Linking Administrative Data to

Understand Effects of Seattle's $15 Minimum Wage Ordinance

Jennifer Romich[1], Anne Althauser[1, 2], Scott W. Allard[2], Mark C. Long[2], and Jacob Vigdor[2]

with major intellectual contributions from members[†] of

THE MINIMUM WAGE STUDY
UNIVERSITY of WASHINGTON

[1]School of Social Work

[2]Evans School of Public Policy & Governance

University of Washington

March 2018

Summary: This manuscript describes how the University of Washington Minimum Wage Study plans to use merged data from multiple state administrative sources to examine policy and social science questions about the implementation and impact of the Seattle $15 Minimum Wage Ordinance.

Key words: Minimum wage, administrative data, poverty and inequality

---

† The team investigators are Jacob Vigdor, Mark C. Long, Jennifer Romich, Scott W. Allard, Heather D. Hill, Jennifer Otten, Robert Plotnick, Scott Bailey, and Anneliese Vance-Sherman**.** Any opinions expressed in this report are those of the authors and not those of the University of Washington, the Washington Employment Security Department, or any supporting or contracted entity. To contact the Seattle Minimum Wage Study Team, email mwage@uw.edu**.**

In 2014, the Seattle City Council passed and its mayor approved a Minimum Wage Ordinance (henceforth 'Ordinance') outlining a series of steps whereby employers would have to, over the course of several years, increase workers' pay to at least $15 per hour. With this measure, Seattle became the first major American city to raise wages to this level for all private-sector employees within its jurisdiction. Other cities, notably San Francisco, Minneapolis, and New York, have since followed.

During the public deliberation over the Ordinance, advocates on both sides of the wage mandate made competing claims that mirror discussion in the ever-growing research literature around minimum wages. Proponents argued that a higher minimum wage would increase workers' well-being, make them better able to care for their children, and reduce growing income inequality. Opponents countered that businesses would struggle to remain profitable, and employers might have to hire fewer workers, raise prices, or leave the city (Thompson and Martinez, 2014).

Adjudication between these counterarguments requires answers to several fundamental policy research questions. For example, how will individuals, households, firms, and the market respond to higher minimum wage rates? How will a higher minimum wage shape the income distribution and broader patterns of well-being in Seattle and the region? To ensure research evidence would be brought to bear on these questions, Seattle's City Council passed a Resolution alongside the Ordinance calling for, "academic researchers who have a proven track record of rigorous analysis of the impacts of minimum wage laws to conduct an evaluation of the economic impacts," (Office of City Clerk, 2014). The Resolution specified areas of particular interest, including standard outcomes of interest in the minimum wage literature such as employment and business openings and closings, as well as more specific topics such as impacts on public assistance and the well-being of immigrant and refugee workers.

In 2015, the Seattle Office of the City Auditor awarded a contract to the University of Washington Minimum Wage Study team (UW-MWS), staffed by seven co-investigators and including several authors here, for the initial phase of the research. A mixture of public and private foundation funds has supported the UW-MWS's work to date (Minimum Wage Study 2017). Apart from examining how wage increases affect employment and earnings, which are standard in the policy research literature around the minimum wage, the UW-MWS is pursuing a range of questions touching on matters such as working families' daily routines, participation in means-tested benefit programs, impacts on the nonprofit human service sector, spillover of the city-level mandate on the surrounding area, and the well-being of vulnerable sub-populations, such as immigrant and refugee workers. These highly practical questions for the workers, employers, and citizens of Seattle correspond to contested or unsettled questions in the economics, social policy, and allied social science literature about the effects and impacts of minimum wage policies on firms, workers, and the distribution of income and well-being across households.

Conventional data sources used to examine minimum wage impacts, however, permit investigation of only a subset of these topics. Therefore, to provide new insight into short-and long-term impacts of the Ordinance and its relationship to worker and family well-being, the UW-MWS is creating an omnibus dataset assembled from the unusually high-quality extant administrative data available in Washington State and combining these administrative data with other publicly-available data. Because of the proliferation of higher state and local minimum wage laws, we expect many researchers also will seek to link administrative data to answer questions about impact. The purpose of this manuscript is to describe how the UW-MWS will use administrative data in combination with other data sources to address some of the limitations present in existing publicly available data sources and explore questions of interest to social science research on poverty and inequality.

The outline of this paper is as follows. We begin with a brief overview of the major questions, methodological approaches, and data sources prominent in the extant minimum wage research evidence base as well as questions not yet well addressed due to data limitations. Next, we describe our data innovation, an omnibus dataset compiled from multiple sources of Washington State administrative data. As of the writing of this draft (December 2017) the merging work is expected to start in early 2018 with initial data transfers to the UW-MWS before mid-year. We describe the component data sources as well as the legal and institutional parameters that shape the planned merging process. In addition, we provide some illustrations of planned analyses using this omnibus merged administrative data. The paper finishes with a discussion of some of the conceptual, practical, and political challenges that emerge when using complex administrative data as part of a multi-method research project.

MINIMUM WAGE EVIDENCE

Debates over the minimum wage reflect debates in the underlying character of labor markets. The standard neoclassical model of "perfect competition" imagines that no firm can hire a worker unless that worker is paid the full value of what they produce for the firm. A worker offered less than this amount will decline the offer given the expectation that another firm can improve on it. In this view, asking the firm to pay more than this amount is tantamount to asking them to lose money, and firms will respond by cutting back on their use of labor, or confining their hiring to those employees whose productivity justifies the higher wage.

Firms may, however, have some ability to pay workers less than the value of what they produce. In this scenario, a minimum wage policy serves to improve workers' bargaining power relative to employers, and if calibrated properly might have no adverse effect on employment – in fact, in a monopolistic labor market a perfectly calibrated minimum wage could lead employers to increase hiring. Monopsony power is only one of several variations on the standard neoclassical model raising doubts about its ability to forecast the impact of the minimum wage.

Given this fundamental theoretical debate over the nature of the labor market, labor economists have turned to empirical analysis to study whether raising the minimum wage causes firms to reduce employment. Empirical studies to date have come to inconsistent conclusions. David Neumark and William Wascher have authored a series of analyses over time reporting negative impacts of higher minimum wages on employment and earnings (Neumark and Wascher 1992; Neumark and Wascher 2000; Neumark and Wascher 2007; Neumark, Salas and Wascher 2014). By contrast, David Card and Alan Krueger's study of fast food restaurants in New Jersey and Pennsylvania, following a minimum wage increase in the former state, found a positive effect (Card and Krueger 1994). These more benign findings have been corroborated by some additional work (Card and Krueger 2000; Dube et al. 2010; Allegretto et al. 2011). Efforts to reconcile divergent findings in the literature suggest that the effects of the minimum wage on labor market outcomes may become more negative after the passage of time from implementation (Neumark and Wascher 2008; Meer and West 2015).

