mathcal{X}_{-G}}$ such that*

$$\begin{aligned}&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{X}_{-G}}(L, X_{-G})) = 1|G = g)-\\&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{X}_{-G}}(L, X_{-G})) = 1|G = g') \leq\\&\mathbb{P}(Y(\mu_{\mathcal{L}}(L)) = 1|G = g) - \mathbb{P}(Y(\mu_{\mathcal{L}}(L)) = 1|G = g')\end{aligned} \tag{6}$$

for all $g, g' \in \mathcal{G}, g \neq g'$ where $\mathbb{P}(Y^0 = 1|G = g) > \mathbb{P}(Y^0 = 1|G = g')$ and there exists a joint distribution $\mathbb{P}$ such that the inequality (6) *is strict for at least one pair g, g'.*

The difficulty of reducing all pairwise outcome disparities lies in the fact that under any policy that does not explicitly use group status, outcomes for each group will have interdependencies. For example, we may reduce the outcome disparity between g and g' but also increase the outcome disparity between g and g'' at the same time. Given this, we show below that a policy explicitly using group status will do better than a policy using all available features except for group status. Let $\mu_{\mathcal{L},\mathcal{G}}$ denote a (possibly randomized) policy that

uses L and G for assignment and emphasize that $\mu_{\mathcal{L},\mathcal{G}}$ does not use the features X.

Proposition 6. *For any given policy $\mu_{\mathcal{L},\mathcal{X}_{-G}}$, and under any distribution $\mathbb{P}$ of $(X, T, A, \{Y^t\}_{t\in\mathcal{T}})$, there exists $\mu_{\mathcal{L},\mathcal{G}}$ such that*

$$\begin{aligned}&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{G}}(L,G))=1|G=g)-\\&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{G}}(L,G))=1|G=g')\leq\\&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{X}_{-G}}(L,X_{-G}))=1|G=g)-\\&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{X}_{-G}}(L,X_{-G}))=1|G=g')\end{aligned} \tag{7}$$

for all $g, g' \in \mathcal{G}, g \neq g'$ where $\mathbb{P}(Y^0 = 1|G = g) > \mathbb{P}(Y^0 = 1|G = g')$ and there exists a joint distribution $\mathbb{P}$ such that the inequality (7) *is strict for at least one pair g, g'.*

While the inequality in Proposition 6 may not always be strict, one advantage of using group status is that if treatment effects are sufficient to reduce baseline outcome disparity, then there exists $\mu_{\mathcal{L},\mathcal{G}}$ that can lead to outcome parity across all groups. To formalize this, let $g_{(0)}, g_{(1)}, \ldots, g_{(|G|-1)}$ denote an ordering of our protected groups such that $\mathbb{P}(Y^0 = 1|G = g_{(0)}) \geq \mathbb{P}(Y^0 = 1|G = g_{(1)}) \geq \cdots \geq \mathbb{P}(Y^0 = 1|G = g_{(|G|-1)})$ so that the groups are ordered from highest to lowest of group outcome under no treatment.

Proposition 7. *If $\mathbb{E}[Y^1 - Y^0|G = g_{(i)}] \geq \mathbb{P}(Y^0 = 1|G = g_{(0)}) - \mathbb{P}(Y^0 = 1|G = g_{(i)})\ \forall i > 0$, then there exists $\mu_{\mathcal{L},\mathcal{G}}$ such that*

$$\begin{aligned}&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{G}}(L,G))=1|G=g)-\\&\mathbb{P}(Y(\mu_{\mathcal{L},\mathcal{G}}(L,G))=1|G=g')=0\\&\forall g,g'\in\mathcal{G},\quad g\neq g'.\end{aligned} \tag{8}$$

Proposition 7 implies that if we could treat all individuals in group $g_{(i)}, i \neq 0$, and the groupwise treatment effect was more than baseline differences under no treatment, then it is possible to allocate resources based on group status to achieve outcome parity.

4 Discussion

In this paper we introduced a framework of measuring fairness in resource allocation systems at various stages of the allocation process. In settings like our motivating example, decision makers may desire a policy that is fair (e.g., imposing statistical parity) both in terms of allocation and outcomes. Our results show this is in general impossible. This incompatibility introduces tension between *equality* (which comes from allocative fairness) and *equity* (which comes from outcome fairness), leaving policymakers to have to choose between the two fairness goals. To add further nuance to this tension, we show that policies that are fair in allocation can result in better outcome fairness when that allocative fairness is achieved conditional on features measuring treatment effects rather than conditional on vulnerability. This provides a compromise of sorts where we could achieve allocative fairness while also inching toward better outcome fairness. However, doing so comes at the cost of not prioritizing the most vulnerable, which is often a desirable goal. Such ethical concerns need to be addressed by policymakers to determine the appropriate metrics of fairness evaluation.

We also show that using contextual information beyond baseline risk and treatment effects – but without considering group status – can lead to lower pairwise group disparities in terms of outcomes. On the other hand, explicit use of group status in addition to baseline risk and treatment effects can produce policies that are most outcome fair, which is often a desirable goal. However, using group status may pose practical and ethical issues since it is usually prohibited and it may be difficult to ensure such policies are not discriminatory even if well intended.

