Can we Use Local Outreach to Improve Equity in Federal Oversight? A Case Study with the H-2A Visa Program

Rebecca A. Johnson,* Elizabeth Shackney, Cassie Davis,
Grant Anapolle, Yuchuan Ma, You-Chi Liu, and Cameron Guage

Prepared for DOL Summer Data Challenge: September 2021

Executive Summary

The Department of Labor's (DOL) H-2A guest worker program plays a critical role in supporting agricultural employment and production in the United States. Under Executive Order 13985, President Joe Biden has provided an opportunity for federal agencies to assess equity challenges under their purview. In this report, we investigate equity issues related to legal oversight of the H-2A program. Currently there are two potential avenues for legal oversight of the H-2A program: (1) oversight at the stage when employers submit a job clearance order seeking workers and (2) oversight after DOL has approved employers for H-2A visas.

Focusing on the latter stage, there are three ways that the DOL Wage and Hour Division (WHD) does or can conduct oversight: (1) rely on workers and advocates to submit complaints about issues, (2) randomly target high-risk employers, and (3) strategic targeting, or using data to try to identify high-risk employers and subject them to heightened oversight. We outline how all forms of enforcement beyond random targeting rely on workers and advocates to submit complaints, but also show how past research has identified barriers employees face in escalating complaints to federal authorities. Due to these barriers, we investigate whether local "trusted messengers"—or organizations that conduct proactive outreach to workers—can uncover issues that are not reported to agencies.

We partner with a legal service provider, Texas Rio Grande Legal Aid (TRLA), which delivers legal services to H-2A workers and conducts extensive outreach to workers. Our research investigates whether TRLA outreach uncovers issues that are not reported federally. If they do, partnership with such organizations could improve both complaint-driven enforcement by supplementing complaints to federal authorities with complaints to local organizations and strategic targeting by providing an additional "high risk" label for models to predict.

We focus on three research questions to explore the role of local outreach in uncovering issues missed by federal enforcement. First, we compare employers investigated by DOL's WHD versus

*This report was prepared for the U.S. Department of Labor (DOL), Chief Evaluation Office (CEO) by the listed authors. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

TRLA. Second, we investigate the feasibility of predicting which employers are likely to be investigated by WHD. Third, we observe if text from employer-drafted job addenda can explain or predict future investigations or violations. In order to address these questions, we use H-2A job certificate data from FY 2008 to Q1 FY 2021, WHD Compliance Action data from FY 2008 to Q1 FY 2021, TRLA intake data for cases on or after January 1, 2014, and American Community Survey tract-level data for 2014-2019. Our sample represents either all employers with an approved H-2A job from FY 2014-FY 2021 (nationwide sample) or all entities located in one of the seven TRLA catchment states: Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Tennessee, and Texas. First, we match (1) these universes of employers to (2) investigations by WHD and investigations by TRLA. We can then calculate descriptive statistics of different categories, most importantly whether an employer-job dyad was investigated by WHD, by TRLA, or both. We then use supervised machine learning (SML) to train a model to predict which employers are investigated and to identify factors that are predictive of either a higher risk of investigation or lower risk. Finally, we perform a computational text analysis of the addendum text.

Our three main findings are (1) that WHD investigations and TRLA intake records investigate *different employers*; (2) predictive modeling is useful at the nationwide scale for predicting WHD investigations, but less useful when restricted to states within TRLA's catchment area or when investigating differences between federal and local enforcement; and (3) the text addendum for employers who applied for H-2A certification suggest that criminal history-focused requirements can predict later investigation. Limitations in our research include the unit of analysis as employer-job dyad rather than employers. This method averts aggregation bias, but it can mask some high-risk employers. In addition, while our models analyze demographics of the surrounding Census tract, we do not have the demographic information of each employer's workforce, an important component of equity. Finally, pooling data over time can mask important year-to-year variation; this could be remedied with a larger sample size.

We propose two data-related recommendations for DOL: (1) Create an internal dataset to compare complaint-derived investigations to proactive investigations for better predictive models, and (2) establish a standardized employer ID to help merge certificate and compliance action data, as well as improve research replicability. On policy and investment fronts, we recommend DOL and WHD incorporate random sampling into their federal oversight strategy. We also recommend DOL and WHD conduct an evaluation of outreach strategies to compare current methods to a method that incorporates trusted community-based organizations.

Contents

1 Background: Legal Enforcement of H-2A Worker Rights

1.1 Background on the H-2A Program and the Equity Executive Order

The H-2A guest worker program seeks to address the hiring needs of agricultural employers throughout the US and provides migrant workers from foreign countries with US employment opportunities. Oversight of this program is provided by the Department of Labor's Wage and Hour Division (WHD), which enforces statutes that protect both U.S. workers and guest workers. More specifically, legal oversight aimed at protecting U.S. workers is meant to ensure that the guest-worker program is not used to lower their wages or degrade their working conditions (Department of Labor, 2021). In addition, the program protects the rights of agricultural guest workers while they are working in the country (Department of Labor, 2021). These protections include monitoring H-2A employers' compliance with the program's set terms, such as workers' rights to pay for work performed, safe work sites, transportation, and habitable living conditions. In addition, the WHD is tasked with investigating noncompliance with wage and hour laws and sanctioning employers who are found to be in violation.

The H-2A program was formally established under the Immigration and Nationality Act and later reorganized by the Immigration Reform and Control Act to separate the program into agricultural and non-agricultural work programs (Bruno, 2020). The overwhelming majority of admitted H-2A workers are from Mexico (approximately 94 percent), followed by Canada (1.5 percent), Jamaica (1.1 percent), South Africa (1.1 percent), Guatemala (0.6 percent) while the remaining workers were citizens of other countries (1.4 percent) (DHS, 2019). In a recent report by the Economic Policy Institute, researchers estimate that H-2A workers made up 10 percent of the agricultural workforce. They face heightened barriers to reporting employer noncompliance due to their immigration status (Costa et al., 2020). Given the particular vulnerabilities of workers who make up the H-2A program, the WHD has an opportunity to advance equity through examining how well the current oversight apparatus protects workers rights and ways to improve.

Opportunities to improve equity are especially salient in light of the Biden administration's efforts to both examine historical inequities in existing federal programs and to implement equity-focused reforms to programs' implementation. In Executive Order 13985, the Biden Administration calls on government agencies to advance equity with a systematic approach in which the "systematic fair, just, and impartial treatment of all individuals, including individuals who belong to under-served communities that have been denied such treatment" (Biden, 2021). In addition, the Evidence Act calls for governmental agencies to develop learning agendas and evaluation plans of their current systems (Dooling, 2020).

In the present report, we focus on equity in legal oversight of the H-2A guestworker program. We ask: How can DOL augment its own data sources used for oversight with data from local, community-based organizations? We focus specifically on a legal services provider (LSP) that does proactive outreach to agricultural workers in seven states with high densities of these workers. Then, we ask what divergences between our two data sources—federal WHD investigations of employers and results from local outreach—reveal about a potential "leaky pipeline" into federal reporting? While the answer to the second question is preliminary, our goal is to prompt a larger research agenda focused on marrying local and federal data sets for H-2A program oversight.

1.2 Avenues for Legal Oversight

There are two avenues for legal oversight of H-2A employers. The first avenue is during the *job clearance process*. This avenue primarily implicates the Office of Foreign Labor Certification (OFLC), from which H-2A employers must seek approval in hiring seasonal labor. There are four steps employers must take to receive temporary labor certification: (1) filing a job order with their state workforce agency to initiate recruitment of U.S. workers, (2) filing an H-2A application to be screened by the OFLC National Processing Center, (3) conduct recruitment of US worker through methods such as advertising, and (4) complete the certification process and receive final determination (Department of Labor, 2021). As part of this screening process, employers present information to DOL that might be relevant for predicting if workers hired for that job, or U.S. workers, will face rights violations. In particular, and as we outline in greater detail in the Data and Methods section, the H-2A applications contain two types of employer-related fields:

1. **Structured characteristics of the employers and jobs:** These include the attorney/agent used to prepare the application, the industry that the job is located in (NAICS code), the occupational title of the job (SOC code), and others. In turn, legal violations, rather than being uniformly distributed across employers with different characteristics, may be higher in certain industries or among employers who solicit the services of certain attorney/agents. We explore whether any of these features predict issues.

2. **Unstructured text of job addendums:** Beginning in FY 2020, for employers who attach an addendum to their submission, we have the full text of the addendum. We explore which themes in addendums are correlated with a higher risk of legal oversight.

The second avenue involves WHD oversight by the WHD of employers currently in H-2A program to ensure they are cited for violations, as past research suggests that issues may be under-reported (Costa et al., 2020). These violations can include 1) willful violation of the work contract, 2) violation of a housing or transportation safety and health prevision, 3) repeat of the previous two violations that causes death or serious injury, and 4) violation for failure to cooperate in an investigation (Department of Labor, 2021). This oversight could supplement the data available at the time the employer undergoes the certification process with data available at the time the work is taking place, including contextual characteristics of the surrounding areas (e.g., unemployment rates) and leads from local outreach.

1.3 Equity Challenges in Complaint-Driven Enforcement & Ways to Improve Equity

Workplace inspections, investigations, and penalties imposed during legal oversight can improve working conditions. The existing body of research on oversight mechanisms under DOL largely focuses on the Occupational Safety and Health Administration (OSHA), which both covers different types of workplace violations and structures oversight in different ways than WHD. Nevertheless, the scholarship on OSHA, and enforcement inequities in that form of oversight, lend insight into possible equity challenges in WHD oversight of agricultural employers. In particular, research on legal oversight of workplace conditions, for both the health and safety violations under

OSHA's purview and the H-2A labor regulations we observe, includes three modes of enforcement: complaint-driven, random targeting, and strategic targeting. In order to understand how we can build equity into H-2A enforcement, we must first examine how these modes of enforcement produce or ameliorate inequity.

1.3.1 Complaint-Driven Enforcement

Complaint-Driven Enforcement is when oversight agencies rely on workers and advocates to (1) perceive problems, (2) weigh the costs and benefits of reporting, and (3) navigate the reporting process. In turn, research focused on OSHA enforcement, but still relevant to H-2A, suggests that complaint-driven enforcement can disparately impact populations that face higher barriers to filing a complaint. Johnson and Grittner show that worker complaints generally lead investigators to worksites with greater hazards (Johnson and Grittner, 2020). However, the data show lower complaint rates and higher rates of injury when the share of the Hispanic population increases andd when immigration enforcement increases. As Hispanic individuals are more likely to live and/or work with undocumented immigrants, a complaint can increase exposure to deportation (Hall et al., 2019; Andersson et al., 2010).

H-2A workers can face even greater barriers to reporting issues, as their visas are tied to their employer, they are less likely to speak or understand English, and they may be less knowledgeable of their rights (Rathod, 2010). Further, H-2A workers are often recruited in friend or family groups, and workers often fear that a complaint from one worker will result in retaliation against not only them, but their entire social group. The cost of this can devastate local economies in rural areas with large quantities of H-2A workers, who send a significant portion of their income home in remittances. These qualitative barriers are associated with quantitative inequities. One study estimated that it took an average of 130 violations to elicit the filing of a single worker complaint to WHD, and public interest legal organizations note that workers will come to them with a case only after they have been fired (Weil, 2003).

1.3.2 Random Targeting for Enforcement

Random Targeting Enforcement is when the oversight agency randomly selects employers and worksites for proactive audits or inspections. This method has potential to reduce biases that arise from complaint-driven enforcement. Research examining random inspections in the OSHA case has shown both that these inspections can (1) uncover evidence of employment issues and (2) have a causal effect in improving subsequent workplace conditions (Levine et al., 2012; Haviland et al., 2012).[1]

Random targeting has two drawbacks. First is *scale relative to resources*. In the case of both OSHA and H-2A-related enforcement, the number of employers and worksites is much larger than resources available for proactive audits. Second is *potential efficiency losses*. Random targeting

[1]The main challenge in examining whether oversight improves working conditions are *sample selection biases* if oversight is based on worker/advocate reporting. More specifically, if researchers observe an empirical correlation—employers who have faced investigations or sanctions have better or worse workplace conditions after the investigation/sanction—the correlation could be spurious and driven by non-random selection into which employers get investigated. In the case of OSHA, researchers have been able to study the impact of enforcement on workplace conditions by leveraging randomly-assigned inspections, which break the link between (1) whether an employer is investigated/sanctioned and (2) underlying issues.

is most effective if all employers have an equal risk of violating labor regulations. But if there are ways to differentiate employer risk, then oversight agencies can make better use of limited enforcement resources by doing a mix of random targeting and strategic targeting of high-risk employers and/or sites (Johnson et al., 2020). This brings us to the next strategy for reducing inequities that can arise in complaint-driven enforcement: data-driven targeting.

1.3.3 Strategic Targeting for Enforcement

Strategic Targeting Enforcement requires employers to change their behavior at the market level, rather than on a case-by case basis, to prevent violations from happening in the first place (Weil, 2010). Given the challenges with both complaint-driven enforcement and random targeting, WHD itself has begun to shift to a higher rate of strategic targeting. In 2010, 75 percent of WHD investigations were triggered by complaint as opposed to "directed" investigations, or those initiated directly by WHD (Weil, 2010). By 2017, David Weil, former WHD Director, reported that directed investigations had grown to 50 percent of all investigations (Weil, 2018).

However, there remain two key challenges with strategic targeting. First: If violations are rare within a large pool of employers, what data can be used to differentiate employers based on risk for the purpose of strategic enforcement? As we outline in greater detail in Section 2, researchers have investigated patterns within enforcement, asking: Among agricultural employers with at least one WHD investigation, what predicts violations and future investigations? No research of which we are aware examines whether we can predict which of the many employers cleared for H-2A job postings will end up in this pool to begin with, which is important given that only a small proportion of employers are ever investigated.

Second are issues of *biased labels*. In particular, if we build a model to predict (1) which employers are investigated by WHD, and (2) which investigated employers have substantiated violations, we end up training a model that picks up, rather than corrects for, inequities in reporting. In other words, the model is trained on investigations and violations that are reported under the predominantly complaint-driven mode of oversight, and so the model can predict where there are complaints, rather than where the most actual issues occur. While one way to address this bias is to restrict our data sources to issues uncovered during random targeting, we can begin to investigate these biases using data from trusted messengers.

1.3.4 Reduce Inequities in Complaint-Driven or Strategic Enforcement: Local Trusted Messengers

Community-based organizations that others regard as credible sources of information, such as labor unions or legal aid organizations, can both reduce biases where the workers most in need of help are least likely to report issues and collect their own data on potential labor law violations (Johnson and Grittner, 2020). We will henceforth refer to these organizations as "trusted messengers." In particular, within the context of complaint-driven enforcement, unions, non-governmental organizations (NGOs), and other locally-based organizations play a role in reducing the barriers to filing a federal complaint or provide resources that substitute for federal oversight. Labor unions provide protections for workers to anonymously file complaints, conduct "know-your-rights" trainings, and facilitate the complaint process (Johnson and Grittner, 2020). Legal services organizations can provide alternatives to the complaint process by representing employees on matters involving unpaid wages, substandard housing, occupational injuries or illnesses, and labor trafficking, among other issues. These organizations often take a proactive approach to identifying legal problems through worker outreach, including visits to workplaces and the perimeters of housing sites.

Here, we partner with a trusted messenger for agricultural workers: Texas RioGrande Legal Aid (TRLA). TRLA is a legal services organization, with its principal office based in Texas' Rio Grande Valley. It is the second largest legal aid organization in the United States. Attorneys and staff provide free legal services to qualifying applicants.[2] Since the 1960s, TRLA has served farmworkers and workers in the logging industry in Texas. Since the 1990s, their services have extended as well to workers in six other southern states, Arkansas, Kentucky, Tennessee, Louisiana, Mississippi, and Alabama, through the Southern Migrant Legal Services (SMLS) project. Potential clients can apply through the organization's Telephone Access to Justice phone hotline, though the farmworker team primarily relies on outreach to farmworker housing and other sites where workers congregate to reach workers who may have legal issues related to their employment.

We use data from TRLA intake to investigate whether data from federal oversight (1) fully encapsulates issues uncovered from local outreach by trusted messengers or (2) can be augmented with data on issues that these local messengers collect. Within a system of complaint-driven enforcement, the trusted messenger data can be potentially be used to investigate "complaint deserts," or sites with high concentrations of employers or workers that rarely report issues to federal oversight agencies. Within a system of data-driven enforcement, the trusted messenger data could be used to correct for potential label biases, though as we show later, this requires local data on a larger scale than in the present project.

2 Gaps in Research on Equitable Enforcement within the H-2A Program

What do we know about either (1) inequities in complaint-driven enforcement or (2) the possibility of data-driven targeting within the H-2A program? The most exhaustive study of both issues is a 2020 report by the Economic Policy Institute (EPI) that analyzed WHD Compliance Action data over two decades to study enforcement within the H-2A program (Costa et al., 2020). Our study builds upon their work in three ways. First, a key goal of data-driven targeting is to (1) begin with the full population of relevant employers and (2) see whether one can prospectively predict which employers are likely to face legal scrutiny for issues. Yet the EPI report limits their analysis to employers with investigations, which is a small proportion of establishments that have job postings. In contrast, our study: (1) begins with the full universe of H-2A related employers and (2) follows those employers to see whether they have investigations or violations corresponding to a particular job posting. As we describe in greater detail in Section 5.4, linking to H-2A certificate data allows our research to include all employers with job postings, which allows us to predict investigations among all employers.