While prior literature has focused much attention on how dollars flow from firms to workers in the wake of a minimum wage increase, only a few studies have asked key questions about how businesses adapt to higher wage rates, or how the lives of workers and their families change as a consequence of the policy. Several studies have analyzed whether businesses, particularly restaurants, raise prices in the wake of minimum wage increases – with evidence suggesting they do. Two analyses show little evidence that minimum wage increases lead to cutbacks in non-monetary compensation including pension and health benefits, perhaps because so few low-wage jobs offer these perquisites in the first place (Acemoglu and Pischke 2003; Simon and Kaestner 2004). There are many other possible avenues of adjustment – altering business hours, the quality of products or services, replacing employees with contractors – which have received little if any attention in scholarly literature.

There is remarkably little evidence on the fundamental question of whether higher minimum wages materially improve workers' lives. Employees in poverty stand to gain much from higher wages, but the marginal tax rates embedded in traditional safety net programs may offset income gains, even if employment opportunities remain plentiful (Hill and Romich 2018). Previous studies report conflicting but overall modest results on the question of whether higher minimum wages lift families out of poverty, though evidence does suggest reduced reliance on public benefits (Neumark, Schweitzer and Wascher 2005; Card and Krueger 1995; Burkhauser and Sabia 2007; Sabia 2008; Dube 2013; West and Reich 2014).

Even if higher minimum wage increases post-tax-and-transfer household income, questions remain regarding whether and how this might translate into improved quality of life for affected families. As noted above, reductions in means-tested benefits will offset some increases in earnings. Some portion of the income gains to families might be further offset by minimum wage-induced price increases. Studies examining fundamental well-being outcomes including access to health care and health status indicators do not present a consensus view (McCarrier et al. 2011; Kim and Leigh 2010; Meltzer and Chen 2011; Cotti and Tefft 2013).

Estimating the Causal Impact of Minimum Wages

A variety of approaches have been used to evaluate the impact of minimum wage ordinances. Most studies use some variant of a difference-in-differences methodology, whereby changes in outcomes in the area that introduces a minimum wage policy change are contrasted with changes in outcomes in a comparison region that did not alter its minimum wage. The changes in outcomes in the comparison region are included as a "counterfactual," thereby estimating what would have likely occurred in the treated area in the absence of a policy change, and are included to capture any macroeconomic trends that may corrupt a simpler "first difference" analysis.

The major threat to validity of using this approach is that the post-policy time trend for outcomes in the control region may not be a reasonable counterfactual. That is, it might not be reasonable to assume that the treated region would follow the path of the control region in the absence of the policy change. Researchers have been particularly concerned with the prospect of divergent pre-policy trends. To address this concern, researchers often conduct falsification tests to evaluate whether there were significant pre-policy divergence, include higher-order polynomial time-trend terms to parametrically capture the pre-policy trends, or use the synthetic control method of Abadie, Diamond, and Hainmueller (2010) to identify control regions with similar pre-policy trends of the treatment region.

These methods still may lead to spurious findings in the presence of endogenous policy changes (i.e., minimum wage policies adopted in anticipation of particular trends in future outcomes) or contemporaneous policy changes or economic changes unrelated to the minimum wage which occur in either the treatment or control regions. Such events would invalidate the assumption of parallel trends between these regions. For detailed discussions of these methodological challenges and critiques of methods used by scholars in this area, see Neumark, Salas, and Wascher (2014a, 2014b) and Allegreto, Dube, Rech, and Zipperer (2017).

Data limitations often make it difficult to distinguish low-wage and higher-wage workers. As a result, most studies in this literature have focused on industries where minimum wage workers are common, such as in fast-food restaurants (e.g., Card and Krueger, 1994), or have focused on subgroups, such as teenagers, who are more likely to contain low-wage workers (e.g., Allegretto, Dube, and Reich, 2011). However, the elasticity of labor demand may be quite different in these industries and/or for these subgroups. If so, results for these subgroups may not generalize to other workers, lessening the capability of such research to answer general policy questions.

Data used in empirical analyses of minimum wage policies

Data on wages and employment can be collected either from workers or businesses, and prior minimum wage studies have utilized both types. While in some cases researchers perform their own data collection – as in Card and Krueger (1994) and the survey/interview components of the Seattle Minimum Wage Study – it is more common to employ secondary datasets collected by government agencies. Table 1 below shows the most common five data sources used across 113 studies reviewed by Belman and Wolfson (2014).

Over half the studies reviewed use Current Population Survey (CPS). The CPS samples between 70,000 and 100,000 households each month – just under 0.1% of the population – and returns to the same street address to re-interview households as many as eight times over 16 months (four months on, eight months off, then four more months on). Twice in each 16-month period, when a household is included in the “outgoing rotation group,” individuals are asked whether they are paid by the hour, and if so their base hourly wage rate. While the CPS includes only a tiny fraction of the workforce at any point in time, the sample is reasonably representative of the population. Moreover, data include not only wages but other demographic and socioeconomic characteristics of individual workers and the other members of their households. These advantages make it a strong choice for looking at national or state-level changes in wage policies, but the sample size precludes examination of local efforts. Two other panel studies commonly used in minimum wage research, the National Longitudinal Survey of Youth (NLSY) and the Study of Income and Program Participation (SIPP) are even more limited in their ability to examine sub-national policy changes.

The Bureau of Labor Statistics conducts repeated cross-sectional surveys of employers that can be used to track economic trends over time. The Current Employment Statistics (CES) data are collected as a monthly survey of business establishments (Bureau of Labor Statistics 2017). CES data are then used to track employment, hours, and earnings nationally and for state and metro areas.

Table 1. Most common data sources used in minimum wage research

	Studies using (%)
Current Population Survey	57%
Bureau of Labor Statistics*	11%
Survey of Income and Program Participation	5%
Quarterly Census of Employment and Wages	5%
National Longitudinal Survey of Youth	6%

Note: Percentages sum to over 100 because some studies use multiple datasets.
Source: Author tabulations of studies (N=113) reviewed by Belman & Wolfson (2014). For inclusion, studies had to appear in B&W tables about microeconomic outcomes and be based on US data. The five most common sources are shown.
*BLS includes BLS, BLS-EEH (Employee Earnings and Hours), and BLS-CES (Current Employment Statistics) series.

