Mathematica
Progress Together

Brief:

Exploring External Data to Enhance Monitoring and Evaluation of WHD's Compliance Strategies

May 29, 2020

Ankita Patnaik

Submitted to:

U.S. Department of Labor
Chief Evaluation Office
200 Constitution Avenue, NW
Washington, DC 20210
Project Officer: Jessica Lohmann
Contract Number: DOLQ129633249

Submitted by:

Mathematica
1100 1st Street, NE, 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Sarah Dolfin
Reference Number: 50519

This page has been left blank for double-sided copying.

CONTENTS

Tables

EXECUTIVE SUMMARY

In order to promote compliance with labor standards and protect the welfare of workers, the Wage and Hour Division (WHD) of the U.S. Department of Labor (DOL) not only pursues enforcement strategies (such as investigations) but also compliance assistance strategies to support employers to comply with labor standards (such as strategic partnerships with leading brands in an industry). Monitoring and evaluation of strategies can inform WHD's decision-making about resource allocation and strategy development. Monitoring involves ongoing tracking of a strategy's components, including activities conducted during implementation and outcomes observed afterwards, and evaluation involves assessing the extent to which a strategy caused outcomes to change. WHD can use monitoring and evaluation not only to assess whether strategies achieved their intended outcomes, but also to identify whether a strategy's expected outcomes did not materialize because of issues with execution or effectiveness. However, the success of monitoring and evaluation depends on the quality of the information and data used. This brief explores whether and how data that are housed outside of WHD (hereafter referred to as external data) could be integrated with WHD's administrative data to enhance the capabilities of monitoring and evaluation activities to deliver timely, precise, actionable insights to inform WHD's decision-making.

WHD's case management system, the Wage and Hour Investigative Support and Reporting Database (WHISARD), tracks the outcomes of WHD's investigations. Some unique strengths of WHISARD data for monitoring and evaluation purposes are that they provide direct measures of violations that incorporate verifications obtained by WHD's investigative staff, describe the results of investigations that are relevant to WHD's goals, and provide context on investigations and violations. However, as a data source for monitoring and evaluation, WHISARD has some limitations: it only includes some establishments and might not be representative of all establishments; it includes a relatively small number of establishments per year; it contains limited descriptive information about establishments; and it rarely contains observations of the same entity at multiple points in time.

Linking external data in conjunction with WHISARD could strengthen monitoring and evaluation activities. External data could add value to monitoring and evaluation activities by providing helpful information for assessing how representative WHISARD data are of all establishments that fall under WHD's purview, identifying entities that are comparable in characteristics to those that received a WHD strategy, accounting for the characteristics of entities that received the strategy, examining whether the strategy was more effective for some subgroups of entities than others, and developing sampling frames for future investigations. WHD is building capacity to conduct these types of analyses in order to continuously improve the evidence available for planning and evaluation.

To that end, we illustrate how external data on descriptive characteristics of establishments could be further integrated with WHISARD by exploring matching two data sets with WHISARD. We focused on the limited-service restaurant (LSR) industry for this illustrative example, because this is one of WHD's priority areas with extensive WHD activity, many vulnerable workers, and relatively more databases available than other priority industries (Dolfin et al.

2018).[1] We selected two sources of external establishment-level data collected by private companies, CHDExpert and FRANdata, which we assessed to be useful for evaluating a WHD strategy in the LSR industry. Next, we describe each data source and discuss our findings from matching WHISARD data to each of the external data sets.

1. CHDExpert contains a rich set of descriptive information about characteristics of restaurants that could influence employers' decisions about compliance. We assessed that our data extract from CHDExpert had good coverage and included nearly all LSRs in the country. We matched 33 percent of the establishments in WHISARD that were investigated from 2017 to 2019 to an extract from CHDExpert from December 2019.

    - To assess the similarities between matched establishments and other establishments in WHISARD, we compared WHISARD establishments that were matched to CHDExpert to those that were not matched. Compared to investigations of unmatched establishments, investigations of matched establishments were less likely to be part of a WHD initiative and more likely to be agency-directed than complaint-based; the compliance outcomes were similar. WHD investigations of LSRs that could be matched to LSRs in CHD-Expert are not representative of all WHD investigations of LSRs.
    - To assess the representativeness of matched establishments with all LSRs, we compared establishments in CHDExpert that were matched to WHISARD to those that were not. Matched establishments have more employees (on average) and are more likely to belong to a large chain, consistent with WHD's targeting of entities in order to impact the greatest number of workers. A larger share of matched establishments than unmatched establishments are franchise-owned (rather than corporate-owned or independently owned)—consistent with WHD's practice of leveraging a brand's franchising structures. There are no stark differences in the other characteristics examined, such as the establishments' market segment, years in business, or locations.

2. FRANdata contains information that can identify establishments with shared ownership. We assessed that our extract had good coverage of franchising activity. We matched 22 percent of the establishments in WHISARD investigated from 2017 to 2019 to an extract from FRANdata from November 2019.

    - To assess similarities between matched establishments and other establishments in WHISARD, we compared the nature and outcomes of investigations of WHISARD establishments that were matched with FRANdata to those that were not. We found minimal differences.
    - To assess the representativeness of matched establishments with franchised LSRs, we also compared establishments in FRANdata that were matched to WHISARD to those that were not matched. We found only a small difference between the two groups in the share of establishments owned by franchisees who own multiple franchised LSRs.

[1] Limited service restaurants are establishments primarily engaged in providing food services, where patrons typically select items and pay before eating; purchases may be consumed on premises, taken out, or delivered to the customer's location (NAICS 2020).

The results of our matching exercise provide two key takeaways:

1. Only a small share of LSRs observed in WHISARD could be matched with external data from CHDExpert and FRANdata. It is likely that a significant share of establishments could not be matched due to differences in data collection and recording processes for the two data sources (for example, WHISARD data is collected by WHD staff during the investigation process while CHDExpert uses third-party data, open sources and surveys) or due to the matching methods. Small sample sizes of matched data for monitoring and evaluation would mean that the potential usefulness of external data in augmenting WHISARD could be limited by low match rates. We recommend further exploring ways to improve matching between WHISARD and external data, such as by prospectively linking establishments to external data before beginning investigations, and determining whether match rates may be higher with other sources of external data. Larger sample sizes will increase the statistical power of an evaluation to detect small effects of a strategy and provide evaluation designs with more options for a comparison group.
2. Establishments in WHISARD differ from other establishments in the external data from CHDExpert and FRANdata—mainly in characteristics that are directly linked to WHD's purview, priorities, and targeting. Because establishments in WHISARD are not representative of all LSRs, we expect that, for many strategies, analyses of WHISARD data cannot provide reliable evidence on the average effect on the target population of the strategy. This has two implications. First, for future investigations, WHD could consider using external data, for example, to develop a sampling frame from which it could select establishments to investigate. Second, when retrospectively planning analyses of WHISARD data, it is important to account for characteristics that are known to increase the likelihood that an establishment enters WHISARD in the design of analyses and interpretation of findings.

This page has been left blank for double-sided copying.

I. INTRODUCTION

The Wage and Hour Division (WHD) of the U.S. Department of Labor (DOL) promotes compliance with labor standards in order to protect the welfare of workers. WHD not only pursues enforcement strategies such as investigations but also undertakes voluntary compliance assistance strategies to encourage employers to comply with labor standards. DOL's Chief Evaluation Office contracted with Mathematica to support WHD in assessing the opportunities and challenges of conducting an impact evaluation to gauge the effectiveness of a WHD compliance strategy, called the "directions for future research" study. The study views monitoring as ongoing assessments that provide information on the progress or delay of a strategy achieving expected outcomes by following implementation oversight and outcomes after a strategy's implementation, while impact evaluation (hereafter referred to as evaluation) establishes whether a strategy caused outcomes to change.

As part of this study, Mathematica's "WHD Compliance Strategies: Directions for Future Research" report (Dolfin et al. 2020), aims to build an understanding of the factors to consider when developing monitoring and evaluation processes. This brief can be considered a supplement to the directions for future research report (Dolfin et al. 2020), as it expands on and provides more detailed discussion around one factor needed for successful monitoring and evaluation: available, appropriate data. Mathematica has developed this brief to explore the data needs and opportunities for monitoring and evaluation of strategies, with a special focus on understanding the applicability of data that are external to WHD and the possibility of matching them to WHD's own administrative data.

