The Department of Labor 2013-2014 Scholars Program

Working Paper Series

This page left intentionally blank

The Department of Labor 2013-2014 Scholars Program

Working Paper Series

These papers were prepared with funding from the U.S. Department of Labor. The views expressed are those of the authors and should not be attributed to the Federal Government or the Department of Labor.

Table of Contents

***Introduction*.....1**

***Meet the Scholars*.....2**

***Unmeasured Effects? Assessing the Impact of Full-day Kindergarten on Maternal Employment*.....3**
Chloe Gibbs, Assistant Professor of Public Policy and Education, Frank Batten School of Leadership and Public Policy, University of Virginia

***Early Effects of the 2010 Affordable Care Act on Labor Market Outcomes*.....3**7
Angshuman Gooptu, PhD Candidate, School of Public and Environmental Affairs, Indiana University

***Effects of Relaxing the Job Search Requirement on Job Match Quality*.....77**
Marta Lachowska, Economist, W.E. Upjohn Institute for Employment Research

***The Relationship between Injury Type and the Likelihood of Underreporting* 163**
Alison Morantz, Professor of Law and John A. Wilson Distinguished Faculty Scholar, Stanford University

***Effects of Unionization on Worker and Workplace Safety* 230**
Aaron Sojourner, Assistant Professor, University of Minnesota

Introduction

The purpose of the Department of Labor (DOL) Scholars Program is to promote and expand labor research. In 2013, scholars from recognized research institutions were selected from a competitive pool of applicants to participate in the DOL Scholars Program. Selection was based on methodological rigor and relevance to DOL policies and programs.

At the conclusion of the program, the scholars submitted a final report describing their research and findings. The final reports were compiled into this paper series.

Meet the Scholars

Chloe Gibbs
Dr. Chloe Gibbs is an Assistant Professor at the University of Virginia's Batten School of Leadership and Public Policy with a joint appointment in the Curry School of Education. She is an affiliated researcher with the University of Virginia's Center on Education Policy and Workforce Competitiveness and the Chapin Hall Center for Children at the University of Chicago. Her research focuses on child and family policy, specifically the effects of early childhood interventions and programming. Prior to completing her Ph.D. at the University of Chicago, Dr. Gibbs worked as a researcher at the American Institutes for Research, Academy for Educational Development, and Chapin Hall, where she led sample frame construction for the National Survey of Early Care and Education.

Angshuman Gooptu
Angshuman Gooptu holds a Master of Arts in Economics from the University of Illinois at Chicago and a Masters in Public Policy from the University of Chicago. Currently, he is a Ph.D. Candidate at Indiana University's School of Public and Environmental Affairs. Previously, he was a short-term consultant at the World Bank, conducting research and analysis on health systems for the Human Development Network and at CVS Caremark's Health Analytics division conducting research and analysis on retail pharmaceutical pricing. In addition, he has worked at the Federal Reserve Bank of Chicago as an Associate Economist and worked as a consultant for a Marketing Analytics firm based in the Greater Chicago area.

Marta Lachowska
Marta Lachowska is an Economist at the W.E. Upjohn Institute for Employment Research. She completed her Ph.D. in Economics at Stockholm University. Prior to coming to Upjohn Institute she spent three semesters at Princeton University's Industrial Relations Section as a visiting doctoral researcher. Her area of interest is mainly in the field of economics of social insurance, with a focus on the impact of public pensions and unemployment insurance systems on labor market outcomes and household behavior.

Alison D. Morantz
Dr. Alison Morantz is a Professor of Law and the John A. Wilson Distinguished Faculty Scholar at Stanford Law School. Her work has explored the law and economics of protective labor regulation, the enforcement of workplace safety laws, and legal history. Dr. Morantz is the principal investigator of multi-year research projects funded by the National Science Foundation and the National Institute of Occupational Safety and Health, and she is the co-director of the Martin Daniel Gould Center for Conflict Resolution's ADR Research Initiative. She earned a Bachelor of Arts summa cum laude from Harvard, a Master of Science from Oxford University on a Rhodes Scholarship, a Juris Doctorate from Yale Law School, and a Ph.D. in economics from Harvard University. Before joining the Stanford Law School faculty, Dr. Morantz clerked for Judge Patti B. Saris of the U.S. District Court for the District of Massachusetts, and worked as a union–side labor lawyer and antidiscrimination advocate in Boston.

Aaron Sojourner
Dr. Aaron Sojourner is an Assistant Professor at the University of Minnesota's Carlson School of Management. His research interests include sources of human capital, impacts of unions in the economic and political arenas, and behavioral consumer finance decisions. Complementing these interests, he has a wide range of policy experience including service as a fellow with the U.S. Senate's Health, Education, Labor and Pensions Committee. Dr. Sojourner completed his Ph.D. in Economics at Northwestern University. He also has a Master of Arts in public policy analysis from the University of Chicago and a Bachelor of Arts in history from Yale University.

A Matter of Time?
Impact of Statewide Full-day Kindergarten Expansions on Later Academic Skills and Maternal Employment

Chloe R. Gibbs*

This paper was prepared with funding from the U.S. Department of Labor. The views expressed are those of the authors and should not be attributed to the Federal Government or the Department of Labor.

*__Contact__: Batten School of Leadership and Public Policy, University of Virginia, P.O. Box 400893, Charlottesville, Virginia 22904-4893, cgibbs@virginia.edu. **WORKING PAPER** prepared for U.S. Department of Labor Scholars Program, Chief Evaluation Office. I thank participants at the University of Virginia's Education Policy Seminar Series and U.S. Department of Labor Scholars Program presentations for helpful comments. David Melnick provided excellent research assistance. I also thank the Indiana Department of Education and Regional Educational Laboratory–Midwest, and gratefully acknowledge funding from the U.S. Department of Labor Scholars Program and the University of Virginia's Bankard Fund for Political Economy. Any remaining errors are my own.

Abstract

In this paper, I leverage variation in response to a statewide full-day kindergarten policy to explore the effects of full-day kindergarten expansions on student academic performance—as measured by school-level standardized test scores—in first and third grade and on women's labor force participation, measured by county-level employment statistics. The policy change, enacted through 2007 legislation, dramatically increased state grant funding availability for the provision of full-day kindergarten, from 8.5 million dollars in the 2006–2007 academic year to 33.5 million dollars in the 2007–2008 academic year. Because the increased funding generated differential responses across schools in full-day kindergarten enrollment rates, I use a triple-differences framework to exploit variation in exposure to full-day kindergarten by geography, time, and cohort, generating intention-to-treat estimates of the impact of the full-day kindergarten expansions. With various specifications of treatment exposure, I find no evidence of full-day kindergarten effects on subsequent student achievement, but suggestive evidence of increases in maternal employment.

Keywords: Kindergarten, early childhood, maternal employment

1 Introduction

The early childhood years have garnered much policy interest as a critical developmental period and opportunity for early intervention, potentially remediating cognitive and noncognitive skill gaps prior to formal schooling and generating high returns on investment in the form of private and social benefits over the life cycle. Several papers have established the early emergence of achievement gaps by race/ethnicity and socioeconomic status (Fryer & Levitt 2004, Fryer & Levitt 2006, Lee & Burkam 2002, Murnane, Willett, Bub & McCartney 2006, Princiotta, Flanagan & Germino Hausken 2006). While there is some dispute about the magnitude of these gaps—and whether and how they can be explained by student characteristics and family background—there is consistent evidence that test-score gaps by race and socioeconomic status are already sizable at the end of the kindergarten year. In addition, these gaps persist and grow throughout the primary grades.

In 2007, the Indiana General Assembly significantly increased state grant funds for the provision of full-day kindergarten, from $8.5 million in the 2006-07 school year to $33.5 million for the 2007-08 school year. The legislation sought to improve access to and availability of full-day kindergarten statewide with grants targeted directly to school districts and funding amounts determined based on kindergarten enrollments. This policy change corresponded to increases in provision and enrollment across the state. According to administrative data, full-day kindergarten enrollment in Indiana rose 20 percentage points from 2006-07, before the legislation, to 2007-08, the first year after the funding increase (Lovell et al., 2009). In addition, the number of school corporations providing full-day to the vast majority of their students (i.e., more than 80 percent) increased by 26 percentage points.

The existing literature on full-day kindergarten impact relies largely on the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) to estimate effects for participants. These studies have generally found significant differences between full- and half-day kindergarten students on measures of literacy and numeracy at the end of the kindergarten year (Cannon, Jacknowitz, & Painter 2006; Lee, Burkam, Ready, Hinigman, & Meisels 2006; Votruba-Drzal, Li-Grining, & Maldonado-Carreño 2008; DeCicca 2007). However, these full-day advantages, as measured by cognitive skills, fade out rapidly over the first grade year and are no longer detectable in third grade (Cannon et al 2006; Votruba-Drzal et al. 2008) or fifth grade (Votruba-Drzal et al. 2008). This literature is limited in its ability to address differential selection by schools and districts into offering full-day kindergarten and differential selection by children and families into full-day kindergarten attendance. Moreover, the data pre-date large expansions across the United States in both the provision of full-day kindergarten and the availability of pre-kindergarten programming.

Recent experimental and quasi-experimental evidence finds sizable effects of full-day kindergarten participation on early literacy skills, particularly for disadvantaged students, but does not speak to the persistence of these effects or the broader systemic effects of offering full-day kindergarten (Gibbs 2012).

Additionally, this study investigates the causal impact of full-day kindergarten availability and participation on women's labor market attachment, a largely unexplored outcome of such full-day kindergarten expansions. Because full-day kindergarten constitutes a wealth transfer to parents in the form of subsidized childcare, and provides approximately three additional hours of public schooling relative to half-day kindergarten, mothers may be made more available for labor market participation as a result of the policy. Additionally, they may avoid a disruption in their labor force participation caused by the different schedule of half-day kindergarten relative to publicly provided preschool, such as Head Start, in the prior year and first grade in the following year. This project leverages the same policy expansion to provide new, quasi-experimental evidence on the impact of full-day kindergarten provision on female labor force participation.

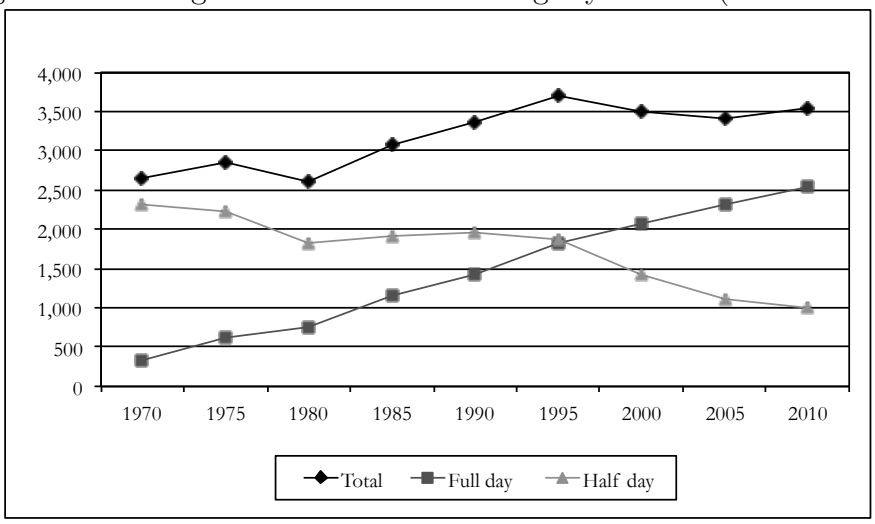
Notably, the state level has been the site of much of the policy conversation and action on full-day kindergarten, and this paper is the first to consider systemic impact with intention-to-treat estimates, rather than participant outcomes. Capitalizing on variation across schools and counties in response to the full-day kindergarten policy change, I explore the impact of full-day kindergarten on first and third grade academic skills to answer the question of whether greater provision of full-day kindergarten results in increased levels of achievement in the providing schools. I also estimate the impact of full-day kindergarten on women's employment at the county level. I use triple-difference estimators to assess the impact of the full-day kindergarten legislation on these outcomes of interest. Importantly, I leverage cross-time, cross-cohort, and cross-geography variation in exposure to full-day kindergarten to assess the effect of this sweeping statewide reform. I specify the full-day kindergarten exposure variable in different ways to allow for non-linearities in the effect of full-day kindergarten enrollment. Findings suggest that full-day kindergarten does not affect subsequent academic achievement in the early adopting, or high-implementing, schools, but is associated with greater female labor force participation at the time of expansions.

2 Study Background

Nearly all students who attend school outside the home participate in kindergarten in the United States and nearly three-quarters of students in kindergarten are in full-day settings (Walston & West 2004). As displayed in Figure 1, less than 20 percent of kindergarten

students were in full-day settings in 1970. Full-day kindergarten enrollment exceeded half-day participation for the first time in 1995 and by the 2000-01 academic year, approximately 60 percent of kindergarteners were in full-day classrooms. Of the 3.5 million kindergarten students enrolled in 2010, over 2.5 million attended full-day kindergarten.

Figure 1: Kindergarten Enrollment Among 5-year-olds (in thousands)



Source: U.S. Census Bureau, Current Population Survey, October School Enrollment Supplement.

States are also considering full-day kindergarten policies and expansions, with Minnesota providing full-day kindergarten for all kindergarten students in the 2014–15 academic year and Kansas considering legislation to do the same. According to the Children's Defense Fund, ten states and Washington, D.C. provided full-day kindergarten at no charge to all children in 2013. In only 16 states is kindergarten attendance mandatory though, so seven of those ten states have both required full-day kindergarten provision and mandatory kindergarten attendance.

2.1 Motivation

Full-day kindergarten provision is rapidly expanding in the United States, but many districts and states are constrained by tight budgets in funding full-day kindergarten at the levels and for the number of children that they would like. Moreover, other early childhood investments may reap larger returns. The case for full-day kindergarten has been made almost entirely on the educational benefits to participants of these programs, but labor force implications are clearly an important consideration in assessing the cost-effectiveness of such policies. In

addition, understanding the longer-term impact on overall school and district performance is crucial to fully capturing program effects. The finding that full-day kindergarten affects later academic achievement or women's labor force attachment—as the existing literature suggests—would constitute an important consideration for policymakers faced with tradeoffs in program provision. Moreover, full-day kindergarten may constitute a relatively cost-effective way to enhance student achievement and incentivize labor force participation as the marginal cost of expanding childcare subsidies or publicly provided preschool generally exceeds the marginal cost associated with expanding half-day kindergarten to full-day.