State administrative data offer many advantages over most publicly available data sources. Most relevant are quarterly employment data businesses are required to submit to state agencies administering Unemployment Insurance (UI) benefit programs. UI provides some unemployed workers with benefits as they seek new jobs. Unemployed workers must qualify for benefits by having held their previous job for a particular duration, and their benefits are based on their earnings at that job. To determine eligibility and compute benefit amounts, the agencies overseeing the program must track employment and earnings for each worker. Apart from their value to the UI system, these data provide insight into employment and earnings within firms. UI administrative data contain some geographic information, which can permit researchers to study local minimum wage ordinances. These data also allow for detailed subgroup analyses by industry.

UI data reports also feed several other data resources relevant to the study of minimum wage laws. For example, the Quarterly Census of Employment and Wages (QCEW) is based on UI data and1624 presents aggregate statistics on employment, with some breakdowns by geography and industry. State agencies also may share individual-level data directly with researchers, or by contributing their data to a Federal resource, the Longitudinal Employer-Household Dynamics file, a non-public use resource available to approved researchers in Federal Statistical Research Data Centers.

Employer wage records are comprehensive but sparse. Only four states – Washington, Oregon, Minnesota, and Rhode Island – require employers to report hours data for each employee. The ability to accurately compute average hourly wages is critical for rigorous assessment of minimum wage law impact. Moreover, UI data cover most of the private sector workforce, but include only information on employment and earnings – no information about household demographics, characteristics, or public program participation. Thus, UI datasets can answer only a limited set of questions about earnings if they are not linked to other data sources. Indeed, much of the minimum wage research to date has relied either on employer records that are comprehensive but reveal little about workers themselves, or household surveys that offer depth at the expense of vastly reduced breadth. We argue that merged administrative records offer the possibility of linking individual workers together into households, necessary to study concerns about household-level income and poverty that rest at the core of the debates about minimum wage laws today.

AVAILABLE ADMINISTRATIVE DATA RESOURCES

Advancing the evidence base on how local minimum wage laws affect poverty and inequality necessarily makes several key requirements of data. First, sufficient numbers of observations in the affected area as well as comparison areas are necessary; major secondary data sets such as the CPS rarely have the former, particularly for subgroups. In the case of evaluating a city-level minimum wage, knowing whether a worker lives or works in the city is relevant to the aspirations of a policy that seeks to make a city's citizens better off. Second, data should contain earnings information as well as unearned income as transfer programs provide important support for lower-income families. Because researchers typically measure poverty at the household level, rather than at the individual level, data should have workers grouped into households along with non-working adults and minor children. Finally, because of the roles that age, race, ethnicity, gender, and nativity play in stratification systems, knowing these demographic characteristics helps provide a more complete picture of policy impact. In this section, we describe the Washington State data resources that, when combined, will meet these data criteria. These data, when combined, have the potential to capture impacts across Washington's 7.3 million residents, of which 0.7 million reside in Seattle (Seattle 2017; Washington (state) 2017).

Washington State UI Data and Findings to Date

The UW-MWS draws upon UI data administered by the Washington State Employment Security Department (ESD) as a principal data source on employment and earnings. Washington State UI data contain quarterly earnings data along with total quarterly hours necessary to calculate hourly wages and distinguish low-wage employees from those who work only short periods. In 2016, these records covered just over 3.1 million of the estimated 3.4 million Washington workers (Washington State Employment Security Department 2017). Workers not in the UI records include federal workers, private household workers, many self-employed or contract workers, and certain publicly paid in-home caregivers.

UI data have several key strengths for studying labor market outcomes. Capturing nearly all wage employment, they provide a census (not a sample) and hence allows for precise estimates with minimal bias. The data are longitudinal and can be analyzed at the worker or employer level. Because only three states other than Washington State maintain UI data with information about hours worked, the UW-MWS is in a unique position to assess the impact of the Ordinance on employment.

Early UW-MWS work on employment outcomes draw on these UI data (Jardim, Long, Plotnick, van Inwegen, Vigdor, and Wething 2017a, 2017b, 2017c; Jardim and van Inwegen, 2017). Jardim et al. (2017a) finds that the Ordinance prompted a reduction in "hours worked in low-wage jobs by around 9 percent, while hourly wages in such jobs increased by around 3 percent. Consequently, total payroll fell for such jobs, implying that the minimum wage ordinance lowered low-wage employees' earnings by an average of $125 per month in 2016" (abstract). Further, the study finds "an effect of zero when analyzing employment in the restaurant industry at all wage levels, comparable to many prior studies" (abstract), which suggests that methods used in prior studies might miss the contraction in hours worked prompted by the minimum wage ordinance. Jardim et al. (2017b) evaluates whether and to what extent the Ordinance had effects on local areas surrounding the city. In particular, the authors evaluate whether it caused a change in average hourly wages and hours worked by low-wage workers in areas outside of the city limits. In preliminary results, Jardim et al. (2017b) find modest and insignificant evidence that there was positive spillover on wages (i.e., wages were increased) in surrounding areas within a 40-minute drive of Seattle. Jardim et al. (2017b) find more robust and statistically significant evidence for a negative spillover on hours. These results suggest that studies using boundary-discontinuity methods might produce attenuated estimates of wage and employment impacts of local minimum wage policies.

Adding Personal, Household, and Geographic Information

As noted, UI data alone cannot answer many key questions about poverty and inequality of interest to social scientists. Such data include only individual-level records so distinguishing primary versus secondary earners is not possible nor are dependents tracked. Without household information, poverty rates remain incalculable. UI data lack information about the demographic or social characteristics of workers beyond wage level and industry. For low-income households who rely on a combination of earnings and means-tested benefits, UI data can only partially describe household economic well-being because the data do not contain public benefit information. Age, race, ethnicity, gender, and nativity play large roles in labor market success, but UI data cannot speak to these dimensions. The lack of these demographic characteristics limits the ability to evaluate distributional effects of minimum wage increases. Furthermore, while UI data show where worksites are for employees at locatable firms (i.e., single-site firms and those multi-site firms opting for separate accounts by location), they do not contain information about where workers live, making it impossible to estimate impacts on residents of Seattle or any other locale.

The UW-MWS omnibus administrative dataset resolves many of these limitations by merging UI data with data from the state voting records, Department of Licensing (Licensing) drivers' license database, public assistance and child support records, vital statistics, and select criminal justice systems. Such linked data can locate workers geographically, assemble workers into households, and assign demographic characteristics. In this section, we describe each of these data sources.

Of critical importance are data from the Research and Data Analysis (RDA) division of Washington's Department of Social and Health Services (DSHS). RDA maintains an integrated client database (ICDB), one of four states with such data (Becker-Green, 2012), which contains information across several agencies within DSHS, including Economic Services Administration, which administers public assistance. DSHS links the individuals in this data to their records in external data from different sources including the Department of Corrections, the Department of Health, ESD, and the Department of Commerce (which administers select housing and homelessness programs) (Mancuso, 2014). This allows DSHS to serve internal and external research needs, preparing reports on topics such as employment status of state Medicaid clients or homelessness among former foster youth (DSHS, n.d.). DSHS also facilitates the ICDB's use by university-based researchers with interests in client populations.