Second, when examining patterns within investigations and violations, we expand upon the set of predictors used in (Costa et al., 2020) in two ways. First, in terms of contextual predictors, Costa et al. (2020) focus on county-level variation, estimated using the zip codes in employer/worksite

[2] The Legal Services Corporation (LSC) requires applicants to have qualifying immigration status, be under 125 percent of the federal poverty level, and for cases to meet specific content guidelines.

addresses. In contrast, we geocode the full address of each employer in the H-2A certificate data (details in Section 5.3), which allows us to examine contextual characteristics at the Census Tract level rather than the county level, a more granular geography. Second, we examine whether the text of the job addendum is a meaningful predictor of issues.

Finally, Costa et al. (2020)'s report focuses on issues that come to the attention of the federal enforcement apparatus, which reflect a mix of complaint-driven enforcement and strategic enforcement. Yet a recent report from Mathematica notes challenges with solely relying on federal data sources to measure the prevalence of violations (Dolfin et al., 2020). Their report highlights the potential for false negatives, or establishments that are never investigated, but have substantial compliance issues that may violate the rights of underserved populations (Dolfin et al., 2020). Our use of "trusted messenger" data from TRLA outreach can potential reveal issues that go unreported to DOL WHD.

3 Present Report

3.1 Research Questions

Given the past research on enforcement actions and the vulnerability of labor violations within the H-2A program, it is clear equity is critical to the future of the program. We use past research to guide us and our partnership with TRLA to develop our research questions. In the present report, we investigate three questions:

1. **Descriptively, how do the employers investigated by DOL WHD compare to the employers investigated by TRLA through its outreach-based intake process?**
2. **How well can we predict (1) which employers are investigated by DOL WHD, (2) which employers in TRLA catchment states appear in TRLA intake records but no DOL WHD investigations, and (3) which employers in TRLA catchment states have DOL WHD investigations but no TRLA intake?**
3. **Does the text of job addendums lend insight into potential issues?**

3.2 Overview of data sources and analytic methods

Table 1 provides an overview of data sources and methods for each of the research questions, linking to the specific sections that have more detail on those data and methods.

Table 1: **Overview of data sources and analysis methods**

Research question	Data sources	Methods
Descriptively, how do the employers investigated by DOL WHD compare to the employers investigated by TRLA through its outreach-based intake process?	H-2A job clearance data (Section 4.1) WHD compliance action data (Section 4.2); TRLA intake records (Section 4.3); American Community Survey (ACS) data on tract demographics (Section 4.4)	Probabilistic record linkage to link certificates to WHD compliance action data and TRLA intake data (Sections 5.4 and 5.5); Geocoding to link certificates to ACS tract-level data (Section 5.3); Descriptive proportions with linked data
How well can we predict (1) which employers are investigated by DOL WHD, (2) which employers in TRLA catchment states appear in TRLA intake records but no DOL WHD investigations, and (3) which employers in TRLA catchment states have DOL WHD investigations but no TRLA intake?	Same as above	Preprocessing of data to prepare data for modeling (Section 5.8); supervised machine learning to predict nationwide issues and logistic regression for issues in TRLA catchment states (Section 5.9)
Does the text of job addendums lend insight into potential issues?	Same as above but focusing on FY 2020 and 1st quarter of FY 2021 and adding data from a FOIA request that has job addendums	Structural topic modeling to uncover themes in addendums and correlate them with outcomes (Section 5.10)

4 Data Sources

4.1 From DOL: H-2A job certificate data

The H-2A disclosure data ranged from FY 2008 to Q1 in FY 2021. This data (https://www.dol.gov/agencies/eta/foreign-labor/performance) captures the details of each H-2A submission. It includes the information about the case's progress, job's description, and employer's and attorney or agent's background. For our research, we scraped the H-2A job disclosure data from FY 2014 to FY 2021 and filtered to approved or partially approved submissions (00_scraping_DOLh2a.ipynb[3]). We found the intersecting columns across these years as well as which new columns are added for each year and rename columns that presented the same information but were named differently (02_RenameCol_Rowbind.py). Notably, the need to reconcile data across the eight years resulted in a dataset with significantly fewer application-related fields than if we had restricted analyses to FY 2020 and FY 2021, when many more fields were added to the public release dataset. Appendix Section 8.1.1 contrasts the fields available in all years/that could be used for identification and prediction of issues with the fields available in some but not

[3]Links like these link out to publicly-available code.

all years. Among the fields consistently collected across years, we focus on three types of fields:

1. **Fields that identify distinct employers, used for probabilistic record linkage to WHD Compliance Action and TRLA intake data:** We use the employer name, city, and state fields for the record linkage we describe in Sections 5.4 and 5.5. Notably, each employer is repeated across multiple approved certificates.

2. **Exact address for linking to census tract:** We use the employer's exact address for the geocoding and census linkage steps we describe in Section 5.3.

3. **Fields used to describe employers/predict issues:** These include the NAICS codes, SOC code, and attorney/agent who helped the employer prepare the application.

Finally, before the record linkage and analyses, we filter to employer-certificate dyads where the `CASE_STATUS`[4] of the certificate is either a certification or partial certification, thus excluding withdrawn or expired certificates.

4.2 From DOL: WHD Compliance Action Data

To measure the outcome of "federal oversight," we use the publicly available WHD compliance action data, which consists of all compliance actions since FY 2005.[5] For our research, we used compliance action data ranging from FY 2014 to early quarters of FY 2021. In order to focus on H-2A employers, we applied the following inclusion criteria, outlined in greater detail in 03_fuzzy_matching.R:

1. **Necessary criteria 1:** the registration act for the compliance action is either H2A, FLSA, or MSPA.

2. **Necessary criteria 2:** either the employer's 2-digit NAICS sector code is 11, indicating "Agriculture, Forestry, Fishing and Hunting" *or* the employer's full NAICS code is in the universe of NAICS codes represented in the H-2A certificate data.

The purpose of this filtering is to reduce the risk of false positive matches to employers who are investigated for issues outside of the H-2A program. In Section 5.6, we discuss how we use these data to define different outcome variables reflecting: (1) a WHD investigation ever, (2) a WHD investigation where the findings overlap with the start and end date of a particular job, and (3) a WHD investigation that reveals evidence of violations.

[4] This is one of the variables in the DOL H-2A data.

[5] From the website linked at the end of this footnote, "The dataset contains all concluded WHD compliance actions since FY 2005. The dataset includes whether any violations were found and the back wage amount, number of employees due back wages, and civil money penalties assessed." https://enforcedata.dol.gov/views/data_summary.php

4.3 TRLA Intake Data

We supplement WHD Compliance Action data with a unique dataset from TRLA's Client Management System (CMS) that identifies potential "false negatives" in the federal compliance data, as well as the equity implications of the false negatives.

TRLA's farmworker team obtains cases through a number of sources. The team conducts extensive direct outreach among farmworker populations, both in-season at worker housing and out-of-season to domestic farmworker populations in their home communities. TRLA staff routinely visit H-2A housing in-season to distribute legal information to workers. While this outreach is one source of cases, TRLA also receives cases from many other sources, including word of mouth and referrals from partner organizations (such as the National Trafficking Hotline[6]) and state and federal agencies. Eighty-two percent of TRLA farmworker cases originate from outside of the complaint hotline, including outreach, clinics, referrals, or office walk-ins.

Regardless of whether TRLA has the resources to take on a case, staff will complete a thorough intake questionnaire with clients who would otherwise qualify for services. One component of the intake questionnaire is identifying adverse parties, or opponents, in a case. In most cases, this is the employer, though it may include government agencies as well. Case types reflected in TRLA opponent data include violations of the H-2A regulations, the Migrant and Seasonal Agricultural Worker Protection Act, the Fair Labor Standards Act, and other state and federal laws that affect H-2A workers and agricultural workers more broadly.

Client and case data are collected and maintained through a Client Management System (CMS). In 2020, TRLA migrated all data to a new CMS. Data can include case status (e.g. "Open," "Closed," "Rejected"), client demographics, case data and time logs, and information about related and opposing parties. For the purposes of this research, TRLA provided us with two datasets on opponents for the farmworker team's cases, due to migration issues during the CMS transition. Opponent data dates back into the early 2000's, though we filter for opponents in cases on or after January 1, 2014.

4.4 American Community Survey (ACS) Tract-Level Data

We used Census Bureau American Community Survey (ACS) 5-year estimates at the tract level to describe contextual characteristics of the locations of employers.[7] To automatically pull large numbers of ACS predictors, we:

1. Used the Census Application Programming Interface (API) to scrape a large number (approximately $N = 155$) of variables, summarized in Appendix Figure 20 through Figure 22 (04_acs_demographics.py). These reflect the demographic characteristics of the population surrounding the employer, the economic situation of the area (which is relevant to whether employers can find American workers), and others.

[6]This organization connects victims of sex and labor trafficking with support and services, and also receives tips about potential instances of trafficking. For more information, go to https://humantraffickinghotline.org/national-hotline-overview

[7]We use employer location, rather than worksite location, because the full employer address was collected more consistently across years than the full worksite address.

2. Transformed the counts to percentages using the relevant denominator for each census table (04_acs_demographics_percentage.py).

3. Auto-named each variable using the Census codebook.

ACS data are currently available for the years 2014 to 2019. We merge these data with the job certificate data outlined in Section 4.1 using calendar year; for jobs posted in 2020 or 2021, we use the ACS 2019 values until newer estimates become available.

5 Methods

5.1 Methods to Address our Research Questions

In Section 6.2, we compare employers with (1) DOL WHD investigations, (2) DOL WHD investigations that return evidence of violations, and (3) that appear in TRLA intake records. The section shows that although both forms of enforcement identify similar *numbers* of employers as having potential issues, they seem to identify distinct sets of employers.

Section 6.3 shows that we can predict which employers go on to face WHD investigations with a relatively high degree of accuracy. However, the limited sample size for the seven TRLA catchment states (Alabama, Arkansas, Kentucky, Louisana, Mississippi, Tennessee, and Texas), and rarity of the outcomes, means that the binary classifications *do not* perform well enough to justify their use for specific predictions of appearance within the TRLA intake process or the WHD process in those states. Instead, in Section 6.4, we use the top predictors from the nationwide prediction and examine, within a standard logistic regression framework, which predict (1) higher risk of both TRLA and WHD oversight, (2) lower risk of both forms of oversight, and (3) higher risk of one form of oversight but lower risk of another.

The previous analyses focus on structured predictors (American Community Survey (ACS) contextual variables at the tract level, North American Industry Classification System (NAICS) and Standard Occupational Classification (SOC) codes) of local and federal oversight. We show that local oversight can be a complement to federal oversight, uncovering issues in a distinct set of employers. To what extent are these issues foreshadowed by details of the employment contract? We use computational text analysis to show some patterns where employment contracts emphasizing certain themes (e.g., criminal history restrictions) are correlated with employers flagged in local outreach.

5.2 Two Units of Analysis: Employer-Certificate Dyads versus Unique Employers

For the methods that follow, we use shorthand to refer to two distinct units of analyses:

1. **Employer-certificate dyads:** The original unit of analysis for the universe of possible entities subject to oversight are rows in the H-2A certificate data, which we refer to as employer-certificate dyads. The same employer can have multiple approved or partially approved certificates over time. Similarly, due to the lack of a unique employer identifier across datasets

(e.g., Employer EIN), different spellings can identify the same employer across their multiple certificates. Therefore, when we refer to jobs we refer to job postings that can be repeated within an employer. When we refer to unique employers or employers, this is a single employer that we believe to be the same even if that employer appears in multiple certificates.

2. Unique employers: these refer to particular employers across all of their certificates. One restriction we impose is that, since names may not be uniquely identifying, we only classify an employer as "the same" if their state stays constant across multiple certificates.

To make this more concrete, Table 2 provides an example of one employer with several approved certificates in the period under study. For the descriptive analyses, we often aggregate up the employer level, and explore whether there are any investigations or violations across any certificate that overlaps with relevant dates in the compliance action date (Section 5.6). For the models, we presently predict issues at the employer-certificate dyad level since, across an employer's multiple certificates, values of the predictors can vary (e.g., changing contextual conditions; different job codes). Section 7.1.1 discusses how results might change if we predict at the employer level.

Table 2: Example of a single employer with several employer-certificate dyads These are rows from the H-2A job certificate data mentioned in Section 4.1.

Employer Name	City	State	Case #	Case Status	Date (start)	Date (end)
Mark Anthony Sadler	Cynthiana	KY	H-300-14091-587932	DETERMINATION ISSUED - CERTIFICATION	2014-06-15	2015-02-15
Mark Anthony Sadler	Cynthiana	KY	H-300-15091-928544	DETERMINATION ISSUED - CERTIFICATION	2015-04-01	2015-10-15
MARK ANTHONY SADLER	Cynthiana	KY	H-300-16050-758310	DETERMINATION ISSUED - CERTIFICATION	2016-02-17	2016-09-16
MARK ANTHONY SADLER	Cynthiana	KY	H-300-16116-648682	DETERMINATION ISSUED - CERTIFICATION	2016-04-25	2016-10-16
MARK ANTHONY SADLER	Cynthiana	KY	H-300-17059-447088	DETERMINATION ISSUED - CERTIFICATION	2017-02-27	2017-08-31
MARK ANTHONY SADLER	Cynthiana	KY	H-300-17125-155620	DETERMINATION ISSUED - CERTIFICATION	2017-05-03	2017-10-08
MARK ANTHONY SADLER	Cynthiana	KY	H-300-18059-391201	DETERMINATION ISSUED - CERTIFICATION	2018-02-27	2018-09-04
MARK ANTHONY SADLER	Cynthiana	KY	H-300-18143-635430	DETERMINATION ISSUED - CERTIFICATION	2018-05-22	2018-10-16
MARK ANTHONY SADLER	Cynthiana	KY	H-300-19023-314454	DETERMINATION ISSUED - CERTIFICATION	2019-01-22	2019-09-03
MARK ANTHONY SADLER	Cynthiana	KY	H-300-19136-949956	DETERMINATION ISSUED - CERTIFICATION	2019-05-13	2019-10-26
Mark Anthony Sadler	Cynthiana	KY	H-300-20140-585083	Determination Issued - Certification	2020-08-01	2021-02-01
Mark Anthony Sadler	Cynthiana	KY	H-300-20048-328346	Determination Issued - Certification	2020-05-01	2021-02-01

5.3 Geocoding Locations of Employers and Merging with American Community Survey Data

Using the exact addresses in the employer-certificate dyad data, we geocode the address of each employer in the certificate data using the `Geocodio` API in the following script (05_geocode_jobs.py).

We then overlay these geocoded latitude and longitudes with Census tract shapefiles to match each certificate to a corresponding census tract (06_h2a_tract_intersections.ipynb).

Of note is that if the employer address varies within an employer's multiple certificates, the same employer may have approved certificates overlapping with multiple Census tracts.

5.4 Fuzzy Matching between Jobs and WHD Investigations

In order to determine which job postings had WHD investigations (or violations) and which did not, we had to match the H-2A job clearance data onto the WHD Compliance Action Data. Given that the rate of false negatives would likely be very high had we only matched (1) an employer-certificate dyad to (2) a compliance action record if the employer name and address matched exactly, we decided to implement a probabilistic, or "fuzzy", matching algorithm[8] using the fastLink package (Enamorado et al., 2020).

After filtering the employer-certificate dyad data to approved or partially approved certificates and filtering the compliance action data based on the logic described in Section 4.2, we cleaned the employer names to filter out issues like extra spaces and variations of INC and LLC, among other steps (03_fuzzy_matching.R).

Then since, as shown in Table 2, employers are repeated across certificates, we assigned a group id value to each observation, which was repeated across instances of the same entity (using the probabilistic deduplication functionality of fastLink to account for spelling differences). After ensuring that all jobs and WHD investigations within the same group id were from the same state, we went ahead with the fuzzy matching process.

To minimize the likelihood of false positive matches, we blocked by state, or restricted matches to cases where: (1) the employer in the certificate data was located in the same state as (2) the employer in the compliance action data. Within a state, the algorithm matched fuzzily on employer name and city. The result was data containing all relevant H-2A job postings, matched with all of their WHD investigations when applicable. Section 5.6 describes how, within these matches, we narrowed down to investigations and findings of violations where the dates within a focal certificate aligned with the dates related to the investigation.

5.5 Fuzzy Matching between Jobs and TRLA Intake Records

We used a similar process to match the H-2A job clearance data to the TRLA intake data, focusing on the seven catchment states covered by the latter dataset (Alabama, Arkansas, Kentucky, Louisana, Mississippi, Tennessee, and Texas). While the process was nearly identical to the one outlined above, we include the script with the exact steps that this fuzzy matching entailed (09_fuzzy_matching_TRLA.R).

[8] A fuzzy matching algorithm helps join together two datasets when there might be discrepancies between "matches" due to errors in the data. Such an algorithm bases matching on probabilities rather than exact matching.