In the context of housing in Los Angeles, there is a desire on the part of community stakeholders to see fairness in all three stages: enrollment, allocation, and outcomes. As we have demonstrated, however, this desire is likely not possible. Furthermore, while HUD seeks to prioritize based on vulnerability scores, we also show this does not guarantee equitable outcomes. Thus, policymakers will have to make difficult decisions as to which stage of the allocation process they value most to impose fairness constraints. It is also incumbent upon the research community to help community stakeholders understand why such decisions must be made (i.e., the logical impossibilities presented above) and the relative trade-offs in opting for fairness in different stages of the allocation process.

Acknowledgements

P. Vayanos, B. Tang, S. Aghaei, and E. Rice gratefully acknowledge support from the Hilton C. Foundation, the Homeless Policy Research Institute, and the Home for Good foundation under the "C.E.S. Triage Tool Research & Refinement" grant. P. Vayanos and K. Dullerud are also grateful for the support of the National Science Foundation under CAREER award number 2046230. B. Tang was also supported by the National Science Foundation Graduate Research Fellowship Program.

References

Agarwal, A.; Beygelzimer, A.; Dudík, M.; Langford, J.; and Wallach, H. 2018. A reductions approach to fair classification. In *International Conference on Machine Learning*, 60–69. PMLR, Stockholm, Sweden: ICML.

Authority, L. A. H. S. 2018. Report and Recommendations of the Ad Hoc Committee on Black People Experiencing Homelessness. https://www.lahsa.org/documents?id=2823-report-and-recommendations-of-the-ad-hoc-committee-on-black-people-experiencing-homelessness. Accessed: 2023-03-24.

Authority, L. A. H. S. 2022a. 2022 Greater Los Angeles Homeless Count. https://www.lahsa.org/documents?id=6545-2022-greater-los-angeles-homeless-count-deck.pdf. Accessed: 2022-11-29.

Authority, L. A. H. S. 2022b. 2022 HOUSING INVENTORY COUNT. https://www.lahsa.org/documents?id=6544-2022-housing-inventory-count.xlsx. Accessed:2022-11-29.

Azizi, M. J.; Vayanos, P.; Wilder, B.; Rice, E.; and Tambe, M. 2018. Designing fair, efficient, and interpretable policies for prioritizing homeless youth for housing resources. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10848 LNCS: 35–51.

Barman, S.; Biswas, A.; Krishnamurthy, S. K.; and Narahari, Y. 2018. Groupwise maximin fair allocation of indivisible goods. In *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 917–924. New Orleans, Louisiana: AAAI.

Barocas, S.; Hardt, M.; and Narayanan, A. 2019. Fairness and Machine Learning: Limitations and Opportunities. http://www.fairmlbook.org. Accessed: 2023-03-24.

Baruah, S. K.; Cohen, N. K.; Plaxton, C. G.; and Varvel, D. A. 1993. Proportionate progress: A notion of fairness in resource allocation. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, 345–354.

Benabbou, N.; Chakraborty, M.; Elkind, E.; and Zick, Y. 2019. Fairness towards groups of agents in the allocation of indivisible items. In *IJCAI International Joint Conference on Artificial Intelligence*, volume 2019-August, 95–101. Macao, China: IJCAI.

Benabbou, N.; Chakraborty, M.; and Zick, Y. 2019. Fairness and Diversity in Public Resource Allocation Problems. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 42(3): 64–87.

Bertsimas, D.; Farias, V. F.; and Trichakis, N. 2013. Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research*, 61: 73–87.

Chouldechova, A. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data*, 5: 153–163.

Conitzer, V.; Freeman, R.; Shah, N.; and Vaughan, J. W. 2019. Group fairness for the allocation of indivisible goods. In *33rd AAAI Conference on Artificial Intelligence*, 1853–1860. Honolulu, Hawaii: AAAI.

Corbett-Davies, S.; Pierson, E.; Feller, A.; Goel, S.; and Huq, A. 2017. Algorithmic decision making and the cost of fairness. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume Part F129685, 797–806. Halifax, Canada: ACM.

Crowson, C. S.; Atkinson, E. J.; and Therneau, T. M. 2016. Assessing calibration of prognostic risk scores. *Statistical methods in medical research*, 25(4): 1692–1706.

Das, S. 2021. Local Justice and the Algorithmic Allocation of Societal Resources. arXiv:2112.01236.

Donahue, K.; and Kleinberg, J. 2020. Fairness and utilization in allocating resources with uncertain demand. In *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 658–668. Barcelona, Spain: ACM.

Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through awareness. In *ITCS 2012 - Innovations in Theoretical Computer Science Conference*, 214–226. New York, USA: ACM.

Elzayn, H.; Kearns, M.; Jabbari, S.; Neel, S.; Schutzman, Z.; Jung, C.; and Roth, A. 2019. Fair algorithms for learning in allocation problems. *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, 2019-January: 170–179.

Fain, B.; Munagala, K.; and Shah, N. 2018. Fair allocation of indivisible public goods. In *ACM EC 2018 - Proceedings of the 2018 ACM Conference on Economics and Computation*, 575–592. Ithaca, NY: ACM.

