Open Problems in (Un)fairness of the Retail Food Safety Inspection Process

Tanya Berger-Wolf [1 2 3] **Allison Howell** [4] **Chris Kanich** [5] **Ian A. Kash** [5] **Moniba Keymanesh** [1]
Barbara Kowalcyk [4] **Gina Nicholson Kramer** [4] **Andrew Perrault** [1] **Shubham Singh** [5]

Abstract

The inspection of retail food establishments is an essential public health intervention. We discuss existing work on roles AI techniques can play in food inspections and resulting fairness and interpretability challenges. We also examine open problems stemming from the complex and dynamic nature of the inspections.

1. Introduction

Foodborne illness is a serious public health issue in the United States, resulting in an estimated 48 million illnesses and 3,000 deaths annually (Scallan et al., 2011). Food served in retail settings have been associated with a significant number of foodborne illness outbreaks. As such, the inspection of retail food establishments, which include restaurants, grocery stores, caterers, school cafeterias, and even vending machines, is an important public health intervention to reduce the spread of foodborne illness.

Every food safety inspection generates data: the properties of the establishment, the *sanitarian* performing the inspection, and the details of the result—what specific violations were found and the inspector's remarks in support of the violations. This data results from a complex process involving multiple stakeholders (policy makers, restaurant owners, the public) and operating via many organizations at a range of scales (from national organizations to local public health departments). There are opportunities to use AI to improve food safety across the system: scheduling inspections for a limited pool of inspectors; training food workers, supervisors, and sanitarians; and aiding regulatory decisions on the inspection process and outcomes.

Fairness considerations emerge immediately. For example, in Chicago, the reported critical violation rate for some sanitarians is 16 times higher than others (Schenk Jr et al., 2015). On the other hand, assignment of sanitarians to food establishments is largely based on location proximity(see §3). Thus, naive inspection scheduling policies that only consider predicted violation rate as a proxy for public health outcomes produce discriminatory outcomes by prioritizing inspection of restaurants in a few regions.

Retail food inspection is a useful application for study of fairness in sequential decision-making. Inspections occur often—Chicago performs tens of thousands of inspections per year—and the generated data is publicly available. States and municipalities are interested in prioritizing limited resources to minimize potential risks to the public health in a fair manner. However, there exist disagreement regarding the fairness objective to account for. Historic decisions may reflect important individual and group fairness considerations that should be understood and integrated into future decisions.

In this paper, we explain the food inspection process, discuss prior computational work on making the process more fair, and outline future research directions. In §2, we describe the regulation and the use of predictive modeling for scheduling food inspections Schenk Jr et al. (2015). §3 describes how such models can lead to unfairness. §4 we discuss three challenges that arise in the real dynamic inspection process: designing fair algorithms for dynamic scheduling, assigning heterogeneous inspectors, and the use of proxies for public health. In §5, we consider potential biases in individual inspectors and describe a previous work on detecting and counteracting bias in written justifications. In §6, we discuss, the multi-stakeholder aspect of fairness in the food inspection process and discuss how fairness notions should account for this.

2. Background on Inspection Scheduling

In the US, the Food and Drug Administration (FDA) releases the FDA Model Food Code to outline food retail

[1]Computer Science and Engineering, Ohio State University, USA [2]Evolution, Ecology and Organismal, Ohio State University, USA [3]Electrical and Computer Engineering, Ohio State University, USA [4]Food Science and Technology, Ohio State University, USA [5]Computer Science, University of Illinois at Chicago, USA. Correspondence to: Ian A. Kash <iankash@uic.edu>, Tanya Berger-Wolf <berger-wolf.1@osu.edu. This material is based upon work supported by the National Science Foundation Program on Fairness in AI in collaboration with Amazon under Grant No. 1939743.">.