Figure 1: **Defining investigations and substantiated violations that overlap with a focal certificate**

```r
matched_data_WHD <- matched_data_WHD %>%
  mutate(outcome_is_any_investigation = is_matched_investigations,
         outcome_is_investigation_aftersd = case_when(is_matched_investigations &
                                                      findings_start_date >= JOB_START_DATE ~ TRUE,
                                                      TRUE ~ FALSE),
         outcome_is_investigation_before_sd = case_when(is_matched_investigations &
                                                      findings_start_date < JOB_START_DATE~ TRUE,
                                                      TRUE ~ FALSE),
         outcome_is_investigation_overlapsd = case_when(is_matched_investigations &
                                                      findings_start_date >= JOB_START_DATE &
                                                      findings_start_date <= JOB_END_DATE ~ TRUE,
                                                      TRUE ~ FALSE),
         outcome_is_viol_overlapsd = case_when(is_matched_investigations &
                                                      findings_start_date >= JOB_START_DATE &
                                                      findings_start_date <= JOB_END_DATE & h2a_violtn_cnt > 0 ~  TRUE,
                                                      TRUE ~ FALSE),
         outcome_is_h2aflsamspaviol_overlapsd = case_when(is_matched_investigations &
                                                      findings_start_date >= JOB_START_DATE &
                                                      findings_start_date <= JOB_END_DATE &
                                                      (h2a_violtn_cnt > 0 | flsa_violtn_cnt > 0 |
                                                      mspa_violtn_cnt) ~  TRUE,
                                                      TRUE ~ FALSE))
```

5.6 Defining the Outcomes

5.6.1 Federal Investigations or Violations

The fuzzy matching tells us, within an employer, whether there are any matches among their certificates to the compliance action data. But important for both descriptive analysis and predictive modeling is ensuring that the investigation *temporally aligns* with a particular job, so that we are not, for instance, using characteristics of a 2019 certificate to predict an investigation of that employer in 2015. Therefore, Figure 1 (an excerpt from 10_construct_outcomes.R) defines a few increasingly narrow definitions of an investigation.

First, and clearly incorrect unless all analyses are aggregated to the employer level, an investigation can be defined by whether that employer had any investigation across any of their certificates. Second, and potentially useful for some purposes, is to define an investigation by whether, for that employer-certificate dyad, there was an investigation where the findings start date occurred on or after the start date of the job. Third, and what we focus on, is a stringent definition where the employer-certificate dyad not only matched to the compliance data, but also: (1) the job start date was on or after the findings start date and (2) the job start date was before or on the findings end date.

For violations, we use this stringent definition of an investigation, and examine for the main outcome whether there was a non-zero count of H-2A violations. The possible violations related to the H-2A statute include wage theft remedied by backwages, violations of housing or transportation safety provisions remedied by civil money penalties, and others outlined at this link: https://www.dol.gov/agencies/whd/agriculture/h2a. For the purposes of this report, we group different types of violations together and predict the presence of any violation. This is a binary variable where 1 represents the presence of any violation and 0 represents no violations. Including FLSA and MSPA only slightly increases the rate of employers with any violations due to the initial registration and NAICS code filtering described in Section 4.2, so our analyses focus on H-2A violations.

5.7 Local TRLA Intake Records

We use similar filtering criteria to construct the "yes TRLA intake" outcome variable. While the federal compliance action data has a findings start date and end date we use for filtering, the local intake records only record an intake date for the issue. We therefore define the outcome as having an intake date on or after the job start date. As such, the filtering criteria and process are the same as earlier, barring the difference in number of timestamps.

5.8 Preprocessing Predictors to Prepare for Models

The previous steps results in a dataset with:

1. **Binary labels:** These are the investigation (local and federal) and violation-related outcomes described in Section 5.6.
2. **Features/predictors:** These are a combination of features from two sources:
   (a) H-2A certificate data (e.g., attorney/agent; SOC code)
   (b) ACS tract-level contextual characteristics (e.g., race/ethnicity breakdowns in the surrounding area; poverty and unemployment)

Preprocessing involves (1) turning categorical predictors into binary numeric representations, and (2) imputing missing values in the predictors (12_mlmodeling_preprocessing.py).

Broadly, the preprocessing went through the following steps:

1. **Separating out features from non-features:** The above data merges between H-2A employer-certificate dyads and WHD compliance action records resulted in many predictors that are *post-treatment*, or that are invalid predictors because they are conditional on there being an investigation of an employer. These include both outcomes of the investigation (CMPs; backwages) and information recorded about the employer in the investigative process. As a result, this step limited our predictors to ACS features and information known about the employer from their certificate applications.
2. **Collapse many-category variables into top category versus other:** Categorical features like the employer's city and the attorney/agent who prepared their application have thousand of distinct values, most of which are likely uninformative for prediction because they are shared by very few observations. We coded levels of categories that appear in fewer than 1 percent of observations to *other*. For example, some city names only appears once in column ATTORNEY_AGENT_CITY and and so they would be replaced by the dummy value "city_other." In this way, we are able to narrow down the number of unique values in categorical features from thousands to less than 100 to reduce the final dimensionality of the feature matrix.
3. **Generate 80% train-20% test split:** We then use GroupShuffler() from Scikit-learn[9] to generate the train test split. We generated the splits so that each employer was *either* in the

[9]This is a functionality used within Python. The interested reader can learn more at https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GroupShuffleSplit.html.

training set *or* the test set, similar to the analysis discussion in Section 5.2, although the modeling unit of analysis was employer-certificate dyads in that case. This helps ensure that when we use the test set to validate the model, the test set is comprised of *employers* the model has not yet seen in the training/estimation set.

4. **Imputation:** Data is imputed after the train/test split to prevent against "leakage" of information from the training set into the test set. We impute numeric feature columns using the mean in the sample and for categorical variables, impute by adding "missing value" as a distinct level.

5.9 Supervised Machine Learning

The preprocessing results in a high-dimensional matrix with 696 predictors where the number of predictors begins to approach the number of rows/observations. Therefore, we turn to flexible, binary classifiers that can handle this high dimensionality and prevent overfitting.

We use these classifiers to predict the different labels: (1) any WHD investigation (nationwide), (2) WHD investigation but no TRLA intake record (seven catchment states), and (3) TRLA intake record but no WHD investigation (seven catchment states).

Table 3 summarizes the models and provides our shorthand abbreviations. With each model, we generated predicted probabilities in the 20% held out test set that we used to evaluate model accuracy.

While the supervised binary classifiers are important for cases where we have a large number of predictors, we also use logistic regression for prediction within the TRLA catchment states, where we use a two-step process: (1) use the nationwide model to filter to highly-relevant predictors and (2) use logistic regression with those highly-relevant predictors.

Table 3: Binary classifiers we compare

Model family	Parameters	Abbreviation	Model details	Parameter definitions
Decision tree	DecisionTreeClassifier (random_state=0, max_depth = 5),	dt_shallow	A decision tree is a classifier that predicts an outcome and where a series of nodes—e.g., % White above or below 50%; NAICS code 111140 or not—eventually ends in a leaf where the outcome is either predicted 1 or predicted 0.	max depth is the maximum depth of the tree; random state controls the ran-domness of the estimator
Decision tree	DecisionTreeClassifier (random_state=0, max_depth = 50),	dt_deep	see above	see above
Random forest	RandomForestClassifier (n_estimators = 100, max_depth = 20),	random_forest_shallow	Consists of a large number of individual decision trees that operate as an ensemble; Each individual tree in the random forest spits out a class prediction and the class/tree with the most votes becomes our model's prediction.	n estimators is the number of trees in the forest, max depth is the maximum depth of the tree.
Random forest	RandomForestClassifier (n_estimators = 1000, max_depth = 20),	random_forest_deep	see above	see above
Gradient boosting	GradientBoostingClassifier (criterion = 'friedman_mse', n_estimators=100),	gb_shallow	Used for classification; it uses an additive model on weak learners (usually regression trees) to fit on neg-ative gradient of the model's loss function, which al-lows for the optimization of arbitrary differentiable loss functions.	n estimators is the number of boosting stages to perform, criterion is the function to measure the quality of a split, 'friedman_mse' is the mean squared error with improve- ment score by Friedman
Gradient boosting	GradientBoostingClassifier (criterion = 'friedman mse', n_estimators=200),	gb_deep	see above	see above
Adaboost	AdaBoostClassifier (),	ada	This is a meta-estimator that begins by fitting a clas-sifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are ad-justed such that subsequent classifiers focus more on difficult cases.	none (default)
Lasso	LogisticRegression (penalty="l1", max_iter=10000, C= 0.01, solver='liblinear')	Lasso	Logistic regression is a classification algorithm used to assign observations to a discrete set of classes. It gives us a set of outputs or classes based on prob-ability when we pass the inputs through a prediction function and returns a probability score between 0 and 1.	"l1" means adding a L1 penalty term, max_iter is the maximum number of iterations taken for the solvers to converge, C is inverse of regularization strength, smaller values specify stronger regular-ization, solver is the algorithm to use in the optimization problem
Ridge	LogisticRegression (penalty = "l2",max_iter=10000, C = 0.01	Ridge	Ridge regression is a model tuning method on logis-tic regression that is used to analyse any data that suffers from multicollinearity; this method performs L2 regularization, and by changing the penalty term in the cost function, the magnitude of coefficients is reduced.	see above

5.10 Text Analysis of Job Order Addendums

Finally, and focusing on FY 2020-2021 and the seven TRLA catchment states, we used computational text analysis to (1) reduce the dimensionality of the unstructured addendums by using topic modeling to inductively learn underlying themes and (2) explore the relationship between each theme and whether there is a TRLA intake record for that observation.

Appendix Table 9 shows the count and proportion of employer-certificate dyads in the TRLA catchment states that have an addendum versus not. Due to the very low count of job orders in catchment states with WHD investigations (5 addendums), we focus on contrasting:

- Cases with TRLA intake records ($N = 38$ with addendums, or 81% of the total intake records in 2020 or 2021)
- Cases with neither a TRLA intake record nor a WHD investigation ($N = 3,756$ with addendums, or 52% of the total cases)

To computationally analyze the addendum text, we go through the following step (outlined in 20_textanalysis_addendums.Rmd).

- **Preprocess the texts:** Following standard practices in text analysis (Denny and Spirling, 2018), we: (1) Convert words to lowercase, (2) remove both standard "stopwords" (the; and; etc.) and a custom set of contract-specific stopwords,[10] (3) remove punctuation and numbers, and (4) implement "stemming," or reduce the words to a common root (e.g., criminal and crime each become "crim"; violate, violator, violations each become "violat"). The purpose of stemming is to reduce trivial differences between words but retain core semantic differences. The result is a vocabulary with 5,619 unique words. Figure 2 presents a random sample of 15 vocabulary words following the preprocessing, which also illustrates some processing challenges such as words combined in the original addendum text that are difficult to separate due to missing spaces and other delimiters (e.g., `theperiod`).
- **Estimate a topic model with k = 10 topics:** In order to reduce the dimensionality of those 5,000+ words, we use a structural topic model (Roberts et al., 2019), which (1) inductively learns clusters or themes in text based on word co-occurrence and (2) correlates those themes with document-level metadata—in this case, we focus on whether there is a TRLA intake record or not.

[10] These were: after; before; employer; employ; job; although; provide; complete; hour; time; begin; list; require; transportation; workers; worker; work; workday; working; workrelated; workplace; worked.

Figure 2: Random sample of addendum words after preprocessing

- sawdust
- seedvarieti
- theperiod
- probe
- generic
- play
- mountain
- potput
- student
- combinesharvest
- pack
- harass
- telephoneii
- show
- neglig

6 Results

6.1 Sample

Before moving to the results, we outline the sample size—both in terms of employer-certificate dyads and unique employers—used in various analyses. The sample represents either all entities from FY 2014-FY 2021 (nationwide sample) or all entities located in one of the seven TRLA catchment states: Alabama, Arkansas, Kentucky, Louisana, Mississippi, Tennessee, and Texas. We see lower sample sizes as we move to studying intake records and WHD investigations in those catchment states, which is likely related to the binary prediction challenges discussed in Section 6.3.3.

- **Nationwide prediction:**
  - 77,759 employer-certificate dyads in *training set* representing 13,838 distinct employers

    * 926 observations, or 1.2% of the sample, have any WHD investigation
    * 755 observations, or 1% of the sample, have at least 1 H-2A-related violation
  - 17,410 employer-certificate dyads in *test set*[11] representing 3,460 distinct employers
    * 206 observations, or 1.2% of the sample, have any WHD investigation
    * 177 observations, or 1% of the sample, have at least 1 H-2A-related violation

- **TRLA catchment states:**
  - 26,393 employer-certificate dyads representing 5,386 unique employers
  - For the modeling, we exclude 32 employer-certificate dyads from 9 employers where there is both a TRLA intake record and a WHD investigation (so that the contrast is between TRLA record and no WHD investigation versus WHD investigation and no TRLA record).
  - 21,407 employer-certificate dyads representing 4,306 employers
    * 861 observations, or 4.1% of the sample, had a *TRLA intake record*
    * 141 observations, or 1.0% of the sample, had a *WHD investigation*
  - 5,314 employer-certificate dyads representing 1,078 unique employers
    * 237 observations, or 4.4% of the sample, had a *TRLA intake record*
    * 44 observations, or 1.0% of the sample, had a *WHD investigation*

6.2 Descriptive Results

6.2.1 Trends Over Time in WHD Investigations

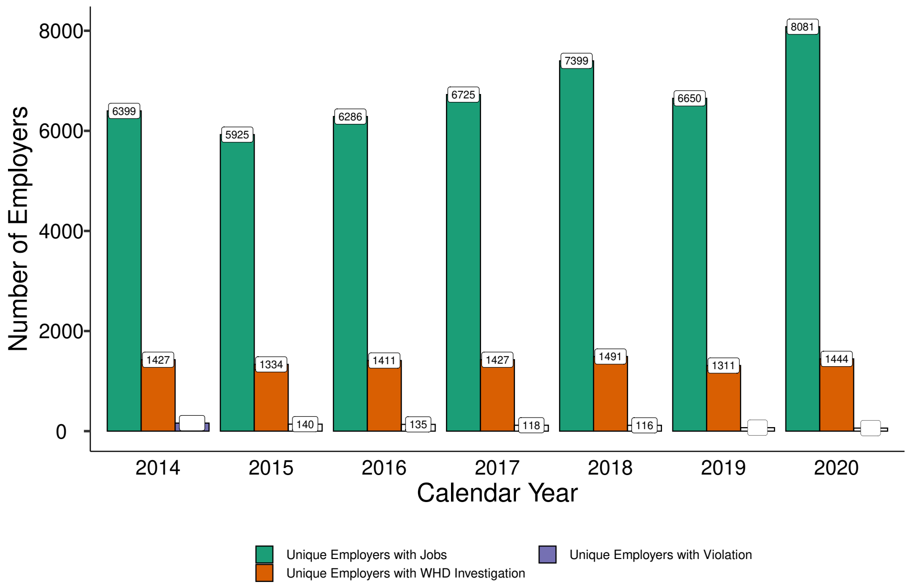
Figure 3 summarizes trends in DOL WHD investigations and findings of H-2A related violations over time (2014-2020). The green bar shows the number of unique employers, which ranges between 6,000 and 8,000. The orange bar shows the number with any WHD investigations, which ranges from 1 to 2.5% depending on the year, a low prevalence that stems from our strict definition that requires an investigation's timing to overlap with the timing of a specific job.[12] Finally, the purple bar shows the number of employers where the WHD investigation reveals a non-zero count of violations, which averages in the 0.5-2% range. Appendix Figure 23 shows the same trends at the level of a *employer-dyad certificate* rather than *employer*, which show higher counts of each reflecting multiple certificates per employer but similar proportions of WHD investigations and violations.

How do these trends of employers facing *federal investigations* compare to trends for TRLA intake records? Focusing on the seven TRLA catchment states, Figure 4 presents the count of employers in four categories: the employer faces no WHD investigations and has had no TRLA intake record; the employer has had a WHD investigation but not a TRLA intake record; the

[11] By *test set*, we mean a random sample of the data that we do not use to estimate the model/predict the label. Instead, we use the test set to validate the accuracy of the model. This prevents us from choosing a model that overfits to noise in the data.

[12] If we relax the timing requirement, the percent of employers with a WHD investigation in a given year jumps from the 1-2.5% range to the 17-22% range.

Figure 3: **Number of unique employers with jobs posted, WHD investigations, or violations***
**Violations are determined using the overlap version of the outcome variable.*



employer has had a TRLA intake record but not a WHD investigation; and the employer faced both a WHD investigation and had a TRLA intake record. Figure 5 presents these as percentages to account for the low counts of investigations relative to overall employers. Overall, the two figures highlight that the two enforcement apparatuses—federal investigations (which reflect a mix of complaint-driven and agency-initiated) and TRLA intake records (which largely derive from direct outreach workers to workers)—capture *different employers* in their enforcement purview, as can be seen by non-zero counts for employers with a WHD investigation but not a TRLA intake record, and employers with a TRLA intake record but not a WHD investigation. In line with our initial motivation, this suggest the potential for local data sources to fill in potential gaps in which employers come under the radar of federal oversight.