In general, monitoring and evaluation effort will ideally use data that meet the following criteria:

- **Have the correct unit of observation.** The data should contain information for entities to which the strategy is targeted. If the data are at a finer level (for example, the data are at the establishment level while the strategy is targeted toward brands), they can often be aggregated up to the appropriate level—but higher-level data generally cannot be disaggregated to more granular units (for example, data on brands cannot be broken down to get information on establishments).
- **Capture all information relevant to the strategy.** Data should contain measures of the long-, short- and intermediate-term outcomes as well as outputs, activities, and inputs of the strategy, and the context in which the strategy is applied. Ideally, data will contain observations of the same entity from before and after a strategy's implementation.
- **Allow linking to other important information.** For example, data must contain identifying information (such as business names or characteristics) that enables WHD to link them to information about which entities received a strategy. Identifying information would also permit linking multiple data sets to obtain a more complete picture of the strategy's unfolding.

- **Contain a representative sample.** Data should not cover only a select or non-random subset of entities because analyses of such data will provide a skewed picture that is not likely to represent the experiences of the broader population of entities that receive the strategy.
- **Are of high quality.** Data should be complete, reliable, and error-free and be accompanied by complete, transparent, and detailed documentation describing the data collection process and any data processing procedures.

We are not aware of a single data set that meets all the ideal criteria for monitoring and evaluation that we have outlined above—however, by combining WHD's administrative data with other data sets that are external to WHD, WHD might be able to create a merged data set that comes closer to achieving this ideal.

As part of the "directions for future research" study, Mathematica has developed this brief to answer the following research questions:

1. What are the strengths and limitations of WHD's administrative data for the purposes of monitoring and evaluation?
2. What is the potential for data that are external to WHD to play a complementary role to WHD's administrative data in monitoring and evaluation activities?
3. What are the challenges and opportunities in integrating these different data sources?

Section II of this brief addresses the first two questions by describing WHD's administrative data and highlighting ways in which external data could support them in monitoring and evaluation. In Section III, we address the third research question using an illustrative example by exploring two external data sources and their potential for integration with WHD's administrative data. We chose to focus on the limited-service restaurant (LSR) industry for this example because this is one of WHD's priority areas with extensive WHD activity, many vulnerable workers, and multiple sources of external data available (Dolfin et al. 2018).

II. WHD'S ADMINISTRATIVE DATA AND HOW EXISTING EXTERNAL DATA CAN SUPPORT MONITORING AND EVALUATION

A. WHD'S administrative data

The Wage and Hour Investigative Support and Reporting Database (WHISARD) is WHD's case management system. It tracks investigations and records case history, including any violations found by the investigators and any penalties assessed. WHISARD data have three key strengths that make them useful for monitoring and evaluation of strategies:

1. **Provide direct measures of the incidence and number of violations found.** These measures do not rely on reports by affected workers or establishments, but incorporate verifications obtained by WHD's investigative staff. They reflect the outcomes of interest to WHD—violations of the laws that WHD is tasked with enforcing.
2. **Track results from an investigation.** In addition to tracking the violations that occurred, WHISARD tracks back wages and any Civil Money Penalties that were agreed upon as a result of an investigation, which indicate the extent to which the violations were remedied.
3. **Provide context about the investigations and their outcomes.** For example, WHISARD records whether the investigation was agency-initiated or complaint-based. Data reported publicly distinguish complaint versus agency-initiated investigations in aggregate statistics, but not at the individual case level.

However, in general, data from WHISARD also have four key limitations for the purpose of monitoring and evaluation of strategies:

1. **Include only some establishments.** An establishment only enters the WHISARD system as the result of WHD beginning an investigation into the establishment, either because of a WHD-initiated strategy or initiative or because a complaint was filed against the establishment.[2] Establishments included in WHISARD might differ from those not included. For example, establishments in WHISARD might be more likely to have violations than the average establishment. As a result, analyses relying on WHISARD data may suffer from selection bias because the sample of entities being analyzed will not be representative of the broader target population of entities for which we wish to measure the effect of a strategy.

[2] Not all complaints result in investigations, and WHISARD does not track all the complaints that WHD receives. If the complaint could relate to a violation of a law or regulation that WHD enforces, the complaint is assigned a complaint identifier and entered into WHISARD. A decision is then made about whether to create a case. Some complaints (for example, those isolated to a single type of violation for a single employee) are conciliated. Others lead to an investigation covering all employees in the establishment.

2. **Include a relatively small number of establishments.** Because an establishment only enters the WHISARD system as the result of a WHD investigation, and because resource constraints limit the number of investigations that WHD can conduct each year, WHISARD contains fairly small samples, relative to the full population of establishments. For example, in 2016 the Bureau of Labor Statistics reported 231,632 establishments operating in the "limited service restaurant" industry, but WHISARD contained only 1,369 cases of limited service restaurants that were investigated in the same year.[3] Small sample sizes provide an evaluation with limited statistical power, that is, the ability to detect an effect when there is one to be detected. Small sample sizes also limit an evaluation's options for comparison groups, which are usually needed for a rigorous assessment of the effectiveness of a strategy. For example, our efforts to retrospectively evaluate the efficacy of partnership strategies raised significant methodological concerns given the relatively small number of observations in WHISARD for comparison.
3. **Have limited follow-up observations.** There is no consistent follow-up protocol to investigate an entity at multiple points in time. Data on the same entity both before the strategy's implementation and at multiple time points afterward are rarely available for a range of reasons, including resource considerations. As a result, WHISARD data do not consistently capture long-term outcomes or facilitate comparisons of outcomes before and after a strategy.
4. **Have limited data on the characteristics of establishments.** WHISARD has limited information on entities beyond that which is immediately relevant to an WHD investigation. For example, WHISARD contains little data, if any, about the investigated entity's ownership, age, management, workforce, or the local context. This lack of descriptive information about establishments constrains the types of questions that can be answered with the data and limits the analytic methods that can be used. These limitations can ultimately result in findings from monitoring and evaluation that are narrower and less precise.

Some of WHISARD's limitations as a data source for monitoring and evaluation of WHD strategies could be addressed through modifications to WHD's data collection processes. For example, WHD has previously addressed the non-representativeness of WHISARD data by randomly sampling establishments to investigate within a specific industry and, in some instances, conducting reinvestigations or investigating another random sample later. However, such tweaks might be logistically and politically challenging to conduct on a large scale or could conflict with WHD's priorities and performance measures if the tweaks lead to a lower share of violators being investigated, identified, and penalized. Instead, WHD could explore addressing these limitations by using external data that complement WHISARD.

[3] The Quarterly Census of Employment and Wages calculates an annual average of 231,632 privately owned establishments falling under the North American Industry Classification System (NAICS) code 722513 (U.S. Bureau of Labor Statistics 2019).

B. The role of external data in monitoring and evaluation

Below, we explore whether external data could provide information for representative samples and supply descriptive information that can support monitoring and evaluation. If so, then by integrating appropriate external data with WHISARD, monitoring and evaluation might be able to leverage the strengths of multiple data sources and ultimately deliver more precise answers to a broader set of questions. External data present two opportunities for WHD's monitoring and evaluation capabilities:

1. External data can facilitate the collection of direct measures of compliance for representative samples of establishments.

No other private organization or public agency collects direct measures of compliance with the laws that WHD enforces.[4] Therefore, the contents of external data are not expected to make up for the first three limitations of WHISARD outlined above as they are unlikely to contain direct, repeated measures of outcomes for large, representative samples of establishments. However, external data could enable WHD to create a sampling frame for investigations, that is, a master list of all entities from which to draw when deciding upon entities to investigate. This would allow WHD to strategically choose a subset of establishments to investigate from amongst the broader population of all entities under its purview. For example, a master list could facilitate random sampling; in other words, WHD could choose to investigate a small subset of all the entities that could be investigated but choose them in a manner such that each entity in the sampling frame has an equal probability of being chosen for investigation. In this way, external data could help improve the data collected in WHISARD so that they were more representative of all entities. In turn, having outcomes data on a representative sample makes WHISARD's small sample sizes less constraining for monitoring and evaluation, since it is better to have data on small, representative samples than large, unrepresentative samples. Using external data to prospectively create a sampling frame would enable WHD to make informed decisions regarding resource allocation before conducting investigations and planning evaluations.