Feldman, M.; Friedler, S. A.; Moeller, J.; Scheidegger, C.; and Venkatasubramanian, S. 2015. Certifying and removing disparate impact. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume 2015-August, 259–268. Sydney, Australia: ACM.

Finocchiaro, J.; Maio, R.; Monachou, F.; Patro, G. K.; Raghavan, M.; Stoica, A.-A.; and Tsirtsis, S. 2021. Bridging Machine Learning and Mechanism Design towards Algorithmic Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 489–503. Virtual: ACM.

Flores, A. W.; Bechtel, K.; and Lowenkamp, C. T. 2016. False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. *Fed. Probation*, 80: 38.

Freeman, R.; Shah, N.; and Vaish, R. 2020. Best of both worlds: Ex-ante and ex-post fairness in resource allocation. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 21–22.

Greater Los Angeles Coordinated Entry System. 2018. Greater Los Angeles Coordinated Entry System For Individuals Survey Packet. http://publichealth.lacounty.gov/sapc/Event/HomelessServices/050318/CES-individuals-survey-packet.pdf, Last accessed on 2022-08-11.

Hardt, M.; Price, E.; and Srebro, N. 2016. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*. Barcelona, Spain: NeurIPS.

Huq, A. Z. 2018. Racial equity in algorithmic criminal justice. *Duke LJ*, 68: 1043.

Kash, I.; Procaccia, A. D.; and Shah, N. 2014. No Agent Left behind: Dynamic Fair Division of Multiple Resources. *J. Artif. Int. Res.*, 51(1): 579–603.

Kleinberg, J.; Ludwig, J.; Mullainathan, S.; and Rambachan, A. 2018. Algorithmic fairness. In *Aea papers and proceedings*, volume 108, 22–27.

Kleinberg, J.; Mullainathan, S.; and Raghavan, M. 2016. Inherent Trade-Offs in the Fair Determination of Risk Scores. *arXiv preprint arXiv:1609.05807v2*.

Kumar, A.; and Kleinberg, J. 2000. Fairness measures for resource allocation. In *Proceedings 41st annual symposium on foundations of computer science*, 75–85. IEEE.

Lan, T.; Kao, D.; Chiang, M.; and Sabharwal, A. 2010. *An axiomatic theory of fairness in network resource allocation*. IEEE.

Mashiat, T.; Gitiaux, X.; Rangwala, H.; Fowler, P.; and Das, S. 2022. Trade-offs between Group Fairness Metrics in Societal Resource Allocation. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1095–1105.

Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; and Galstyan, A. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54: 1–35.

Milburn, N. G.; Edwards, E.; Obermark, D.; and Rountree, J. 2021. Inequity in the Permanent Supportive Housing System in Los Angeles: Scale, Scope and Reasons for Black Residents' Returns to Homelessness. https://www.capolicylab.org/wp-content/uploads/2021/10/Inequity-in-the-PSH-System-in-Los-Angeles.pdf. Accessed: 2023-03-24.

Rahmattalabi, A.; Vayanos, P.; Dullerud, K.; and Rice, E. 2022. Learning Resource Allocation Policies from Observational Data with an Application to Homeless Services Delivery. *arXiv preprint arXiv:2201.10053*.

Raman, N.; Shah, S.; and Dickerson, J. 2021. Data-Driven Methods for Balancing Fairness and Efficiency in Ride-Pooling. In *IJCAI*, 363–369. Montreal, Canada: IJCAI.

Rodolfa, K. T.; Lamba, H.; and Ghani, R. 2021. Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy. *Nature Machine Intelligence*, 3(10): 896–904.

Rodolfa, K. T.; Salomon, E.; Haynes, L.; Mendieta, I. H.; Larson, J.; and Ghani, R. 2020. Case study: predictive fairness to reduce misdemeanor recidivism through social service interventions. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 142–153.

Scarlett, J.; Teh, N.; and Zick, Y. 2021. For One and All: Individual and Group Fairness in the Allocation of Indivisible Goods. In *IJCAI 2021 Workshop on AI for Social Good*. IJCAI.

U.S. Department of Housing and Urban Development. 2014. Notice on Prioritizing Persons Experiencing Chronic Homelessness and Other Vulnerable Homeless Persons in Permanent Supportive Housing and Recordkeeping Requirements for Documenting Chronic Homeless Status. https://www.hud.gov/sites/documents/14-12CPDN.PDF, Last accessed on 2022-08-11.

Ustun, B.; and Rudin, C. 2019. Learning Optimized Risk Scores. *Journal of Machine Learning Research*, 20(150): 1–75.

Vayanos, P.; Ye, Y.; McElfresh, D.; Dickerson, J.; and Rice, E. 2021. Robust Active Preference Elicitation. arXiv:2003.01899.

Wager, S.; and Athey, S. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

Zafar, M. B.; Valera, I.; Rodriguez, M. G.; and Gummadi, K. P. 2017. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. *26th International World Wide Web Conference, WWW 2017*, April-2017: 1171–1180.