This study improves upon the existing literature in a few ways. First, the research design leverages the availability of state funding for full-day kindergarten and the resulting expansions in provision as plausibly exogenous changes to estimate the effect of full-day kindergarten. Thus, this work complements the existing work which uses nationally representative data by using more limited samples, but an identification strategy that reasonably addresses the selection problem. In addition, the policy change took place in 2007, so the study employs more recent data than that on which much of the existing work relies.

Previous work in this vein certainly establishes the plausibility of full-day kindergarten impact on maternal employment, and in fact, the only existing paper on full-day kindergarten and this outcome finds positive effects on maternal labor supply (Cannon, Jacknowitz & Painter 2006). With a more current policy context, quasi-experimental approaches, and more nuanced outcome measures, the work represents an important extension and contribution to our collective understanding of the impact of childcare on women's employment. By adding maternal labor market outcomes to the conversation on full-day kindergarten and relatedly, incorporating full-day kindergarten into our broader discourse around childcare, this work could affect the way policymakers think about how best to incentivize and support labor force participation among mothers.

2.2 Prior Research

The existing literature on full-day kindergarten takes two forms: observational studies using nationally representative data and district- and school-level evaluations. In observational studies using the ECLS-K, researchers have found significant differences between full- and half-day kindergarten students on literacy and mathematics assessments at the end of the kindergarten year (Cannon et al. 2006, DeCicca 2007, Lee et al. 2006, Votruba-Drzal et al. 2008). These full-day kindergarten advantages failed to persist, however, over the first-grade year. In one study, marginally significant differences were found in the spring of first grade (Cannon et al. 2006). DeCicca (2007) found significant differences in mathematics and

reading in the fall of first grade, but only for white children, which faded but continued to be significant in spring literacy performance. No significant differences were found between full- and half-day kindergarten students in the ECLS-K in third grade (Cannon et al. 2006, Votruba-Drzal et al. 2008) or fifth grade (Votruba-Drzal et al. 2008).

Additional smaller-scale evaluations have supported the ECLS-K findings of short-term outcomes in the kindergarten year, but no significant long-term effects (Zvoch, Reynolds & Parker 2008, Hall-Kenyon, Bingham & Korth 2009). In general, findings on the impact of full-day kindergarten relative to half-day kindergarten suggest some positive associations, particularly in the early schooling years. Results related to the impact of full-day kindergarten over time, or the persistence of these positive findings, are more mixed. Recent experimental and quasi-experimental evidence on the impact of full-day kindergarten for participating children finds positive effects, but does not speak to longer-term effects nor the implications of greater provision of full-day kindergarten for schools and districts (Gibbs 2012). From the perspective of policymakers, the effects on overall academic achievement in schools and districts offering full-day kindergarten may be of greatest interest. The intention-to-treat effect may better capture—as opposed to participant-level treatment on the treated effects—the return on a school or district's investment in full-day kindergarten provision.

While some previous research has investigated the relationship between other types of childcare and women's employment (Blau & Robins 1988, Connelly 1992, Karoly et al. 1998, & Ribar 1992), there is more limited work exploring public education as a form of childcare with potential effects on maternal labor supply (Gelbach 2002). One paper uses the Early Childhood Longitudinal Study–Kindergarten Cohort (ECLS-K) to look specifically at full-day kindergarten impact on maternal employment (Cannon et al. 2006). Using state full-day kindergarten policies to explain maternal labor supply, they find that mothers are more likely to work full-time in the kindergarten year when full-day kindergarten is provided (*ibid*), but the identification strategy is not sufficiently rigorous to rule out potential confounding variables that may be related to both provision of full-day kindergarten and women's employment.

Most studies suggest that childcare subsidies and access to other subsidized early childhood programs increase maternal employment, though estimates vary based on the study and the policy being analyzed. Childcare subsidies, with an explicit intention of facilitating parental labor force participation, have been studied most extensively. In accordance with theoretical predictions, higher childcare prices reduce labor supply of mothers (Blau 2003). Estimated childcare cost elasticities of maternal employment range from 0 to approximately -1, with the most credible estimates varying between -0.1 and -0.5 (Anderson & Levine 2000). Overall, subsidization of childcare contributes to increased maternal employment (Connelly

& Kimmel 2003, Blau & Currie 2006, Tekin 2007, Blau & Tekin 2007, Herbst 2010).

Relatedly, researchers have also explored the availability of Head Start and other subsidized forms of preschool in relation to maternal employment. A study of single mothers who had received public assistance in Massachusetts found that both the availability of Head Start and state funding for preschool education for low-income children increased maternal employment (Lemke, Witt & Witte 2007). Interestingly, Fitzpatrick (2010) finds no increases in maternal labor supply when examining the effects of the introduction of a universal and near-universal preschool provision for four-year-olds in Georgia and Oklahoma in the 1990s.

There is a growing body of evidence to speak to the effects of public education on women's labor force attachment. Using a novel approach, Gelbach (2002) employs quarter-of-birth as an instrument for enrollment in kindergarten in 1980 and finds evidence of increases in maternal labor supply on the order of six to 24 percent. Barua (2007) finds similar effects as Gelbach (2002), but concentrated among married women whose youngest child is of kindergarten age. Similarly, Cascio (2009a)—exploiting variation in timing of public funding for kindergarten in the 1960s and 1970s—shows evidence of an increase in the labor supply of single mothers without other young children, but no effects for single mothers with younger children or for married mothers. Fitzpatrick (2012) finds that the labor supply of single mothers of five-year-olds without additional young children increases as a result of a child's enrollment in public kindergarten, but is unchanged for other mothers.

To date, the only study looking at the effects of full-day kindergarten on maternal employment is the aforementioned Cannon, Jacknowitz, and Painter (2006) paper. They use data from ECLS-K to study effects of full-day K policy enactment and find that mothers with students in full-day kindergarten are more likely to work compared to their half-day counterparts, although the effect is no longer significant when the child moves beyond kindergarten. Again, this study is observational and, while based on nationally representative data in the 1990s, cannot address the endogeneity of full-day kindergarten availability. While the existing literature certainly suggests that public schooling and subsidized childcare may relate to maternal employment—particularly for single women and for those with no younger children in the household—there is room to augment this research base with rigorous, causal research on a particularly understudied and relatively inexpensive intervention, full-day kindergarten.

2.3 Full-day Kindergarten in Indiana

This study leverages a policy change in Indiana that led to dramatic and relatively immediate changes in the availability of full-day kindergarten across the state. The study capitalizes on this plausibly exogenous shock in the provision of publicly funded child care to explore

the impact on female labor force participation using a triple-differences research design that accounts for underlying trends in labor force participation.

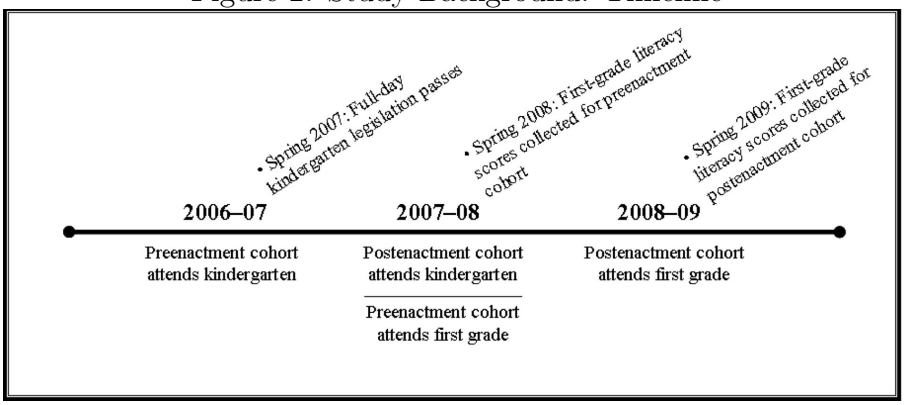
In 2007, the Indiana legislature passed legislation which provided funding to increase access and availability of full-day kindergarten in the 2007–08 school year, with grants targeted directly to school districts and charter schools (Indiana Public Law 234-2007). The stated goal of the bill was to allow, "school corporations [districts] and charter schools to provide full-day kindergarten programs to improve the academic and social development of children in kindergarten." A policy initiative of Governor Mitch Daniels' administration and supported by the Indiana State Board of Education and the Indiana Department of Education, the legislation expanded state grant funds for full-day kindergarten from $8.5 million in the 2006-07 school year to $33.5 million in the 2007-08 school year (Indiana General Assembly 2007). School districts and charter schools, operating as autonomous school districts in the state, were eligible to apply to the state for full-day kindergarten funding. [1] Grant funding was then dispersed to all applicants on a per-pupil basis, allocated based on kindergarten enrollment in the school or district in the 2007-08 school year.

Figure 2 provides the timeline of the legislation relevant for this study's approach. Changes in full-day kindergarten availability and participation were realized quickly across the state. Full-day kindergarten enrollment in the state increased by 20 percentage points from 41 percent of kindergarten students in 2006-07 to 63 percent in 2007-08. The number of school districts and schools offering full-day kindergarten also increased with a 26 percentage-point increase in the number of school districts enrolling more than the vast majority of their kindergarten students in full-day settings, and a 21 percentage-point increase in the number of schools with nearly all of their kindergarten students in full-day settings (Lovell, Kochanek, Mathers & Burke 2009).[2] With the increased funding availability from the state, the full-day kindergarten grant became the primary means for supporting full-day kindergarten enrollment, coupled with federal Title I funds, school district general funds, and parent fees in some cases. School district data indicated that 91 percent of full-day kindergarten students in the state were funded, at least in part, by the state full-day kindergarten grant monies in 2007-08 (Lovell et al. 2009).

[1] "Application" for full-day kindergarten funding consisted of indicating interest to the Indiana Department of Education and doing so by a deadline in order to establish enrollment numbers and make grant allocations.

[2] The definition employed for the purposes of reporting change in school district and school provision is that more than 80 percent of enrolled kindergarten students in the district or school participate in full-day kindergarten.

Figure 2: Study Background: Timeline



Participation rates responded to the policy change, and this result was driven through changes in school-level provision or availability of full-day kindergarten. Figures 3 and 4 display the distributions of the proportion of kindergarten students in full-day settings at the school level for the pre- and post-policy change school years.

Figure 3: School-level Provision of Full-day Kindergarten, 2006–07

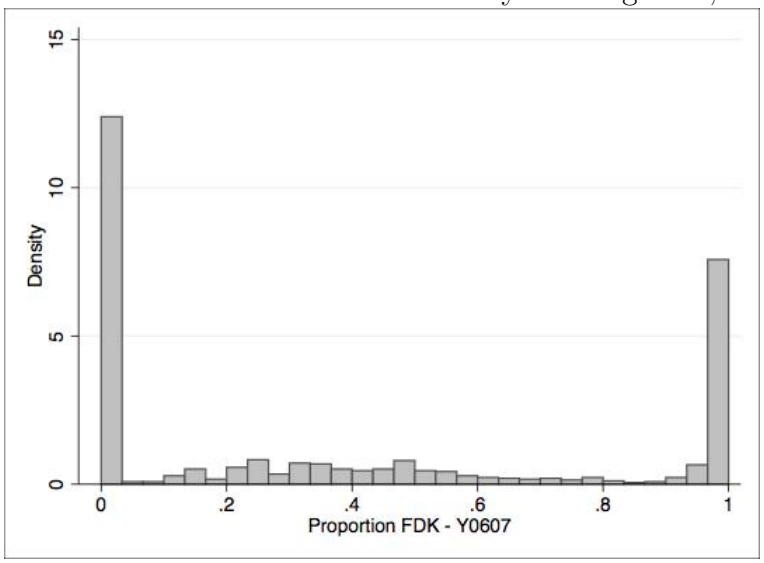
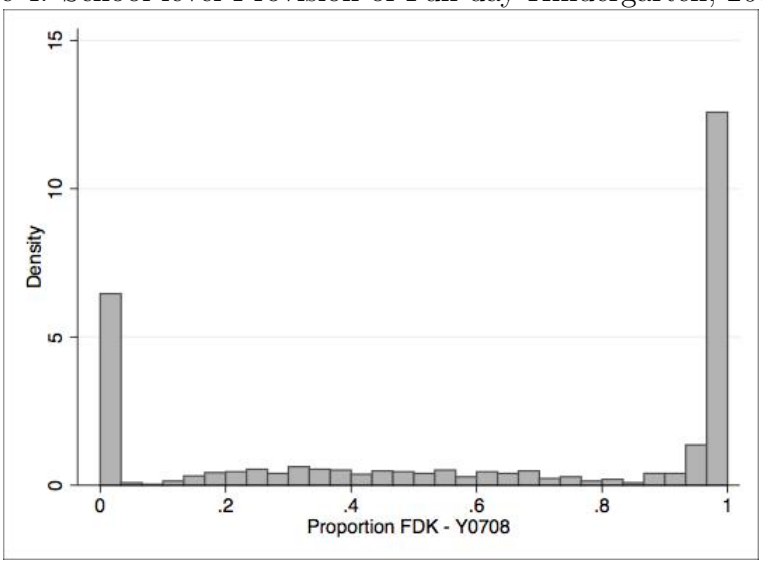
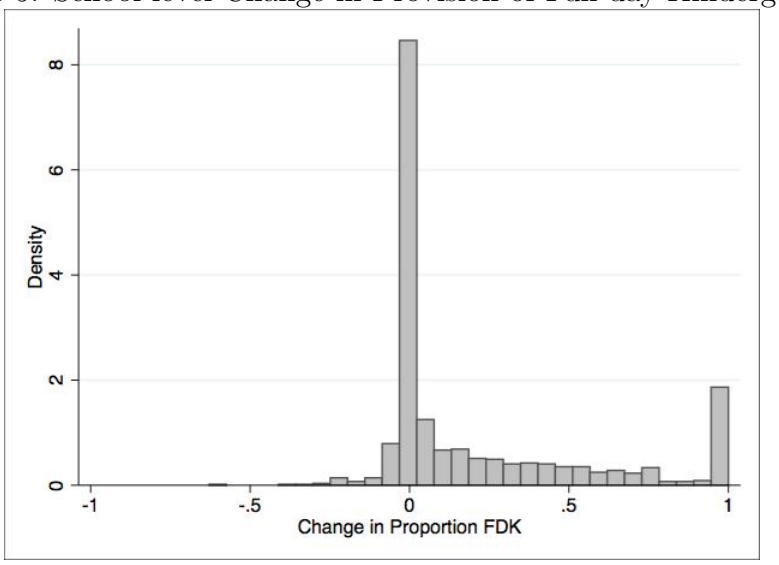