In order to understand impacts of the Ordinance on labor market and family outcomes for DSHS clients and non-clients, the research team made independent data requests to the other state agencies. The UW-MWS will augment UI and ICDB data with other public sources, summarized in Table 2 and Figure 1.

Voting records and identification records from the Department of Licensing provide demographic and address data for a large proportion of state residents. Voter registration data from the Washington Secretary of State includes date of birth and address. As of the November 2016 election, 4.3 million Washingtonians were registered to vote, representing an estimated 77 percent of the 5.6 million residents of voting age (Washington Office of the Secretary of State, n.d.). These data are available on a monthly basis, with updated address information for any registered voter who files a change of address with the U.S. Post Office.

Licensing data in the form of driver licenses and state identification card records contain gender, date of birth, eye color and address data. Washington had 5.6 million licensed drivers, over 619,000 state ID card holders, and just under 90,000 persons holding instructional driving permits in 2016 alone (Washington State Department of Licensing, 2016). Washington drivers are supposed to update their driver's license when they move (although many do not) and they are required to get a new driver's license every six years.

The economic assistance and other program data contained in the ICDB will make an important contribution to understanding low-income families' financial well-being. Approximately 28 percent of Washington residents are clients of the DSHS Economic Services Administration. Data include participation in and monthly amounts for major means-tested programs, most notably the Basic Food, which serves 1.4 million Washington residents with federal Supplemental Nutrition Assistance Program (SNAP) benefits or comparable state-funded benefits (DSHS 2015). DSHS also serves just over a million clients with child support enforcement. Smaller percentages of residents receive other assistance such as Temporary Assistance to Needy Families (TANF), child care, and other programs. DSHS also maintains records on the almost 2.1 million residents who qualify for the state Apple Health program, which includes Medicaid, expanded Medicaid, and the State Children Health Insurance Program (SCHIP).

Our initial data request will also include selected birth and criminal records which the team will use for initial explorations of effects on health and crime. Birth statistical files from the Department of Health provide birth and health information on both children and parents. Birth information provides valuable information on parents that is unavailable in other administrative data such as parent's race, parent's birth country, parent's education, and current residential address. For health information, birth files provide birth vital statistics such as birth weight, gestation, and hypertension, as well as mother's behaviors such as smoking, drinking, and prenatal visits. While we will only receive birth data back to 2010, these data will provide information on around 700,000 children born and 1.4 million adults (author calculation using Washington State Department of Health, 2017).[1]

Finally, data from the Washington State Patrol will include select criminal history, including state convictions and prison terms. We plan to assess the quality and

[1] Per DSHS staff, health records prior to 2010 do not contain sufficient data to allow for reasonably certain linkages.

comprehensiveness of this data in early analyses. If such analyses suggest additional data would be of helpful for understanding the interplay between wage regimes and criminal system involvement, we may seek additional data from county-level arrest or jail records.

Table 2. Data by Agency

STATE AGENCY/ DATASET	KEY INFORMATION
Employment Security Department	**Hours and wages for each individual job in WA State based on Unemployment Insurance (UI) records**
Voter Registration	**Names and addresses of individual registered voters.**
Department of Licensing	**Name, birth, sex, and address information for WA State driver's license and state ID holders**
Economic Services Administration	**Integrated Client Database (ICDB) containing poverty reduction and self-sufficiency public assistance programs**
Child Support Services (CSS)	Collection and distribution of child support. Data includes arrears owed, monthly order amount, and case status.
Temporary Assistance for Needy Families (TANF) State Funded Assistance (SFA) WorkFirst	Temporary cash assistance and job search programs. Data includes monthly amount paid.
Supplemental Nutrition Assistance Program (SNAP) Food Assistance Program for Legal Immigrants(FAP)	Food stamps and nutrition assistance. Data includes monthly amount paid.
Child Care Subsidy Program	Helps families pay for quality childcare. Data includes monthly amount paid.
Supplemental Security Income (SSI)	Federal income supplement to help aged, blind, and people living with disabilities. Data includes monthly amount paid.
Basic Food Employment and Training (BFET)	Employment readiness opportunities. Data includes monthly amount paid.
Refugee and Immigrant Assistance (RIA)	Cash and program aid to refugees and immigrants. Data includes monthly amount paid.
Health Care Authority	**Individuals using Medicaid and Apple Health insurance.**
Department of Health	**Birth and death records including some statistical information on children and parents.**
Washington State Patrol	**Individual criminal history including charges and prison terms.**

In sum, the available Washington State administrative data resources will allow us to assemble close to a census of potentially affected Washingtonians. Adults in the data will include employees covered by the UI system; residents who hold state identification cards, notably driver licenses; registered voters; parents of children born in in-state in this decade; and clients of DSHS. We believe this is a reasonably comprehensive list of all adults potentially affected by the Ordinance. Children in the data will include those born in-state in 2010 or later and those in households that receive DSHS services. Many school-age and older children from moderate- to high-income households will be absent; a limitation that we may address using state educational data in a future iteration. In the next sections we overview key legal parameters for using this data as well as the practical steps involved in merging across systems and generating the desired household, geographic, and demographic indicators.

LEGAL PARAMETERS

Understandably, a set of legal restrictions and privacy concerns set parameters around how and to whom the state agencies named above can release data. The Washington State Institutional Review Board, which oversees human subjects research involving DSHS client records or staff, reviewed the research study and is responsible for assembling the data sharing agreement across the seven different state agencies or administrations. Data use must align with relevant regulations, including state and federal statutes. The study team and state agency partners had to work creatively to develop merging protocols (summarized in Figure 2) consistent with regulations and research goals.

We negotiated a tri-party agreement between UW, Licensing, and DSHS in order for Licensing to share data with DSHS to merge with the assistance data. Per federal regulations on the protection of human subjects (45 CFR, Part 46), state law on the release of records for research (e.g., Chapter 42.48, Revised Code of Washington), and with rules for the protection of human subjects codified in the Washington Administrative Code (e.g., Chapter 388-04 WAC), UW is not permitted to receive identifiable records from DSHS. Given these regulations, UW is unable to merge the Licensing and DSHS data ourselves (DSHS 2003). Furthermore, according to the Driver's Privacy Protection Act (DPPA), Licensing data can be used "in research activities, and for us in producing statistical reports, so long as the personal information is not published, redisclosed, or used to contact individuals" so we are not permitted to share the Licensing records with DSHS (NY Department of Motor Vehicles, n.d.). Using these laws and restrictions as guidance, we wrote a tri-party agreement to enable the merging of two datasets that UW is unauthorized to receive separately with identifiable data. In the agreement, we state the following procedures:

> Licensing will provide Data to DSHS directly. DSHS will add their client data and remove all Licensing-provided Personal Information before forwarding the Data to UW. UW will reimburse Licensing for all costs associated with researching, analyzing, compiling and providing the Data. Licensing will apprise UW of approximate costs in advance of performing work.