ICML 2022 Workshop on Responsible Decision Making in Dynamic Environments, Baltimore, Maryland, USA, 2022.

practices and provide guideline for retail food safety inspection. The adaption and enforcement of the Food Code is left to state and local governments. This results in a considerable variability in the inspections conducted in different states. This includes differences in the nature and format of data gathered, what constitutes a violation, and the frequency and nature of inspections (which include routine *canvass inspections* where establishments are inspected once or twice a year, inspections triggered by complaints or illnesses, follow-up inspections, and additional state-specific inspection types). Data generated from retail food safety inspections has been used to evaluate and enhance inspection policies and training requirements across jurisdictions (Jin et al., 2020). In the most relevant prior work, Schenk Jr et al. (2015) built a predictive model to prioritize scheduling food inspections conducted by the Chicago Department of Public Health (CDPH).[1] The model predicts the outcome of a canvass inspection, using historical inspections as supervision. The features used include details about the establishment (e.g., whether it has an alcohol license), the sanitarian who conducted the inspection, and information about the outcomes of past inspections. To protect individual identities, the three dozen sanitarians were clustered by their past *critical violation rate*, the empirical percentage of conducted inspections that resulted in a critical violation. These are violations which, if left uncorrected, are more likely than other violations to directly contribute to food contamination or illness.

Schenk Jr et al. (2015) seek to detect critical violations as quickly as possible. Thus, the model predictions are used to prioritize inspections in decreasing order of risk. Under the assumption that critical violations are detectable at any time during the 60-day testing window, the average time to detect a critical violation can be computed for any schedule. Risk-prioritized schedules were found to detect critical violations one week faster, on average, when applied to the two-month test set. The Schenk Jr et al. (2015) analysis has been revisited by Kannan et al. (2019) and, specifically from a fairness perspective, by Singh et al. (2021).

3. Fair Schedules

Singh et al. (2021) quantify and assess the fairness of a schedule for food inspections. We focus on their interpretation of Equal Opportunity (Hardt et al., 2016) which is defined as:

$$\mathbb{E}[T|A = a^i, Y = 1] = \mathbb{E}[T|A = a^{i+1}, Y = 1] \quad (1)$$
$$\text{s.t. } 0 \leq i < n \, .$$

[1]While we are aware of other work on predicting inspection results and the biases this may introduce (Altenburger & Ho, 2019; Kang et al., 2013; Liu, 2020), this is the only example we are aware of where predictions have been linked to inspection schedules.

Sanitarian Cluster	**Critical Violation Rate**
Purple	40.83%
Blue	25.53%
Orange	13.76%
Green	9.68%
Yellow	5.94%
Brown	2.5%

Table 1. The critical violation rate for different inspection clusters. Clusters are named after the lines of Chicago rail transit system similar to (Singh et al., 2021).

In the equation above, T denotes the time to complete a food inspection, Y indicates the result of a food inspection, and A indicates a categorical protected attribute.Equation (1) defines a schedule as fair if the expected detection time for a critical violation ($Y = 1$) is same across different protected groups.

Singh et al. (2021) use restaurants' Zip codes to assign them to nine regions of the city. Their analysis shows that the schedule suggested by the CDPH's substantially violates fairness based on Equal Opportunity. Moreover, for several regions of the city, this schedule, results in detecting the critical violations even later than the default setting. A closer look at coefficients of the logistic regression model used by CDPH (Schenk Jr et al., 2015) shows that cluster label of the sanitarian conducting the food inspection is a key feature (Kannan et al., 2019) for the model. Significant variation in the detected critical violation rate is observed for different sanitarians groups (see Table 1). The detection rate is as much as 40.8% for the Purple cluster and as little as 2.5% for the Brown cluster! Singh et al. (2021) demonstrate that this difference is due to sanitarian behavior rather than other possible confounding factors and using sanitarian cluster as an input feature causes the resulting unfairness. They explore two broad classes of approaches to remedy the unfairness. First, they explore approaches that involve retraining the model without using sanitarian as a feature and using various fairness-aware classification algorithms (Zafar et al., 2017; Rezaei et al., 2020; Krishnaswamy et al., 2021). These approaches only offer a partial solution since they mitigate the unfairness associated with the sanitarian clusters to an extent. Second, they explore using approaches based on how the risk prediction is used. They find that using the prediction scores to reorder inspections within sanitarian clusters offers the fairest schedules.