2. External data can provide descriptive information on entities

External data can directly alleviate the fourth limitation of WHISARD by providing important descriptive information about the characteristics of establishments as well as the context in which strategies are being implemented—both of which may relate to compliance. Many types

[4] External data can provide proxy measures of minimum wage and overtime violations by inferring compliance from reports of wages and hours. Data sources such as the Current Population Survey (CPS) and Survey of Income and Program Participation can be used to derive estimates of minimum wage and overtime violation prevalence, as WHD has done in the past (Eastern Research Group 2014), but indirect proxy measures can pose challenges for an evaluation of a specific strategy. First, the CPS collects data from households as opposed to establishments. It is not possible to observe where the potential violation occurs. Second, there is a limited scope of geographic granularity to assist investigators. For example, it would be difficult to identify which survey respondents worked at an establishment that was targeted by the strategy and, moreover, there would likely be only a small number of such respondents, limiting the ability of an evaluation to provide precise estimates of the strategy's effects.

of descriptive data about establishments could be valuable to WHD (see sidebar). Notably, other data might be relevant to analyses focused on certain industries; for example: a study of restaurants could use data on health code violations, while a study of factories could use data on Occupational Safety and Health Administration violations. Descriptive and contextual data can play important roles in monitoring and evaluation. They enable WHD to:

1. **Strategically choose entities for investigation based on their characteristics.** Information on entity characteristics would facilitate stratified sampling, wherein WHD would divide the sampling frame into non-overlapping groups based on certain characteristics (for example, whether corporate-owned, franchise-owned, or independent) called strata, and then select entities for investigation such that each entity within a stratum has an equal probability of being chosen. WHD has successfully used this methodology previously—for example, in 2012 WHD developed a representative, stratified sample of 300 hotel properties for agency-initiated directed investigations and used the data to estimate a baseline level of compliance. Stratified sampling could be particularly helpful in situations where WHD expects the target population to be heterogeneous and wants to explore and compare the impact of a strategy on specific subgroups.
2. **Assess the extent of WHISARD's selection bias among establishments that have been investigated** If a large share of establishments in WHISARD can be matched to external data on establishment characteristics, then WHD could assess how representative establishments in WHISARD are of an industry by comparing their characteristics to those of establishments that are not in WHISARD. If the WHISARD sample appears to be very different from the broader population of establishments that are eligible for a strategy, then this could inform WHD's thinking about whether and how it may be possible to design monitoring and evaluation activities to produce thoughtful evidence on the effectiveness of a strategy.

Examples of descriptive characteristics of establishments

Ownership
- Corporate-owned or franchise-owned or independently owned
- Years of ownership
- Multiple establishment owner
- Owner's socioeconomic characteristics
- Ownership turnover

Business structure
- Brand-name status
- Franchisee status
- Publicly traded
- Contract status
- Subcontract status
- Subcontract tasks

Workforce
- Number of employees
- Employee occupations
- Management structure
- Workers' socioeconomic characteristics
- Pay structure
- Union membership rates
- Turnover rates
- Shift work
- Use of contingent workers, including staffing services, independent contractors, or temporary employees

Finances
- Annual sales
- Annual profit
- Average check size

Local context

3. **Identify entities similar to those that received the strategy in order to create comparison groups for an evaluation design.** In order to produce evidence on whether a strategy caused outcomes to change, an evaluation must be able to construct a counterfactual condition; that is, an approximation of what would have happened if the strategy had not been implemented. A comparison group (for example, of similar establishments) that has not been exposed to the strategy is one way to capture the counterfactual. While WHD has a long track record of randomly sampling entities for investigations in order to determine compliance rates within industries, the agency has less experience designing evaluations using comparison groups.[5] By providing data on entity characteristics, external data can help WHD construct a comparison group by identifying entities with similar characteristics to those that received the strategy. Further, if WHD's development of strategies can be coupled with the design of evaluation and monitoring activities, this could facilitate experimentation with different strategies or variants of a strategy across comparison groups.
4. **Account for the characteristics of entities that received the strategy.** Sometimes, it is difficult to select a comparison group of entities that is similar across an array of relevant characteristics to the entities that received the strategy. Using data on entities' characteristics, evaluators can use a statistical method called covariate adjustment to isolate how much of the difference in outcomes between the two groups can be attributed to the strategy rather than to differences in characteristics. Once evaluators "subtract" differences in characteristics between the groups from the difference in outcomes between the groups, they can more accurately estimate the true effect of the strategy in changing outcomes. Such adjustment is important if the treatment and comparison groups differ in characteristics that are correlated with compliance, for example, the age of establishments.
5. **Account for differences in implementation and results.** WHD regional and district offices operate in diverse economic, demographic, and political conditions. As a result, implementation of strategies may differ to account for those local conditions. The external factors most relevant to understanding implementation, however, are not necessarily the factors that define the boundaries within WHD regions. From an evaluation perspective, while offices adhere to the guidance and protocols of their WHD region, in practice, their local conditions may more closely approximate areas outside their WHD region. For example, WHD offices operating in different regions may be characterized by a significant number of employers in agriculture, by thousands of restaurants within an urban area, or a large immigrant population vulnerable to wage violations. Each of these offices face unique considerations, opportunities, and challenges when designing and implementing a strategy. Differences in results can also be explained by these characteristics and should be accounted for when assessing the effectiveness of specific strategies or results. Developing a more comprehensive, systematic approach to describing the characteristics of WHD offices and the

[5] The majority of WHD studies have focused on measuring changes in overall compliance and recidivism rates. For example, a prior internal evaluation in the hotel industry compared compliance rates across hotels with different owner and operator characteristics organized around the parent companies of nationally recognized brands. and independent establishments. While the study produced compliance rates, it did not estimate the impacts of WHD strategies within that industry.

factors influencing implementation and results could significantly strengthen analysis and expand options for evaluation design.

6. **Examine whether the strategy was more effective for some subgroups of entities than others.** Subgroup analyses could reveal whether the strategy was more effective in increasing compliance at some types of establishments than others, which could help WHD improve targeting of the strategy. By clearly defining the subgroups, WHD is better positioned to estimate the potential effects of strategies and under what conditions they show better results.

III. INTEGRATING EXTERNAL DATA SOURCES WITH WHISARD TO EXAMINE THE LIMITED-SERVICE RESTAURANT INDUSTRY

WHD may be able to use external data to complement WHISARD for monitoring and evaluation, but in order to take the fullest advantage of the external data, those data need to be linked to WHISARD data. In this section, we examine the applicability of two potential sources of external data and explore the challenges and opportunities in integrating these different data sources. We illustrate how external data on descriptive characteristics can be integrated with WHISARD by exploring matching two data sets with WHISARD. We begin by discussing our rationale for selecting the external data sets used in this exercise and describe the broad challenges in matching establishments in WHISARD to the external data. Next, for each of the external data sets, we describe the descriptive characteristics available and the success of the matching procedure. We then assess whether the WHISARD establishments that were matched to the external data differ in compliance outcomes from those that were not matched, in order to determine the similarities between the matched establishments and all the establishments investigated by WHD. We also explore whether the external data sets' establishments that were matched to WHISARD differ in characteristics from those that were not matched, in order to assess the representativeness of the characteristics of matched establishments with the broader sample in the external data. Finally, we discuss key findings and recommendations.

Our exercise focuses on a single industry because an evaluation within one industry can help WHD develop deeper insights about the association between WHD strategies and compliance and can better isolate the influence of WHD strategies (Dolfin et al. 2019). We chose the LSR industry for this exercise, because this is one of WHD's priority areas with extensive WHD activity, many vulnerable workers, and relatively more databases available than other priority industries (Dolfin et al. 2018).

A. Selection of external data sources

For our matching exercise, we considered several external data sources, paying attention to their coverage, unit of observation, data elements, frequency of updates, and key identifying information. We selected two sources of external data for use in this matching exercise: CHDExpert and FRANdata. Table 1 introduces the characteristics of the two data sets being examined, as well as those of WHISARD. We selected these data sources from among the relevant available data sources that Mathematica explored in earlier work for this study (Dolfin et al. 2018), because we assessed these data sources to be useful for evaluating a compliance

strategy in the restaurant industry, complementary to WHISARD, and resource-effective.[6] The two data sources and their strengths are described below:

- CHDExpert is a data collection and analysis company that specializes in data on the food service industry. Their data include a rich set of descriptive information about restaurants' characteristics that could influence employers' decisions about compliance and establishments' exposure to WHD and labor regulations.
- FRANdata is a data collection and analysis company that specializes in data on U.S.-based publicly traded companies that operate under the franchise business model. Their data make it possible to group franchisees with shared ownership of multiple units, thereby identifying multiple entities owned by a single franchisee, even across multiple brands. With these data, an evaluation could focus on franchising business models and potentially gauge spillover effects.