Figure 4: School-level Provision of Full-day Kindergarten, 2007–08



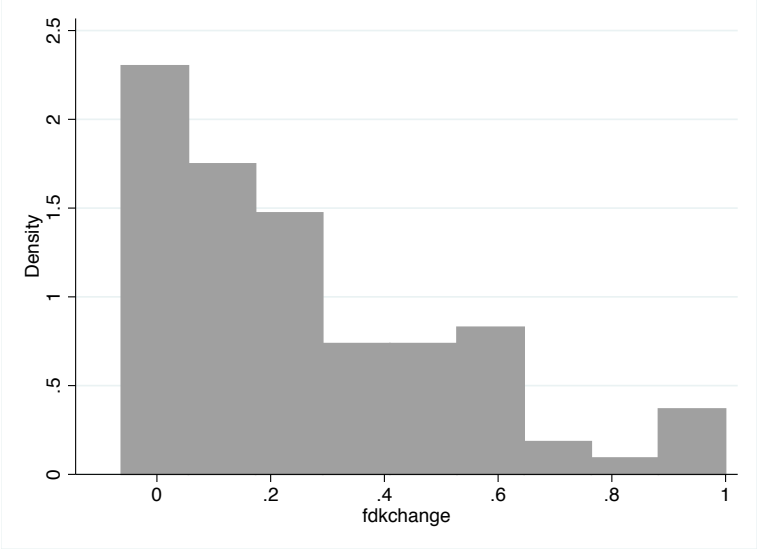
In 2006–07, 160 schools provided all full-day kindergarten, that is, kindergarten was only provided in full-day settings. That number increased to 306 schools in 2007–08. Relatedly, 428 schools provided only half-day kindergarten before the policy change, which decreased to 225 schools in the following academic year. School-level changes in full-day kindergarten provision are displayed in Figure 5. Notably, 317 schools experienced no change in proportion enrolled in full-day and 149 schools experienced a decline in full-day kindergarten participation over that same timeframe. 712 schools had an increase in full-day kindergarten participation with 75 schools moving from no full-day to all full-day in the pre- and post-policy change academic years.

Figure 5: School-level Change in Provision of Full-day Kindergarten



While the analysis of impact on achievement will leverage these school-level changes, models that explore impact on women's employment will be conducted at the county level. Figure 6 provides a similar distribution for county-level change in full-day kindergarten provision from the pre-policy year to the post-policy year.

Figure 6: County-level Change in Proportion of Kindergarten Students in Full-day Settings



3 Methods

The study will employ rigorous approaches to identifying the causal effect of full-day kindergarten participation and availability on school achievement and women's employment. First, this work leverages these changes in full-day kindergarten provision and participation to look at effects on later academic performance at the school level. Because of data availability, the design relies on "treated" and "untreated" cohorts of students to identify effects of the expansions on student performance. Figure 6 provides the relevant cohorts for comparison in the triple-differences design described in more detail in the sections that follow.

Figure 7: Cohort Comparisons

	2006-07	2007-08	2008-09	2009-10	2010-11
Kindergarten	*pre*	post	post	post	post
1st Grade	*pre*	*pre*	post	post	post
2nd Grade	*pre*	*pre*	*pre*	post	post
3rd Grade	*pre*	*pre*	*pre*	*pre*	post
4th Grade	*pre*	*pre*	*pre*	*pre*	*pre*

Secondly, a triple-differences strategy compares employment among women of mothering age in areas that experienced rapid growth in full-day kindergarten provision to employment among women of mothering age in low or no change school areas, before and after the expansion of full-day kindergarten, netting out the changes in employment among women of older and younger age groups. That is, how much schools responded to full-day kindergarten policy changes in the expansion of their own programs will again be used to generate the comparison though at the county level for employment.

3.1 Data

This section first describes data used to quantify the treatment, or independent variable, in the proposed triple-differences analysis in order to gauge responsiveness of the outcomes of interest to a sweeping policy reform in Indiana to expand full-day kindergarten availability and access. The researcher has data on the number of children enrolled in kindergarten, and the number of those children in full- and half-day settings, in each school for both the 2006–07

and 2007–08 school years, also matched from schools and schools districts to counties and aggregated to the county level. The 2006–07 school year pre-dates the legislative reform, and its increased funding availability to school districts statewide for the provision of full-day kindergarten, and the 2007–08 school year is the first year in which the new policy was in place.

The Indiana Department of Education provided kindergarten enrollment records for the 2006–07 and 2007–08 academic years (i.e., the pre- and post-policy change years). These kindergarten enrollment records include a flag at the individual-student level for full-day kindergarten participation. These records were aggregated to generate school-level, and in the employment analyses county-level, full-day kindergarten provision rates. Four versions of measuring within-school change from the 2006–07 to the 2007–08 academic year are employed in the analysis:

- Continuous measure of change in full-day kindergarten participation as a proportion of all kindergarteners within a school,
- Binary measure of a positive change, or increase, in full-day kindergarten participation as a proportion of all kindergarteners within a school,
- Binary measure of above average change, called high change, in full-day kindergarten participation as a proportion of all kindergarteners within a school, and
- Binary measure of change to provide all full-day kindergarten in schools that previously served fewer than all kindergarteners in full-day settings.

Table 1 displays the descriptive statistics for these four versions of the treatment variable across the 92 counties in the dataset, the level used in maternal employment models.

Table 1: Full-day Kindergarten Availability Variables at the County Level

Variable	Mean (SD)
Change in Proportion of County's Kindergarten Students in Full-day	0.2564 (0.2563)
Proportion of Counties with Positive Change	0.8804 (0.3262)
Proportion of Counties with Above Mean Change in Full-day	0.3913 (0.4907)
Proportion of Counties with 90% or More Kindergarten Students in Full-day	0.3478 (0.4789)

The school achievement analyses rely on administrative data on schools and districts in Indiana to capture school and district characteristics (e.g., per pupil expenditures, average daily attendance, average teacher salary, number of students, number of full-time equivalent teachers) and third and fourth grade reading and math performance. Student achievement

data and many school and district characteristics were collected from state-administered, publicly available datasets. This information was also supplemented by school and district characteristics in the Common Core Data maintained by the National Center for Education Sciences at the U.S. Department of Education.

I employ administrative data and specially collected student assessment data from a representative, statewide sample of first grade classrooms to assess the impact of the statewide policy change in the subsequent grade across cohorts. As mentioned, student achievement as measured by standardized assessments in reading and math in third and fourth grades was collected from state databases. The assessment in use over this time period was the Indiana Statewide Testing for Educational Progress–Plus (ISTEP+). I also utilize first grade literacy scores on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) collected on a statewide representative sample of schools offering kindergarten and first grade to the pre-policy change (2007–08 for first graders) and post-policy change (2008–09) cohorts. Randomly selected first grade classrooms were tested with the sampled schools, drawn by NORC at the University of Chicago to be representative on the following dimensions:

- Poverty status proxy (proportion of kindergarten students receiving free or reduced-price lunch in the 2006–07 school year): low, high
- Locale: suburban, rural, urban
- Region: central, northeast, northwest, southeast, southwest
- Full-day kindergarten enrollment (proportion of kindergarten students in full-day kindergarten in the 2006–07 school year): low, high

For the maternal employment analyses, data on full-day kindergarten provision were linked by geography from the school and school district level to county to allow for merging with employment data. The workforce data for women in the relevant age range was obtained from data that Indiana contributes to the U.S. Bureau of Labor Statistics' Quarterly Census of Employment and Wages. While data from 2000 through 2013 is included, the focal pre- and post- quarters are Q1-Q3 in 2007, and Q4 in 2007 and Q1-Q2 in 2008, respectively. Moreover, to conduct the triple-differences analysis, employment data for different age groups are included in the merged data. The use of these different subgroups to generate causal estimates is described in further detail in the following section. The employment data contains approximately 50,000 observations of total employment, new hires, and job separations aggregated by sex and age group at the quarter by year by county level. For each county, there are observations for each quarter in each year 2000 through 2013 for men and women in each of the following age groups:

- 19-24 years old,
- 25-34 years old,
- 35-44 years old, and
- 45-54 years old.

For the purposes of this analysis, women ages 25-44 are considered "at risk" for treatment in that they are the most likely age group affected by full-day kindergarten expansions as child care.

3.2 Empirical Strategy

To account for district and school selection into greater availability of full-day kindergarten and to account for the threat of history in making pre-post comparisons of policy impact, I utilize a triple-difference approach to leverage cross-school (or cross-county), cross-time, and cross-cohort variation in exposure to full-day kindergarten availability. The triple-difference approach avoids concerns about time-varying, within-school or within-county characteristics as I difference out from the estimation of full-day kindergarten impact changes in outcomes for unaffected cohorts (i.e., older students in the same schools).

A simple specification would leverage pre-post variation within schools or counties as follows:

$$y_s = \alpha + \beta X_s + \theta POST_s + \delta_s + \varepsilon_s \tag{1}$$

The triple-differences empirical approach, leveraged in this paper, employs the following model:

$$y_{gjst} = \alpha + \beta X_{jst} + \theta FDK_{gjst} + \Gamma_g + \lambda_j + \delta_s + \gamma_t + \varepsilon_{gst} \tag{2}$$

where *FDK* captures change in full-day kindergarten enrollment, continuously or as a binary variable, turning on for affected groups, g, in treated schools or counties, s, in the post-policy time period (quarter j and year t). In the achievement analyses, the outcome is standardized test scores, and all pairwise interactions of school, year, and grade level are included as well as school and year dummies. Time-varying controls are also included in some specifications of the model. In the employment analyses, y is the employment outcome of total employment or new hires per 100 people in the specific subgroup. Age group, quarter, year, and county fixed effects are included. Regressions are population-weighted, and heteroskedasticity-robust standard errors are clustered on school or county.

3.3 Descriptive Statistics

Tables 1 through 3 in the appendix display summary statistics for the binary measures of full-day kindergarten change at the school level. Notably, the dimensions on which the treated and untreated group differ statistically change with the different specifications of treatment. When considering any increase in full-day kindergarten enrollment ("positive change"), the treatment schools are larger, with more pre-treatment kindergarten students and full-time equivalent teachers. They are also in districts that spent more per student on average before the policy change, and generally performed worse on measures of third and fourth grade reading and math achievement. This type of negative selection into the provision of full-day kindergarten, possibly as a means to improve subsequent test scores, is common at the national level.

When comparing high change (i.e., above average increases in full-day kindergarten participation) schools to other schools, they differ only on demographic composition–more white and less black and Hispanic than their non-high change counterparts–and district-level average daily attendance and teacher salary, both of which are lower in high change schools.

Finally, those schools that changed to full provision of full-day kindergarten are more disadvantaged and smaller (e.g., fewer students, fewer teachers, lower student-teacher ratio) than other schools. Their districts also had lower average teacher salaries before the policy change, and they exhibit worse student performance at the school level. The selection story is most pronounced with this cut at "treated" and "untreated" schools.

4 Results

Differences in achievement and maternal employment associated with the pre- and post-policy change time period point to the need for the triple-differences analyses conducted in this paper. Figure 8 displays the pre-/post- changes in achievement in schools that exhibited low and high starting points in full-day kindergarten availability. The figure indicates that effects might be more concentrated in places that did not have significant full-day kindergarten availability prior to the policy changes, though the noisiness in the estimates should also be noted.

Figure 8: Pre-/Post- Changes in Achievement in Schools with Low and High Baselines

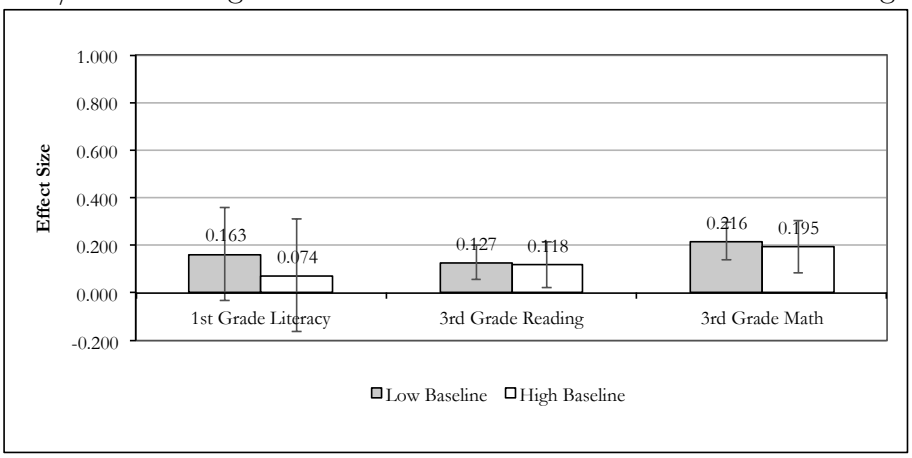
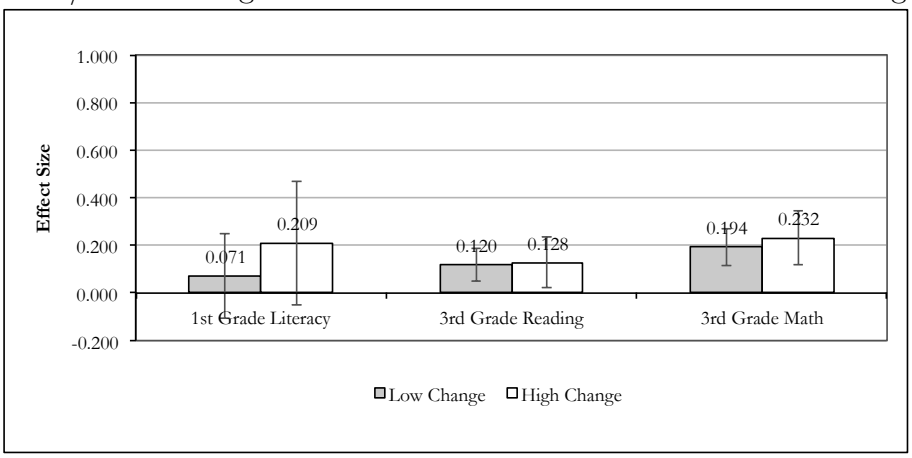


Figure 9 similarly displays pre-/post- differences in places with low and high *change* in full-day kindergarten provision over this time period. Again, the differences are suggestive of a concentration of effects in high change places, though the estimates are again imprecise.