The observed variation in how the inspection is conducted in different regions is not unique to Chicago (Jones et al., 2004). This variation presents a number of challenges and opportunities for fairness research that will be discussed in §4 and §5.

4. Dynamic Inspection Scheduling

Prior work has looked at food inspection scheduling as a static problem in which fixed number of inspections are conducted in a fixed time window. However, this problem has a dynamic nature. This poses at least three distinct (though interacting) open challenges. First, how can we design fair dynamic scheduling algorithms? Second, how should we assign sanitarians to inspections? Third, since violations are proxies for public health outcomes, how can we avoid feedback loops?

Schenk Jr et al. (2015) trained classification models and used the posterior probability of a critical violation to rank food establishments based on potential risk for public health. Their model prioritized inspecting establishments that pose a higher risk. However, they did not account for equal distribution of benefits obtained by this rescheduling. Better results may be possible by using ideas from the literature on fair rankings (Singh & Joachims, 2019; Zehlike et al., 2021) or the literature on fair classification in the context of larger systems (Dwork & Ilvento, 2018). However, this neglects opportunities to adjust the frequency and timing of inspections over longer periods of time. Ideally, we could inspect riskier establishments more frequently as well as sooner. Existing inspection regimes do this in coarse ways, with a small number of "risk levels" based on the nature of the food service and differing inspection frequencies. This could be done in a finer-grained manner using automated and data-driven methods. However, the objectives, both for efficiency and fairness, are unclear.

An alternative approach might be to treat scheduling as a sequential decision problem and apply, e.g., bandit algorithms. Previous work has looked at designing fair algorithms of this flavor (Joseph et al., 2016; Wen et al., 2021; Hossain et al., 2021) and studied issues of overall fairness in comparison to local or immediate fairness (D'Amour et al., 2020; Emelianov et al., 2019; Dwork et al., 2020; Liu et al., 2018). However, it is not clear how retail food safety inspections can be modeled as a bandit. There is a significant gap between existing theory and a deployment-ready model.

Regardless of the approach, an implementable solution to the dynamic scheduling problem should consider the limited capacity for a sanitarian to conduct inspections in a day both in terms of the time needed to conduct the inspections themselves and the time needed to travel from inspection to inspection. Efficiency or fairness gains which do not respect these constraints may be illusory.

Table 1 indicates there are substantial differences among the critical violation rates of different inspectors. The differences between sanitarians could be modeled by different severity thresholds that lead to citation of a violation. In this approach, what a violation "means" varies between sanitarians. Therefore, we might estimate a counterfactual "sanitarian-independent" violation probability, as is done when predicting clicks in search advertising (Graepel et al., 2010) and has been explored in the literature on causal models in fairness (Kilbertus et al., 2017). More detail may be desired—prior work has found that factors such as the outcome of a previous inspection and the position of an inspection in an inspector's daily schedule may significantly impact the detection of violations in an inspection (Ibanez & Toffel, 2020). Furthermore, violations related to keeping food at a proper temperature may be more prone to occur on warm days (Kannan et al., 2019).

A key challenge in planning inspections is that critical violations are only a proxy for the true goal of identifying and fixing risks to public health. The use of such proxies is common and has caused notable issues in other domains (for example the use of arrests as a proxy for crime (Ensign et al., 2018; Keymanesh et al., 2020)). The risk of feedback loops has been pointed out in both this and other domains (Kannan et al., 2019; Chouldechova et al., 2018). In addition to leading to sub-optimal public health outcomes, this can be viewed as an issue of unfairness to restaurants (Kannan et al., 2019). Similar concerns arise from the use of other proxies such as customer complaints, which reflect both consumer biases (Altenburger & Ho, 2019) and biases due to under-reporting (Liu & Garg, 2022). Data from surveillance of foodborne outbreaks can potentially provide a signal of inspection effectiveness that is separate from inspection outcomes.