The United States Social Security Administration (SSA) puts restrictions on how Social Security Numbers (SSNs) are shared or disclosed without consent. Given we are using data on the entire state of Washington; we included a waiver for signed consent in our IRB given the infeasibility of collecting consent from a population of over seven million people. Due to the sensitivity of the SSN as a unique personal identifier, Licensing is only able to share the last four digits of SSN in the data they send to DSHS. While full SSN makes linking persons across datasets much more reliable, we are instead linking people using only the last four digits of SSN that Licensing can share, with birthdate (month/day/year) and name (U.S. Social Security Administration 2013). UW signed a confidentiality agreement with all participating agencies to ensure we would not have or recreate a key that would allow us to reverse engineer the identity code and thereby have the ability to figure out identities. Since merging will happen in house at DSHS, there won't be any reason that UW will ever need any identifiers.

Health Insurance Portability and Accountability Act (HIPPA) limits the amount and type of identifiable health information the Department of Health (DOH) can share. Identified birth files are included in these restrictions; therefore, we designed a two-step matching process to ensure researchers never received protected health information. The identifiable birth records will first be linked to Licensing and DSHS identifiable records using the following variables: parents' names, parents' date of birth, and parents' address. Next, DSHS will send DOH a file of records that matched with Licensing or DSHS records. DOH will then extract and send the medical/statistical information for these matched records. This protection ensures UW will only receive confidential birth information for the study's population of interest.

DATA ASSEMBLY

The legal constraints on releasing identifying data elements conflict with research data management best-practices designed to capture and bound ambiguity in merge matches, household construction, and other data uses. For instance, while DSHS could deliver data with a "best guess" household definition, the UW-MWS would like to be able to test if any empirical results are sensitive to particular matching or merging choices. In light of legal constraints, the UW-MWS team and DSHS have together designed a data merging and cleaning process that will allow the team to receive flexible but non-identifying information about confidential variables such as address and household membership. As shown in Figure 2 and detailed in the "Scope of work" appendix, the data assembly process contains three major phases. First, each state agency or data source (listed in Table 2 above) sends IRB approved data to RDA. RDA will receive the data from each source and then go through a series of steps for address management and ID management. Lastly, RDA will send the process datafiles to UW for analysis.

Assembling Quasi-Households

Residential address information is contained in Licensing, DSHS, and voter registration data. DSHS will clean and standardize this data, for example by converting "St." into "Street." Each uniquely identified street address will be assigned a pseudo-randomly generated nine-digit "ADDRESS_ID." For multi-unit structures (e.g., apartment buildings), unit numbers will be converted to a pseudo-randomly generated four-digit number, and this four-digit number will be appended to ADDRESS_ID after a decimal place. So, for example, ADDRESS_ID = 178622896.3979 will indicate unit "3979" at street address "178622896".

The de-identified datasets that we will receive from DSHS will strip the actual street address, but include ADDRESS_ID. With this information, and the date of the information (e.g., the month associated with the voter registration information), we can sequence each individual's known addresses. That is, for each individual, we merge all observed data from Licensing, DSHS, and voter registration, sequence those data by date of observation, and then form a time series of known ADDRESS_IDs. We then will use various methods to predict the ADDRESS_ID of the individual in months where we lack observed ADDRESS_ID from Licensing, DSHS, and voter registration data. A simple method would be to (a) assume the individual does not move during the intervening period between which the person is observed at the same ADDRESS_ID in non-consecutive months with no other ADDRESS_ID observed in the intervening month(s); (b) the person is assumed to continue at his/her last observed ADDRESS_ID for each month until observed at new ADDRESS_ID or until the end of the received data; and (c) that the person does not reside in the state of Washington until the person has an observed ADDRESS_ID from Licensing, DSHS, or voter registration data. Once the data is received, we will use the data to evaluate whether this simple method makes accurate predictions of the known data. We will additionally explore more sophisticated, probabilistic methods, to allocate individuals to ADDRESS_IDs during months that lack observed ADDRESS_ID.

After assigning each person an observed or assumed ADDRESS_ID for each month, we will then assemble quasi-households. A "quasi-household" will be defined as the set of individuals who are observed or assumed to be living at a particular ADDRESS_ID during a particular month. We will evaluate the combined earnings and social services receipt of the individuals in the quasi-household during the month. This assembly of individuals will allow us to evaluate the impact of the minimum wage on household earnings and social services receipt. We should note, however, that we will have limited ability to compute a household-based measure of poverty as we will lack sufficient data on the number of children in the household.

Geocoding the Location of Residence

In addition to having an ADDRESS_ID, we will obtain information on the 2010 Census block associated with that street address. Census blocks are "the smallest geographic area for which the Bureau of the Census collects and tabulates decennial census data" and "are formed by streets, roads, railroads, streams and other bodies of water, other visible physical and cultural features, and the legal boundaries shown on Census Bureau maps" (Bureau of the Census, 1994, p. 11-1). Washington contains almost 200,000 block groups, thus having this Census block provides a fine degree of resolution on the individual's location. We will append to our file aggregated data on the Census block (e.g., poverty rate of the block). With this information, we will be able to roughly compute the distance between individual's residence and the exactly geocoded address of the individual's employer.[2] We will then be able to address questions such

2 Jardim et al. (2017a) note the limitations of locating employers in the UI data. "Firms with multiple locations have the option of establishing a separate account for each location, or a common account. Geographic identification in the data is at the account level. As such, we can uniquely identify business location only for single-site firms and those multi-site firms opting for separate accounts by location…. [In] Washington State as a whole, locatable businesses comprise 89% of firms, employ 62% of the entire workforce … and 63% of all employees paid under $19 per hour" (p. 13).

as: (1) did the Ordinance cause changes in the distance traveled by worker's to low-wage jobs; (2) did the Ordinance cause changes in worker's residential choices; and (3) what is the average of minimum wages available to the worker locally (weighted by the probability that worker *i* residing in Census block *j* would be employed in Census block *k* given the distance between *j* and *k*) and did changes in local minimum wages have impacts on worker and quasi-household outcomes.