Figure 4: **Comparing employers focused on by DOL WHD vs. TRLA (Counts)**

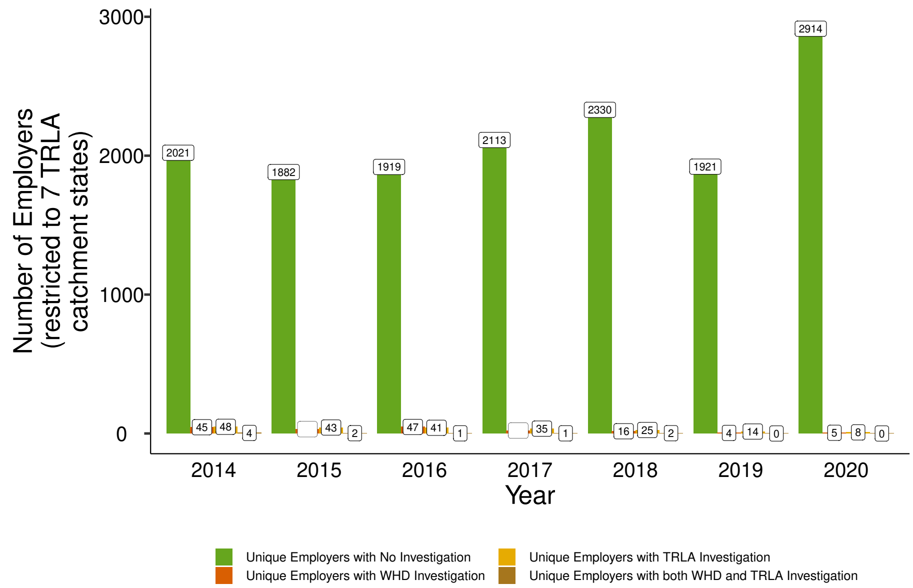
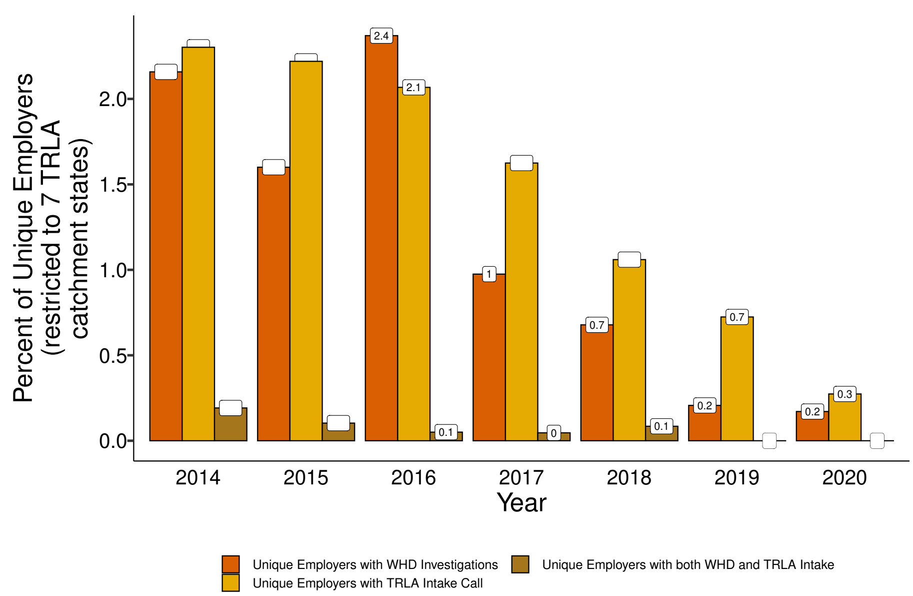


Figure 5: **Comparing employers focused on by DOL WHD vs. TRLA (Percentages)** *The number of unique employers by year (denominators) are 2,085 (in 2014), 1,937 (in 2015), 1,983 (in 2016), 2,154 (in 2017), 2,359 (in 2018), 2,933 (in 2019), and 2,923 (in 2020).*



6.2.2 How Did COVID-19 Impact these Data Sources?

Beginning in March of 2020, COVID-19 might have impacted oversight, especially the outreach-based activities of the TRLA intake process, which are obscured in counts aggregated to the year level. Figure 6, focusing on WHD investigations in the seven catchment states and aggregating counts of unique employers to the month rather than the year level, shows that although there are seasonal patterns in the numbers of jobs posted, we do not see a sharp decline in jobs posted or investigations during the COVID-19 period (red shaded area). In contrast, Figure 7 shows some declines during COVID-19 in TRLA outreach, with some months having no new intakes.

Figure 6: **Monthly patterns pre- and during COVID-19: WHD investigations aggregated to the employer level** *The shaded areas represent time after the beginning of the COVID-19 pandemic*

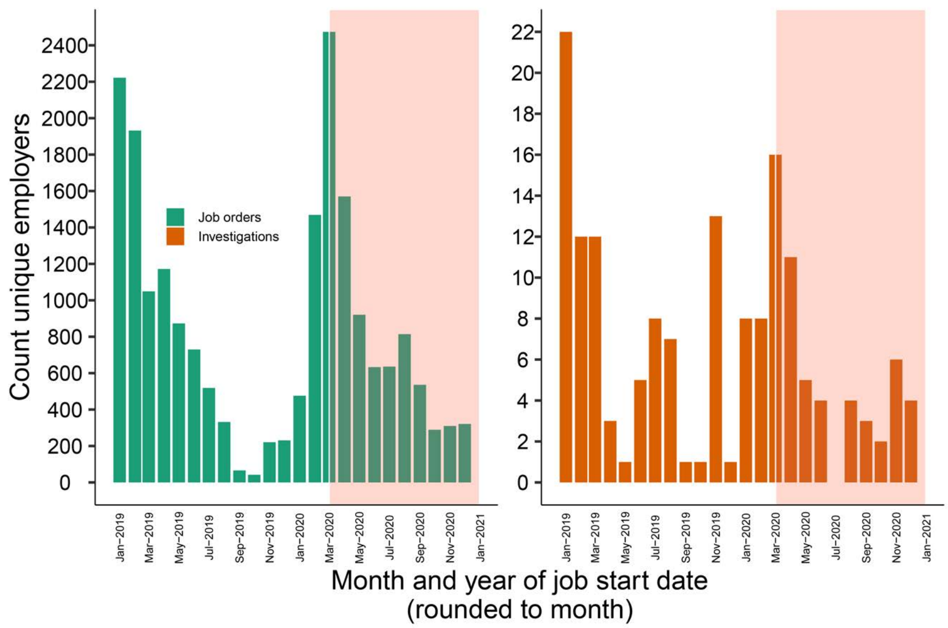
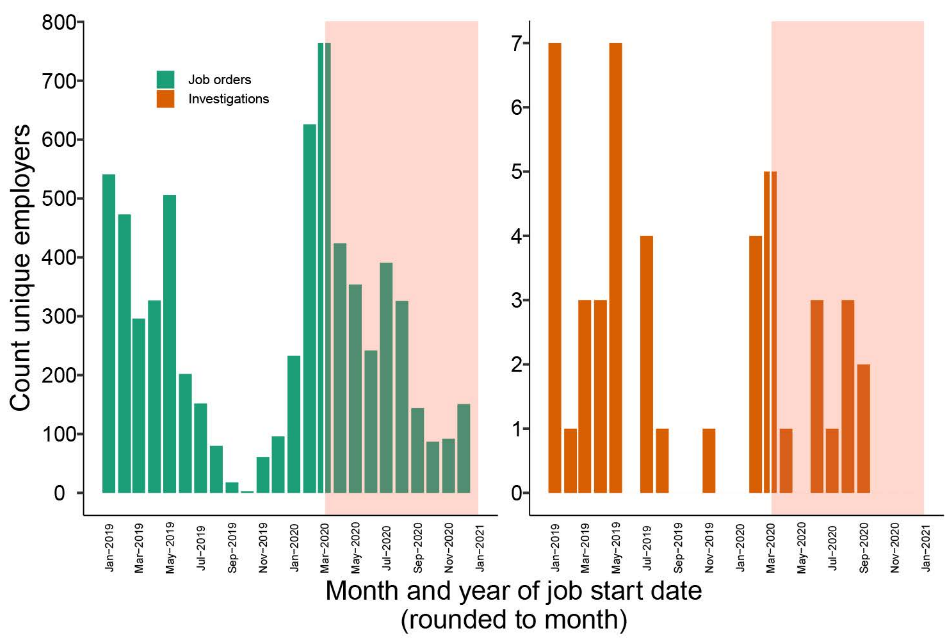


Figure 7: **Monthly patterns pre- and during COVID-19: TRLA intake records aggregated to the employer level** *The shaded areas represent time after the beginning of the COVID-19 pandemic*



6.2.3 What Distinguishes Employers who End Up with TRLA Investigations or TRLA Intake Records?

While Figure 5 showed that TRLA intake records capture a distinct set of employers relative to WHD investigations, important to examine is whether the *general types* of employers that come under the purview of each form of oversight are similar or different. Here, we focus on a few key factors from the certificate data, with Section 6.4 using a model-based comparison of ACS contextual factors.

Attorney/agent who prepares an employer's H-2A job clearance order

Attorneys play an important role in helping agricultural employers prepare applications. In turn, there are economies of scale where particular attorneys help prepare applications for many employers. Due to the need to reconcile fields across years, this analysis focuses on individual attorney/agents since the law firm they are employed by is a more recent field. Future analyses could focus on firms rather than attorneys.

While 324 of the about 770 unique attorney agents in the nationwide data are only linked to one employer, 28% of these attorney agents represent 10 or more employers. Therefore, there may be patterns of attorney/agents who tend to represent employers who later face WHD investigations, TRLA oversight, or both.

Focusing on the first, Figure 8 shows, among attorney/agents representing at least 10 employers, the fifteen attorneys whose employers had the highest rates of WHD investigations among employers they represented, with some having one in five employers they represented later facing an investigation. Other attorney/agents representing similar numbers of employers had no employers they prepared applications for facing WHD investigations. The labels on the bars reflect the total number of employers represented, with those representing more employers having less uncertainty in this proportion.

Figure 9, which focuses on the seven TRLA catchment states, shows that although TRLA intake records and WHD investigations focused on distinct sets of *employers*, there were attorney/agents representing these employers who had both a WHD investigation and a TRLA intake record. The figure shows the rate of WHD investigations and TRLA intake calls for attorney/agents represented in each dataset.

Finally, moving from investigations to confirmed violations, Figure 10 depicts: Of the employers investigated by WHD, how many had 1 or more confirmed H-2A-related violations? The data is filtered for attorney/agents in TRLA catchment states with at least 2 employers investigated. The figure shows that most of the high-violation attorney/agents (conditional on their employers being investigated) had one or more TRLA records calls about issues. This shows that although TRLA outreach captures a distinct set of employers compared to federal oversight, these employers may share common legal representation and other characteristics.

Figure 8: **Attorney/agents represent employers with high rates of WHD investigations (Nationwide)** *The number of unique employers represented by each attorney/agent (denominators) are shown directly on each bar in the plot*

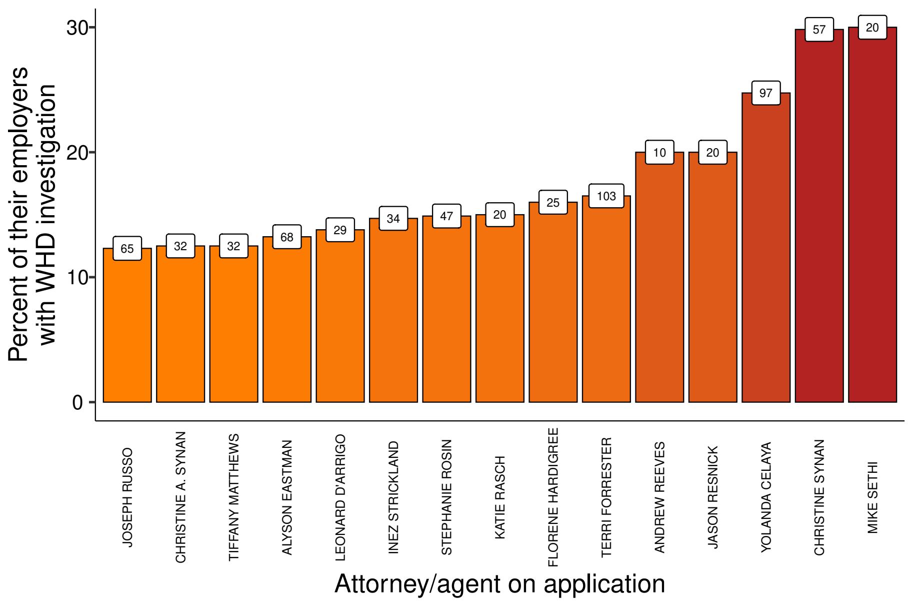


Figure 9: **Attorney/agents representing employers with high rates of investigations (TRLA catchment states and employers in WHD investigations and TRLA intake records)** *The number of employers represented by Robert Kershaw is 16, by Steve Mckay is 11, by Gloria Roriguez is 12, by Mayra Ballard is 46, by Christi Tabaretti is 77, by Ginny Muilenburg is 31, by Melissa Green is 245, and by Patricial Hall is 206.*

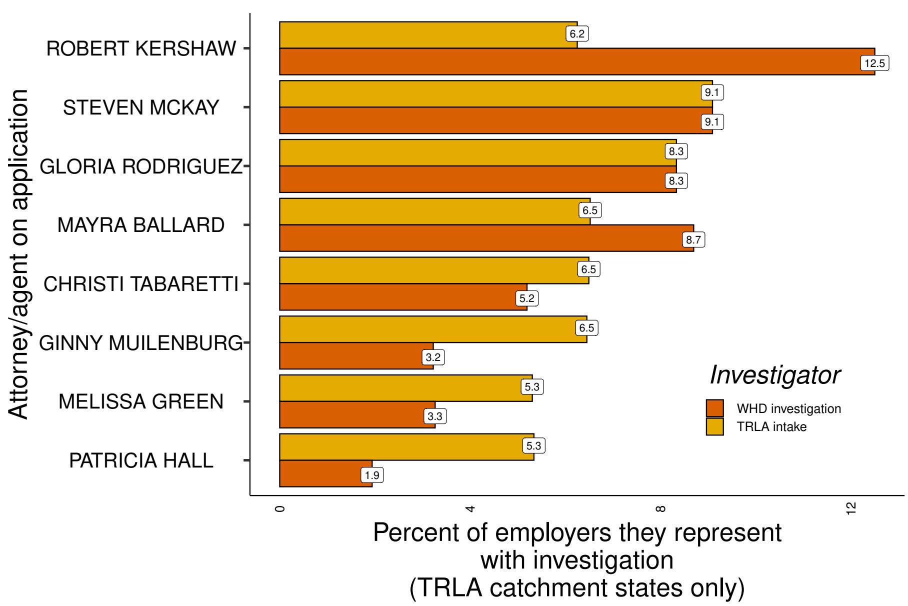
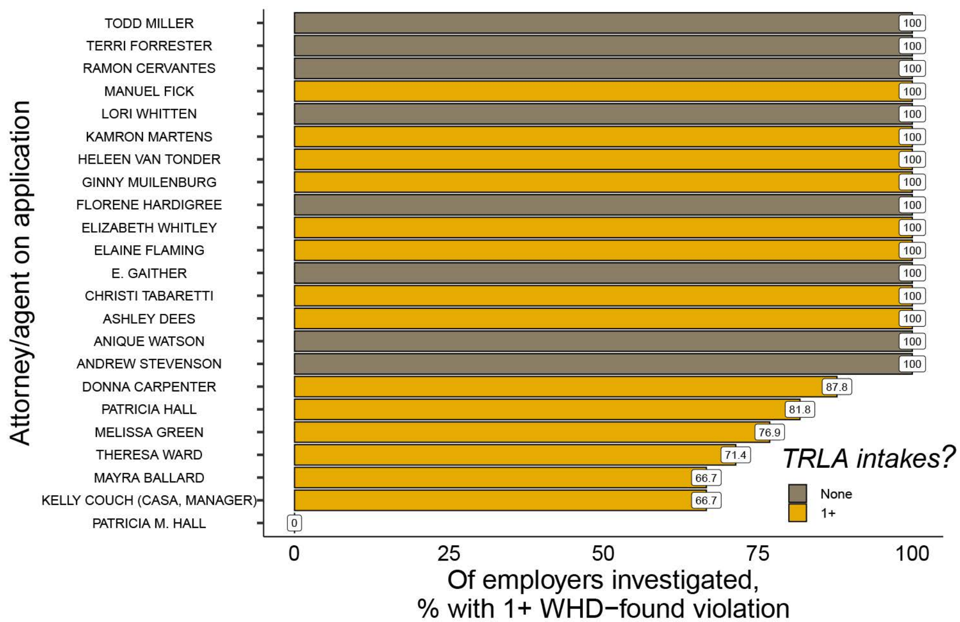


Figure 10: **Overlap between attorney/agents whose employers had confirmed WHD-found violations and TRLA intake records** *To clarify, the "None" category represents attorney/agents where 100% of the attorney/agent's employers had a WHD violation but where there were no H-2A investigations related to that attorney/agent. We can see from this figure that many high-WHD attorney/agents also had a TRLA intake call. The number of investigated employers for Todd Miller is 4, for Terri Forrester is 8, for Ramon Cervantes is 2, for Manuel Fick is 3, for Lori Whitten is 2, for Kamron Martens is 3, for Heleen Van Tonder is 2, for Ginny Muilenburg is 2, for Florence Hardigree is 4, for Elizabeth Whitley is 2, for Elaine Flaming is 4, for E. Gaither is 3, for Christi Tabaretti is 5, for Ashley Dees is 3, for Anique Watson is 2, for Andrew Stevenson is 2, for Donna Carpenter is 41, for Patricia Hall is 11, for Melissa Green is 13, for Theresa Ward is 7, for Mayra Ballard is 3, for Kelly Couch (Casa, Manager) is 3, and for Patricia M. Hall is 2*

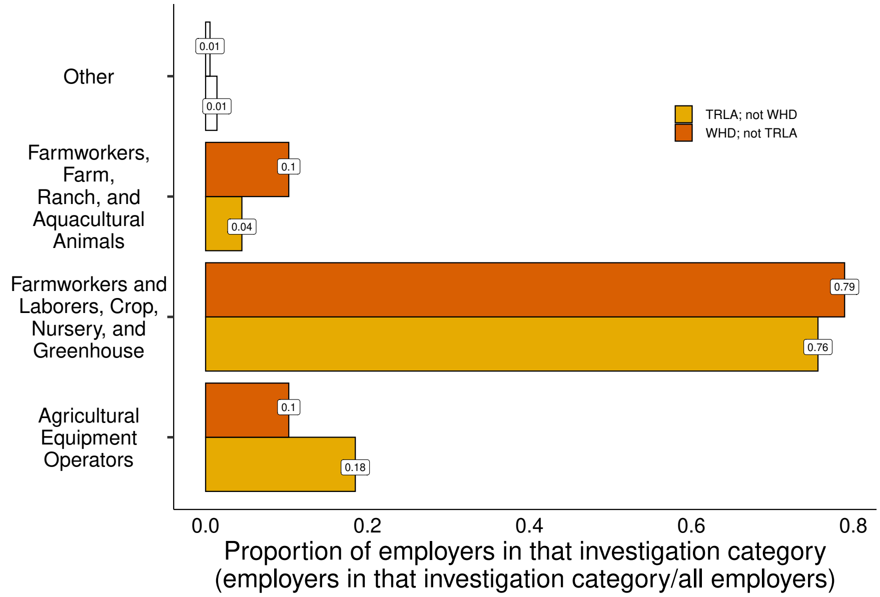


Job characteristics

The previous section focused on attorney/agents who, in the application process, represent employers who go on to have high rates of investigations or violations. Here, we focus on whether there is variation in the types of agricultural jobs among employers who face WHD investigations versus TRLA intake records.