Table 1. Comparing WHISARD, CHDExpert, and FRANdata

	WHISARD	CHDExpert	FRANdata
Characteristics of the data sources			
Coverage	Establishments investigated by WHD[7]	Food service establishments	Franchise-owned establishments
Unit of observation	Establishment	Establishment	Establishment
Key content	Compliance outcomes	Descriptive characteristics	Ownership characteristics
Update frequency	Continuously	Monthly	At least annually
Establishment identifiers	Name, address	Business name, address	Brand name, address
Characteristics of the extracts used for matching			
Time period	January 2017 – December 2019	December 2019	November 2019

[6] The other data sources we considered were the Chain Store Guide's Restaurant Franchisee PREMIER database, Restaurant Data, WorldFranchising, the Franchisor Database, and Dun & Bradstreet. We did not believe these data sources were as useful for an exploratory study as FRANdata and CHD Expert, for several reasons. Some are at the level of the franchisor and not the establishment (for example, the Franchisor Database and WorldFranchising), meaning we could only match franchisor-level information with WHISARD, not establishment-level information. Some are at the level of franchisee and not establishment (for example, The Chain Store Guide's Restaurant Franchisee Only PLUS database), and little if any establishment-level information would be available. Some (for example, Restaurant Data) focus on data such as restaurant locations and contact information, which are less relevant than descriptive information about the establishments for the purpose of evaluating a WHD strategy. Some data sources (such as Dunn and Bradstreet) were too expensive for a full sample of LSRs, which is necessary to fairly assess match rates with WHISARD.

[7] The WHISARD extract comprised only concluded cases of full investigations that had begun between January 2017 and December 2017.

	WHISARD	CHDExpert	FRANdata
Industry segment	Limited service restaurants (NAICS code 725113)	Limited service restaurants (Global Foodservice Classification codes 2310, 2320, and 2330)	Limited service restaurants (NAICS code 725113)

Source: WHISARD, CHDExpert, and FRANdata.

Note: The WHISARD extract comprised only concluded cases of full investigations that had begun between January 2017 and December 2017.

NAICS = North American Industry Classification System.

B. Challenges with matching establishments in WHISARD to external data

In matching observations in WHISARD to external data, we encountered three major challenges. We describe these challenges and our approaches to resolving them below:

1. **Matching by name and address can be messy.** The key identifying information on which we can match observations are text fields: the name, address, and zip code of each establishment. Matching on text fields can be difficult as those fields can include typographical errors, alternative permutations, missing or extra spaces, lower- or upper-case letters, homonyms, spelling mistakes or alternatives, and other inconsistencies. Matching text fields despite these inconsistencies requires significant data cleaning that is resource intensive. Further, establishments' names might legitimately differ from one data set to another, especially when they are part of a franchise operation—for example, one data set might use establishments' trade names while the other uses legal names. To ensure that minor inconsistencies did not prevent us from matching the same establishment between WHISARD and external data, we did not require these fields to match exactly across data sets. Instead, we undertook a process of approximate string matching (colloquially referred to as "fuzzy" matching), a technique of finding strings that match a pattern approximately rather than exactly.[8] Notably, compared to requiring an exact match, fuzzy matching may result in

[8] We used the statistical software Stata and employed the *matchit* command (Raffo 2015) to apply the three-gram method for decomposing a vector of text, which splits text into pieces of three moving characters. For example, a business name such as Jungle Cafe would be split into "Jun," "ung," "ngl," "gle," "le c," "e ca," "caf," and "afe." We inversely weighted each vector's elements according to the frequency of its occurrence; for example, an element such as "Cafe," which appeared very frequently in the data, would be given lower weight in the matching process than the word "Jungle," which appeared more infrequently in the data and is therefore more likely to be unique to an establishment. Compared to requiring an exact match, fuzzy matching might result in some false positives, that is, matches of observations across two data sets that are not in reality the same establishment. It is not possible to assess the extent of false positives without manually reviewing all matched observations, which is labor intensive. However, we attempted to minimize the rate of false positives by setting a conservative threshold for accepting a match. Stata's *matchit* command provides a similarity score between two different text strings by performing many different string-based matching techniques and returning a numeric variable containing the similarity score, which ranges from 0 to 1. We required a similarity score of 0.9 or greater for each match; we chose this threshold in order to be relatively conservative and minimize false positives. Our illustrative matching exercise utilized one matching method and did not examine alternative matching methods, such as matching based on edit distances, where strings are matched based on the minimum number of single-character edits (insertions, deletions or substitutions) that would be required to change one string into the other (Levenshtein 1966).

some false positives, that is, matches of observations across two data sets that are not in reality the same establishment, and the rate of false positives may vary based on the matching method and parameters selected. For our illustrative example, we selected relatively conservative parameters for fuzzy matching, which enabled us to find matches across the two data sets despite minor inconsistencies across data sources.

2. **WHISARD data are historical whereas data from CHDExpert and FRANdata are a current snapshot.** WHISARD data might include establishments that have closed since WHD's investigation and are not present in the external data, which present a snapshot of establishments that were operating in late 2019. The inclusion of older WHISARD data would mechanically result in a lower match rate and could lead us to underestimate the ability to match current WHISARD data to current external data. Further, this could lead to survival bias in the matched data if some establishments' closing was non-random and influenced by WHD activities (see box below). To maximize the match rate and minimize the risk of such bias, we restricted our WHISARD data to investigations that began in 2017 or later, resulting in a sample of 2,173 establishments.

Survival bias

Many LSRs investigated before 2017 may not have survived until late 2019, as the median lifespan of an LSR is about 3.75 years (Luo and Stark 2014). Consider the following example of how survival bias can be a threat in matched data. We might expect an establishment's profit to be correlated with not only compliance (because the establishment has more incentive to stay in business and more resources to develop procedures for compliance) but also business survival. Thus, the higher the profit of an establishment, the more likely it is to be compliant and the less likely it is to close. Accordingly, a WHD investigation with back wage findings is more likely to contribute to the closing of low-profit establishments than high-profit establishments. When WHISARD data from multiple years are merged with a current snapshot of external data, high-profit establishments that were investigated are more likely than low-profit establishments that were investigated to have survived, be captured in the current snapshot, be matched, and end up in the merged data. The final merged data would then be non-representative of all establishments investigated by WHD; the data would skew toward LSRs that have higher profits and better compliance than the average LSR.

3. **The three extracts have different coverage.** First, WHISARD only includes FLSA-covered establishments, while CHDExpert and FRANdata contain data on establishments regardless of whether they are covered by the FLSA. Thus, we expect CHDExpert and FRANdata to include establishments that would never be included in WHISARD.[9] Second, the data sources use different systems of industry classification. WHISARD and FRANdata use the North American Industry Classification System (NAICS), and CHDExpert uses the Global Foodservice Classification (GFC) system—so it is not straightforward to define the extracts

[9] We are unable to whittle down the CHDExpert and FRANdata extracts to only FLSA-covered establishments; CHDExpert provides only categorical information on the number of employees and annual sales, while FRANdata has no information relevant to FLSA coverage.

to perfectly overlap with one another. We used our judgement to similarly define the extracts in order to cover all LSRs.

C. Example: Matching establishments in WHISARD to CHDExpert

1. Introduction to CHDExpert

CHDExpert collects data on all types of food service operators.[10] We used our judgement to define an extract that covered LSRs based on GFC codes.[11] We requested an extract of restaurants comprising "quick service restaurants, fast food, snacks"; "fast casual"; and "delivery and takeaway only" by CHDExpert. [12] Our extract contained 296,818 restaurants that could be uniquely identified through a combination of business name, main address, and zip code.

The CHDExpert data offer two key advantages for monitoring and evaluation. They:

1. **Offer a rich set of descriptive information on restaurants.** The data capture restaurant characteristics that might influence employers' decisions about compliance and employers' exposure to WHD and wage and hour regulations. For example, the data indicate whether an establishment has been in business for many years, which suggests that it has more experience with labor regulations. The data also capture characteristics of the area in which the establishment is located, which might also influence employers' decisions. For example, hypothetically, employers might be more inclined to commit violations in areas where there is a higher degree of competition from other restaurants as indicated by a high operator density. Table 2 illustrates the types of restaurant characteristics captured in the CHDExpert data and describes our requested extract of LSRs.
2. **Cover nearly all restaurants.** Our extract contained 296,818 establishments while the quarterly census of employment and wages reported about 247,000 establishments in the first quarter of 2019 (U.S. Bureau of Labor Statistics 2019). This suggests that CHDExpert covers nearly all (if not all) entities in the universe of LSRs.

[10] CHDExpert uses multiple data collection methods. They purchase comprehensive data from sources such as Dunn & Bradstreet, InfoGroup, Database USA, and Data.gov; use open sources such as Yelp, Google, and OpenTable; and conduct surveys. They exclude food trucks from their data.

[11] We assessed that GFC codes 2310, 2320, and 2330 (LSR–Quick service restaurants, fast food, snacks; LSR–Fast casual; and LSR–Delivery and takeaway only, respectively) correspond to NAICS code 722513 (in the NAICS 2017 classification system), while GFC codes 2341 (LSR–Ice cream parlors, frozen desserts), 2342 (LSR–Coffee shops, teahouses), 2343 (LSR–Smoothie, Juice), 2350 (LSR-Self-service restaurants) and 2360 (LSR–Buffet restaurants), correspond to NAICS codes 722514 and 722515.