Imputing Race and Ethnicity

In order to examine differences in the impact of the minimum wage by race and ethnicity, the UW-MWS will provide DSHS with code to assign probability of racial group membership and Hispanic ethnicity based on surnames and addresses, available in the UI, Voter Registration, and Licensing data. This code estimates and assigns a discrete probability distribution of race for each individual in the combined Washington State data. We will use two sources of information to estimate missing values of race. First, we will use proportions of each race associated with surnames as developed from 2010 Census Bureau data that contains over 150 thousand surnames matched to race distributions. Second, we will use proportions of each race associated with individual residential Census block obtained from geocoded addresses matched to 2010 Census information. Employing the Bayesian Improved Surname Geocoding Algorithm (BISG) developed by Elliot and colleagues (2009), as well as the data sources described, we will estimate a final discrete probability distribution for race among those included in the Washington State data. For more information on the BISG algorithm, Fremont et al. (2016) contains a brief description, while a report from the Consumer Financial Protection Bureau (2014) offers a more detailed account. Stata (2015) syntax to implement BISG is publicly available on GitHub courtesy of the Consumer Financial Protection Bureau (updated April 2017). In addition, we propose a sensitivity analysis among a subset of individuals with known race, as available in DSHS ICDB and DOH birth records. Among these observations, we will systematically compare each known race to its corresponding estimated discrete probability distribution, assessing for levels of agreement.

ANTICIPATED ANALYSES

The UW-MWS will be able to use the resulting data to examine the policy effects and create new knowledge about the interplay among wages, employment, and different family and social outcomes. Planned analyses include both descriptive examinations of individual, family, or employer-level conditions during the observation period as well as impact studies, which will use the methods described above to determine causal effects of the Ordinance on outcomes. We will use the omnibus administrative data as one element in mixed-method inquiries, to examine hypotheses or questions arising in other parts of the project but unanswerable without the merged data.

As an example of one descriptive analysis, we will examine the interplay between means-tested benefit receipt (available in the DSHS data), parental employment, and family income. The phase-out of benefits as earnings increase is a potential limit on any anti-poverty effect of a minimum wage law (Romich & Hill 2018). In longitudinal family interviews conducted as a separate part of the UW-MWS, low-wage workers voiced fears about losing public assistance as incomes rise. We will be able to use the merged administrative data to describe the changes in SNAP benefits and construct household data to estimate changes in Earned Income Tax Credit (EITC) receipt and other tax credits using the NBER's TAXSIM program.

This same household income data will enable us to examine the causal impact of the Ordinance on household poverty rates. Past research on the minimum wage and poverty relies almost exclusively on the Official Federal Poverty Measure (OPM), but we believe this measurement choice may fail to capture real changes in material well-being brought about by minimum wage laws. The OPM, or poverty threshold, counts pre-tax cash income, including earnings and cash transfers such as child support or Social Security. However, the OPM does not count two major supports for low-income working families, SNAP and the EITC. Both SNAP benefits and EITC amounts vary by earnings and family size. Omitting taxes and transfers may mean that researchers are setting the bar for the minimum wage's anti-poverty effectiveness too high. To lower the official poverty rate, minimum wage laws would have to increase earnings and child support enough to raise families over the poverty line. Using the Supplemental Poverty Measure (SPM), which counts the EITC, SNAP, and other sources of support, may show different results, even withstanding reductions in means-tested benefits for those who do receive raises. To our knowledge, our omnibus administrative data will allow the first rigorous study of the anti-poverty impact of the minimum wage using the SPM.[3]

Employment and residential dynamics complicate the choice of population for whom poverty will be tracked. Seattle workers are not necessarily Seattle residents, and both workplaces and places of residence may change over the observation period. In particular, a sharp increase in Seattle's housing costs coincided with the first three years of the Ordinance implementation, leading to at least anecdotal concerns that lower income residents are being rapidly "priced out" of the city. Hence analyses will focus on changes in poverty among four different but overlapping populations: those who lived in Seattle prior to the Ordinance taking effect (baseline Seattleites); those who worked in Seattle prior to the Ordinance taking effect (baseline workers); Seattle workers over the observation period (point-in-time workers); and Seattle residents over the observation period (point-in-time Seattleites).

These studies using the household income measures are only two of dozens of potential analyses using the omnibus administrative data. Other team priorities include tracking changes in and impacts on child support collections and income of families involved with the state child support enforcement system; using the geo-coded data to examine the geography of opportunity relative to local wage regimes and spatial mismatch between low-wage workers, low-wage jobs, and local minimum wage laws; and looking at labor market and business ownership trajectories of refugees, a small but socially important sub-population identifiable here through initial contact with state refugee services system.

[3] One limitation of this analysis is that the data will not contain information on all dependent children. Households will only contain data on children born in 2010 or later unless the household is among the 40% of Washington households that receive assistance from DSHS. Because both the state expanded Medicaid program and the SNAP program allows assistance units with income up to or above 200% of the Federal Poverty Guidelines, we believe this limitation will not unduly restrict our ability to examine poverty impacts among the most likely affected population, but we will be less able to look at movements into or out of near poverty.

CONCLUSION AND DISCUSSION

This paper has outlined the plan for an omnibus dataset compiled from multiple public administrative data sources. This data, once assembled, hold great promise for generating new and important insights into the effects of a local minimum wage policy on inequality and the distribution of financial well-being among the working poor and near-poor. However, our experience also provides insight into some of the complexities of using administrative data for evidence-based policy. While the idea of local researchers informing local policy has a certain intuitive appeal, our experience suggests that the impact of our work will be on the larger and slowly-evolving scholarly literature. The majority of Seattle City Councilmembers voice strong and continued support of the minimum wage, and we have not received any indication of interest in revising or revisiting the overall policy or implementation schedule. Despite this path dependence, we believe interest in impacts on demographic subgroups will be high, and that the new research made possible by this data may help inform targeted mitigation effects, such as employment programs for any populations experiencing concentrated adverse effects. While our project provided the impetus for assembling an interesting new data set, we think future work will benefit from one of the types of ongoing data merging and data-sharing models emerging in other locations and domains of work.

BIBLIOGRAPHY

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105(490): 493–505.

Acemoglu, Daron and Jorn-Steffen Pischke. 2003. “Minimum Wages and On-the-Job Training. Research in Labor Economics: Worker Well-Being and Public Policy.” Elsevier: 159-202.

Allegretto, Sylvia A., Arindrajit Dube, and Michael Reich. 2011. “Do Minimum Wages Really Reduce Teen Employment? Accounting for Heterogeneity and Selectivity in State Panel Data.” *Industrial Relations* 50(2): 205-240.

Allegretto, Sylvia A., Arindrajit Dube, Michael Reich, and Ben Zipperer. 2017. “Credible Research Designs for Minimum Wage Studies: A Response to Neumark, Salas, and Wascher.” *ILR Review* 70(3): 559-592.