Figure 11 focuses on `SOC` occupational codes used by at least 1% of employers in TRLA catchment states, with Appendix Table 8 containing the full set of codes. We see that although the two types of oversight focus on generally similar types of employers, the TRLA intake records reflect a slightly higher proportion (0.18 vs. 0.1) of agricultural equipment operators than WHD investigations.

Figure 11: **SOC occupational codes in WHD investigations versus TRLA intake records** *The denominators are 185 (WHD; not TRLA) and 357 (TRLA; not WHD).*



6.3 Nationwide View: Predicting DOL WHD Investigations

We now move from descriptive comparisons to investigating factors predictive of federal oversight, pooling data from all states across 2014-2021. As outlined in Table 1, these predictive models pool three data sources: (1) data from the job certificates to form the analytic universe of jobs and employers, and that contain features like attorney/agent and occupational codes, (2) data from the American Community Survey (ACS) on demographics of the area surrounding employers, and (3) data from the WHD compliance action database on employers that face investigations.

6.3.1 Accuracy of Predictive Models

After estimating each of the machine learning models outlined in Section 5.9, we generate predictions in the held-out test set that contains employers fully distinct from the training/estimation set employers. In particular, we separate employers into two categories. First are employers whose jobs and associated investigation outcomes are used to build the predictive model (training/estimation set employers). Second are employers whose jobs and associated investigation outcomes are used to validate the model's accuracy (test/validation set employers). We split employers randomly into one of two categories, which is common practice in machine learning to prevent models overfit to idiosyncrasies in the data. We (1) rank these test set observations from the highest to lowest predicted probability of having an investigation, (2) in line with but slightly higher than the about 1.5 % rate of investigations, coded the top 5% as "predicted yes investigation" and the remaining 95% as "predicted no investigation." We then categorized each test set observation into one of four categories:

1. **True negative (TN):** no WHD investigation for that job and we correctly predict no investigation

2. **False negative (FN):** WHD investigation for that job and we incorrectly predict no investigation

3. **True positive (TP):** WHD investigation for that job and we correctly predict investigation

4. **False positive (FP):** no WHD investigation for that job but we incorrectly predict investigation

Figure 24 shows the results. In line with the high threshold for classifying an observation as "predicted yes investigation," we see that most test set observations are correctly classified as true negatives. More informative for distinguishing between models are: (1) the false negative rate (with a lower false negative rate helping produce higher recall, or capturing all issues) and (2) the false positive rate (with a lower false positive rate helping precision, or correctly ruling out non-issues). Given the present policy context—an investigative process that focuses on *potential* issues, and only sanctions employers if these issues are substantiated—we focus on minimizing false negatives. The two best models for doing so are `gradient-boosting` (a shallower form with 100 estimators) and `LASSO` (a logistic regression that penalizes many coefficients to have zero input).[13] Gradient boosting is an *ensemble classifier* that takes a series of shallow decision trees ("weak learners") and successively upweights observations poorly predicted in the previous round. Lasso is a form of penalized/regularized logistic regression that shrinks the coefficients on many of the > 600 predictors to zero.

Focusing on these two models, the next two figures compare: (1) the distribution of predicted probabilities of any investigation to (2) the actual investigation status. Figure 12 shows the separation for gradient boosting; Figure 13 shows the results for `LASSO`. In line with the accuracy summaries in Figure 24, we see clearly separated peaks where those not investigated have peaks closer to zero and those investigated have peaks closer to one. However, in line with the false positives noted in that graph, we see false positives, or those not investigated that nevertheless have high predicted probabilities.

[13]For more on `gradient-boosting`, see https://stackabuse.com/gradient-boosting-classifiers-in-python-with-scikit-learn/, and for more on `LASSO`, see https://www.publichealth.columbia.edu/research/population-health-methods/least-absolute-shrinkage-and-selection-operator-lasso

Figure 12: **Predicted probability of investigation versus actual investigation: gradient boosting** This graph shows predicted outcomes versus actual outcomes for units in the held-out test set. The x axis shows an employer-job dyad's predicted probability of an investigation, which is higher if we predict the entity has a higher risk of investigation. The two densities and color shading show how those predicted probabilities vary across two groups: those that did not face an investigation and those that did face an investigation. As expected if the model is working, those that did face an investigation are predicted to have higher risk.

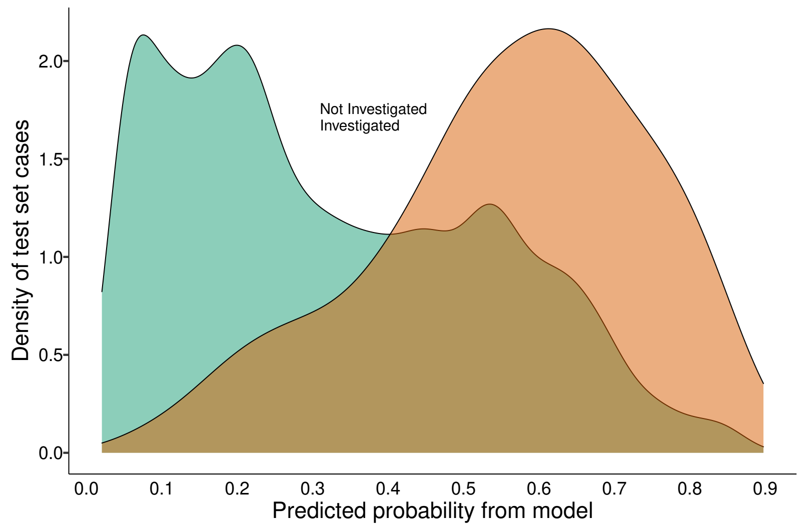
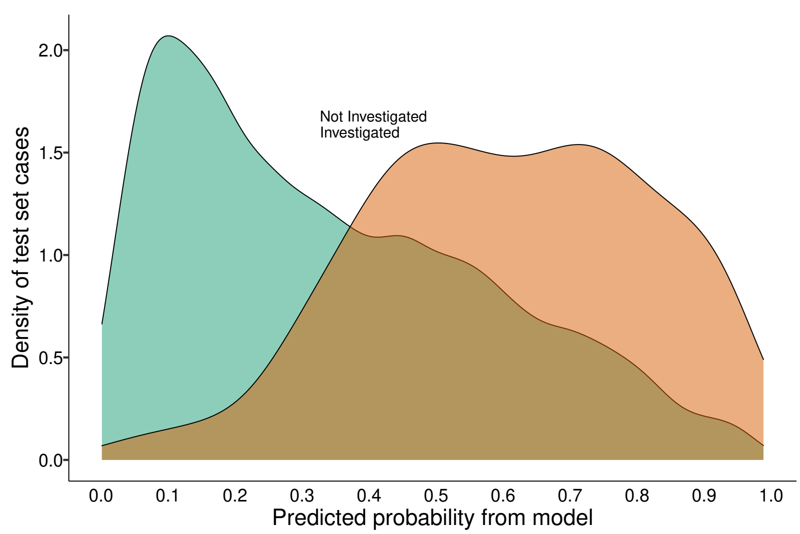


Figure 13: **Predicted probability of investigation versus actual investigation: LASSO** This shows the same patterns as described in Figure 12 but for the LASSO model.



6.3.2 Feature importances

The previous section compared the accuracy of different models at predicting WHD investigations. Importantly, and as we note in the introduction, these investigations reflect a mix of agency-initiated strategic enforcement and complaint-initiated enforcement. Therefore, the most predictive factors likely reflect a mix of the results of strategic targeting strategies and contexts where workers and advocates are better able to navigate reporting processes. In addition, highly predictive features should not be mistaken for causal features—for instance, if certain industries are more likely to face investigation, that could be the industry code serving as a proxy for other, unmeasured predictive factors. With these caveats in mind, we examine the features most predictive of either having an investigation or having no investigation.

Table 4, focusing on the `LASSO` model that has signed coefficients, shows the twenty features most highly predictive of investigations (EMPLOYER_CITY_VAS having the highest value of 1.23). We see largely location-related features, such as locations in Vass, North Carolina and certain states, as well as certain attorney/agents.

Table 5 shows features from the American Community Survey (ACS) predictive of a higher risk of investigation. We see that employers located in areas with more either White or White, non-Hispanic residents have a higher risk of facing investigations. Since this is unlikely to be the result of industry-based strategic enforcement, it could arise either due to reporting patterns or due to that serving as a proxy for other area-level characteristics.

Finally, Table 6 shows features predictive of a lower risk of investigation. In general, we see certain states and small occupational categories (e.g., 45-2091, agricultural equipment operators) facing lower risk, which again could reflect a mix of lower actual issues and reporting differences.

Table 4: **Twenty features predicting higher risk of investigation: LASSO model**. These are features/predictors from the `LASSO` model described in Table 3. We rank predictors from the largest/most positive coefficient (predictive of higher risk) to the smallest/most negative coefficient (predictive of lower risk) and these are the top 20 features predicting higher risk. The sample size is 77,759 employer-certificate dyads, representing 13,838 unique employers (Section 6.1 has more details).

coef_name_upper	value	is_acs
EMPLOYER_CITY_VASS	1.23	
SOC_CODE_45-2092	1.19	
EMPLOYER_CITY_SNOW HILL	1.15	
EMPLOYER_CITY_SANTA MARIA	1.07	
EMPLOYER_STATE_FL	1.05	
WORKSITE_CITY_ARCADIA	1.03	
EMPLOYER_STATE_KS	1.01	
EMPLOYER_STATE_PA	1.01	
ATTORNEY_AGENT_NAME_SANDRA STUNTEBECK	0.94	
EMPLOYER_STATE_NJ	0.79	
ATTORNEY_AGENT_NAME_WILLIAM L. CARLSON	0.69	
SOC_CODE_45-2093	0.65	
EMPLOYER_STATE_WI	0.64	
EMPLOYER_STATE_GA	0.59	
SOC_CODE_45-2091	0.59	
WORKSITE_STATE_FL	0.56	
ATTORNEY_AGENT_STATE_CA	0.56	
ATTORNEY_AGENT_STATE_MT	0.54	
EMPLOYER_STATE_IA	0.52	
ATTORNEY_AGENT_NAME_KELLI GRIFFITH	0.50	

Table 5: **Top ACS features predicting *higher* risk of investigation**

coef_name_clean	value
% Non-hispanic White	0.35
% Male Hispanic aged 45-54	0.34
% from Latin America/Central America	0.33
% from Caribbean	0.30
% earning $75,000 or more	0.20
% earning $1 to $9999	0.16
% White	0.16
% Non-citizen	0.05

Table 6: **Top features predictive of *lower* risk of investigation**

coef_name_upper	value
ATTORNEY_AGENT_CITY_HEYBURN	-1.88
EMPLOYER_CITY_TWIN FALLS	-1.76
SOC_CODE_45-2091.00	-1.70
WORKSITE_CITY_VASS	-1.26
EMPLOYER_STATE_MT	-1.08
ATTORNEY_AGENT_NAME_MELISSA CARSON	-0.92
WORKSITE_STATE_WA	-0.92
WORKSITE_STATE_LA	-0.80
EMPLOYER_STATE_CO	-0.79
WORKSITE_STATE_ID	-0.72
EMPLOYER_CITY_WINTER HAVEN	-0.71
EMPLOYER_STATE_KY	-0.69
WORKSITE_STATE_ND	-0.59
ATTORNEY_AGENT_NAME_KAMRON MARTENS	-0.56
SOC_CODE_45-2093.00	-0.56
SOC_TITLE_FARMWORKERS AND LABORERS, CROP	-0.53
JOB_TITLE_RANGE LIVESTOCK WORKER	-0.50
ATTORNEY_AGENT_STATE_MO	-0.48
ACS_MALE!!35 TO 44 YEARS_SEX_BY_AGE__HISPANIC_OR_LATINO_	-0.45
EMPLOYER_STATE_WY	-0.44

6.3.3 How Do the Models Perform in the TRLA Catchment States?

We applied the same models to the "TRLA intake record but no WHD investigation" and "WHD investigation but no TRLA intake record" labels/samples discussed in Section 6.1. Unfortunately, and likely due to the much smaller sample size, the machine learning models were not accurate enough at this point in time to justify using for this federal versus local prediction task. In the next section, we describe the results from a parametric logistic regression that (1) pares features down from the $N = 696$ features in the initial modeling to $N = 95$ features that the nationwide `LASSO` model left as significant, non zeroed out predictors and (2) estimates two standard logistic regression models. The first predicts which jobs face TRLA oversight but no WHD investigation. The second predicts which jobs face a WHD investigation but no TRLA oversight.

6.4 Predictors of Federal versus Local Oversight in Seven Catchment States

Using the results from the logistic regression described in Section 6.3.3, the following figures contrast three sets of features/predictors, focusing on the ACS contextual variables.

1. Figure 14 shows predictors that reflect both a high risk of a TRLA intake record and high risk of a WHD investigation (from two models, each of which contrasts the focal outcome to employer-certificate dyads with neither type of investigation). We see that jobs located in areas with higher levels of working age Hispanic/Latino males are more likely to face oversight than jobs located in areas with lower concentrations (with male showing up in 3 of the 7 tract-level characteristics, and Hispanic/Latino/Latin America showing up in 6 of them). For equity purposes, this is potentially a positive finding where employers located in areas with higher levels of linguistic and other forms of vulnerability are subject to both local oversight and federal oversight.

2. Figure 15 shows predictors that reflect both a low risk of a TRLA intake record and low risk of a WHD investigation. These are less interpretable than the high risk ones, but could reflect higher concentrations of recent foreign-born residents (with mentioning of foreign or moving from abroad occurring in 4 of the 6 tract-level characteristics).

3. Finally, and most relevant to our questions of whether and how local outreach complements federal oversight, Figure 16 shows features that have an opposite relationship with coming to the attention of local versus federal oversight. Perhaps due to its outreach strategy focusing on large and visible farms, we see that employers located in areas with high concentrations of Hispanic/Latino males aged 45-54 have a *higher* risk of a TRLA intake record relative to other areas (coefficient > 3) but a *lower* risk of a WHD investigation (coefficient < 0). Again, this variable could be a proxy for other aspects of employers and areas, but in general, the figure is consistent with the descriptive evidence of TRLA outreach potentially "widening the net" of which employers are investigated.

Examining the spatial variation that underlies these patterns, and focusing on Census tracts in Texas, Figure 17 contrasts characteristics of tracts containing employers with: (1) only a TRLA intake record versus (2) only a WHD investigation. This figure shows TRLA outreach efforts occurring in areas with high concentrations of Hispanic/Latino residents, potentially reflecting locations of larger employers.