[12] CHDExpert decides upon the industry classification of the food operator based on its primary product (as defined by sales volume).

However, the data from CHDExpert also have some drawbacks. Namely, they:

1. **Only include establishments currently in operation.** CHDExpert data represents a snapshot of establishments that are currently in business. CHDExpert data would not capture any restaurants that have closed since they were investigated by WHD; this reduces the potential for matching observations from WHISARD to CHDExpert and poses a risk of survival bias in the matched data (see above text box). This risk could be mitigated by using additional external data that are longitudinal in nature, such as the National Establishment Time-Series database, which can identify establishment openings and closings.[13]
2. **Capture the current characteristics of establishments.** A restaurant's current characteristics might differ from its characteristics at the time of the WHD investigation. Evaluators should be careful in interpreting such characteristics in matched data if they might have been influenced by the investigation—that is, if they might be outcomes rather than independent descriptive characteristics. For example, an establishment that owed significant back wages in the past for overtime violations might decide to hire new workers to avoid overtime liabilities; the current number of employees would therefore be an outcome of the prior investigation and should not be considered an independent characteristic.
3. **Have missing data.** A handful of variables in the data have a high degree of missing information. For example, the variable capturing the number of employees at the establishments was uncoded for more than 45 percent of establishments.

Table 2. Characteristics of LSRs in the CHDExpert extract from 2019

Establishment characteristics	Share of sample (percentage)
Market segment	
Quick service restaurants	78.2
Fast casual	20.5
Delivery and takeaway	1.3
Annual sales	
$500,000 or less	45.3
$500,001–$1,000,000	29.6
$1,000,001–$2,500,000	21.8
More than $2,500,000	3.4

[13] NETS is a time-series database, constructed from Dun and Bradstreet (D&B) archival establishment data, that provides longitudinal data on various dynamics of the U.S. economy that include establishment job creation and destruction, survivability of business startups, and other items.

	Share of sample (percentage)
Number of employees[a]	
1–4	9.1
5–9	11.1
10–19	14.5
20–49	17.1
50 or more	2.1
Number of units	
Independent (1–9 units)	43.7
10–50	3.8
51–100	2.4
101–250	3.3
251–500	3.8
More than 500 units	42.9
Ownership type	
Corporate-owned	16.3
Franchise-owned	41.0
Independent	42.6
Years in business	
Less than 1 year	4.8
1 year	5.8
2–4 years	21.1
5 or more years	68.3
Average check	
Under $5	5.9
$5–$6	32.6
$7–$9	54.4
$10 or more	7.1
Zip code characteristics	
Average household income	
Less than $30,000	0.8
$30,001–$60,000	42.8

	Share of sample (percentage)
$60,001–$80,000	32.0
$80,001–$100,000	16.8
More than $100,000	7.6
Density of other restaurants (Index; 100 = national average)	
Less than 50	8.4
51–100	11.1
101–150	12.7
More than 150	67.8
Population	
Less than or equal to 10,000	13.2
10,001–20,000	18.0
20,001–30,000	22.6
30,001–50,000	31.9
More than 50,000	14.2
Food away from home spend (Index; 100 = national average)	
Less than 50	0.2
51–100	48.1
101–150	43.0
151–200	7.2
More than 200	1.5
Sample size	**296,818**

Source: CHDExpert.

[a] This information is missing for more than 45 percent of establishments.

4. **Use the GFC system instead of the NAICS.** It is difficult to define industry-specific extracts to overlap with other data sources, which primarily use NAICS codes. We used our judgement to identify LSRs using GFC codes.[14]

[14] We assessed that GFC codes 2310, 2320, and 2330 (LSR–Quick service restaurants, fast food, snacks; LSR–Fast casual; and LSR–Delivery and takeaway only, respectively) correspond to NAICS code 722513 in the NAICS 2017 classification system (Limited Service Restaurants), while GFC codes 2341 (LSR–Ice cream parlors, frozen desserts), 2342 (LSR–Coffee shops, teahouses), 2343 (LSR–Smoothie, Juice), 2350 (LSR-Self-service restaurants) and 2360 (LSR–Buffet restaurants) correspond to NAICS codes 722514 (Cafeterias, Grill Buffets, and Buffets) and 722515 (Snack and Nonalcoholic Beverage Bars.).

5. **Cannot be used to definitively identify FLSA-covered establishments.** CHDExpert does not record granular information on annual revenue and number of employees, which would allow us to identify which establishments are covered by the FLSA and which therefore fall under WHD's purview. We conservatively estimate that, at a minimum, 31 percent of the establishments in this extract would be covered by the FLSA.[15]

2. Results from matching cases in WHISARD to CHDExpert

We were able to match data from 33 percent of establishments in WHISARD (investigated from 2017 to 2019) to CHDExpert data using fuzzy matching of the names, addresses, and zip codes of establishments (Table 3).[16] The majority of matches (18 percent of establishments) were achieved using the trade name, address, and zip code of establishments, while a significant portion (12 percent of establishments) of matches were based on only the trade name and zip code of establishments.[17] More than 65 percent of establishments in our WHISARD sample could not be matched to CHDExpert data, with several possible explanations as to why. It is likely that a significant share of establishments could not be matched due to differences in data collection and recording processes for the two data sources (for example, WHISARD data is collected by WHD staff during the investigation process while CHDExpert uses third-party data, open sources and surveys) or due to the process of approximate string matching. We do not assess that the match rate was significantly dampened due to some establishments' closing or moving since they were observed in WHISARD, because the match rates did not vary considerably across cases in WHISARD across years.[18]

To explore the extent to which our matching methods determined the rate of matches, we conducted two sensitivity checks that illustrated the potential match rates through more or less stringent matching protocols. First, we attempted to directly match WHISARD data to CHDExpert data, that is, to search for matches that had exactly the same business name, address, and zip code. This stricter matching protocol resulted in a lower match rate of about 17 percent (not shown in Table 1). This suggests that using less strict methods such as fuzzy matching can be a valuable means to boost the rate of matches between WHISARD and external data. Second, to illustrate the potential for matching using less stringent matching methods, we assessed how sensitive the match rate was to our threshold for the similarity score required by the matching program, which can range from 0 (no similarity required) to 1 (a perfect match). When we lowered the threshold for the score from 0.9 to 0.8, the match rate increased from 33 percent to

[15] This estimate is based on the fact that the data indicate that 31 percent of the establishments have more than five employees and have annual sales of at least $500,000. This estimate is conservative for two reasons. First, for about 22 percent of the sample, we are unable to determine FLSA coverage due to missing data on the number of employees. Second, we expect that some of the establishments that are recorded as having 1–4 employees in the CHDExpert data would not represent self-employment and therefore would also be covered by the FLSA.

[16] We used the variables ER_TRADE_NAME, ER_LEGAL_NAME, ER_ADDRESS, and ER_ZIP from WHISARD, and the variables BUSINESSNAME, MAINADDRESS, and ZIPCODE from CHDExpert.

[17] ER_TRADE_NAME provided substantially more matches than ER_LEGAL_NAME; the latter's marginal contribution was less than 2 percent of matches.

[18] The match rates for establishments investigated in 2017, 2018 and 2019 were 33%, 32% and 34% respectively.

52 percent. This represents a significant improvement in the match rate, but some share of the additional matches may represent false positives. This illustrates one trade-off that should be considered for future matching efforts: it is possible to improve the rate of matches by tweaking the parameters of the matching process, but this may have implications for the quality of the matches. To address the risk of false positives, future matching efforts can consider manual inspection of matched cases to estimate the rate of false positives and can explore whether combining fuzzy matching with machine learning can mitigate this rate.

Table 3. Match statistics from matching establishments in WHISARD to CHDExpert

Count of eligible WHISARD establishments to be matched (NAICS 722513; 2017–2019)	2,173
Percentage of eligible WHISARD establishments that were matched	32.63
Shares of WHISARD establishments matched using the following sets of variables (percentages):[a]	
WHISARD: ER_TRADE_NAME, ER_ADDRESS, ER_ZIP CHDExpert: BUSINESSNAME, MAINADRESS, ZIPCODE	18.36
WHISARD: ER_LEGAL_NAME, ER_ADDRESS and ER_ZIP CHDExpert: BUSINESSNAME, MAINADRESS, ZIPCODE	0.97
WHISARD: ER_TRADE_NAME, ER_ZIP CHDExpert: BUSINESSNAME, ZIPCODE	12.43
WHISARD: ER_LEGAL_NAME, ER_ZIP CHDExpert: BUSINESSNAME, ZIPCODE	0.87
Percentage of eligible establishments in WHISARD that were not matched	67.37

Source: WHISARD and CHDExpert.