Becker-Green, Jody. 2012. “Revealing Patterns of Behavior and Outcomes Using Advanced Analytical Capabilities and Predictive Modeling.” *Planning, Performance and Accountability, RDA* report no. 11.180. Olympia, WA: Department of Social and Health Services.

Belman, Dale, and Paul J. Wolfson. 2014. What Does the Minimum Wage Do? Kalamazoo: W.E. Upjohn Institute for Employment Research.

Bureau of the Census. 1994. “Geographic Areas Reference Manual.” Accessed December 15, 2018. https://www.census.gov/geo/reference/garm.html

Bureau of Labor Statistics. 2017. Current Employment Statistics. Accessed December 20, 2017 at https://www.bls.gov/web/empsit/cesprog.htm. U.S. Department of Labor.

Burkhauser, Richard V., and Joseph J. Sabia. 2007. “The Effectiveness of Minimum-Wage Increases in Reducing Poverty: Past, Present, and Future.” Contemporary Economic Policy. https://search.proquest.com/docview/274268202?pq-origsite=gscholar.

Card, David, and Alan Krueger. 1994. “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review* 84(4): 772–93.

Card, David, and Alan B Krueger. 1995. “Time-Series Minimum-Wage Studies:A Meta-Analysis.” The American Economic Review 85 (2): 238–43.

Card, David, and Alan B. Krueger. 2000. "A Reanalysis of the Effect of the New Jersey Minimum Wage Increase on the Fast-Food Industry with Representative Payroll Data." National Bureau of Economic Research working paper 6386. Cambridge, MA: NBER Working Paper Series.

Cotti, Chad, and Nathan Tefft. 2013. “Fast Food Prices, Obesity, and the Minimum Wage.” Economics & Human Biology 11 (2): 134–47. doi:10.1016/j.ehb.2012.04.002.

Dube, Arindrajit. 2013. “Minimum Wages and the Distribution of Family Incomes.” University of Massachusetts, Amherst Working Paper. …. https://labor.gov/whd/flsa/75event/presentations/Arindrajit_Dube.pdf.

Dube, Andrajit, William T. Lester, and Michael Reich. 2010. “Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties.” The Review of Economics and Statistics 92(4): 945-964.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma van Inwegen, Jacob Vigdor, Hillary Wething. 2017a. “Minimum Wage Increases, Wages and Low-Wage Employment: Evidence from Seattle.” *NBER* working paper no. 23532. Cambridge, MA: National Bureau of Economics Research.

———. 2017b. “The Extent of Local Minimum Wage Spillovers.” *University of Washington* working paper. Seattle, WA: University of Washington.

———. 2017c. “Job Creation, Job Destruction, and the Minimum Wage: A Longitudinal Analysis of Workers.” *University of Washington* working paper. Seattle, WA: University of Washington.

Jardim, E., van Inwegen, E. 2017. **“**Payroll, Revenue, and Labor Demand Effects of the Minimum Wage.” *University of Washington* working paper. Seattle, WA: University of Washington.

Kim, Daehwan, and John Paul Leigh. 2010. “Estimating the Effects of Wages on Obesity.” Journal of Occupational and Environmental Medicine 52 (5): 495–500. doi:10.1097/JOM.0b013e3181dbc867.

Mancuso, David. 2014. “DSHS Integrated Client Database.” *Research and Data Analysis*, *RDA* report no. 11.205. Olympia, WA: Department of Social and Health Services. https://www.dshs.wa.gov/sesa/research-and-data-analysis

McCarrier, Kelly P., Frederick J. Zimmerman, James D. Ralston, and Diane P. Martin. 2011. “Associations between Minimum Wage Policy and Access to Health Care: Evidence from the Behavioral Risk Factor Surveillance System, 1996-2007.” American Journal of Public Health 101 (2): 359–67. doi:10.2105/AJPH.2006.108928.

Meer, Jonathan and Jeremy West. 2015. "Effects of the Minimum Wage on Employment Dynamics." Journal of Human Resources 51(2): 500-522.

Meltzer, David O., and Zhuo Chen. 2011. “The Impact of Minimum Wage Rates on Body Weight in the United States.” Economic Aspects of Obesity, no. April: 17–34.

Minimum Wage Study. 2017. https://evans.uw.edu/policy-impact/minimum-wage-study

Neumark, David. 2014a. “More on Recent Evidence on the Effects of Minimum Wages in the United States.” *NBER* working paper no. 20619. Cambridge, MA: National Bureau of Economic Research.

Neumark, David, J. M. Ian Salas, and William Wascher. 2014b. “Revisiting the Minimum Wage and Employment Debate: Throwing Out the Baby with the Bathwater? *ILR Review* 67(Supplement): 608–48.

Neumark, David, Mark Schweitzer, and William Wascher. 2005. "The Effects of Minimum Wages on the Distribution of Family Incomes A Nonparametric Analysis." Journal of Human Resources 40 (4): 867–94. doi:10.2307/4129543.

Neumark, David, and William Wascher. 1992. "Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws?" Industrial and Labor Relations Review 46(1): 55-81.

Neumark, David, and William Wascher. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment." The American Economic Review 90(5): 1362-396.

Neumark, David, and William Wascher. 2007. "Minimum Wages, the Earned Income, Tax Credit, and Employment: Evidence from the Post-Welfare Reform Era." IZA Discussion Paper 2610.

Neumark, David, and William L. Wascher. 2008. Minimum Wages. Cambridge, MA: MIT Press.

NY Department of Motor Vehicles.*Driver's Privacy Protection Act, (18 U.S.C §2721 et seq., Prohibition on Release and Use of Certain Personal Information from State Motor Vehicle Records.* New York, NY. Accessed December 15, 2017 https://dmv.ny.gov/forms/mv15dppa.pdf

Office of City Clerk. 2014. "City of Seattle Legislative Information Service." Seattle, WA. http://clerk.seattle.gov/~scripts/nph-brs.exe?d=RESF&s1=31524.resn.&Sect6=HITOFF&l=20&p=1&u=/~public/resn1.htm&r=1&f=G

Romich, Jennifer and Heather D. Hill. 2018. "Coupling a Federal Minimum Wage Hike with Public Investments to Make Work Pay and Reduce Poverty." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 4(3).

Sabia, J J. 2008. "Minimum Wages and the Economic Well-Being of Single Mothers." Journal of Policy Analysis and Management 27 (4): 848–66. doi:10.1002/pam.20379.