Figure 14: **ACS contextual characteristics predicting higher risk of both TRLA intake record and WHD investigation**

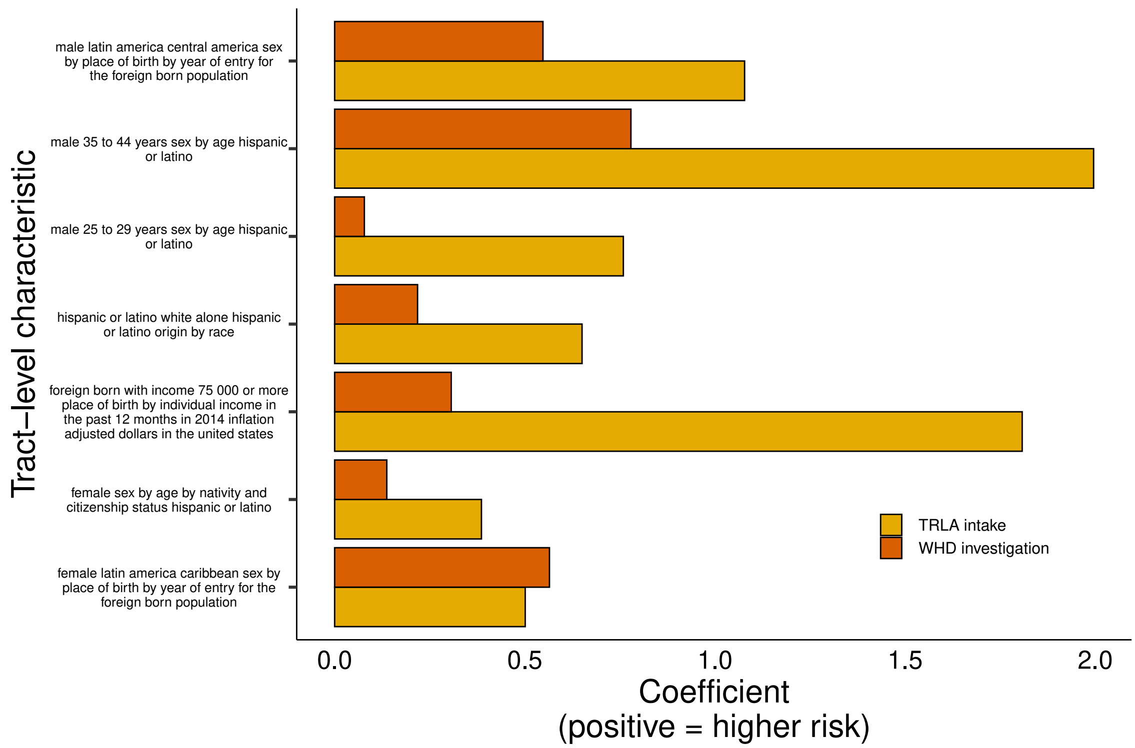


Figure 15: **ACS contextual characteristics predicting lower risk of both TRLA intake record and WHD investigation**

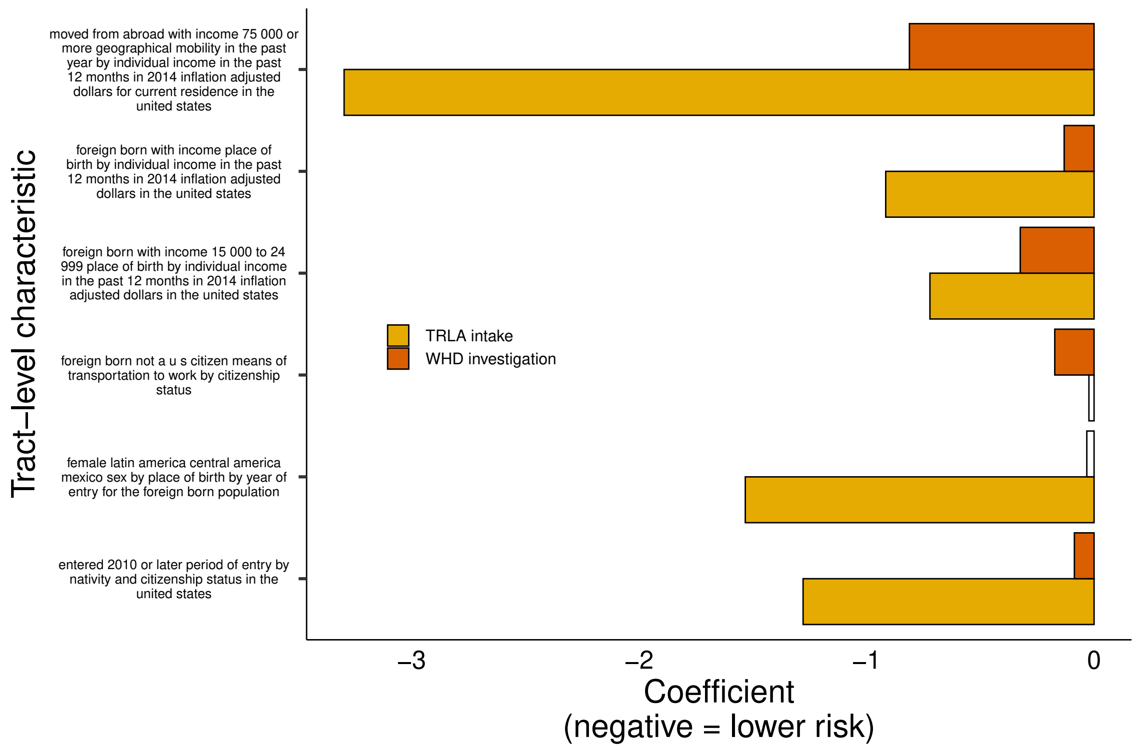


Figure 16: **ACS contextual characteristics that have opposite predictions about risk of TRLA intake record versus WHD investigations**

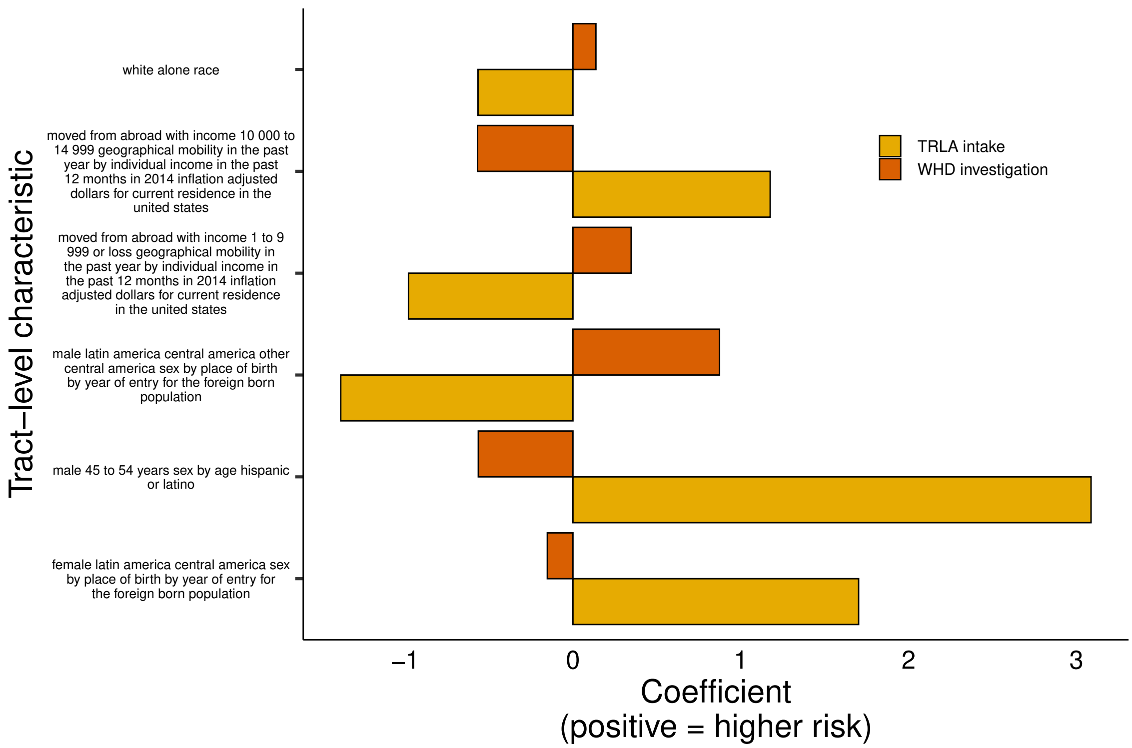
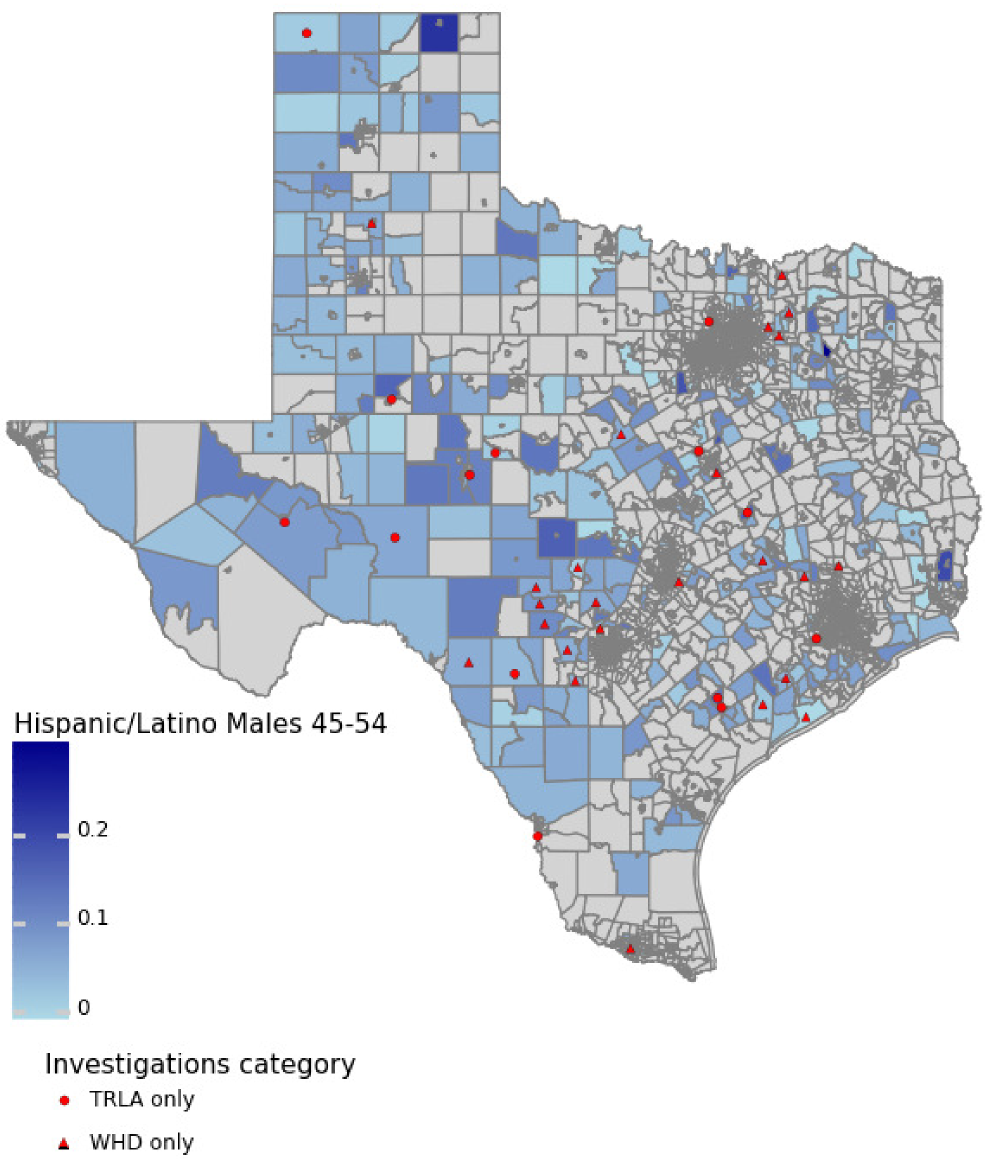


Figure 17: **Tract-level demographics surrounding employer and local versus federal oversight for Hispanic/Latino males ages 45-54** Tracts with any H-2A jobs geocoded to a site in the tract are shaded using their ACS percentages, with those containing no data shaded to gray for missing. Darker gray areas occur near the clustering of borders, but this only occurs because of space constraints here and has no independent meaning.



6.5 Text Analysis of Addendums

The previous section showed that: (1) At the nationwide level, we can prospectively predict employers with potential issues with a relatively high degree of accuracy but (2) the smaller sample size of the TRLA catchment states inhibits similar machine learning-based predictions, though we compare outputs of a more parsimonious model with top features from the nationwide analysis. We now turn to a data source available beginning in FY 2020: the free text of job addendums, obtained by TRLA from a FOIA request.

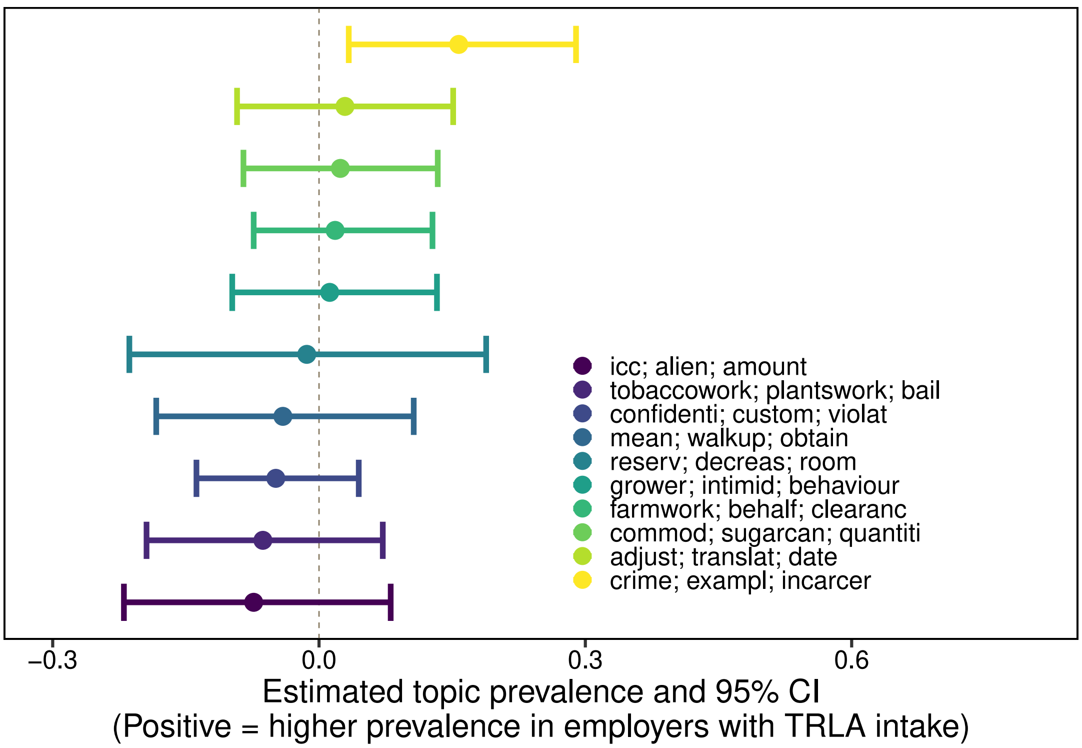
Figure 18 shows top words from the ten themes from our inductively-estimated topic modeling of the text of job addendums. We see themes related to transportation/interstate commerce (*icc; alien; amount*, where icc refers to the Interstate Commerce Commission), themes related to behavioral expectations for workers (*grower*; *intimid*; *behaviour*); and themes related to logistics (*adjust*; *translat*; *date*). (It is worth reiterating that we have "translat" instead of "translate," for example, because of stemming, as discussed in Section 5.10.) Most themes have similar contracts in addendums for jobs that generate TRLA intake records and ones that do not (point estimates close to zero). The one theme that differs significantly at the $p < 0.05$ level is addendums reflecting the topic *crime*; *example*; *incarcer*, which corresponds to discussions of criminal history restrictions on employment offers. These addendums are associated with a significantly higher risk of TRLA intake.

Why might contracts emphasizing criminal history restrictions on employment, measured at the time of an employer's application to OFLC, be correlated with later issues? Employers are only allowed to use pre-employment background checks if they are a normal and accepted practice in the industry. Table 10 contains addendums where this topic occupies a significant proportion of the addendum—we see that the criminal history restrictions use standard language indicating that these screenings are in compliance with normal and accepted practices. Yet despite their facially neutral nature, some employers use background checks as an implicit barrier to employment for U.S. workers who apply for H-2A jobs. For instance, in 2017, the Justice Department filed suit against a Colorado-based H-2A employer, alleging discrimination against U.S. citizens for whom they "imposed more burdensome requirements...For instance, the complaint alleges that whereas U.S. citizens had to complete a background check and a drug test before being permitted to start work, H-2A workers were allowed to begin working without completing them and, in some cases, never completed them."[14] Contracts that emphasize criminal background checks may be equally enforced against both U.S.-based workers and guest workers, but they may also be unequally enforced in ways that harm vulnerable U.S. residents.[15] Therefore, this theme can be seen as highly associated with potential future issues in job addendums. In contrast, a theme associated with slightly lower risk of TRLA intake records (*icc*; *alien*; *amount*, examples depicted in Table 10) emphasizes the rights of guestworkers to transportation reimbursement.

[14] DOJ press release, September 28, 2021 (https://www.justice.gov/opa/pr/justice-department-files-lawsuit-against-crop-production-services-alleging-discrimination)

[15] As U.S. background checks are unlikely to produce any information about a foreign worker, H-2A workers may have an easier time navigating such requirements.

Figure 18: **Inductively-estimated addendum topics associated with higher versus lower risk of TRLA intake (relative to no TRLA or WHD intake) through text analysis** *Certain words such as "violat" or "decreas" may look strange, but are the result of "stemming" as discussed in Section 5.10*



7 Discussion

In the present report, we asked: Can local outreach from "trusted messengers" supplement federal oversight of the H-2A program? Taken together, our analyses suggest three key takeaways. First, we show that although WHD investigations and TRLA intake records uncover a similar prevalence of potential issues to investigate, the two forms of oversight flag distinct entities. This can be seen by Figure 4's non-zero counts for the categories of employers with a WHD investigation but no TRLA intake record and, conversely, employers with a TRLA intake record and no WHD investigation. Second, while predictive modeling is useful at a nationwide scale for using fields available at screening to predict later issues, it was less useful to predict which entities are likely to face local, TRLA-based oversight of activities versus federal oversight. In particular, we do not present results because our predictive models either returned predictions of 0 (no investigation) for all entities or predictions of 1 (yes investigation) for all entities, indicating that there was insufficient data to distinguish between the two labels. Third, despite the limitations of predictive modeling for comparing local and federal oversight, we are able to find distinguishing characteristics of employers that come under local scrutiny; for instance, the text analysis of addendums showing that criminal history-focused contracts at the time of applying for a certificate portend later issues. Concluding, we review limitations of our analysis, limitations of the DOL public use data and suggestions for collection, and recommendations on policies and investments to promote equity in H-2A oversight.

7.1 Limitations and Future Directions

There are several limitations, which we group into two categories. First are *addressable limitations*, or additional analyses we could perform as we continue work with the present data. Second are *intrinsic limitations*, or challenges that require new forms of data to answer.