[a] Each row shows the additional share of establishments in WHISARD that were successfully matched by using the variables indicated. These numbers may not sum to the total percentage of eligible WHISARD establishments in WHISARD that were matched due to rounding.

It is important to understand the similarities between matched establishments and other establishments in WHISARD—that is, are the matched establishments representative of the establishments investigated by WHD? To address this question, we compared the establishments in WHISARD that were matched to those we were unable to match with CHDExpert data, in terms of the nature of their WHD investigations and their outcomes (Table 4). We tested whether the two groups differed in the characteristics of their investigations and found two statistically significant differences. First, investigations of matched establishments were less likely to be part of a WHD initiative and, second, they were more likely to be agency-initiated than complaint-based, compared to unmatched establishments. However, the two groups had similar compliance outcomes. The average matched and unmatched establishment had about three FLSA violations, with a miniscule share of establishments having repeat or recurring violations. About 82 and 85 percent of matched establishments and unmatched establishments agreed to pay some back wages, respectively, with small differences in the average amount of back wages that establishments in each group agreed to pay for violations, per employee with violations.

Table 4. Investigations and outcomes of establishments, by whether they were successfully matched between WHISARD and CHDExpert

	Matched establishments in WHISARD	Unmatched establishments in WHISARD
Type of investigation		
Part of a WHD initiative (percentage)	16.88	11.67*
Source of investigation		†
Agency-initiated (percentage)	22.73	16.78
Complaint-based (percentage)	77.27	83.22
Violation characteristics		
Number of violations	2.65	2.86
Repeated violations (percentage)	0.42	0.41
Recurring violations (percentage)	0.00	0.07
Agreed to pay any back wages (percentage)	81.73	84.81
Back wages agreed to pay for all violations	$242	$271
Back wages agreed to pay for all violations per employee with violations	$126	$142
Sample size	717	1,448

Source: WHISARD.

† Chi-squared test indicates statistically significant difference across categories at the .05 level.

* Two-tailed t-test indicates statistically significant difference at the .05 level.

To assess the representativeness of the matched establishments with all establishments in the LSR industry, we compared the characteristics of establishments in CHDExpert that were matched to WHISARD to those we could not match with WHISARD (Table 5). We tested whether the two groups differed in the characteristics of LSRs and found some statistically significant differences. Several differences were found in characteristics linked to WHD's purview, priorities, and targeting. First, matched establishments were less likely to have annual revenue of under $500,000, which is expected because the FLSA does not apply to establishments with revenue under $500,000 and so WHD would not investigate such establishments for FLSA violations.[19] Second, matched establishments have more employees on average, and they are more likely to belong to a chain with 500 or more units and are less likely to belong to a chain with fewer than 10 units—consistent with the idea that WHD targets its enforcement activities in order to impact the greatest number of workers. Third, matched establishments are more likely to be franchise-owned rather than corporate-owned or independently owned.

[19] We expect no establishments investigated by WHD to have annual sales under $500,000. The non-zero share of matched establishments with annual sales under $500,000 suggests discrepancies in the timing of observation or methods of data collection or variable definitions between WHISARD and CHDExpert.

Table 5. Characteristics of establishments, by whether they were successfully matched between CHDExpert and WHISARD data

	Matched establishments in CHDExpert (percentage)	Unmatched establishments in CHDExpert (percentage)
Characteristics of the restaurant		
Market segment		†
Quick service restaurants	75.45	78.19
Fast casual	24.13	20.52
Delivery and takeaway	0.42	1.29
Annual sales		†
$500,000 or less	27.75	45.30
$500,001–$1,000,000	32.08	29.57
$1,000,001–$2,500,000	37.66	21.77
More than $2,500,000	2.51	3.37
Number of employees		†
1–4	9.49	16.84
5–9	14.84	20.57
10–19	27.25	26.92
20–49	43.31	31.73
50 or more	5.11	3.94
Number of units		†
Independent (1–9 units)	27.34	43.77
10–50	3.63	3.85
51–100	2.23	2.44
101–250	4.04	3.30
251–500	3.63	3.80
More than 500	59.14	42.84
Ownership type		†
Corporate-owned	11.16	16.36
Franchise-owned	62.48	40.97
Independent	26.36	42.66
Years in business		
Less than 1 year	3.07	4.84
1 year	5.30	5.83
2–4 years	20.36	21.08
5 or more years	71.27	68.24

	Matched establishments in CHDExpert (percentage)	Unmatched establishments in CHDExpert (percentage)
Average check		†
Under $5	6.456	5.94
$5–$6	39.47	32.58
$7–$9	47.28	54.40
$10 or more	6.69	7.08
Characteristics of the location (zip code)		
Average household income		†
Less than $30,000	0.70	0.76
$30,001–$60,000	50.63	42.81
$60,001–$80,000	30.96	31.98
$80,001–$100,000	12.13	16.82
More than $100,000	5.58	7.63
Density of other restaurants		
Less than 20	1.12	2.33
21–50	4.74	6.07
51–100	11.58	11.10
101–150	11.72	12.72
More than 150	70.85	67.78
Population		
Less than or equal to 10,000	12.13	13.21
10,001–20,000	18.55	18.03
20,001–30,000	21.90	22.63
30,001–50,000	34.17	31.89
More than 50,000	13.25	14.24
Food away from home spend (Index; 100 = national average)		†
Less than 50	0.14	0.21
51–100	54.39	48.08
101–150	39.19	43.00
151–200	5.30	7.25
More than 200	0.98	1.45
Sample size	**717**	**296,101**

Source: CHDExpert data.

† Chi-squared test indicates statistically significant difference across categories at the .05 level.

Due to the low match rate, it is not possible to assess the extent to which differences in characteristics between matched and unmatched establishments are explained by factors that determine selection for investigation by WHD or factors that influence the likelihood of being successfully matched across data sources. For example, there are two possible explanations for why the share of franchise-owned establishments is higher among matched establishments in CHDExpert than unmatched establishments. It may be that franchise-owned establishments are more likely to be investigated by WHD than independently owned establishments, because WHD considers franchising characteristics of entities when targeting investigations. However, it is also possible that franchise-owned establishments are easier to match across data sources because they are more likely to follow consistent naming patterns that are suggested or required by the brand or franchisor.

Nonetheless, Table 5 suggests that the LSRs that we matched between WHISARD and CHDExpert (that is, LSRs that we know have definitely been investigated by WHD) differ in characteristics from other LSRs (that may or may not have been investigated by WHD). This provides suggestive evidence that some types of establishments are more likely to be investigated by WHD (and thus observed in WHISARD) than other types of establishments. Therefore, any evaluation of a strategy must carefully consider whether the entities observed in WHISARD are representative of all the entities eligible for the strategy. For strategies that are targeted very broadly, our findings suggest that the current setup of WHISARD may not enable WHD to provide a rigorous answer to the question, "*Can this strategy improve compliance at the average eligible establishment?*"

Ideally, WHD could explore how the selection of establishments for investigation can be designed to make WHISARD data more suitable for evaluation purposes. For example, since CHDExpert appears to cover nearly all (if not all) entities in the universe of LSRs, it could be used to develop a sampling frame of all LSRs from which WHD can select LSRs to investigate using random sampling. Alternatively, WHD could ask investigators to link LSRs to CHDExpert data before beginning their investigation; this would boost match rates and also enable verification of data (such as the address of the LSR) in WHISARD and CHDExpert. If it is not possible to change these processes, then WHD could carefully select analytic methods that can be used retrospectively to account for LSR characteristics that influence the likelihood of being selected for investigation. Table 5 suggests that the matched and unmatched establishments in WHISARD primarily differ in characteristics that are linked to WHD's purview, priorities, and targeting of strategies. This suggests that it should be possible to collect data on the characteristics that determine selection into WHISARD and then apply analytic methods such as covariate adjustment. However, other factors might determine selection into WHISARD that are not observed in the CHDExpert data or are even unobservable, so this may not fully resolve concerns around the non-representativeness of WHISARD data as it is currently collected.