Seattle. N.D. In *Wikipedia*, December 15, 2017. https://en.wikipedia.org/wiki/Seattle

Simon, Kosali I., and Robert Kaestner. 2004. "Do Minimum Wages Affect Non-Wage Job Attributes? Evidence on Fringe Benefits." ILR Review 58(1):52-70.

Social Security Administration. 2013. "GN 03325.002 Disclosure of Social Security Numbers (SSN) Without Consent." Accessed December 15, 2017. http://policy.ssa.gov/poms.nsf/lnx/0203325002

Thompson, Lynn, and Amy Martinez. 2014. "Seattle City Council Approves Historic $15 Minimum Wage." *The Seattle Times*, June 2, . 2014. https://www.seattletimes.com/seattle-news/seattle-city-council-approves-historic-15-minimum-wage/

Washington Secretary of State. N.D. "Voter Participation Statistics." U.S. Social Security Administration. 2013. Accessed December 15, 2017. https://www.sos.wa.gov/elections/voter-participation.aspx

Washington (state). N.D. In *Wikipedia*, December 15, 2017. https://en.wikipedia.org/wiki/Washington_(state)#cite_note-Census1July2016-2

Washington State Department of Health. N.D. "All Birth Tables by Year". Access December 13, 2017. https://www.doh.wa.gov/DataandStatisticalReports/HealthStatistics/Birth/BirthTablesbyYear

Washington State Employment Security Department. 2017. *Labor Market and Performance Analysis*. Olympia, WA. https://esdorchardstorage.blob.core.windows.net/esdwa/Default/ESDWAGOV/labor-market-info/Libraries/Economic-reports/Annual-Report/2016%20Labor%20Market%20and%20Economic%20Report.pdf

Washington State Employment Security Department. 2016. *FTE Employment by Hourly Wage.* Olympia, WA. https://fortress.wa.gov/esd/employmentdata/docs/economic-reports/fte-hourly-wages-2016-summary.xlsx

Washington State Department of Licensing. 2016. *Statistics At-A-Glance.* Olympia, WA. http://www.dol.wa.gov/about/docs/2016-CY-stats-at-a-glance.pdf

Washington State Department of Social and Health Services (DSHS). 2003. *Washington State Agency Policy on Protection of Human Research Subjects*. Olympia, WA: Department of Social and Human Services. https://www.dshs.wa.gov/sites/default/files/SESA/hrrs/documents/guideandpolicy.pdf

Washington State Department of Social and Health Services (DSHS). 2015. "DSHS Client Services of All Ages." Accessed December 15, 2017. http://clientdata.rda.dshs.wa.gov/Home/ShowReport?reportMode=0

Washington State Department of Social and Health Services (DSHS). 2015. "DSHS Client Services of All Ages." Accessed December 15, 2017. http://clientdata.rda.dshs.wa.gov/Home/ShowReport?reportMode=0Washington State Department of Social and Health Services (DSHS). N.D. "Research Reports." Accessed June 12, 2017. https://www.dshs.wa.gov/sesa/rda/research-reports

Washington State Secretary of State. N.D. Accessed December 15, 2017. https://www.sos.wa.gov/elections/voter-participation.aspx

West, Rachel, and Michael Reich. 2014. "The Effects of Minimum Wages on SNAP Enrollments and Expenditures." doi:10.1111/j.1477-9552.1995.tb00748.x.

Figure 1. Data sources and elements

*Indicates variables that will be provided for DSHS clients only

Figure 2. Data Assembly Process

1 **Each data source sends approved data to RDA**

- ESD
- DOL
- DOH
- DSHS/ESA
- WSP

2 **RDA cleans and de-identifies datasets**

Research & Data Analysis (RDA):

1. Receives data from each source
2. Cleans & geocodes residential addresses from DSHS, DOL and Voter Registration (UW supplies this data) to yield residential Census block and ADDRESS_ID.
3. Creates a PERSON_ID for each unique person (based on information such as SSN, name, date of birth) in each dataset
4. Creates race imputation for each dataset using program provided by UW researchers
5. Sends UW 6 distinct, de-identified datasets, stripping off name, SSN, day of birth, and residential address.

3 **UW receives 6 de-identified datasets**

University of Washington (UW)

1. Receives 6 distinct datasets: ESD, DOL, DSHS, DOH, WSP, Voter Registration. All datasets are de-identified, but include PERSON_ID, ADDRESS_ID, CENSUS_BLOCK, and six race probabilities

Note abbreviations above: Employment Security Department (ESD), Department of Licensing (DOL), Department of Health (DOH), Department of Social and Health Services (DSHS), Economic Services Administration (ESA), Washington State Patrol (WSP).

APPENDIX – SCOPE OF WORK

The following Scope of Work details the process whereby DSHS's Research and Data Analysis Division (RDA) will prepare the analysis dataset and deliver to UW.

1. *Address Management*. RDA standardizes addresses by cleaning residential addresses associated with Licensing, VR and ICDB records. Then RDA creates separate Department of Licensing (Licensing), voter registration (VR), and Integrated Client Database (ICDB) tables with standardized addresses. Next, RDA geocodes addresses and assigns Census block (based on 2010 assignments) to each record with an address. RDA then creates ADDRESS_ID by creating a unique "ADDRESS_ID" for all observed addresses in the combined Licensing, VR and ICDB records, then adds ADDRESS_ID to these records. Finally, RDA sends UW a technical memo describing address management process and data quality, noting any issues that arise.
2. *ID Management.* RDA, in collaboration with UW, develops two alternative (one conservative and one liberal) matching solutions to be included in a single output file if possible. Next, RDA assigns unique identifiers based on these two alternative matching solutions, un-duplicates and links IDs across all data sets, and assigns each person a unique PERSON_ID. RDA then attaches selected ICDB month arrays for individuals matched with the ICDB. For the race/ethnic probability, RDA uses a UW-provided SAS code that merges probabilities of being a member of various racial/ethnic groups for each last name. Finally, RDA sends UW a technical memo describing PERSON_ID management process and data quality, noting any issues that arise.
3. *Data files sent to UW*. Lastly, RDA de-identifies all data files and sends the following to UW:
    a. Licensing de-identified records with the following added: PERSON_ID, ADDRESS_ID, CENSUS_BLOCK, and imputed race.
    b. Voter Registration de-identified records with the following added: PERSON_ID, ADDRESS_ID, CENSUS_BLOCK, and imputed race.
    c. ESD de-identified records with the following added: PERSON _ID
    d. DOH de-identified records with the following added: PERSON_ID
    e. ESA de-identified records with the following added: PERSON_ID
    f. WSP de-identified records with the following added: PERSON_ID
    g. ICDB de-identified records with selected month arrays and with the following added: PERSON_ID
    h. Summary: RDA sends UW a technical memo describing individual data files management process and individual data file quality, noting any issues that arise.