7.1.1 Addressable Limitations

The first limitation and future direction is exploring how the results change with different pre-modeling, analytic choices. The current results use a unit of analysis—employer-certificate dyads—and definition of the outcome variable—the dates corresponding to the intake call or WHD investigation must overlap with the job start and end dates of the job—that leads to a very low prevalence of investigations. At the national level, we are able to predict investigations with a high degree of accuracy. But when we restrict analyses to the seven TRLA catchment states, the combination of a rare outcome and limited observations creates challenges for modeling. As Section 6.2 highlights, the rates of investigations are much higher when we examine *any* investigation across all of an employer's distinct jobs. While predicting issues at the employer level creates questions about the best way to aggregate job-level features within an employer—e.g., if multiple industries are represented, should we use the modal industry? As ACS contextual characteristics change over time, should we take the mean or the max?—these challenges may be outweighed by benefits of analyses at the employer level.

Second, our linkage between H-2A employer-certificate dyads and compliance actions is based on fuzzy/probabilistic record linkage. While we set high string distance thresholds in requiring names of employers and cities to be very similar, and requiring states to match exactly, there are likely still some false negatives and false positives. Testing the results' robustness to different record linkage methods would improve the reliability.

A third limitation is that, for sample size reasons, we pool all analyses over time, which could obscure important changes within the period under study as both WHD and TRLA adapt their oversight strategies to changing contexts and conditions. While sample size prevents year-specific modeling, future analyses could use temporal cross-validation where, instead of pooling across years, we use data from 2014-2018 to predict 2019 issues, data from 2014-2019 to predict 2020 issues, and so on.

Fourth, for the computational text analysis of addendums, many of the themes in the text have now been incorporated into the H-2A certificate data as structured fields. Future analyses could add two sets of predictors—the topic modeling themes from the addendum texts and the structured fields reflecting addendum elements—to the predictive model to compare the relative predictive power.

7.1.2 Intrinsic limitations

First, we have noted advantages of pooling data on issues uncovered in local outreach with issues uncovered during federal oversight. Yet the TRLA intake records have their own biases. Limitations on TRLA data include the organization's limited resources (both time and money), which leads the organization to focus on larger employers during outreach. In addition, industries such as herding and livestock have more difficult-to-reach housing, which produces an undercount of open range herders and livestock employers generally in the TRLA data. Finally, TRLA does not have investigatory authority, which would legally entitle them to access worksites. As a result, TRLA must rely on workers to report violations from worksites they are unable to access for outreach or other legal business.

Second, our anchoring of the analysis in predictive questions, such as asking what predicts a TRLA intake record versus WHD investigation, rather than a causal framework, such as asking if a specific characteristic causes an increase or decrease in issues, means that when we compare the characteristics of employers subject to local versus federal scrutiny, these characteristics may be proxies for other attributes. In addition, these characteristics are limited to the few fields collected in certificate applications (before the 2020 expansion of fields) and ACS tract-level characteristics that reflect demographics of residents who are typically disconnected from the H-2A program.

Third, we have drawn on TRLA's own observations of barriers to reporting among guestworkers. But we lack individual-level data that would allow us to correlate (1) employee characteristics (e.g., race/ethnicity; are family members also participating in the H-2A program?) with (2) measures of awareness of rights and willingness to report. Therefore, our analysis is limited to showing *that* local outreach efforts uncover issues not reported federally but lends limited insight into *why*.

7.2 DOL Data Limitations and Suggestions for Collection

The DOL data in its current form yielded many valuable insights.

To enhance the data's usability, we recommend two additional forms of data collection.

First, and potentially in internal analyses rather than public release datasets to ensure employers can't "game" the oversight system, our analyses could be replicated splitting the DOL compliance action data into two sets:

1. **Investigations triggered by complaints or worker/advocate reporting**
2. **Investigations triggered as part of strategic enforcement priorities**

Examining similarities and differences between contexts and employers in the two groups could yield insight into which group of employers is more similar to the employers in the TRLA intake records.

Second is that we have performed our own entity resolution of employers using name, address, and other fields. Since many research projects aim to follow employers across several job certificates and compliance actions, a stable, employer-level identifier that helps link H-2A certificate data to compliance actions would greatly lower the uptake costs of requiring knowledge of fuzzy matching/entity resolution methods. This would also enhance comparability across research, since different matching algorithms can produce different variations of what counts as the same employer.

7.3 Policies and Investments to Promote Equity in H-2A Oversight

We have two recommendations for policies and investments to enhance equity in H-2A oversight. First, for federal oversight, we recommend randomly sampling employers for audits to obtain a less biased measure of the prevalence of issues, echoing recommendations in (Dolfin et al., 2020). Random sampling provides a less biased measure of the prevalence of issues than current WHD enforcement records, which reflect a mix of complaint-driven and strategic enforcement. This random sampling, rather than giving each employer an equal odds of selection, could use a risk-based sampling strategy,[16] with adjustments then made to the resulting estimates to reflect unequal probabilities of being sampled. Random sampling, in addition to obtaining a less biased measure of the prevalence of H-2A-related workplace issues, can also then be used to examine the causal impact of enforcement on various outcomes—e.g., hiring outcomes for low-wage workers in the surrounding area or future issues by the same employer, following the work of (Johnson et al., 2020) with OSHA.

Yet the legal violations that random sampling helps uncover often still require workers to disclose issues to investigators. Therefore, our second recommendation is to evaluate different ways of structuring the worker outreach process with equity in mind. For instance, an evaluation could compare two outreach strategies: (1) one where DOL investigators use their normal strategies for worker outreach and (2) one where DOL investigators take a "trusted messenger" approach to outreach, working with local community-based organizations and legal service providers. The present report shows how these local actors can uncover additional issues unreported to federal entities, and future research could more systematically explore both federal-local data linkages through a Memorandum of Understanding (MOU) and federal-local partnerships for worker outreach.

[16] For an example of another governmental department already employing risk-based sampling, see https://www.aphis.usda.gov/aphis/ourfocus/planthealth/import-information/agriculture-quarantine-inspection/rbs

8 Appendix

8.1 Additional Details on Data Cleaning

8.1.1 H-2A Certificate Data

Figure 19 shows the columns that we are able to preserve in the rowbound FY 2014-2021 certificate data. Other variables were inconsistently collected across years and could not be included. The majority of these came from a large expansion in 2020 to add structured fields associated with contractual requirements (e.g., meal reimbursement policies, housing type, criminal background checks, pay deductions).

Figure 19: **Columns included as IDs/predictors since consistent across years** These are raw variable names from the H-2A certificate data. Acronyms are SOC (Standard Occupational Classification). Otherwise, the variable name is the spelled out description (e.g., JOB refers to job).

ATTORNEY_AGENT_CITY
ATTORNEY_AGENT_NAME
ATTORNEY_AGENT_STATE
CASE_NUMBER
CASE_STATUS
DECISION_DATE
EMPLOYER_ADDRESS1
EMPLOYER_CITY
EMPLOYER_NAME
EMPLOYER_POSTAL_CODE
EMPLOYER_STATE
JOB_END_DATE
JOB_START_DATE
JOB_TITLE
REQUESTED_END_DATE_OF_NEED
REQUESTED_START_DATE_OF_NEED
SOC_CODE
SOC_TITLE
WORKSITE_CITY
WORKSITE_STATE

We filter to approved or partially approved certificates. Due to the reliance on fuzzy matching, we filter out four certificates that are blank for the `EMPLOYER_NAME` field, with the case numbers shown in Table 7.

Table 7: **Cases excluded due to missing employer name**

CASE_NUMBER	DECISION_DATE	EMPLOYER_NAME
H-300-15174-219030	2015-06-30 15:43:17.999996	
H-300-19051-164654	2019-03-12 11:16:35.999999	
H-300-18361-395193	2019-02-11 17:34:04.999999	
H-300-18361-395193	2019-02-11 17:34:04.999999	
H-300-19023-360692	2019-02-25 15:27:48.000004	
H-300-19028-021894	2019-03-01 11:55:24.000004	

8.2 ACS Data

The following tables show the predictors we use from the ACS.

Figure 20: **ACS predictors (1 of 3)**

10 to 14 minutes sex of workers by travel time to work
10 to 14 minutes travel time to work
10 00 a m to 10 59 a m sex of workers by time leaving home to go to work
10 00 a m to 10 59 a m time leaving home to go to work
100 to 149 percent of the poverty level geographical mobility in the past year by poverty status in the past 12 months for current residence in the united states
100 to 149 percent of the poverty level place of birth by poverty status in the past 12 months in the united states
11 00 a m to 11 59 a m sex of workers by time leaving home to go to work
11 00 a m to 11 59 a m time leaving home to go to work
12 00 a m to 4 59 a m sex of workers by time leaving home to go to work
12 00 a m to 4 59 a m time leaving home to go to work
12 00 p m to 3 59 p m sex of workers by time leaving home to go to work
12 00 p m to 3 59 p m time leaving home to go to work
15 to 19 minutes sex of workers by travel time to work
15 to 19 minutes travel time to work
20 to 24 minutes sex of workers by travel time to work
20 to 24 minutes travel time to work
25 to 29 minutes sex of workers by travel time to work
25 to 29 minutes travel time to work
25 to 29 years geographical mobility in the past year by age for current residence in the united states
30 to 34 minutes sex of workers by travel time to work
30 to 34 minutes travel time to work
30 to 34 years geographical mobility in the past year by age for current residence in the united states
35 to 39 minutes sex of workers by travel time to work
35 to 39 minutes travel time to work
35 to 39 years geographical mobility in the past year by age for current residence in the united states
4 or more vehicles available household size by vehicles available
4 00 p m to 11 59 p m sex of workers by time leaving home to go to work
4 00 p m to 11 59 p m time leaving home to go to work
40 to 44 minutes sex of workers by travel time to work
40 to 44 minutes travel time to work
40 to 44 years geographical mobility in the past year by age for current residence in the united states
45 to 49 years geographical mobility in the past year by age for current residence in the united states
45 to 59 minutes sex of workers by travel time to work
45 to 59 minutes travel time to work
5 to 9 minutes sex of workers by travel time to work
5 to 9 minutes travel time to work
5 00 a m to 5 29 a m sex of workers by time leaving home to go to work
5 00 a m to 5 29 a m time leaving home to go to work
5 30 a m to 5 59 a m sex of workers by time leaving home to go to work
5 30 a m to 5 59 a m time leaving home to go to work
50 to 54 years geographical mobility in the past year by age for current residence in the united states
55 to 59 years geographical mobility in the past year by age for current residence in the united states
6 00 a m to 6 29 a m sex of workers by time leaving home to go to work
6 00 a m to 6 29 a m time leaving home to go to work
6 30 a m to 6 59 a m sex of workers by time leaving home to go to work
6 30 a m to 6 59 a m time leaving home to go to work
60 to 64 years geographical mobility in the past year by age for current residence in the united states
60 to 89 minutes sex of workers by travel time to work
60 to 89 minutes travel time to work
65 to 69 years geographical mobility in the past year by age for current residence in the united states

Figure 21: **ACS predictors (2 of 3)**

7 00 a m to 7 29 a m sex of workers by time leaving home to go to work
7 00 a m to 7 29 a m time leaving home to go to work
7 30 a m to 7 59 a m sex of workers by time leaving home to go to work
7 30 a m to 7 59 a m time leaving home to go to work
8 00 a m to 8 29 a m sex of workers by time leaving home to go to work
8 00 a m to 8 29 a m time leaving home to go to work
8 30 a m to 8 59 a m sex of workers by time leaving home to go to work
8 30 a m to 8 59 a m time leaving home to go to work
9 00 a m to 9 59 a m sex of workers by time leaving home to go to work
9 00 a m to 9 59 a m time leaving home to go to work
agriculture forestry fishing and hunting and mining industry by median earnings in the past 12 months in 2014 inflation adjusted dollars for the civilian employed population 16 years and over
agriculture forestry fishing and hunting and mining industry by median earnings in the past 12 months in 2014 inflation adjusted dollars for the full time year round civilian employed population 16 years and over
american indian and alaska native alone race
asian alone race
at or above 150 percent of the poverty level geographical mobility in the past year by poverty status in the past 12 months for current residence in the united states
at or above 150 percent of the poverty level place of birth by poverty status in the past 12 months in the united states
below 100 percent of the poverty level geographical mobility in the past year by poverty status in the past 12 months for current residence in the united states
below 100 percent of the poverty level place of birth by poverty status in the past 12 months in the united states
black or african american alone race
entered 2010 or later period of entry by nativity and citizenship status in the united states
entered 2010 or later foreign born not a u s citizen period of entry by nativity and citizenship status in the united states
estimate median family income in the past 12 months in 2014 inflation adjusted dollars median family income in the past 12 months in 2014 inflation adjusted dollars
estimate median income in the past 12 months total median income in the past 12 months in 2014 inflation adjusted dollars by place of birth in the united states
estimate total median age by nativity and citizenship status by sex
estimate total median earnings in the past 12 months in 2014 inflation adjusted dollars by sex by educational attainment for the population 25 years and over
estimate total median household income in the past 12 months in 2014 inflation adjusted dollars by receipt of food stamps snap in the past 12 months
female sex by age by nativity and citizenship status hispanic or latino
female 18 years and over foreign born not a u s citizen sex by age by nativity and citizenship status hispanic or latino
female latin america sex by place of birth by year of entry for the foreign born population
female latin america caribbean sex by place of birth by year of entry for the foreign born population
female latin america caribbean entered 2010 or later sex by place of birth by year of entry for the foreign born population
female latin america central america sex by place of birth by year of entry for the foreign born population
female latin america central america mexico sex by place of birth by year of entry for the foreign born population
female latin america central america mexico entered 2010 or later sex by place of birth by year of entry for the foreign born population
female latin america central america other central america sex by place of birth by year of entry for the foreign born population
female latin america central america other central america entered 2010 or later sex by place of birth by year of entry for the foreign born population
female latin america entered 2010 or later sex by place of birth by year of entry for the foreign born population
female latin america south america sex by place of birth by year of entry for the foreign born population
female latin america south america entered 2010 or later sex by place of birth by year of entry for the foreign born population
foreign born place of birth by nativity and citizenship status
foreign born place of birth by poverty status in the past 12 months in the united states
foreign born 100 to 149 percent of the poverty level place of birth by poverty status in the past 12 months in the united states
foreign born at or above 150 percent of the poverty level place of birth by poverty status in the past 12 months in the united states
foreign born below 100 percent of the poverty level place of birth by poverty status in the past 12 months in the united states
foreign born no income place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born not a u s citizen means of transportation to work by citizenship status
foreign born not a u s citizen median age by nativity and citizenship status by sex
foreign born not a u s citizen place of birth by nativity and citizenship status
foreign born not a u s citizen female median age by nativity and citizenship status by sex
foreign born not a u s citizen male median age by nativity and citizenship status by sex

Figure 22: **ACS predictors (3 of 3)**

foreign born with income place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 1 to 9 999 or loss place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 10 000 to 14 999 place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 15 000 to 24 999 place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 25 000 to 34 999 place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 35 000 to 49 999 place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 50 000 to 64 999 place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 65 000 to 74 999 place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
foreign born with income 75 000 or more place of birth by individual income in the past 12 months in 2014 inflation adjusted dollars in the united states
hispanic or latino hispanic or latino origin by specific origin
hispanic or latino white alone hispanic or latino origin by race
income in the past 12 months below poverty level 18 to 59 years poverty status in the past 12 months by age
less than 5 minutes sex of workers by travel time to work
less than 5 minutes travel time to work
male sex by age hispanic or latino
male sex by place of birth by year of entry for the foreign born population
male 18 and 19 years sex by age hispanic or latino
male 18 years and over foreign born not a u s citizen sex by age by nativity and citizenship status hispanic or latino
male 20 to 24 years sex by age hispanic or latino
male 25 to 29 years sex by age hispanic or latino
male 30 to 34 years sex by age hispanic or latino
male 35 to 44 years sex by age hispanic or latino
male 45 to 54 years sex by age hispanic or latino
male latin america sex by place of birth by year of entry for the foreign born population
male latin america caribbean sex by place of birth by year of entry for the foreign born population
male latin america caribbean entered 2010 or later sex by place of birth by year of entry for the foreign born population
male latin america central america sex by place of birth by year of entry for the foreign born population
male latin america central america mexico sex by place of birth by year of entry for the foreign born population
male latin america central america mexico entered 2010 or later sex by place of birth by year of entry for the foreign born population
male latin america central america other central america sex by place of birth by year of entry for the foreign born population
male latin america central america other central america entered 2010 or later sex by place of birth by year of entry for the foreign born population
male latin america entered 2010 or later sex by place of birth by year of entry for the foreign born population
male latin america south america sex by place of birth by year of entry for the foreign born population
male latin america south america entered 2010 or later sex by place of birth by year of entry for the foreign born population
male worked in the past 12 months usually worked 1 to 14 hours per week sex by work status in the past 12 months by usual hours worked per week in the past 12 months by weeks worked in the past 12 months for the population 16 to 64 years
male worked in the past 12 months usually worked 15 to 34 hours per week sex by work status in the past 12 months by usual hours worked per week in the past 12 months by weeks worked in the past 12 months for the population 16 to 64 years
male worked in the past 12 months usually worked 35 or more hours per week sex by work status in the past 12 months by usual hours worked per week in the past 12 months by weeks worked in the past 12 months for the population 16 to 64 years
male worked in the past 12 months usually worked 35 or more hours per week 48 and 49 weeks sex by work status in the past 12 months by usual hours worked per week in the past 12 months by weeks worked in the past 12 months for the population 16 to 64 years
male worked in the past 12 months usually worked 35 or more hours per week 50 to 52 weeks sex by work status in the past 12 months by usual hours worked per week in the past 12 months by weeks worked in the past 12 months for the population 16 to 64 years
moved from abroad no income geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 1 to 9 999 or loss geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 10 000 to 14 999 geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 15 000 to 24 999 geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 25 000 to 34 999 geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 35 000 to 49 999 geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 50 000 to 64 999 geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 65 000 to 74 999 geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
moved from abroad with income 75 000 or more geographical mobility in the past year by individual income in the past 12 months in 2014 inflation adjusted dollars for current residence in the united states
native hawaiian and other pacific islander alone race
no vehicle available household size by vehicles available
not hispanic or latino hispanic or latino origin
not hispanic or latino hispanic or latino origin by specific origin
speak other languages means of transportation to work by language spoken at home and ability to speak english
white alone race