D. Example: Matching establishments in WHISARD to FRANdata

1. Introduction to FRANdata

FRANdata collects data on all franchise brands that are publicly or privately held. We requested an extract of its data on all LSRs, corresponding to the NAICS industry code 725113.[20] This provided a sample of 140,158 establishments that were uniquely identified by a combination of brand name, business address, and zip code.[21]

Coverage of FRANdata extract

- 350 brands
- 140,158 establishments
- 45,428 franchisees
- 19,069 franchisees who own more than one franchised LSR

Data from FRANdata have several key advantages. Consider that they:

- **Use NAICS codes for industry classification.** It is easy to define industry-specific extracts to overlap with WHISARD and other external data that use the NAICS codes.
- **Include important contextual information about establishments.** FRANdata covers a narrow set of descriptive information about establishments, but what it covers is important because franchising structure and characteristics are important context for understanding establishments' compliance outcomes. Multiple research studies (Weil 1996, 2005, 2008, 2009, 2010, 2012, 2014; Ji and Weil 2015, 2009; Weil and Mallo 2007; Weil and Pyles 2005) and WHD experience indicate that industry structures and business models create incentives and opportunities for employers that can influence compliance. For example, one can use these data to account for the fact that franchisors who own multiple franchised establishments tend toward better compliance (Ji 2011). Further, if a WHD strategy's targeting or implementation is differentiated by any aspect of the franchise business model, WHD could use data from FRANdata to identify franchisees or account for such differentiation.
- **Cover nearly all franchised establishments.** All actively franchising brands are required to publish a franchise disclosure document (FDD), in which they must provide contact information for current franchisees; FRANdata collects data from the FDDs. We assess our

[20] FRANdata uses its own internal industry classification system, which is more incremental than NAICS. A brand can be in more than one industry, but FRANdata assigns a primary classification using its judgement based on the brands' self-description of the product offerings.

[21] FRANdata's data exist at the level of establishment-franchisee. If, for example, an establishment is co-owned by three franchisees, that establishment will appear as three rows in the data, with one row representing each establishment-franchisee combination. Our extract includes 161,675 establishment-franchisees.

extract has good coverage: it includes 141,372 LSRs, whereas the CHDExpert data included 121,757 franchise-owned LSRs.[22]

However, data from FRANdata also have some limitations. Consider that they:

1. **Do not contain the business names of establishments.** Establishments can usually be uniquely identified by a combination of brand name, business address, and zip code.[23] However, matching to WHISARD is more difficult without the business name; matching must be conducted based on business address and the name of the franchisor brand.
2. **Only include establishments currently in operation.** This reduces the potential for matching observations from WHISARD to FRANdata and poses a risk of survival bias in the matched data (see above for a discussion of this issue for CHDExpert).
3. **Capture the current characteristics of establishments.** An establishment's current franchising characteristics might differ from its franchising characteristics at the time of the WHD investigation (see above for a discussion of this issue for CHDExpert).
4. **Cannot be used to definitively identify FLSA-covered establishments.** FRANdata has no information on annual revenue and number of employees that would allow us to identify establishments that are covered by the FLSA, and so fall under WHD's purview.

2. Results from matching cases in WHISARD to FRANdata

We were able to match 22 percent of establishments in WHISARD (investigated between 2017 and 2019) to the FRANdata extract using fuzzy matching of the brand names, addresses, and zip codes of establishments (Table 6).[24] The majority of matches (14 percent of establishments) were achieved using the brand name, address, and zip code of establishments, while a portion (8 percent of establishments) of matches were based on only the brand name and zip code of establishments. [25] About 78 percent of establishments in our WHISARD sample could not be matched to the FRANdata extract. We expect the match rate of WHISARD to FRANdata to be lower than the match rate of WHISARD to CHDExpert, because a significant share of establishments in WHISARD are not franchise-owned and would not be covered by FRANdata.

[22] The discrepancy might be due to different data collection methods for the two data sources or because the extracts were defined using different industry classification systems.

[23] We found fewer than 100 combinations of BUSINESS_ADD1 BUSINESS_ZIP BRANDNAME that were associated with more than one UNITID. In some of these cases, BUSINESS_ADD2 could help to uniquely identify separate establishments.

[24] We used the variables ER_TRADE_NAME, ER_LEGAL_NAME, ER_ADDRESS, and ER_ZIP from WHISARD, and the variables BRANDNAME, BUSINESS_ADD1, and BUSINESS_ZIP from FRANdata.

[25] ER_LEGAL_NAME contributed no matches.

Table 6. Match statistics from matching establishments in WHISARD to FRANdata

Count of eligible WHISARD establishments to be matched (NAICS 722513; 2017–2019)	2,173
Percentage of eligible WHISARD establishments that were matched	21.95
Shares of WHISARD establishments matched using the following sets of variables (percentages):[a]	
WHISARD: ER_TRADE_NAME, ER_ADDRESS, ER_ZIP FRANdata: BRANDNAME, BUSINESSADD1, BUSINESSZIP1	13.99
WHISARD: ER_LEGAL_NAME, ER_ADDRESS and ER_ZIP FRANdata: BRANDNAME, BUSINESSADD1, BUSINESSZIP1	0.18
WHISARD: ER_TRADE_NAME, ER_ZIP FRANdata: BRANDNAME, BUSINESSZIP1	7.69
WHISARD: ER_LEGAL_NAME, ER_ZIP FRANdata: BRANDNAME, BUSINESSZIP1	0.09
Percentage of eligible establishments in WHISARD that were not matched	78.05

Source: WHISARD and FRANdata.

[a] Each row shows the additional share of establishments in WHISARD that were successfully matched by using the variables indicated. These numbers may not sum to the total percentage of eligible WHISARD establishments in WHISARD that were matched due to rounding.

It is important to understand the similarities between matched establishments and other establishments in WHISARD—that is, are the establishments that we matched between WHISARD and FRANdata representative of the establishments investigated by WHD? We compared the nature and outcomes of investigations of establishments in WHISARD that were matched with FRANdata to those of establishments we could not match (Table 7). We tested whether the two groups differed in the characteristics of LSRs and found minimal differences between the two groups. Further, because FRANdata only includes franchised establishments, matched establishments must be franchise-owned but unmatched establishments need not be franchise-owned; therefore, differences between the two groups could also reflect disparities in outcomes of establishments that are franchise-owned rather than independent or corporate-owned.

To assess the representativeness of the matched establishments with all franchised LSRs, we compared the characteristics of establishments in FRANdata that were matched to WHISARD to those that did not match with WHISARD (Table 8). We tested whether the two groups differed significantly in the share of establishments owned by franchisees who also own other franchised LSRs—which we know is a characteristic that tends to be associated with better compliance outcomes (Ji and Weil 2018). We found a small but statistically significant difference in this characteristic between the matched and unmatched establishments. This suggests that franchise-owned LSRs included in WHISARD are likely to have a larger share of establishments owned by multiple-franchise owners, as the broader group of all franchise-owned LSRs.

Table 7. Investigations and outcomes of establishments, by whether they were successfully matched from WHISARD to FRANdata

	Matched establishments in WHISARD	Unmatched establishments in WHISARD
Type of investigation		
Part of a WHD initiative (percentage)	14.79	13.02
Source of investigation		
Agency-initiated (percentage)	18.54	18.82
Complaint-based (percentage)	81.46	81.18
Violation characteristics		
Number of violations	2.56	2.85
Repeated violations (percentage)	0.00	0.53
Recurring violations (percentage)	0.21	0.00
Agreed to pay any back wages (percentage)	83.75	83.79
Back wages agreed to pay for all violations	$218	$273*
Back wages agreed to pay for all violations per employee with violations	$123	$140
Sample size	480	1,690

Source: WHISARD.

* Two-tailed t-test indicates statistically significant difference at the .05 level.

Table 8. Characteristics of establishments, by whether they were successfully matched from FRANdata to WHISARD

	Matched establishments in FRANdata	Unmatched establishments in FRANdata
Establishment is owned by a franchisee who owns multiple other franchises	86.67	83.33*
Sample size	480	139,678

Source: FRANdata

* Two-tailed t-test indicates statistically significant difference at the .05 level.

E. Discussion

Overall, the examples in Sections C and D above demonstrate that it is possible to find data sources external to WHD that contain information that can complement WHD's administrative data. The extract from CHDExpert provides a glimpse of how some external data can provide a rich set of descriptive characteristics that can help WHD assess the extent of WHISARD's selection bias, construct comparison groups for evaluations, and conduct covariate adjustment and subgroup analyses. The extract from FRANdata shows the potential for franchising data to be used to identify establishments with certain ownership characteristics and to establish the inter-connectedness of establishments. Next, we summarize three key findings from this exercise.

Integrating external data with WHISARD is resource intensive. The costs of integrating external data will depend on the costs of three key components: the data extracts, integration procedures, and labor. External data sources vary in their pricing and pricing structure. Although our exercise highlighted the cost of a one-time pull of extracts from two external data sources, it would likely take considerably more resources for WHD to purchase, clean, and integrate external data sources for a more comprehensive monitoring or evaluation exercise, for example, one that examined multiple industries or multiple years. Costs can vary depending on the specifics of data processing, for example, the methods used for cleaning data, matching observations, and conducting quality checks. However, once the matching algorithm has been developed it can be used many times and for many applications at low cost.