Figure 9: Pre-/Post- Changes in Achievement in Schools with Low and High Change



Figures 10 and 11 show pre-/post- changes in employment outcomes for all counties in the state and then within counties that experienced positive or above average change in full-day kindergarten provision, in counties that were above the mean in provision after the policy change, in counties that provided full-day kindergarten for all students after the policy change, and in counties that shifted to full provision of full-day kindergarten in this timeframe. As is evident in the tables, employment differences are positive in these places while flat overall.

Figure 10: Pre-/Post- Differences in Total Employment

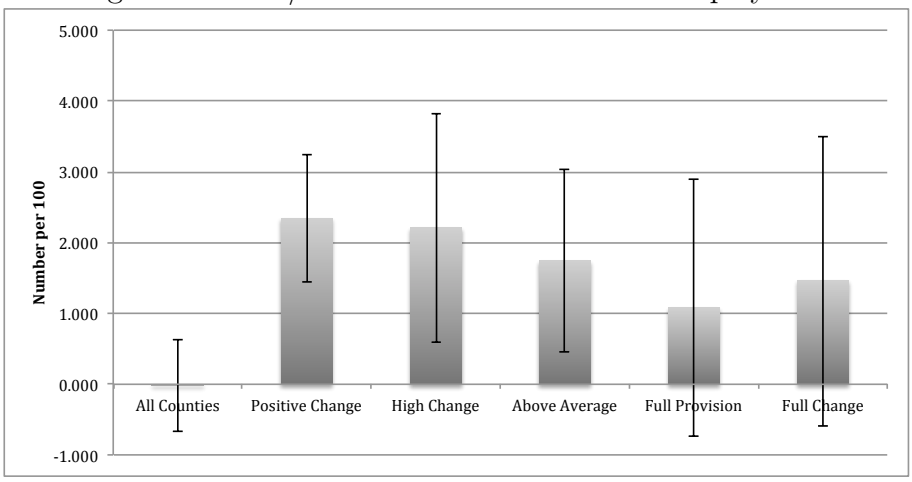
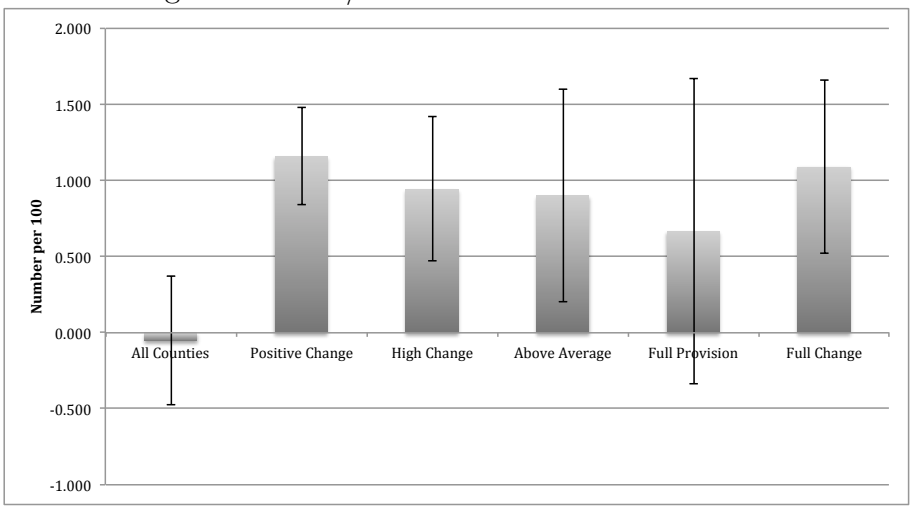


Figure 11: Pre-/Post- Differences in New Hires



For the reasons outlined previously and in the summary statistics, this type of approach is insufficient in capturing the endogeneity of full-day kindergarten provision. This study design employs an additional attempt at dealing with threats to the internal validity of impact estimates by including untreated cohorts within the same schools. The pre-post changes for older students within the same schools, or older women within the same counties, may capture the other time-varying programs, resources, and events affecting student performance or female labor force participation over this time period. Tables 4 through 9 present the results of the triple-difference models, beginning with the continuous measure of full-day kindergarten change and then leveraging the alternative specifications of the treatment dummy.

The models of impact on school achievement are presented in Tables 4 through 7. While

the point estimates for the coefficient of interest in the first grade outcomes models suggest positive relationships, they are not statistically significant. Because the first grade literacy assessment was collected on only a subsample of schools in the state, these estimates lack precision. Notably, the post-policy change dummy is consistently positive and significant across models suggesting a secular upward trend in academic performance in the state across this same time period. Consistent with patterns of fade out in both the observational full-day kindergarten literature and in the broader evidence base on early childhood interventions, there are no effects in third grade and the coefficients are consistently zero. The evidence suggests that increases in full-day kindergarten provision have no school-level effects on later academic performance.

The analyses of maternal employment outcomes displayed in Tables 8 and 9, however, suggest a correspondence between increased full-day kindergarten provision and female employment, as measured both by total employment and new hires. The results indicate an increase of two to three new hires per 100 women associated with changes in full-day kindergarten availability resulting from the policy expansion. These results are particularly strong for new hires, and are consistent across specification of the treatment exposure variable.

This paper speaks to an as-yet-unanswered question in the literature about full-day kindergarten impact, that of the systemic return on the early investment in the form of subsequent student achievement. Rather than focusing on participant-level effects, this work sought to quantify the intention-to-treat effect, on both later school performance and female labor force participation, of making full-day kindergarten more available to students within schools. It could be argued that this is, in fact, the policy relevant parameter of interest.

References

[1] Anderson, P.M. & Levine, P.B. (2000). Child care and mothers' employment decisions. In Finding Jobs: Work and Welfare Reform, D.E. Card & R.M. Blank (Eds.). New York: Russell Sage Foundation.

[2] Angrist, J.D. & Pischke, J. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press.

[3] Becker, G. (1964). Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. New York: Columbia University Press.

[4] Berger, M.C. & Black, D.A. (1992). Child care subsidies, quality of care, and the labor supply of low-income, single mothers. Review of Economics and Statistics, 74(4), 635-642.

[5] Blau, D.M. (2003). Do child care regulations affect the child care and labor markets? Journal of Policy Analysis and Management, 22(3), 443-465.

[6] Blau, D.M. & Hagy, A.P. (1998). The demand for quality in child care. Journal of Political Economy, 106(1), 104-146.

[7] Blau, D.M. & Robins, P.K. (1988). Child care costs and family labor supply. The Review of Economics and Statistics, 70(3), 374–381.

[8] Blau, D.M. & Tekin, E. (2007). The determinants and consequences of child care choices of single mothers in the USA. Journal of Population Economics, 20(4), 719-741.

[9] Cannon J.S., Jacknowitz, A., & Painter, G. (2006). Is full better than half? Examining the longitudinal effects of full-day kindergarten attendance. Journal of Policy Analysis and Management, 25(2), 299–321.

[10] Carneiro, P.M. & Heckman, J.J. (2003). Human capital policy. IZA Discussion Paper No. 821. Available at: http://ssrn.com/abstract=434544.

[11] Cascio, E.U. (2009a). Maternal labor supply and the introduction of kindergartens into American public schools. Journal of Human Resources, 44(1), 140-170.

[12] Cascio, E.U. (2009b). Do investments in universal early education pay off? Long-term effects of introducing kindergartens into public schools. NBER Working Paper No. 14951.

[13] Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W. & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. Quarterly Journal of Economics, 126(4), 1593-1660.

[14] Connelly, R. (1992). The effect of child care costs on married women's labor force participation. The Review of Economics and Statistics, 74(1), 83–90.

[15] Connelly, R. & Kimmel, J. (2003). The effect of child care costs on the employment and welfare recipiency of single mothers. Southern Economic Journal, 69(3), 498-519.

[16] Currie, J. (2001). Early childhood education programs. Journal of Economic Perspectives, 15(2), 213-238.

[17] DeCicca, P. (2007). Does full-day kindergarten matter? Evidence from the first two years of schooling. Economics of Education Review, 26(1), 67-82.

[18] Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. American Economic Journal: Applied Economics, 1(3), 111-134.

[19] Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., & Duckworth, K. (2007). School readiness and later achievement. Developmental Psychology, 43(6), 1428-1446.

[20] Education Commission of the States (2005). How states fund full-day kindergarten. Denver, CO: Author.

[21] Education Commission of the States (2008). State statutes regarding kindergarten: Policies concerning district offering of and student attendance in full- and half-day kindergarten programs. Denver, CO: Author. Available at: http://www.ecs.org/clearinghouse/78/60/7860.pdf

[22] Elicker, J. & Mathur, S. (1997). What do they do all day? Comprehensive evaluation of a full-day kindergarten. Early Childhood Research Quarterly, 12, 459-480.

[23] Fitzpatrick, M. (2010). Preschoolers enrolled and mothers at work? The effects of universal pre- kindergarten. Journal of Labor Economics, 28(1), 51-85.

[24] Fitzpatrick, M. (2012). Revising our thinking about the relationship between maternal labor supply and preschool. Journal of Human Resources, 47(3), 583-612.

[25] Fryer, R.G. & Levitt, S.D. (2004). Understanding the black-white test score gap in the first two years of school. Review of Economics and Statistics, 86(2), 447-464.

[26] Fryer, R.G. & Levitt, S.D. (2006). The black-white test score gap through third grade. American Law and Economics Review, 8(2), 249-281.

[27] Fusaro, J.A. (1997). The effect of full-day kindergarten on student achievement: A meta- analysis. Child Study Journal, 27, 269-279.

[28] Gelbach, J.B. (2002). Public schooling for young children and maternal labor supply. American Economic Review, 92(1), 307–322.

[29] Gibbs, C., Ludwig, J. & Miller, D.L. (2011). Does Head Start do any lasting good? NBER Working Paper No. 17542. Cambridge, MA: National Bureau of Economic Research.

[30] Gibbs, C., Ludwig, J. & Miller, D.L. (2013). Head Start: Origins and impacts. In Legacies of the War on Poverty, M. Bailey & S. Danziger (Eds.), New York: Russell Sage Foundation Press.

[31] Gibbs, C.R. (2012). Experimental and Quasi-experimental Evidence on the Impact of Full-day Kindergarten. Ph.D. diss., University of Chicago.

[32] Gibbs, C.R. (2014). Experimental evidence on early intervention: The impact of full-day kindergarten. Frank Batten School of Leadership and Public Policy Working Paper No. 2014–04. Charlottesville, VA: University of Virginia.

[33] Heckman, J.J. (2000). Policies to foster human capital. Research in Economics, 54(1), 3-56.

[34] Heckman, J.J. & Lochner, L. (2000). Rethinking myths about education and training: Understanding the sources of skill formation in a modern economy. In Securing the Future: Investing in Children from Birth to College, S. Danziger & J. Waldfogel (Eds.). New York: Russell Sage.

[35] Heckman, J.J. & Masterov, D.V. (2007). The productivity argument for investing in young children. Review of Agricultural Economics, 29(3), 446-493.

[36] Indiana General Assembly. (2007). Public Law 234–2007. Indianapolis, IN: Author. Retrieved August 12, 2009 from http://www.in.gov/legislative/pdf/acts_2007.pdf.

[37] Karweit, N. (1989). Effective kindergarten programs and practices for students at risk. In Effective Programs for Students at Risk, R.E. Slavin, N.L. Karweit, & N.A. Madden (Eds.). Boston: Allyn and Bacon.

[38] Karoly, L., Greenwood, P.W., Everingham, S.S., Hoube, J.M., Kilburn, R.C., Rydell, C.P., et al. (1998). Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions. Santa Monica, CA: RAND Corporation.

[39] Kauerz, K. (2005). Full-day kindergarten: A study of state policies in the United States. Denver, CO: Education Commission of the States.

[40] Knudsen, E.I., Heckman, J.J., Cameron, J. & Shonkoff, J.P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. Proceedings of the National Academy of Sciences, 103(27), 10155-10162.

[41] Krueger, A. B. (1999). Experimental estimates of education production functions. Quarterly Journal of Economics, 114(2), 497-532.

[42] Krueger, A.B. (2003). Economic considerations and class size. Economic Journal, 113, F34-F63.

[43] Lee, V.E. & Burkam, D.T. (2002). Inequality at the starting gate: Social background differences in achievement as children begin school. Washington, DC: Economic Policy Institute.

[44] Lee, V.E., Burkam, D.T., Ready, D., Honigman, J., & Meisels, S.J. (2006). Full-day versus half-day kindergarten: In which program do children learn more? American Journal of Education, 112(2), 163-208.

[45] Lemke, R., Witte, A., Queralt, M. & Witt, R. (2000). Child care and the welfare to work transition. NBER Working Paper No. 7583. Cambridge, MA: National Bureau of Economic Research.

[46] Lovell, R., Kochanek, J., Mathers, C. & Burke, M. (2009). Snapshots of Indiana's full-day kindergarten programs before and after the state's funding increase for the program (REL Technical Brief, REL 2009–No. 013). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest.

[47] Plucker, J.A., Eaton, J., Rapp, K.E., Lim, W., Nowak, J., Hansen, J., & Bartleson, A. (2004). The effects of full day versus half day kindergarten: Review and analysis of national and Indiana data. Mimeo. Center for Education Evaluation and Policy, Indiana University.

[48] Princiotta, D., Flanagan, K.D., & Germino Hausken, E. (2006). Fifth grade: Findings from the fifth- grade follow-up of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99. (NCES 2006-038) U.S. Department of Education. Washington, DC: National Center for Education Statistics.