5. Fairness of Justification

Violations are not the only outcomes produced by retail food safety inspections and the scheduling of inspections is not the only part of the process in which AI can play a role. In the current inspection form, sanitarians are required to report their observations and justifications for any reported violation. A fair inspection should both have a fair outcome and be fairly justified (Carvalho et al., 2019)—the justification should include enough information to explain the outcome and should be consistent across establishments. Keymanesh et al. (2021) focused on the problem of (un)fairness of justifications. Examples of unfair justifications include those that do not reflect the guidelines, use racially coded language, display implicit bias, or apply different standards to different establishments (Keymanesh et al., 2021).

To detect and counteract the unfairness in justification, Keymanesh et al. (2021) propose a text pre-processing approach called FAIRSUM. Given a collection of n potentially biased inspections $\{(X_i, Y_i, A_i)\}_{i=1}^{n}$ where X denotes a textual report written by a sanitarian to provide evidence or justify an outcome Y and A indicates one or more protected variables, they extract a fairly-justified summary $\{{X_i}'\}_{i=0}^{n}$

such that X' provides sufficient information to predict and justify Y and X' is independent of protected variable A. To achieve this, they use a multi-task neural model and an attribution mechanism to compute a utility score u_i and a discrimination score d_i for each sentence in the justification.

First the model is trained for decision prediction and membership identification. Next, they measure the salience of each sentence in predicting outcome $\hat{Y}_i$ and protected attribute $\hat{A}_i$ using integrated gradients (Sundararajan et al., 2017). They include high-utility and low-bias sentences in the fairly-justified summary of the explanations. The final inclusion score s_i for each sentence is computed as

$$s_i = \sigma(u)_i - \alpha \times \sigma(d)_i, \quad (2)$$

where σ indicates the Softmax function and is applied to generate a distribution over input sentences, and α is a hyper-parameter that controls the utility-discrimination trade-off. Higher values of α correspond to removing more information about the protected attribute from the input justification. The goal is to identify and remove arguments that are not useful for decision prediction, except through the prediction of the protected attribute. The subtraction operation ensures that such arguments get a small inclusion score s_i. Finally, only sentences with a positive final attribution score are included in the summary X'.

Enhancing the fairness of justification for retail food inspections could improve inspection quality in multiple ways, including addressing several challenges discussed in §4:

- **Bias detection:** Auditing sanitarians for bias to ensure the inspection process is consistent across different neighborhoods and sanitarians.
- **Better training to avoid future biases:** Improving the training of sanitarians by explaining cases of unfair historical decisions or justifications in a data-driven manner.
- **Better explanation for stakeholders:** Attributing the inspection outcomes to specific violations reported by sanitarians. This can help explain the outcomes to stakeholders and assist restaurant owners in prioritizing resources in order to reduce the risk to public health.
- **Better data quality:** Removing unjustified claims or biased reasoning from the data facilitates training fair automated models for decision-making.

6. Multi-Stakeholder Fairness

In contrast to much of the literature that focuses on the fair treatment of individuals, the establishments that are inspected are typically businesses or other entities in which many individuals have a stake. So far, we have discussed the challenges associated with fair scheduling of inspections with respect to people living nearby the restaurants (§4) and fair justification for the inspection results with respect to restaurant owners (§5). These are examples of a broader challenge of establishing definitions and metrics for fairness in settings with multiple stakeholders.

Fair classification can be thought of as the problem of learning a classifier satisfying appropriate constraints from samples of the joint distribution of (X, Y, A), where X indicates the input features, Y indicates the target variable (outcome of a decision), and A is the protected attribute. The case where the support of A is finite has been well-studied in the literature (Calders et al., 2009; Hardt et al., 2016), and some recent works have also proposed improvement to the fairness notions for the real-valued case (Mary et al., 2019; Jiang et al., 2021).

One way to capture the scheduling of retail food safety inspections would be to take the support of A to be probability distributions. For example, consider a restaurant located in a neighborhood with demographic distribution of 50% White, 30% Black, and 20% Hispanic. Existing fair classification algorithms are generally not compatible with such heterogeneous protected attributes. This also poses thorny challenges for the definition of fairness, including how to define the relevant set of stakeholders, how to handle overlapping sets of stakeholders, and how to capture fairness with respect to subgroups of stakeholders. There is also a dynamic aspect when viewed on longer timescales since the population of a census block, ward, or neighborhood changes over time (as people move or as the result of a redistricting process).