8.3 Additional Descriptives

Figure 23: **Number of job applications total, with WHD investigations, or with violations** *This draws on the WHD Compliance Action data*

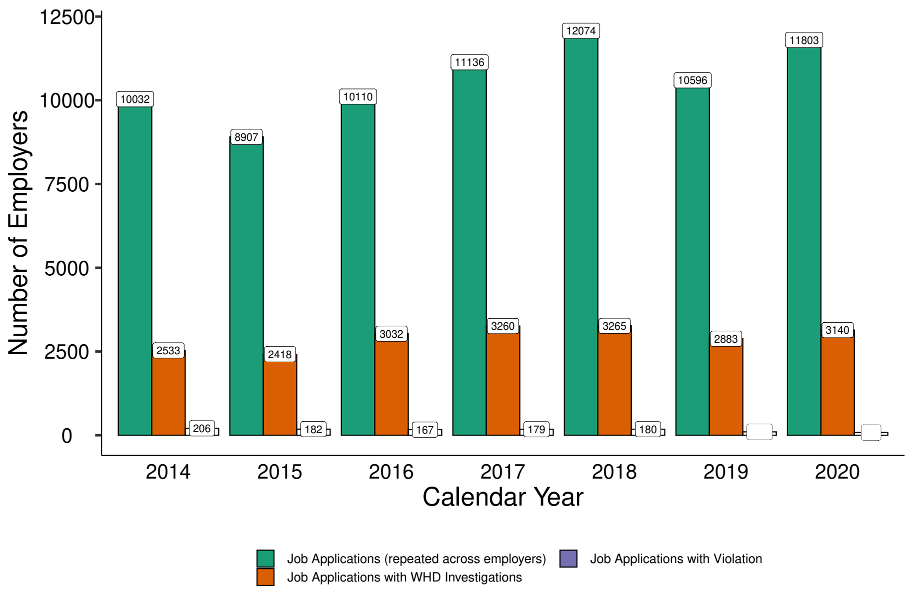


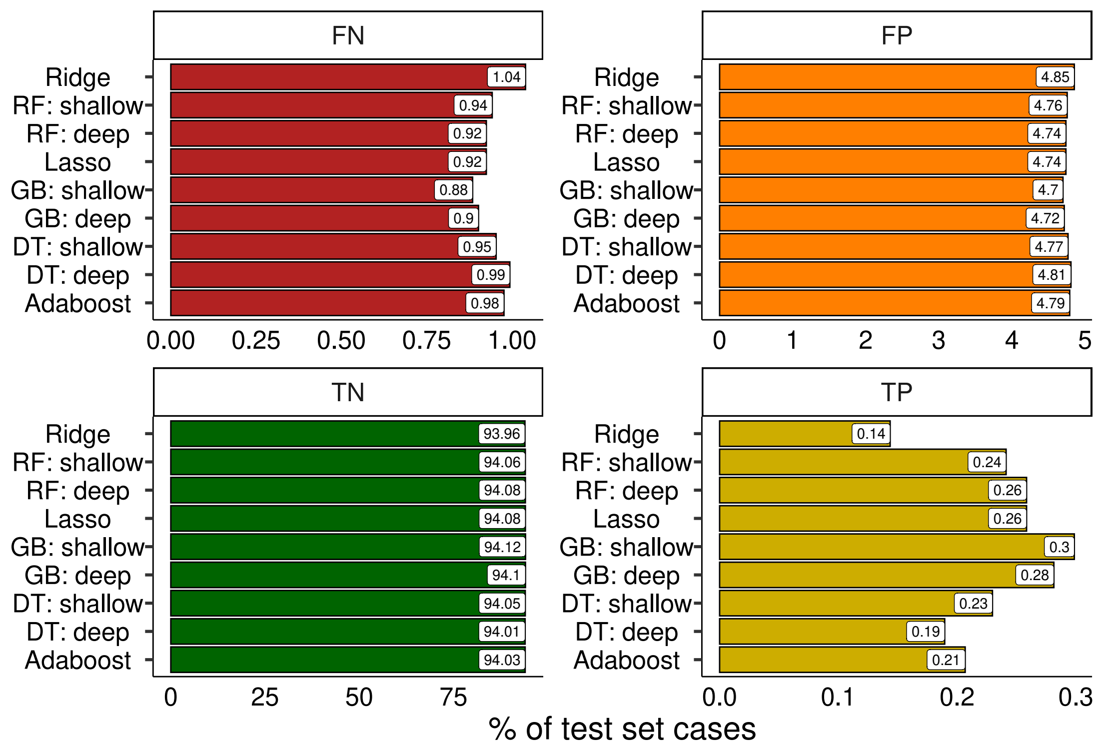
Table 8: **Full list of SOC occupational codes (consolidated to the 6-digit summary) in TRLA catchment states**

soc_code_6dig	soc_title_consolidated
11-9013	Farmers, Ranchers, and Other Agricultural Managers
11-9021	Construction Managers
45-1011	First-Line Supervisors of Farming, Fishing, and
45-2021	Animal Breeders
45-2041	Graders and Sorters, Agricultural Products
45-2091	Agricultural Equipment Operators
45-2092	Farmworkers and Laborers, Crop, Nursery, and Greenhouse
45-2093	Farmworkers, Farm, Ranch, and Aquacultural Animals
45-2099	Agricultural Workers, All Other
45-3011	Fishers and Related Fishing Workers
45-4011	Forest and Conservation Workers
45-4022	Logging Equipment Operators
45-4029	Logging Workers, All Other
47-1011	Supervisors of Construction and Extraction Workers
47-2061	Construction Laborers
49-3041	Farm Equipment Mechanics and Service Technicians
51-3022	Meat, Poultry, and Fish Cutters and Trimmers
51-6041	Shoe and Leather Workers and Repairers
51-9111	Packaging and Filling Machine Operators and
51-9111	Packaging and Filling Machine Operators and Tenders
51-9198	Helpers–Production Workers
53-3032	Heavy and Tractor-Trailer Truck Drivers
53-7021	Crane and Tower Operators

8.4 Additional Details on Modeling Process

The following figure describes error rates across models.

Figure 24: **Model performance: any WHD investigation**. The four categories correspond to (1) false negatives (FN), or cases where we do not predict issues but where issues are found; (2) false positives (FP), or cases where we predict issues but none are found; (3) true negatives (TN), or cases where we correctly predict no issues; (4) true positives (TP), or cases where we correctly predict issues. Section 6.3.1 explains these categories in greater depth.



8.5 Additional Details on Text Analysis

The following tables present additional details of the text analysis.

Table 9: **Coverage rates of addendums versus categories of job orders: 2020-2021**

Category	Job addendum?	N with addendum	N in category	Prop. with addendum
Neither WHD nor TRLA	No	3518	7274	0.48
Neither WHD nor TRLA	Yes	3756	7274	0.52
TRLA; not WHD	No	9	47	0.19
TRLA; not WHD	Yes	38	47	0.81
WHD; not TRLA	No	1	6	0.17
WHD; not TRLA	Yes	5	6	0.83

Table 10: **Example of topic "icc; alien; amount" (boilerplate reimbursement language; associated with lower risk of intake)**

topic	document
Topic 1	Workers who qualify for inbound and/or outbound travel reimbursements are entitle to reimbursements up to $55.00 per day determining the appropriate amount of reimbursement for meals for less than a day, the employer may provide for meal expenses reimbursement, with receipts, up to 75% of the maximum reimbursement for meals, or $41.25 based on the GSA per diem schedule. If a worker cannot produce receipts they will be reimbursed $12.46 per day.The employer will provide advance transportation for reasonable (most economical) common carrier or other transportation which conforms to the Interstate Commerce Commission (ICC) inbound transportation (if it is the prevailing practice.) If not the prevailing practice, the employer will reimburse the worker for transportation costs and subsistence to the employer?s work site when the worker completes 50% of the work period. The employer will also provide advance subsistence at a minimum amount of $ N/A per 24 hour period of travel from the place of recruitment to the place of employment (if it is the prevailing practice.) Workers who provide receipts for meals and nonalcoholic beverages in excess of $_N/A will be reimbursed during the first pay periods, up to the maximum amount of $ N/A per 24 ?hour period of travel from place of recruitment to the place of employment (if it is the prevailing practice.) Workers who voluntarily quit or are terminated for cause prior to completing 50% of the contract period will be required to reimburse the employer for the full amounts of transportation and subsistence which were advanced and/or reimbursed to the worker. After worker has completed 50% of the work contract period, employer will reimburse worker for the cost of transportation and subsistence from the place of recruitment (travel reimbursement subsistence will be the minimum amount of $ 12.46 per 24hour period of travel and the maximum amount will be $ 55.00 per day from the place of employment to the place of recruitment. Due to subsequent employment with another employer who agrees to pay such costs, in which the employer will only pay for the transportation and subsistence to the next job. The amount of the transportation payment will be equal to the most economical and reasonable similar common carrier transportation charges for the distance involved. Upon completion of the work contract, employer will pay reasonable costs of return transportation and subsistence in accordance with current rates published in the Federal Register (currently no less than $12.46 per day without receipts and up to $55.00 per day with receipts as the maximum amount to be reimbursed. (Per 20 CFR 655.173.)The employer will not be responsible for providing the cost of return transportation and subsistence from the place of employment to the place of recruitment if the worker voluntarily abandons the job or is terminated for just cause.

Table 11: **Example of topic "crime; example; incarcer" (criminal history-focused language; associated with higher risk of intake)**

topic	document
Topic 9	Worker must be able to withstand working in the direct sunlight and weather conditions ranging from hot and humid weather, moderate rain and cold while performing their required job duties. The employer may conduct a criminal background checks on all new applicants for employment. Seasonal Employees seeking rehire will not be required to submit a new background check. For purposes of this policy, "rehires" shall be defined consistently with IRCA's employment eligibility reverification requirements for former hires. As a general rule, absent compelling circumstances, qualified applicants with criminal records will not be considered for employment if any of the following criteria are met: The conviction was for a violent crime against one or more persons or property, (eg., battery, assault, lewdness, sexual battery, molestation, arson or criminal mischief); The conviction was for any felony committed or which resulted in the applicant's incarceration at any time within the past 5 years (i.e., a crime which subjects the individual convicted to imprisonment for longer than a year); or The conviction was for a crime committed or which resulted in the applicant's incarceration at any time within the past 5 years involving theft or disorderly conduct. Employer has identified these limited categories of recent criminal convictions as those which raise an unnecessary risk of further criminal conduct and the potential of injury to coworkers due to the physical strenuous work being offered with communal temporary living quarters and daily transportation to and from the place of employment which is being provided. For purposes of this policy, a plea of nolo contenderé to a disqualifying criminal record as described above shall be deemed to be a disqualifying event for employment purposes, irrespective of whether adjudication was withheld. Employer will pay all fees associated with conducting a criminal background check on any applicants.
Topic 9	Raises and end of the season bonuses may be offered to any seasonal worker, at the employer's discretion, based on individual factors including but not limited to, performance, experience, number of hours worked in the season, number of seasons worked with the company, adherence to work rules and ability to follow supervisor's instructions. common carrier transportation charges for the distances involved. Daily subsistence is subject to change with the publication of new rates by the Office of Foreign Labor Certification in the Federal Register. Workers who provide receipts for meals and nonalcoholic beverages in excess of $12.68 will be reimbursed up to the maximum amount of $55.00 per 24hour period of travel per 20 CFR 655.122(h)(1). Due to possible Date of Need changes, worker may be required to purchase travel insurance, if available. Worker will be reimbursed for this expense.Outbound: Employer will provide and pay for transportation by charter bus/van/public transportation and daily subsistence at end of contract period to place of recruitment. During slow, time, workers may choose to go home at their own expense. All Transportation (bus/van) will meet State, Local and Federal requirements regarding Vehicle Insurance and Workers' Compensation regulations and are all DOT inspected. employment if any of the following criteria are met: The conviction was for a violent crime against one or more persons or property, (e.g., battery, assault, lewdness, sexual battery, molestation, arson or criminal mischief); The conviction was for any felony committed or which resulted in the applicants incarceration at any time within the past 5 years (i.e., a crime which subjects the individual convicted to imprisonment for longer than a year); or the conviction was for a crime committed or which resulted in the applicant's incarceration at any time within the past 5 years involving theft or disorderly conduct. Employer has identified these limited categories of recent criminal convictions as those which raise an unnecessary risk of further criminal conduct and the potential of injury to coworkers due to the physical strenuous work being offered with communal temporary living quarters and daily transportation to and from the place of employment which is being provided. For purposes of this policy, a plea of nolo contenderé to a disqualifying criminal record as described above shall be deemed to be a disqualifying event for employment purposes, irrespective of whether adjudication was withheld. Employer will pay all fees associated with conducting a criminal background check on any applicants. Lifting requirement may have a range of 5 75 lbs. Work in 0 degree to 100degree temperatures and possibly rain. Most of the workday is spent in agricultural fields and involves exposure to sun, wind, rain, soil, mud, dust, heat, cold, humid and other natural elements. Workers may stand in one place for long periods of time and must be able to climb, stand, sit, stoop, squat, kneel, crouch, bend (from the waist), push, pull, reach and lift while performing their required job duties. If the employee is unable or unfit to perform the duties listed after the 14 day pretrial, the employee will receive warnings, hours may be reduced to the minimum allowed in the certified petition or terminated. Workers should expect periods of little/no work and hours and days will vary, due to weather conditions beyond employers' control. Proper work attire is required. Workers must wear long sleeve shirts, long pants, no shorts, hard sole shoes, preferably boots and no tennis shoes. No cell phone usage during working hours except during breaks and for emergencies. All tools supplies and equipment will be provided at no cost to workers.

References

Andersson, F., García-Pérez, M., Haltiwanger, J. C., McCue, K., and Sanders, S. (2010). Workplace Concentration of Immigrants. Working Paper 16544, National Bureau of Economic Research.

Biden, J. (2021). Executive Order On Advancing Racial Equity and Support for Underserved Communities Through the Federal Government.

Bruno, A. (2020). H-2A and H-2B Temporary Worker Visas: Policy and Related Issues. Technical report, Congressional Research Service.

Costa, D., Martin, P., and Rutledge, Z. (2020). Federal labor standards enforcement in agriculture. *Economic Policy Institue*.

Denny, M. J. and Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, 26(2):168–189.

Department of Labor (2021). H-2A: Temporary Agricultural Employment of Foreign Workers.

DHS (2019). Table 32 Nonimmigrant Temporary Worker Admissions (I-94 Only) by Region and Country of Citizenship Fiscal Year 2019.

Dolfin, S., Maxwell, N., and Patnaik, A. (2020). WHD Compliance Strategies: Directions for Future Research. Technical Report b7a5ca876e0b448f9b9c0850e920e283, Mathematica Policy Research. Publication Title: Mathematica Policy Research Reports.

Dooling, B. (2020). Agency learning agendas and regulatory research. Technical report, Brookings Institution.

Enamorado, T., Fifield, B., and Imai, K. (2020). fastlink: Fast probabilistic record linkage with missing data.

Hall, M., Musick, K., and Yi, Y. (2019). Living Arrangements and Household Complexity among Undocumented Immigrants. *Population and Development Review*, 45(1):81–101.

Haviland, A. M., Burns, R. M., Gray, W. B., Ruder, T., and Mendeloff, J. (2012). A new estimate of the impact of OSHA inspections on manufacturing injury rates, 1998–2005. *American Journal of Industrial Medicine*, 55(11):964–975.

Johnson, M. S. and Grittner, A. (2020). When Labor Enforcement and Immigration Enforcement Collide: Deterring Worker Complaints Worsens Workplace Safety. SSRN Scholarly Paper ID 3756974, Social Science Research Network, Rochester, NY.

Johnson, M. S., Levine, D. I., and Toffel, M. W. (2020). Improving Regulatory Effectiveness Through Better Targeting: Evidence from OSHA. SSRN Scholarly Paper ID 3443052, Social Science Research Network, Rochester, NY.

Levine, D. I., Toffel, M. W., and Johnson, M. S. (2012). Randomized Government Safety Inspections Reduce Worker Injuries with No Detectable Job Loss. *Science*, 336(6083):907–911. Publisher: American Association for the Advancement of Science Section: Report.

Rathod, J. (2010). Beyond the 'Chilling Effect': Immigrant Worker Behavior and the Regulation of Occupational Safety & Health. *Employee Rights and Employment Policy Journal*, 14(2).

Roberts, M. E., Stewart, B. M., and Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, 91(1):1–40.

Weil, D. (2003). Individual Rights and Collective Agents: The Role of Old and New Workplace Institutions in the Regulation of Labor Markets. Working Paper 9565, National Bureau of Economic Research.

Weil, D. (2010). Improving Workplace Conditions Through Strategic Enforcement. Technical report, Department of Labor.

Weil, D. (2018). Creating a strategic enforcement approach to address wage theft: One academic's journey in organizational change. *Journal of Industrial Relations*, 60(3):437–460.