[49] Morgan, S.L. & Winship, C. (2007). Counterfactuals and Causal Inference: Methods and Principles for Social Research. New York: Cambridge University Press.

[50] Murnane, R.J., Willett, J.B., Bub, K.L. & McCartney, K. (2006). Understanding trends in the black-white achievement gaps during the first years of school. Brookings-Wharton Papers on Urban Affairs, 97-135.

[51] Ribar, D. (1992). Child care and the labor supply of married women: reduced form evidence. Journal of Human Resources, 27(1), 134–165.

[52] Schweinhart, L.J., Montie, J., Xiang, Z., Barnett, W.S., Belfield, C.R. & Nores, M. (2005). Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40. Ypsilanti, MI: High/Scope Press.

[53] Shonkoff, J.P. & Phillips, D.A., editors. (2000). From Neurons to Neighborhoods: The Science of Early Childhood Development. Washington, DC: National Academy Press.

[54] Tekin, E. (2004). Child care subsidy receipt, employment, and child care choices of single mothers. IZA Discussion Paper No. 1121.

[55] U.S. Department of Commerce, Bureau of the Census, Current Population Survey (CPS), October supplement, selected years 1977–2001.

[56] U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), Survey of local government finances, school systems (F-33), 2007-08 (FY 2008).

[57] Votruba-Drzal, E., Li-Grining, C.P. & Maldonado-Carreño, C. (2008). A developmental perspective on full- versus part-day kindergarten and children's academic trajectories through fifth grade. Child Development, 79(4), 957-978.

[58] Walston, J. & West, J. (2004). Full-day and half-day kindergarten in the United States: Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99. (NCES 2004-078) U.S. Department of Education. Washington, DC: National Center for Education Statistics.

[59] Yan, W. & Lin, Q. (2005). Effects of class size and length of day on kindergartners' academic achievement: Findings from the Early Childhood Longitudinal Study. Early Education & Development, 16, 49-68.

[60] Zvoch, K., Reynolds, R.E., & Parker, R.P. (2008). Full-day kindergarten and student literacy growth: Does a lengthened school day make a difference? Early Childhood Research Quarterly, 23(1), 94-107.

Tables

Table 1. Summary Statistics of Positive Change and Other Schools.

	Positive Change Schools n=579	Other Schools n=465	p-value
Panel A. School characteristics			
Title I schoolwide	0.64 (.482)	0.64 (.480)	0.859
Title I eligible	0.84 (.371)	0.79 (.405)	0.079
Total students	450.35 (164.08)	414.15 (164.20)	0.000**
Pre-K students	6.04 (12.94)	5.90 (17.95)	0.883
K students	69.62 (29.33)	62.87 (28.69)	0.000**
Student/teacher ratio	17.24 (4.60)	17.56 (2.77)	0.181
Full-time equivalent teachers	26.30 (8.67)	23.70 (8.85)	0.000**
% Black and Hispanic	0.21 (.268)	0.18 (.271)	0.174
% White	0.73 (.282)	0.76 (.283)	0.111
% Free or reduced-price lunch	0.48 (.228)	0.45 (.242)	0.110
Panel B. School district characteristics			
K-4th grade students	239.59 (299.94)	223.38 (289.54)	0.382
Per-pupil expenditures	11189.19 (1841.59)	10911.07 (1876.87)	0.017*
Daily attendance	8990.10 (10096.22)	8427.48 (9623.02)	0.366
Teacher salary	47209.54 (4140.19)	47624.48 (4605.82)	0.130
Panel C. Pre-policy test scores			
1st grade reading (2007-08)	-.087 (.525)	-.079 (.483)	0.931
3rd grade reading (2009-10)	-.041 (.927)	.091 (1.026)	0.032*
3rd grade math (2009-10)	-.051 (.935)	.110 (1.026)	0.010*
4th grade reading (2010-11)	.214 (.925)	.344 (.968)	0.036*
4th grade math (2010-11)	.130 (.953)	.252 (.984)	0.054

Notes: Standard deviations are provided in parentheses. P-values from the statistical tests of mean differences are displayed.

Table 2. Summary Statistics of High Change and Other Schools.

	High Change Schools n=335	Other Schools n=709	p-value
Panel A. School characteristics			
Title I schoolwide	0.62 (.486)	0.65 (.478)	0.492
Title I eligible	0.84 (.368)	0.81 (.395)	0.212
Total students	432.98 (170.47)	434.82 (162.53)	0.867
Pre-K students	4.85 (10.50)	6.51 (17.18)	0.103
K students	65.98 (28.93)	66.91 (29.38)	0.633
Student/teacher ratio	17.04 (2.42)	17.54 (4.42)	0.051
Full-time equivalent teachers	25.34 (9.06)	25.05 (8.74)	0.611
% Black and Hispanic	0.17 (.235)	0.21 (.283)	0.013*
% White	0.78 (.253)	0.73 (.295)	0.007**
% Free or reduced-price lunch	0.46 (.194)	0.47 (.252)	0.418
Panel B. School district characteristics			
K-4th grade students	206.73 (275.17)	244.33 (303.71)	0.057
Per-pupil expenditures	11114.98 (11042.17)	11042.17 (1847.36)	0.556
Daily attendance	7792.33 (9384.60)	9183.98 (10090.05)	0.036*
Teacher salary	46952.74 (3869.08)	47602.70 (4556.36)	0.026*
Panel C. Pre-policy test scores			
1st grade reading (2007-08)	-.118 (.529)	-.063 (.495)	0.547
3rd grade reading (2009-10)	.023 (.790)	.015 (1.051)	0.905
3rd grade math (2009-10)	-.008 (.837)	.033 (1.040)	0.534
4th grade reading (2010-11)	.267 (.808)	.273 (1.006)	0.924
4th grade math (2010-11)	.189 (.870)	.182 (1.013)	0.915

Notes: Standard deviations are provided in parentheses. P-values from the statistical tests of mean differences are displayed.

Table 3. Summary Statistics of Schools that Changed to Full-day Only and Others.

	Change to All Full-day Schools n=202	Other Schools n=854	p-value
Panel A. School characteristics			
Title I schoolwide	0.71 (.455)	0.62 (.486)	0.025*
Title I eligible	0.86 (.352)	0.81 (.396)	0.094
Total students	384.54 (150.13)	446.18 (166.06)	0.000**
Pre-K students	5.28 (10.61)	6.14 (16.24)	0.000**
K students	57.55 (26.00)	68.69 (29.65)	0.000**
Student/teacher ratio	16.12 (2.46)	17.70 (4.10)	0.000**
Full-time equivalent teachers	24.03 (9.11)	25.39 (8.76)	0.048*
% Black and Hispanic	0.19 (.278)	0.20 (.268)	0.864
% White	0.75 (.291)	0.74 (.282)	0.543
% Free or reduced-price lunch	0.54 (.213)	0.45 (.236)	0.000**
Panel B. School district characteristics			
K-4th grade students	246.44 (324.39)	227.91 (286.68)	0.424
Per-pupil expenditures	11191.69 (2077.43)	11035.46 (1802.00)	0.285
Daily attendance	8996.66 (11074.21)	8642.54 (9543.71)	0.649
Teacher salary	46086.17 (4721.848)	47686.22 (4214.91)	0.000**
Panel C. Pre-policy test scores			
1st grade reading (2007-08)	-.208 (.448)	-.058 (.516)	0.206
3rd grade reading (2009-10)	-.191 (.901)	.066 (.987)	0.001**
3rd grade math (2009-10)	-.176 (.914)	.068 (.988)	0.002**
4th grade reading (2010-11)	.058 (.857)	.321 (.959)	0.001**
4th grade math (2010-11)	-.030 (.889)	.233 (.980)	0.001**

Notes: Standard deviations are provided in parentheses. P-values from the statistical tests of mean differences are displayed.

Table 4. Triple-Difference Model with Continuous Full-day K Change Measure.

	Standardized Test Score Outcome				
	1st Grade Reading	3rd Grade Reading		3rd Grade Math	
Full-day K Exposure	0.100	-0.020	0.001	0.031	0.056
	(1.266)	(0.181)	(0.179)	(0.184)	(0.180)
Post	0.973**	0.121**	0.159**	-0.009	0.027
	(0.316)	(0.033)	(0.035)	(0.035)	(0.037)
1st Grade	0.574	—	—	—	—
	(0.316)				
3rd Grade	—	-0.318**	-0.318**	-0.079**	-0.079**
		(0.000)	(0.000)	(0.000)	(0.000)
Post*Grade	0.238	-0.119	-0.125	0.048	0.043
	(0.490)	(0.073)	(0.073)	(0.077)	(0.078)
School fixed effects	X	X	X	X	X
School*Post interactions	X	X	X	X	X
School*Grade interactions	X	X	X	X	X
Controls			X		X
Adj. R^2	0.781	0.796	0.796	0.786	0.787
n	2,462	4,219	4,191	4,220	4,192

Notes: Robust standard errors in parentheses, clustered on school. Time-varying controls include number of students, student-teacher ratio, full-time equivalent teachers, and student demographic characteristics. Because of limited sample size and collinearity, 1st grade reading models cannot be estimated with controls.

* $p < 0.05$, ** $p < 0.01$

Table 5. Triple-Difference Model with Binary Measure of Increase in Full-day K.

	Standardized Test Score Outcome				
	1st Grade Reading	3rd Grade Reading		3rd Grade Math	
Positive Change in Full-day K	0.153	0.017	0.020	0.084	0.089
	(0.800)	(0.120)	(0.121)	(0.127)	(0.127)
Post-policy	0.960**	0.119**	0.132**	-0.029	-0.013
	(0.246)	(0.039)	(0.037)	(0.040)	(0.038)
1st Grade	0.561*	—	—	—	—
	(0.246)				
3rd Grade	—	-0.318**	-0.318**	-0.079**	-0.079**
		(0.000)	(0.000)	(0.000)	(0.000)
Post*Grade	0.173	-0.133	-0.136	0.008	0.006
	(0.630)	(0.092)	(0.093)	(0.098)	(0.099)
School fixed effects	X	X	X	X	X
School*Post interactions	X	X	X	X	X
School*Grade interactions	X	X	X	X	X
Controls			X		X
Adj. R^2	0.777	0.791	0.791	0.780	0.780
n	2,318	3,908	3,888	3,909	3,935

Notes: Robust standard errors in parentheses, clustered on school. Time-varying controls include number of students, student-teacher ratio, full-time equivalent teachers, and student demographic characteristics. Because of limited sample size and collinearity, 1st grade reading models cannot be estimated with controls.

* $p < 0.05$, ** $p < 0.01$

Table 6. Triple-Difference Model with Binary Measure of Above Average Change in Full-day K.

	Standardized Test Score Outcome				
	1st Grade Reading	3rd Grade Reading		3rd Grade Math	
High Change in Full-day K	0.063	0.047	0.058	0.074	0.088
	(0.803)	(0.127)	(0.127)	(0.132)	(0.132)
Post-policy	0.972**	0.130**	0.146**	0.001	0.019
	(0.319)	(0.036)	(0.037)	(0.038)	(0.039)
1st Grade	0.573	—	—	—	—
	(0.319)				
3rd Grade	—	-0.318**	-0.318**	-0.079**	-0.079**
		(0.000)	(0.000)	(0.000)	(0.000)
Post*Grade	0.240	-0.139	-0.144*	0.031	0.027
	(0.487)	(0.072)	(0.073)	(0.076)	(0.077)
School fixed effects	X	X	X	X	X
School*Post interactions	X	X	X	X	X
School*Grade interactions	X	X	X	X	X
Controls			X		X
Adj. R^2	0.776	0.791	0.791	0.780	0.781
n	2,318	3,908	3,888	3,909	3,889

Notes: Robust standard errors in parentheses, clustered on school. Time-varying controls include number of students, student-teacher ratio, full-time equivalent teachers, and student demographic characteristics. Because of limited sample size and collinearity, 1st grade reading models cannot be estimated with controls.

* $p < 0.05$, ** $p < 0.01$

Table 7. Triple-Difference Model with Binary Measure of Change to Offering Only Full-day K.

	Standardized Test Score Outcome				
	1st Grade Reading	3rd Grade Reading		3rd Grade Math	
Full-day K Only	0.248	-0.045	-0.034	0.009	0.026
	(1.127)	(0.162)	(0.162)	(0.164)	(0.162)
Post-policy	0.891	0.117**	0.132**	-0.010	0.007
	(0.524)	(0.032)	(0.032)	(0.034)	(0.034)
1st Grade	0.492	—	—	—	—
	(0.524)				
3rd Grade	—	-0.318**	-0.318**	-0.079**	-0.079**
		(0.000)	(0.000)	(0.000)	(0.000)
Post*Grade	0.217	-0.113	-0.116	0.053	0.050
	(0.411)	(0.064)	(0.065)	(0.068)	(0.069)
School fixed effects	X	X	X	X	X
School*Post interactions	X	X	X	X	X
School*Grade interactions	X	X	X	X	X
Controls			X		X
Adj. R^2	0.779	0.790	0.791	0.779	0.780
n	2,342	3,954	3,934	3,955	3,935

Notes: Robust standard errors in parentheses, clustered on school. Time-varying controls include number of students, student-teacher ratio, full-time equivalent teachers, and student demographic characteristics. Because of limited sample size and collinearity, 1st grade reading models cannot be estimated with controls.

* $p < 0.05$, ** $p < 0.01$

Table 8. Triple-Difference Models with Female Employment Outcomes.

	Female Employment Outcome			
	Total Employment		New Hires	
Full-day K Exposure	4.287*	4.161*	5.157**	9.289**
	(1.688)	(1.674)	(0.597)	(1.406)
County fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Quarter fixed effects	X	X	X	X
Population-weighted		X		X
Adj. R^2	0.933	0.957	0.464	0.556
n	2,208	2,208	2,208	2,208
Positive Change in Full-day K	2.594**	1.802**	3.661**	4.465**
	(0.587)	(0.579)	(0.206)	(0.336)
County fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Quarter fixed effects	X	X	X	X
Population-weighted		X		X
Adj. R^2	0.934	0.957	0.529	0.623
n	2,208	2,208	2,208	2,208

Notes: Robust standard errors in parentheses, clustered on county.