We may also wish to consider particular subgroups. Individuals from lower SES are disproportionately affected by foodborne illnesses. They are also more likely to live in food deserts and/or shop at different types of food establishments. If these establishments are over-prioritized for inspection, this may result in retail closures (due to violations or increased compliance costs) that contribute to increased food insecurity. If they are under-prioritized, violations may go unnoticed, causing illness in a vulnerable population.

7. Conclusion

In this work, we examined retail food safety inspections as a key application domain for fair AI techniques. We discuss several open challenges including (dynamic) inspection scheduling, fairness of justification, and multiple stakeholders. Governments conduct a number of other types of inspections including structural inspections of buildings, fire safety, business licensing, and enforcement of environmental and accessibility regulations. Thus we believe retail food safety inspections are an application domain that merits substantial attention in its own right; as an example of a number of related domains; and as opportunity to explore fundamental fairness challenges related to deficiency of information, dynamic decisions, and disagreement among objectives.

References

Altenburger, K. M. and Ho, D. E. When algorithms import private bias into public enforcement: The promise and limitations of statistical debiasing solutions. *Journal of Institutional and Theoretical Economics*, 175(1):98–122, 2019.

Calders, T., Kamiran, F., and Pechenizkiy, M. Building Classifiers with Independency Constraints. In *2009 IEEE International Conference on Data Mining Workshops*, pp. 13–18, December 2009. doi: 10.1109/ICDMW.2009.83. ISSN: 2375-9259.

Carvalho, D. V., Pereira, E. M., and Cardoso, J. S. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8):832, 2019.

Chouldechova, A., Benavides-Prado, D., Fialko, O., and Vaithianathan, R. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency*, pp. 134–148. PMLR, 2018.

D'Amour, A., Srinivasan, H., Atwood, J., Baljekar, P., Sculley, D., and Halpern, Y. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 525–534, 2020.

Dwork, C. and Ilvento, C. Fairness under composition. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

Dwork, C., Ilvento, C., and Jagadeesan, M. Individual fairness in pipelines. *arXiv preprint arXiv:2004.05167*, 2020.

Emelianov, V., Arvanitakis, G., Gast, N., Gummadi, K., and Loiseau, P. The price of local fairness in multistage selection. *arXiv preprint arXiv:1906.06613*, 2019.

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., and Venkatasubramanian, S. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency*, pp. 160–171. PMLR, 2018.

Graepel, T., Candela, J. Q., Borchert, T., and Herbrich, R. Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine. In *ICML*, 2010.

Hardt, M., Price, E., and Srebro, N. Equality of opportunity in supervised learning. *arXiv preprint arXiv:1610.02413*, 2016.

Hossain, S., Micha, E., and Shah, N. Fair algorithms for multi-agent multi-armed bandits. *Advances in Neural Information Processing Systems*, 34, 2021.

Ibanez, M. R. and Toffel, M. W. How scheduling can bias quality assessment: Evidence from food-safety inspections. *Management science*, 66(6):2396–2416, 2020. ISSN 0025-1909.

Jiang, Z., Han, X., Fan, C., Yang, F., Mostafavi, A., and Hu, X. Generalized demographic parity for group fairness. In *International Conference on Learning Representations*, 2021.

Jin, C., Bouzembrak, Y., Zhou, J., Liang, Q., Van Den Bulk, L. M., Gavai, A., Liu, N., Van Den Heuvel, L. J., Hoenderdaal, W., and Marvin, H. J. Big data in food safety-a review. *Current Opinion in Food Science*, 36:24–32, 2020.

Jones, T. F., Pavlin, B. I., LaFleur, B. J., Ingram, L. A., and Schaffner, W. Restaurant inspection scores and foodborne disease. *Emerging Infectious Diseases*, 10(4):688–692, April 2004. ISSN 1080-6040. doi: 10.3201/eid1004.030343.

Joseph, M., Kearns, M., Morgenstern, J. H., and Roth, A. Fairness in learning: Classic and contextual bandits. *Advances in neural information processing systems*, 29, 2016.