Not all cases in WHISARD could be matched with external data. Only a small share of LSRs observed in WHISARD could be matched with external data from CHDExpert and FRANdata. The resulting matched data had small sample sizes and would offer limited statistical power for monitoring and evaluation. As a result, at present, reliable lessons from merging WHISARD data with external data from these two sources are limited. If the match rates from the current exercise are like those that could be obtained with other external data, the potential usefulness of external data in augmenting WHISARD could be limited by low match rates. We recommend further exploration of means by which WHD might be able to improve the match rate between WHISARD and external data. The following are a few illustrative examples:

- Request "snapshot" extracts of external data at multiple points in time, so that WHISARD data and external data could have the same coverage in terms of time period. If WHD plans to collect cross-sectional data such as CHDExpert, we recommend that WHD consider requesting multiple years of data or developing multiyear agreements to get repeated cross-sections of data.
- Record trade name, establishment owner, and franchise owner information on establishments in WHISARD, which would provide more identifying information on which one could attempt matching establishments to external data.
- Explore variations in the matching process, such as lowering the similarity score requirement to 0.75, using other methods for approximate string matching, or recognizing name variants (for example, Elizabeth, Beth, Liz, and Lizzy) and resulting differences in match rates. In addition, we recommend exploration of the trade-off between match rate and quality for each of these variations, for example, by estimating the rate of "false positive" matches by conducting manual inspections of a sample of matched cases.

Establishments in WHISARD differ from other establishments. Currently**,** some types of establishments are more likely to be investigated by WHD (and thus observed in WHISARD) than other types of establishments. As a result, depending on the strategy being tested and how it was targeted, WHISARD data may not be representative of the entities eligible for the strategy. This makes the current setup of WHISARD unsuitable for supporting rigorous evaluation efforts

that can answer broad questions, such as, "Can this strategy improve compliance at the average establishment?"

The findings from our matching exercise underscore the importance of appropriately accounting for the characteristics that determine selection into WHISARD when designing monitoring and evaluation activities. Ideally, WHD could consider making changes to the planning process for investigations to help make WHISARD data more suitable for evaluation purposes. For example, WHD could use external data such as CHDExpert to develop a sampling frame of all restaurants, from which it could select establishments to investigate using random or stratified sampling. Alternatively, WHD could establish a protocol of pro-actively linking establishments to external data before an investigation begins, which would facilitate efforts to validate both the external data and WHISARD data and improve the eventual match rate.

If it is not possible to change WHD's processes for investigation and data collection, then WHD could carefully select analytic methods that can account for the characteristics that currently influence the likelihood that an establishment will be selected for investigation. One positive finding from our matching exercise is that the establishments in WHISARD that we were able to match to external data mainly differed from other establishments in WHISARD in characteristics that are directly linked to WHD's purview, priorities, and targeting. This suggests that it should be possible to identify and collect data on the characteristics that determine selection into WHISARD. The following are examples of characteristics that appear to predict selection for investigation and how WHD could account for them:

- Establishments in WHISARD tend to have more revenue and more employees because these characteristics determine FLSA coverage. If WHD can use external data on revenue and employees to identify an appropriate comparison group, an evaluation of a specific strategy can yield important, precise, and actionable insights about the effectiveness of that strategy in improving compliance among establishments under WHD's purview.
- Establishments in WHISARD are more likely to be franchise-owned because WHD often uses the franchise network to implement certain strategies. If WHD can use external data on franchising to identify an appropriate comparison group, an evaluation of a specific strategy can yield actionable insights about the effectiveness of that strategy in improving compliance among establishments typically targeted for that strategy.

However, as we noted earlier, it is not possible to account for unobservable characteristics that may determine selection for investigation by WHD. Retrospectively accounting for observed characteristics that determine selection for investigation by WHD can provide only a second-best solution to this problem. Changes to WHISARD's data generating processes could result in more representative and comprehensive data to support monitoring and evaluation of its strategies.

REFERENCES

Dolfin, Sarah, Nan Maxwell, and Ankita Patnaik. "WHD Compliance Strategies: Directions for Future Research." Washington, DC: U.S. Department of Labor, January 3, 2020.

Dolfin, Sarah, and Nan Maxwell. "WHD Compliance Evaluation (CE) Study Priorities Memo." Memorandum to Jessica Lohmann and Karen Livingston, U.S. Department of Labor, July 23, 2018.

Dolfin, Sarah, Nan Maxwell, Alix Gould-Werth, Armando Yañez, Jonah Deutsch, and Libby Hendrix. "Compliance Strategies Evaluation Literature and Database Review." Report submitted to U.S. Department of Labor. Washington, DC: Mathematica Policy Research, October 18, 2018.

Eastern Research Group. "The Social and Economic Effects of Wage Violations: Estimates for California and New York." Washington, DC: U.S. Department of Labor, 2014.

Levenshtein, Vladimir I. "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady, vol. 10, no. 8, 1966, pp. 707–710.

Ji, Min Woong, and David Weil. "Does Ownership Structure Influence Regulatory Behavior? The Impact of Franchising on Labor Standards Compliance." Working paper. Boston, MA: Boston University, 2009. Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.422.6544. Accessed November 14, 2019.

Ji, Min Woong, and David Weil. "The Impact of Franchising on Labor Standards Compliance." *ILR Review,* vol. 68, no. 5, 2015, pp. 977–1006.

Ji, MinWoong. "Impacts of Multi-Unit Franchising on Workplace Compliance Behavior." Working paper. Boston, MA: Boston University, 2011.

Ji, MinWoong, and David Weil. "Does Ownership Structure Influence Regulatory Behavior? The Impact of Franchisee Free-Riding on Labor Standards Compliance." Working paper. Boston, MA: Boston University, 2012.

Luo, T., and P. B. Stark. "Only the Bad Die Young: Restaurant Mortality in the Western US." Ithaca, NY: Cornell University Library, 2014. Available at https://arxiv.org/ftp/arxiv/papers/1410/1410.8603.pdf. Accessed November 14, 2019.

Raffo, Julio. "MATCHIT: Stata Module to Match Two Datasets Based on Similar Text Patterns." Statistical Software Components S457992. Boston College Department of Economics, 2015, revised April 2019.

U.S. Bureau of Labor Statistics. "Quarterly Census of Employment and Wages." Washington, DC: U.S. Bureau of Labor Statistics, 2019. Available at https://www.bls.gov/cew/. Accessed January 2, 2020.

Weil, David. "A Strategic Approach to Labour Inspection." *International Labour Review,* vol. 147, no. 4, 2008, pp. 349–375. Available at http://onlinelibrary.wiley.com/doi/10.1111/j.1564-913X.2008.00040.x/full. Accessed March 14, 2018.

Weil, David. "Examining the Underpinnings of Labor Standards Compliance in Low Wage Industries: Final Report." Report to the Russell Sage Foundation, 2012. Available at http://www.russellsage.org/sites/all/files/Weil.Final20Report202012.pdf. Accessed October 9, 2018.

Weil, David. *The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done to Improve It*. Cambridge, MA: Harvard University Press, 2014.

Weil, David. "If OSHA Is So Bad, Why Is Compliance So Good?" *RAND Journal of Economics*, vol. 27, no. 3, 1996, pp. 618–640. Available at http://www.fissuredworkplace.net/assets/Weil.If-OSHA-So-Bad.RAND.1996.pdf. Accessed March 14, 2018.

Weil, David. "Improving Workplace Conditions Through Strategic Enforcement." Boston University School of Management Research Paper No. 2010-20, 2010. Available at https://www.dol.gov/whd/resources/strategicEnforcement.pdf. Accessed March 14, 2018.

Weil, David. "Public Enforcement/Private Monitoring: Evaluating a New Approach to Regulating the Minimum Wage." *ILR Review,* vol. 58, no. 2, 2005, pp. 238–257. Available at http://journals.sagepub.com/doi/pdf/10.1177/001979390505800204. Accessed March 14, 2018.

Weil, David. "Rethinking the Regulation of Vulnerable Work in the USA: A Sector-Based Approach." *Journal of Industrial Relations,* vol. 51, no. 3, 2009, pp. 411–430. Available at http://journals.sagepub.com/doi/pdf/10.1177/0022185609104842. Accessed March 14, 2018.

Weil, David, and Carlos Mallo. "Regulating Labour Standards via Supply Chains: Combining Public/Private Interventions to Improve Workplace Compliance." *British Journal of Industrial Relations,* vol. 45, no. 4, 2007, pp. 791–814. Available at http://www.fissuredworkplace.net/assets/Weil_Mallo.Regulating-Supply-Chains.BJIR.2007.pdf. Accessed March 14, 2018.

Weil, David, and Amanda Pyles. "Why Complain? Complaints, Compliance, and the Problem of Enforcement in the U.S. Workplace." *Comparative Labor Law & Policy Journal,* vol. 27, no. 59, 2005, pp. 59–92.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org