* $p < 0.05$, ** $p < 0.01$

Table 9. Triple-Difference Models with Female Employment Outcomes (continued).

	Female Employment Outcome			
	Total Employment		New Hires	
High Change in Full-day K	1.983*	1.179	2.387**	2.872**
	(0.852)	(0.957)	(0.207)	(0.313)
County fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Quarter fixed effects	X	X	X	X
Population-weighted		X		X
Adj. R^2	0.933	0.956	0.451	0.524
n	2,208	2,208	2,208	2,208
Full-day K Only	2.778**	2.858**	2.351**	3.063**
	(0.776)	(0.796)	(0.244)	(0.424)
County fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Quarter fixed effects	X	X	X	X
Population-weighted		X		X
Adj. R^2	0.934	0.957	0.448	0.518
n	2,208	2,208	2,208	2,208

Notes: Robust standard errors in parentheses, clustered on county.

* $p < 0.05$, ** $p < 0.01$

Early Effects of the 2010 Affordable Care Act Medicaid Expansion on Labor Market Outcomes

Angshuman Gooptu
The School of Public and Environmental Affairs (SPEA), Indiana University

Asako S. Moriya[1]
The School of Public and Environmental Affairs (SPEA), Indiana University

and

Kosali Simon
The School of Public and Environmental Affairs (SPEA), Indiana University and NBER

This paper was prepared with funding from the U.S. Department of Labor. The views expressed are those of the authors and should not be attributed to the Federal Government or the Department of Labor.

Abstract

We test for early labor market effects in terms of eased job-lock from the Affordable Care Act Medicaid expansion of January 2014 that targeted non-elderly low-income adults. An expansion of health insurance options not tied to employment could increase job turnover among newly eligible low-income populations, enabling them to move to preferred jobs (measured here as higher wage jobs). We use a differences-in-differences (DD) strategy, comparing rates of job turnover and wages, after the policy implementation relative to the outcomes before the implementation, among the treatment group (low-educated populations in Medicaid expansion states) relative to the control group (similar individuals in non-expansion states). We use educational level rather than income to define groups because of the potential endogeneity of income, but caution that since education is only a crude proxy for Medicaid eligibility, measurement error may affect our results. However, we also use alternative estimation strategies and find our conclusions are unchanged. We examine triple-differences (DDD) models with an additional within-state control group of those who have higher education. We also find our results are unchanged when we use potentially endogenous measures of income; in future drafts we plan to instrument for actual eligibility with a simulated policy measure, and to use a one year lagged income measure. We conduct tests to verify that our relevant DD and DDD comparisons satisfy the common trends assumption before proceeding with our analysis, using Current Population Survey (CPS) Basic Monthly data from January 2005 through August 2014. We use these data because of large sample sizes and quick release dates. However, the CPS Basic Monthly data do not contain information on health insurance status itself. We find no statistically significant evidence that the ACA Medicaid expansion increased job turnover rates or affected wages in either our base DD or DDD models. We caution that these are early results, and come from a data set in which we cannot estimate insurance impacts to estimate the elasticity of job transition with respect to health insurance.

[1] Present address for corresponding author: Asako S. Moriya, Service Fellow, the Center for Financing, Access, and Cost Trends (CFACT), the Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, Asako.Moriya@ahrq.hhs.gov, (301) 427-1428. The majority of this work was completed while the corresponding author held her position as a postdoctoral fellow at the SPEA, Indiana University. The paper represents the views of the authors, and no official endorsement by the Agency for Healthcare Research and Quality or the Department of Health and Human Services is intended or should be inferred.

I. Introduction

The Patient Protection and Affordable Care Act (ACA) of 2010 is projected to extend health insurance coverage to 32 million people (CBO, 2012). The ACA expands the number with health insurance through many features including ones that encourage employer-sponsored insurance (ESI), as exemplified by the (currently postponed) introduction of an employer mandate. The predominant ways the ACA currently expands insurance options are not tied to one's own employment, as they occur through Medicaid expansion, subsidized Marketplace insurance plans, and coverage for young adults through parental employer policies. These new health insurance options could potentially affect the behaviors of workers and employers, leading to consequences for labor market outcomes.

The purpose of this paper is to analyze the early effects of the 2014 ACA Medicaid expansion on "job lock", a phenomenon in which individuals are less likely to move to preferred jobs (possibly measured as higher wage jobs) because of the traditional tie between employment and health insurance. The 2014 Medicaid expansion varies by state, enabling a more robust study design than otherwise. Before the ACA, Medicaid eligibility for childless low-income adults (and less so for parents) was extremely limited. The ACA initially intended to expand Medicaid to all non-elderly Americans with household incomes less than 138% of the federal poverty level (FPL). In June 2012 the U.S. Supreme Court ruled the ACA's Medicaid expansion to be unconstitutionally coercive of states' rights, and as of June 10, 2014 27 states decided to expand Medicaid for 2014, 21 states have decided not to implement the expansion, and three states are still debating whether to expand Medicaid (Kaiser Family Foundation, 2014). This state-by-state variation in Medicaid expansion for 2014 provides us with a potential source of variation to identify causal impacts of Medicaid on labor market outcomes.

We use Current Population Survey (CPS) Basic Monthly data collected by the United States Bureau of Labor Statistics as the main data source for testing the job-lock impacts of Medicaid expansions. Due to its large sample size and almost immediate release of data, the CPS Basic Monthly dataset is a valuable data source for timely analysis of labor market behavior. The CPS Basic Monthly dataset includes rich information on point-in-time job characteristics as well as demographic, geographic (state) and socioeconomic characteristics of individuals.

The implementation of the ACA will have significant impacts on the welfare of Americans, and these impacts could extend beyond access to health insurance and health care. In our paper we test the potential for effects on job transitions from one particular ACA provision, the Medicaid expansion. A large literature exists on the impact of health insurance on job-lock and labor markets in general, and several new papers find varying results from recent state Medicaid expansions prior to the ACA (Dague et al., 2014; Garthwaite et al., 2014; and Baicker et al., 2014).

By reducing job-lock, health insurance not tied to employment could increase job mobility and increase wages (Monheit and Cooper, 1994; and Gruber and Madrian, 2004). While the theory of job-lock release applies to those with ESI as a counterfactual, our data set does not contain information on health insurance held during 2013.[2] Moreover, those who held ESI in 2013 are not necessarily the group whose counterfactual insurance status in 2014 would involve ESI as there is substantial movement from year to year in insurance status. On the other hand, those eligible for Medicaid benefits may not display signs of released job lock after insurance

[2] Prior papers show evidence of non-trivial ESI prevalence among low-income adults. For example, Hamersma and Kim (2009) show using Survey of Income and Program Participation data from the 1996 and 2001 panels that the rate of ESI in own name is 38.4% among men below 100% FPL, 29.8% among married women below 100%FPL, and 28.3% among unmarried women below 100% FPL among working parents aged 20-54 years. We find in ASEC CPS data from years prior to 2013 that the rates are somewhat lower among childless adults, and that there has been a secular decline in ESI during the recessionary years, but that ESI rates are nonetheless still non-trivial for low-income adults.

expansions if their labor market behavior is largely independent of health insurance. For example, they may have a low marginal willingness to pay for health benefits and/or not work for employers who offer health insurance. This is consistent with recent evidence observed through an experimental design in Oregon (Baicker et al., 2014) where Medicaid expansion has little effect on work behavior, although other estimates point to much larger labor market reactions to Medicaid expansions (Garthwaite et al., 2013). Given the theoretical ambiguity and conflicting evidence from past Medicaid expansions, our research will provide timely and preliminary information on how the ACA's Medicaid expansions affect labor market outcomes and can help inform discussions related to future steps in implementation of the ACA.

This paper makes contributions to two strands of the literature. The first strand is the literature on job lock and other labor market effects of insurance expansion in general. The second strand is the growing body of literature that analyzes the effects of the ACA implementation. Our preliminary analysis of the early effects of the Medicaid expansion prepares us to evaluate the first-year effects more rigorously when the full-year 2014 data on labor market and insurance outcomes from the post-implementation period become available.

II. Prior Literature

Many prior papers estimate effects of job lock but find mixed results (See Gruber and Madrian (2004), Fairlie et al. (2013), and Bailey and Chorniy (2014) for comprehensive literature reviews). Three prior papers are particularly relevant to our work because they use expansions in Medicaid-related public health insurance to investigate job lock; all three find evidence of released job lock following the insurance expansions. Bansak and Raphael (2008) found that the Children's Health Insurance Program (CHIP) expansion of Medicaid or Medicaid-

like programs to children lead to reduced job lock among their parents. Hamersma and Kim (2009) find that the expansions in Medicaid to parents after welfare reform of 1996 lead to reduced job lock in one population (unmarried mothers) but not in others (married mothers and fathers). Dave et al. (2013) examine the case of pregnancy-related Medicaid expansions of the late 1980s and early 1990s, finding that transitions out of current jobs occur with the receipt of new Medicaid insurance. However, Dave et al. (2013) only study transitions out of employment, rather than between jobs, which is relevant as the period studied is pregnancy and childbirth. Thus, there is evidence that health insurance has inhibited job movement among those adults affected directly or indirectly by prior Medicaid expansions.

Previous studies on the effects of public insurance expansion on labor market outcomes in general and specifically on job-lock have mainly focused on the expansion to subpopulations other than low-income childless adults, those targeted by the ACA Medicaid expansion. The subpopulations studied in the literature include single mothers (Moffitt and Wolfe, 1992; Meyer and Rosenbaum, 2001) and pregnant women (Dave et al., 2013), and these studies reported mixed findings. In particular, Moffitt and Wolfe (1992) used data from the 1984 wave of the Survey of Income and Program Participation (SIPP) and found that single mothers were more likely to reduce labor supply and increase participation in welfare (Aid to Families with Dependent Children, AFDC) when welfare offered greater expected Medicaid benefits. In addition, Meyer and Rosenbaum (2001) concluded that the increases in female labor supply observed from data in the March CPS supplements between 1984 and 1996 were a result of a concurrent policy implementation—the Earned Income Tax Credit (EITC)—rather than the increases in Medicaid eligibility income thresholds. Dave et al. (2013) used CPS March supplement data from 1986 to 1997 and found analogous results in the labor supply decisions of

pregnant women (between 18 and 39 years of age) eligible for Medicaid benefits; they estimated a 6% to 7% (13% to 16% among unmarried women with less than a high school degree) decline in labor supply among pregnant women associated with a 20 percentage point increase in Medicaid coverage.

More recent studies investigate the impact of health insurance coverage on the labor supply decisions of low-income childless adults and the general low income population (see Garrett and Kaestner (2014), and Dave et al. (2013), for reviews of these papers). Garthwaite et al. (2014) analyzed the effects of abrupt disenrollment of low- to moderate-income adult Medicaid recipients in Tennessee. The authors found that both job search behavior and employment increased tremendously after the disenrollment. They concluded that the ACA may reduce the labor supply of low-income adults, and their estimates have been influential in projections of the employment effects of the ACA. Dague et al. (2014) found a negative effect - between 0.9 to 7.2 percentage points in a differences-in-differences specification and between 6.1 to 10.6 percentage points in a regression discontinuity specification - of Wisconsin's BadgerCare Plus Core Plan's health insurance expansion on the labor supply of eligible childless adults (childless adults under 200% of the Federal Poverty Line, FPL). Despite the differences in the range of the estimates under two different identification strategy specifications, significant declines in labor supply were observed. However, since Dague et al. (2014) studied a Section 1115 program specific to the state of Wisconsin, stricter institutional rules regarding when the eligible population in Wisconsin can enter or exit the insurance program could create a “lock-in” enrollment effect. This can overstate the effects on labor markets relative to the ACA Medicaid expansions where the eligible population can transition between private and Medicaid coverage based on changes in their eligibility status.

In contrast to the statistically significant findings in Dague et al. (2014) and Garthwaite et al. (2014), Baicker et al. (2014) report that Medicaid coverage of traditionally uninsured adults in Oregon did not result in any significant reductions in labor supply. Moreover, the magnitude of the statistically insignificant point estimates are small, suggesting that Medicaid coverage is associated with a 1.6 percentage point decline in employment and a statistically insignificant reduction of $195 in annual income, both of which represent 3% (Baicker et al., 2014). Baicker et al.'s (2014) intent to treat (ITT) model estimates meet the gold standard for evaluation since it is based on results from the Oregon Health Insurance trial, in which low-income adult applicants - with incomes at or below the FPL and assets of $2000 - were chosen randomly from a lottery and assigned to control and treatment groups conditional on household size.

The credibility of the research design and the findings of Baicker et al. (2014) suggest that quasi-experimental designs may overstate the effects of health insurance coverage on labor supply. However, Baicker et al.'s (2014) experiment is specific to the Oregon population; state and regional differences may therefore drive some of the trends in labor supply among the low-income population who have recently become eligible for health insurance under the ACA. Our work adds to the literature by being the first to consider the early effects of the ACA Medicaid expansion in terms of eased job-lock and wages.

III. Conceptual Framework

The availability of Medicaid, an insurance option that is not tied to employment, may enable workers to switch to jobs that offer a better productivity match (as measured by higher wages) but do not provide health insurance. This type of transition would be less likely to occur in labor markets that have high rates of unemployment and low demand for labor. Higher wages could also result because of the extra wage compensation that an employer who does not provide

health insurance is able to give their worker. These theories apply to the extent that the worker would have been in a job that provided ESI absent the Medicaid expansion. We also expect that the effects on labor market outcomes are larger among those with higher health insurance demand, such as those who are older, given the correlation between age and health. General equilibrium effects may also be present in large-scale public health insurance expansions. Employers may increase their demand for Medicaid eligible workers, in the face of an employer mandate that would hold them responsible for health insurance provision to higher income workers, leading to higher job turnover. This would occur under the assumption that workers do not bear the full cost of their employer provided benefits.

In summary, we expect to find that Medicaid expansion reduces job lock, thereby increasing job transitions and increasing reported wages. To the extent that ESI is not relevant for the job markets of those affected by Medicaid expansions, or that the demand for health insurance is low for other reasons, we would be less likely to find these effects. We are also less likely to find the expected result of released job lock in job markets that are less flexible and characterized by low demand for labor.