Kang, J. S., Kuznetsova, P., Luca, M., and Choi, Y. Where not to eat? improving public policy by predicting hygiene inspections using online reviews. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1443–1448, 2013.

Kannan, V., Shapiro, M. A., and Bilgic, M. Hindsight analysis of the chicago food inspection forecasting model. *arXiv preprint arXiv:1910.04906*, 2019.

Keymanesh, M., Gurukar, S., Boettner, B., Browning, C., Calder, C., and Parthasarathy, S. Twitter watch: Leveraging social media to monitor and predict collective-efficacy of neighborhoods. In Barbosa, H., Gomez-Gardenes, J., Gonçalves, B., Mangioni, G., Menezes, R., and Oliveira, M. (eds.), *Complex Networks XI*, pp. 197–211, Cham, 2020. Springer International Publishing. ISBN 978-3-030-40943-2.

Keymanesh, M., Berger-Wolf, T., Elsner, M., and Parthasarathy, S. Fairness-aware summarization for justified decision-making. *arXiv e-prints*, pp. arXiv–2107, 2021.

Kilbertus, N., Rojas-Carulla, M., Parascandolo, G., Hardt, M., Janzing, D., and Schölkopf, B. Avoiding discrimination through causal reasoning. *arXiv preprint arXiv:1706.02744*, 2017.

Krishnaswamy, A., Jiang, Z., Wang, K., Cheng, Y., and Munagala, K. Fair for all: Best-effort fairness guarantees for classification. In Banerjee, A. and Fukumizu, K. (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 3259–3267. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/krishnaswamy21a.html.

Liu, L. T., Dean, S., Rolf, E., Simchowitz, M., and Hardt, M. Delayed impact of fair machine learning. In *International Conference on Machine Learning*, pp. 3150–3158. PMLR, 2018.

Liu, P. Making the grade: Rethinking the us food retail inspection and rating regulatory system. *Calif. L. Rev.*, 108:1639, 2020.

Liu, Z. and Garg, N. Equity in resident crowdsourcing: Measuring under-reporting without ground truth data. *arXiv preprint arXiv:2204.08620*, 2022.

Mary, J., Calauzènes, C., and Karoui, N. E. Fairness-Aware Learning for Continuous Attributes and Treatments. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 10. PMLR, 2019. URL https://proceedings.mlr.press/v97/mary19a.html.

Rezaei, A., Fathony, R., Memarrast, O., and Ziebart, B. Fairness for robust log loss classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):5511–5518, Apr. 2020. doi: 10.1609/aaai.v34i04.6002. URL https://ojs.aaai.org/index.php/AAAI/article/view/6002.

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M.-A., Roy, S. L., Jones, J. L., and Griffin, P. M. Foodborne illness acquired in the united states—major pathogens. *Emerging infectious diseases*, 17(1):7, 2011.

Schenk Jr, T., Leynes, G., Solanki, A., Collins, S., Smart, G., Albright, B., and Crippin, D. Forecasting restaurants with critical violations in chicago. Technical report, Technical report, 2015.

Singh, A. and Joachims, T. Policy learning for fairness in ranking. *arXiv preprint arXiv:1902.04056*, 2019.

Singh, S., Shah, B., and Kash, I. A. Fair decision-making for food inspections. *CoRR*, abs/2108.05523, 2021. URL https://arxiv.org/abs/2108.05523.

Sundararajan, M., Taly, A., and Yan, Q. Axiomatic attribution for deep networks. In *International conference on machine learning*, pp. 3319–3328. PMLR, 2017.

Wen, M., Bastani, O., and Topcu, U. Algorithms for fairness in sequential decision making. In *International Conference on Artificial Intelligence and Statistics*, pp. 1144–1152. PMLR, 2021.

Zafar, M. B., Valera, I., Rogriguez, M. G., and Gummadi, K. P. Fairness Constraints: Mechanisms for Fair Classification. In Singh, A. and Zhu, J. (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 962–970. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/zafar17a.html.

Zehlike, M., Yang, K., and Stoyanovich, J. Fairness in ranking: A survey. *arXiv preprint arXiv:2103.14000*, 2021.