IV. Data

Our main data source is CPS Basic Monthly data, which interviews around 60,000 households each month to collect basic demographic and labor force status information. The rotational structure of the CPS interviews households monthly for four months, then ceases to interview the household for eight month, and finally returns to the same households for four additional months of monthly interviews. These Basic Monthly data become publicly available approximately one month after the interviews, making it a valuable real-time resource for studying labor market results of the ACA.

Our outcome variables are a measure of job transitions, and the weekly wage at the current job. Job transitions are captured by a job change variable available in the CPS, which indicates whether workers stay employed by the same employer or changed employers between two consecutive months. Prior studies that use this variable include Bailey and Chorniy (2014) and Fairlie et al. (2013). The weekly wage is asked only of those in the "outgoing rotation group" of the CPS, which represent about one quarter of all respondents in a given month. For descriptive purposes, we also use the monthly family income variable in the CPS (asked of everyone), to estimate the fraction of FPL that is represented by the respondent's family income.[3] Table 1 shows descriptive statistics of the CPS Basic Monthly data from January 2005 through August 2014. The data set consists of 8,559,950 person-month observations for most variables. The number of valid observations is lower for job transitions (4,206,703) as this measure is available only for the subset of individuals who were employed in the month prior to the survey. Only about one third of these individuals (1,425,723) are asked the question about the weekly wage. About one tenth of our sample has less than high school completion and about one third have only a high school diploma.

We supplement the CPS data with information on the status of state Medicaid expansion decisions for 2005-2014, which we obtain through the Kaiser Family Foundation website as well as our own investigation of news reports and legislative records that adds details of income thresholds used by states over time and for different adult populations.

V. Identification Strategy

[3] In future drafts we will use a version of this data set that is linked to previous Annual Social and Economic Supplement (March) of the CPS in order to have measures of family size, family income, and prior year annual earnings and health insurance, for those individuals who can be matched across years.

Starting in January 2014, Medicaid was expanded to non-elderly individuals with household incomes less than 138 percent—including an income disregard—of the Federal Poverty Level (FPL) in 27 states (Kaiser Family Foundation, 2013).[4] In the 24 states that decided not to implement the Medicaid expansions, individuals with household incomes less than 100 percent of the FPL but above any existing state Medicaid income eligibility standard are ineligible for either Medicaid or for premium subsidies through the Marketplace.[5]

Using current income to measure eligibility poses problems because income may have responded to the policy through labor supply changes. For example, someone whose earnings placed them at 140% of FPL prior to 2014 in an expansion state may have retired early because of the possibility of receiving Medicaid and earn far less than 138% FPL in 2014. Similarly, someone with 95% FPL in 2013 in a non-expansion state may now work longer hours and earn more than 100% or 138% FPL, thereby qualifying for Exchange subsidies in that state.

To overcome endogeneity in current income to define the treatment group, we utilize education status as a proxy for Medicaid eligible individuals. This is similar to other work that faces endogeneity problems in identifying treatment and control groups based on income or wage, or does not have access to income data (such as Currie and Gruber (2001) in the case of Medicaid expansions on birth certificate outcomes, or Simon and Kaestner (2004) in the case of minimum wage laws and health insurance). We use the logic that education status is likely to be exogenous to Medicaid policy (at least in the short run) but that those with low education are more affected by Medicaid expansions, relative to those with higher education.

[4] Among the 27 states, Michigan implemented the ACA expansion in April 2014, and New Hampshire implemented the expansion in July 2014.

[5] An exception is Wisconsin, which amended its Medicaid state plan and existing Section 1115 waiver to cover adults up to 100% FPL in Medicaid but did not adopt the expansion (Kaiser Family Foundation, 2014). We include Wisconsin as an expansion state in our analysis.

In Table 2, we show that a larger share of non-elderly adults in the CPS who have at most a high-school diploma have family incomes below 138% of the FPL (between 23 and 20 percent, depending on the year) than do those in higher educational categories (between 10 to 12 percent, depending on the year).[6] We use non-elderly adults (aged 19 to 64 years) with a high school diploma or less as our main study sample. We examine the behavior of our study sample who reside in states that expanded Medicaid (treatment group), compared to our control group of non-elderly individuals with low-education status who reside in states that decided not to implement the expansion. We also use another level of difference to estimate a triple-differences (DDD) specification where we compare the behavior of our study sample (low-educated non-elderly adults), to non-elderly adults with higher educational status—more than high school diploma. We find that this DDD estimate leads to the same conclusion as our DD model. We also find that the conclusion does not change when we compare the behavior of childless low-educated non-elderly adults in expansion states to low-educated non-elderly parents; one might expect that childless adults would display more of the hypothesized effects of job lock release since parents have had access to Medicaid at more generous levels than childless adults, prior to 2014. The fact that our results are qualitatively the same reduce concerns that are results are affected by the particular choice of educational status as the proxy for control and treatment groups.

Before turning to our DD analysis, we provide results of tests of the assumption that prior to the policy change, the control group and treatment group (low-educated individuals in states with and without expansions) followed time trends that were not statistically different in relevant

[6] In unreported results, we confirm that our results using educational categories hold when we use the potentially endogenous (but more accurate as a measure of Medicaid eligibility) current income measure for defining the treatment and control groups. We have also instrumented for Medicaid eligibility that is calculated from current income using a simulated eligibility measure, and find results are not qualitatively different from the ones presented here.

outcomes in years prior to 2014. We find that these conditions are satisfied for the most part; there are some marginally significant effects, but the implied percent effects are extremely small (close to one thousandth of a percent effect).

Our empirical specification for the DD (estimated only among those aged 19-64 with low –education status) takes the form:

$$[1]\ Y_{igst} = \alpha + \beta \text{Post} + \gamma\ MedicaidState_g + \eta(Post_t * MedicaidState_g) + \mathbf{X}_{igst}\boldsymbol{\beta} + \tau_t + \zeta_s + \varepsilon_{igst},$$

where Y_{igst} represents job transition (an indicator variable for the month in question) and logged weekly wages that month for individual *i* in the treatment (or control) group *g*, state *s*, and time *t*. $Post_t$ represents a dummy for the period after the ACA Medicaid Expansion enactment (January 2014). $MedicaidState_g$ represents a dummy for being in a state that expanded, and the interaction of $Post_t$ and $MedicaidState_g$ captures the average effect of the policy. Individual-level control variables, $\mathbf{X}_{igst}$, include demographic characteristics such as age, gender, marital status, educational attainment (indicator for high school diploma vs. high school dropout), and race/ethnicity. We include year-specific and month-specific fixed effects in τ_t. We also include state fixed effects, state monthly unemployment rates, and state linear time trends, ζ_s. Because of the inclusion of year and state fixed effects, the terms *Pos*t and *MedicaidState* on their own are written out only for expositional reasons in Equation [1]—during estimation, they are subsumed in the state and year fixed effects. We cluster standard errors at the level of the state and use a linear probability model as our main specification for its ease of interpretation and computation of marginal effects of interacted variables.

Our main model [1] also includes an indicator for presence of a state-level Medicaid expansion prior to January 2014, which takes a value of 1 if individuals reside in states that have

a Medicaid program for childless adults through 1115 waiver or a market-based Medicaid plan if the program (or plan) is neither closed nor capped. This variable takes a value of 0 once the ACA Medicaid expansion took effect in January 2014. The state expansions are rather heterogeneous and, on average, are much weaker than the ACA expansion; thus, we expect them to have less of an effect and indeed they consistently show no statistically significant effects in our models. The study period ranges from January 2005 to the latest month of available data (currently, August 2014). In sub analysis, we stratify the sample by age and by parental status to observe heterogeneity of the results.

Last, we construct a triple-differences estimate (DDD) and use higher educated groups as a further control group. We estimate Model 2 on the overall CPS Basic monthly sample from 2005-2014 (August 2014).

$$[2]\ Y_{igst} = \alpha + \gamma\, Post_t + \eta(MedicaidState_g * Post_g) + \rho\big(MedicaidState_g * Post_t * LowEdAdult_i\big) + \phi(LowEdAdult_i * Post_t) + \omega\big(MedicaidState_g * LowEdAdult_i\big) + \lambda(LowEdAdult_i) + \mathbf{X}_{igst}\boldsymbol{\beta} + \tau_t + \zeta_s + \varepsilon_{igst}$$

where Y_{igst} represents outcomes for individual *i* (can be childless adult or a parent) in the treatment (or control) group *g*, state *s*, and time *t*. The outcome variable, Y_{igst}, includes job switch, and log of weekly wages. $Post_t$ represents a dummy for the period after the ACA enactment. $MedicaidState_g$ represents a dummy for residing in a state that expanded Medicaid benefits in 2014, $LowEdAdult_i$ represents a dummy variable to indicate if individual *i* has low education, and the triple interaction of $Post_t$, $MedicaidState_g$, and $LowEdAdult_i$ captures the average effect of the policy. Interactions between $Post_t$, and $MedicaidState_g$; $LowEdAdult_i$ and

$MedicaidState_{g;}$ and $LowEdAdult_i$ and $Post_t$ act as controls for unobserved factors. Individual-level control variables, $\mathbf{X}_{igst}$, include demographic characteristics such as age, gender, marital status, educational attainment, and race/ethnicity. It also includes an age variable, its squared term, and monthly state unemployment rate to control for the state of the economy.

In unreported preliminary results, we also investigated the use of an instrumental variables strategy, in which we impute Medicaid eligibility due to the expansion using point-in-time income measures, and instrument for it with a simulated eligibility measure based on the fraction of individuals in their age and race/ethnic group who would be eligible if they lived in a certain state and month (following Currie and Gruber 2001). Because of the binary nature of the 2014 Medicaid expansion (all states either expanding or not, and if expanding, using similar income threshold (138%FPL), we find that this analysis does not provide added insight beyond the identification provided through DD or DDD methods above and have not included it in this version of the paper. Future versions will also use data from the ASES component of the CPS to estimate models restricted to those who held ESI in prior years and to create treatment and control groups based on prior year income.

VI. Results

Validity of Identifying DD Assumption on Less-Educated, Non-elderly Adults

Before proceeding to our main analysis, we first verify that time trends in outcomes prior to the Medicaid expansion are graphically similar between the treatment and control groups as would be appropriate for the DD analysis; we also conduct tests of statistical differences in these trends. This specification controls for all variables included in the main DD regression.

Figure 1a shows that the pre-trends of the outcomes are visually similar between the treatment and control groups (those in states that expanded vs. did not expand) among those with less education (high school diploma or less). The differences in the trends of the probability of job switch between the treatment and control groups are statistically insignificant (Appendix Table 1a). However, an extremely small (less than 3/1000th of a percent of the base value), but statistically significant at the 10 percent significance level difference exists in logged weekly wages where the less-educated, non-elderly population earned a very trivial amount less in weekly wages in Medicaid expanding states prior to expansion relative their counterfactuals in non-Medicaid expanding states. Given the weakness of the statistical evidence and the small magnitude, we interpret this as largely showing that time trends were very similar for the treatment and control groups, prior to the policy.

We also verify that pre-trends for the Medicaid expansion analysis are graphically and statistically similar between the treatment and control groups for both the low-educated childless adult and the parental population separately. Figures 1b and 1c show that the pre-trends of the outcomes are similar between the treatment and control groups (those in states that expanded vs. did not expand) among childless adults and parents with less education (high school diploma or less) respectively. Appendix Table 1b and 1c show that the corresponding differences in trends between the treatment and control groups are very small and statistically insignificant for all outcomes, except that the pre-trends of logged weekly wages are not statistically the same prior to the policy, for the parental population. This estimate is extremely small (less than 6/1000 percent of the base value) although it is statistically significant at the 5% level.

DD Results

The 2014 Medicaid expansions primarily impact childless adults, since low-income parents and pregnant women were able to receive some state-level benefits to qualify for Medicaid benefits prior to the ACA Medicaid expansion implementation. Therefore, we construct DD estimates outlined in Model 1 for the less-educated (high school diploma or less) for the non-elderly population as a whole, and then for non-elderly childless adult and the non-elderly parental populations separately.

First, Table 3a displays the difference-in-differences (DD) estimates for less educated non-elderly adults. The expansion displays no statistically significant impact on the labor market outcomes we study. The (statistically not different from zero) magnitudes of the job switching point estimate is consistent with the likelihood of a job switch in Medicaid expanding states reduced by 0.10 percentage points (4.8 percent) in Medicaid expanding states, along with a 95% confidence interval of an increase of 0.1 to a decrease of 0.3 percentage points. The effects on log weekly wages are also statistically insignificant. These coefficients are consistent with weekly wages reduced by 0.57 percent in Medicaid expanding states; the associated confidence intervals are also large and include 0. Thus, in both cases, the point estimates are wrong-signed relative to our hypothesis as well as being statistically insignificantly different from zero.

Table 3b estimates the corresponding Model 1 for those with higher education. This could be viewed as a placebo test as we expect many fewer in our high-education sample are affected by Medicaid policy relative to those in our low-education sample. We find no statistically significant effects of Medicaid policy in this population, which is reassuring given the lack of findings for the lower educated sample.

In Tables 4a and 5a, we rerun Model 1 on the childless adult population and the parental population separately (among those with low education). Most results are statistically insignificant in these models. As one exception, we find a marginally significant result consistent with that job switching *reducing* significantly (at the 10 percent significance level) for the parental population by 0.4 percentage points (19%). Since our hypothesis is that Medicaid expansion would ease job lock, the negative coefficient for the parental population is an unexpected result. Tables 4b and 5b show the DD estimates for non-elderly childless adults with high education, and non-elderly parents with high education. The results of Tables 4b and 5b—placebo tests corresponding to our estimates on the less educated childless adult and parental population in Table 4a and 5a—show no statistically significant effects. The results in Tables 4b and 5b are also consistent with the results Table 3b which did not separate out the childless adults and the parents.

DDD Results

Before we estimate a triple-differences model, we compare whether the differences in time trends between the low-educated sample and the higher-educated sample changed differently between Medicaid-expansion states and non-expansion states before the Medicaid expansion. The results in Appendix Table 2 shows statistically significant differences in the changes in trends in job switching and in log weekly wages with magnitudes of 5 percent and 1 percent. Even though the differences in changes in trends in job switching and log weekly wages are significant only at the 10% significance level, we emphasize that the pre-trends are not ideal for interpreting the results of the DDD strategy. The baseline DDD estimates in Table 6 show a statistically significant result, but opposite in sign to the expected direction. The difference in job switching between the less educated (high school diploma or less) population relative to the more

educated (more than high school diploma) population reduced significantly at the 5% significance level for states expanding Medicaid in 2014 by 0.3 percentage points. However, the baseline DDD estimates in Table 6 shows that differences in logged weekly wages between the less educated and more educated in states that expanded Medicaid were not statistically significantly different from differences in states that did not expand Medicaid for the non-elderly population. In the additional rows in Table 6, we show separately estimated DDD models among those of different ages. In the estimate for 19-34 yr olds, we find a marginally significant negative effect on wages of 0.4 percent. This is again opposite to our expected findings. We find no statistically significant results for the older population.

VII. Discussion and Conclusion

In this paper, we estimate the early labor market consequences of the 2014 Medicaid expansion using CPS Basic Monthly data, which offer large samples and timely access but are limited because they do not contain health insurance information. Because of the possible endogeneity of actual income, we use educational category as a proxy for Medicaid expansion. However, this results in measurement error that could affect our results. However, we also use alternative estimation strategies and find our conclusions are unchanged. We examine triple-differences (DDD) models with an additional within-state control group of those who have higher education. We also find our results are unchanged when we use potentially endogenous measures of income; in future drafts we plan to instrument for actual eligibility with a simulated policy measure, and to use a one year lagged income measure.

Our results indicate no evidence of strong effects from ACA Medicaid expansions on our hypothesis of released job-lock or our associated hypothesis of higher wages. Most of the outcomes we study show no statistically significant changes in either DD or DDD models. There are some instances of small and marginally statistically significant effects, but in all cases those are of the opposite sign than expected. Our estimates are consistent with evidence from the closest experimental variation in Oregon that show no statistically significant effects on labor market outcomes from Medicaid. However, while the Oregon experiment estimates are related to actual receipt of Medicaid whereas ours is only an "intent to treat" estimate as we do not have health insurance information.

References

Baicker, K., A. Finkelstein, , J. Song, and S. Taubman. 2014. "The Impact of Medicaid on Labor Force Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment". *American Economic Review: Papers and Proceedings*, 104(5): 322-328.

Bailey, J., and A. Chorniy, 2014. "Employer-Provided Health Insurance and Job Mobility: Did the Affordable Care Act Reduce Job Lock?" Working paper, Temple University and Clemson University.

Bansak, C., and S. Raphael. 2008. "The State Children's Health Insurance Program and Job Mobility: Identifying Job Lock among Working Parents in Near-poor Households." *Industrial and Labor Relations Review*, 61(4): 564-579.

Congressional Budget Office (CBO). 2012. "Updated Estimates for the Insurance Coverage Provisions of the Affordable Care Act." Available at http://www.cbo.gov/publication/43076 (last accessed on August 16, 2013).

Currie, J. and J. Gruber. 2001. "Public Health Insurance and Medical Treatment: the Equalizing Impact of Medicaid Expansions." *Journal of Public Economics* 82(1): 63-89.

Dague, L., T. DeLiere, and L. Leininger. 2014. "The Effect of Public Insurance Coverage for Childless Adults on Labor Supply". *NBER Working Paper* No. w20111.

Dave, D., S. Decker, R. Kaestner, and K. Simon. 2013. "The Effect of Medicaid Expansions in the Late 1980s and Early 1990s on the Labor Supply of Pregnant Women," *NBER Working Paper*, No. w19161.

Fairlie, R. W., K. Kapur, and S. Gates. 2013. "Job Lock: Evidence from a Regression Discontinuity Design", *IZA Discussion Paper Series*, No. 7785.

Garrett, B. and R. Kaestner. 2014. "Best Evidence Suggests the Effects of the ACA on Employment Will Be Small". Urban Institute publication, available at http://www.urban.org/publications/413109.html

Garthwaite, C., T. Gross, and M. J. Notowidigdo. 2014. "Public Health Insurance, Labor Supply, and Employment Lock." *The Quarterly Journal of Economics*, *129* (2), 653-696.

Gruber, J., and B.C. Madrian. 2004. "Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature". In C. McLaughlin. *Health Policy and the Uninsured*. Washington, DC: Urban Institute Press. pp. 97–178. Earlier version available at http://www.nber.org/papers/w8817

Hamersma, S., and M. Kim. 2009. "The Effect of Parental Medicaid Expansions on Job Mobility." *Journal of Health Economics* 28(4): 761-770.

Kaiser Family Foundation. 2014. “Status of State Action on the Medicaid Expansion Decision, as of June 10, 2014.” Available at http://kff.org/medicaid/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/ (last accessed on July 24, 2014).

Kaiser Family Foundation. 2013. “A Look at Section 1115 Medicaid Demonstration Waivers Under the ACA: A Focus on Childless Adults.” Available at http://kaiserfamilyfoundation.files.wordpress.com/2013/10/8499-a-look-at-section-1115-medicaid-demonstration-waivers.pdf (last accessed on September 30, 2014).

Monheit, A. C. and P. F. Cooper. 1994. “Health Insurance and Job Mobility: Theory and Evidence,” *Industrial and Labor Relations Review* 48(1): 68-85.

Meyer B. D., and D.T. Rosenbaum. 2001. “Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers.” *Quarterly Journal of Economics*, 116(3): 1063-1114.

Moffitt R., and B. Wolfe. 1992. “The Effect of the Medicaid Program on Welfare Participation and Labor Supply.” *The Review of Economics and Statistics* 74(4): 615-626.

Simon, K. and R. Kaestner. 2004. “Do Minimum-Wage Hikes Reduce Non-wage Job Attributes? Evidence on Fringe Benefits and Working Conditions”. *Industrial Labor Relations Review*, 58(1): 52-70.

Summary Tables, Results Tables, and Figures:

Table 1. Descriptive Statistics—Current Population Survey, Non-Elderly Adults Aged 19-64

	Mean	Std. Dev.	N
Demographic Characteristics			
Age	41.6	12.88	8,559,940
Indicator: male	0.483	0.499	8,559,940
Indicator: white	0.708	0.455	8,559,940
Indicator: African-American	0.094	0.292	8,559,940
Indicator: Hispanic	0.124	0.3297	8,559,940
Outcome variables			
Indicator: switched a job this month	0.021	0.144	4,206,703
Weekly wages	825.472	601.384	1,425,723
Education Level			
Indicator: High School Drop Out	0.101	0.302	8,559,940
Indicator: High School Diploma	0.295	0.456	8,559,940
Indicator: High School Diploma or Less Education	0.396	0.489	8,559,940
Indicator: More than a High School Level of Education	0.604	0.489	8,559,940
Income			
Yearly Family Income	40,000 TO 49,999	3.9412	8,559,940

Note: Sample estimates from the CPS Basic Monthly data from January 2005 to August 2014.

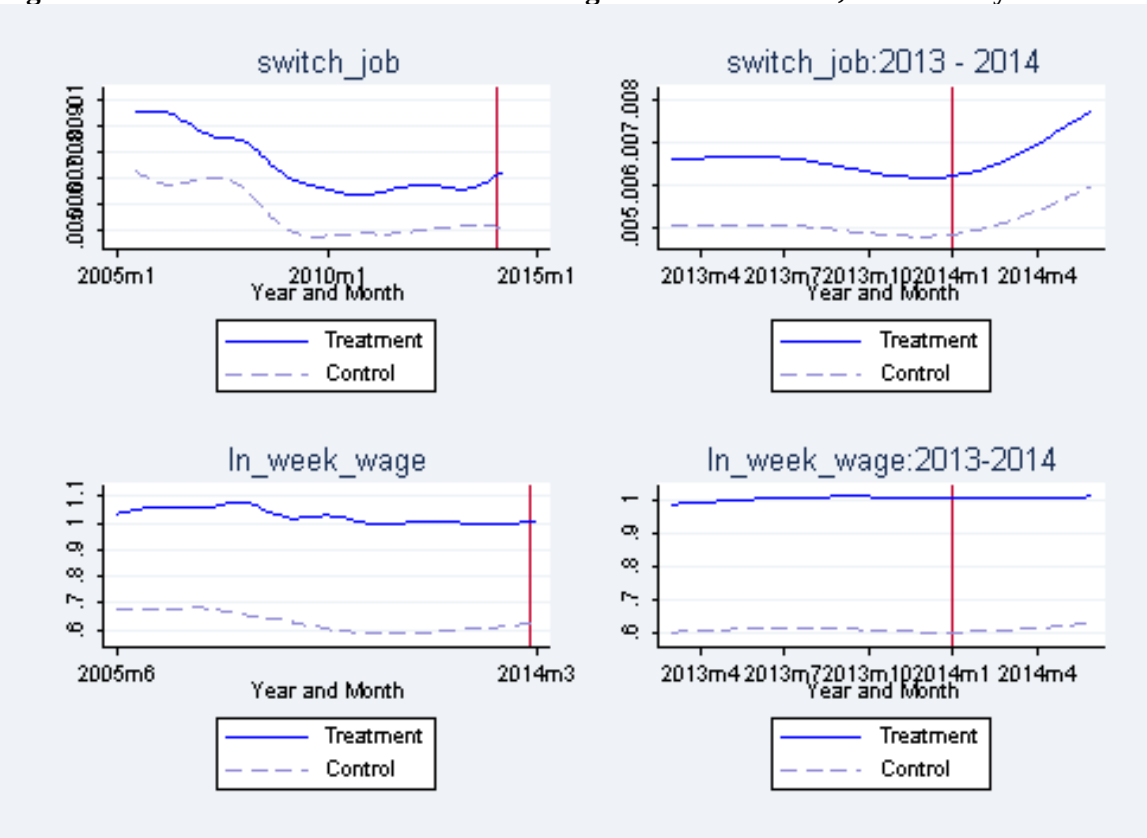
Table 2: Correspondence Between Educational Level and Income, Non-Elderly Adults in CPS Basic Monthly, January 2005-August 2014

	High School Drop outs	High School Diploma	High School Diploma or Less	More than High School	High School Drop outs	High School Diploma	High School Diploma or Less	More than High School
Year	Income: At or Below 100%FPL				Income: Below 138%FPL			
2005	0.221	0.089	0.124	0.061	0.377	0.173	0.227	0.115
2006	0.207	0.086	0.118	0.057	0.357	0.162	0.215	0.107
2007	0.197	0.084	0.113	0.055	0.343	0.156	0.205	0.102
2008	0.190	0.080	0.108	0.053	0.335	0.151	0.198	0.098
2009	0.204	0.086	0.117	0.058	0.355	0.163	0.213	0.107
2010	0.215	0.096	0.126	0.066	0.365	0.179	0.226	0.121
2011	0.212	0.102	0.130	0.068	0.366	0.185	0.230	0.123
2012	0.213	0.098	0.127	0.065	0.362	0.180	0.225	0.119
2013	0.208	0.097	0.124	0.063	0.355	0.176	0.219	0.115
2014	0.199	0.093	0.118	0.062	0.350	0.173	0.216	0.113

(1): The values of the cells show the fraction of those in a certain educational group whose income is below 100% or 138% of FPL. The CPS Monthly data do not contain household size, thus we make the following simplifying assumption for our calculations when taking reported monthly income and translating it into fraction of FPL for the relevant year.

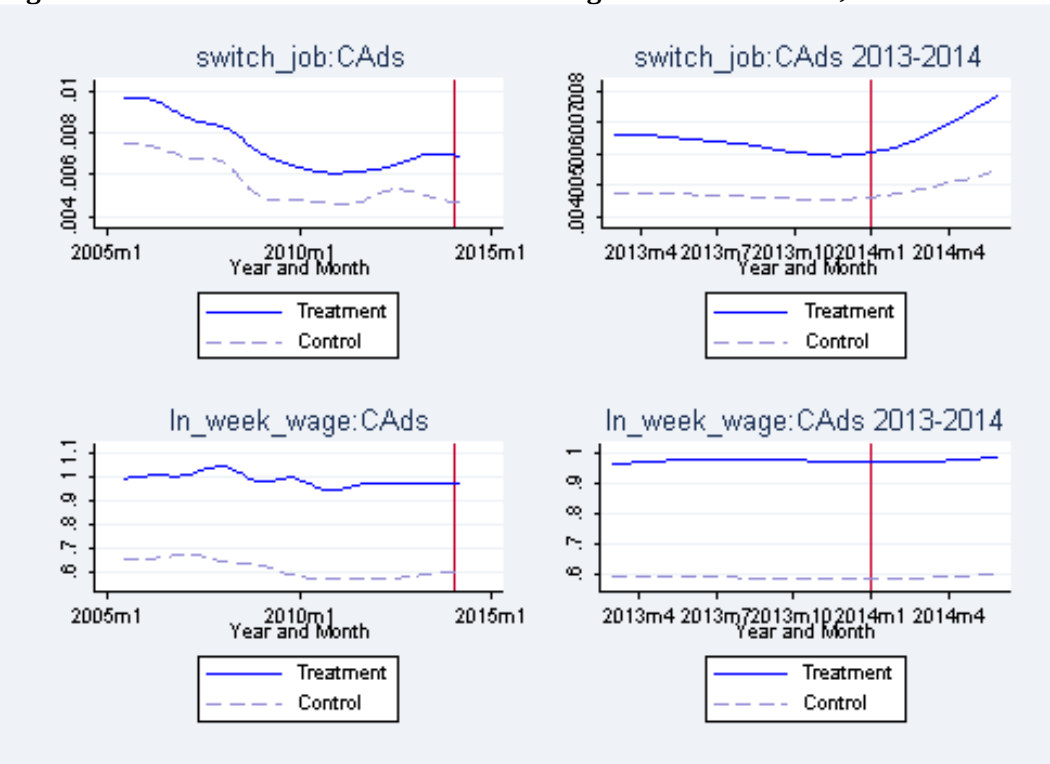
Childless adults are assumed to have household size of 1 while married parents are assumed to have household size of 3.
(2) The third column, High School Diploma or Less, is a combination of the first two columns, in each of the two panels of this table.

Figure 1a: Trends in Outcome Variables among the Less-Educated, Non-Elderly Adults



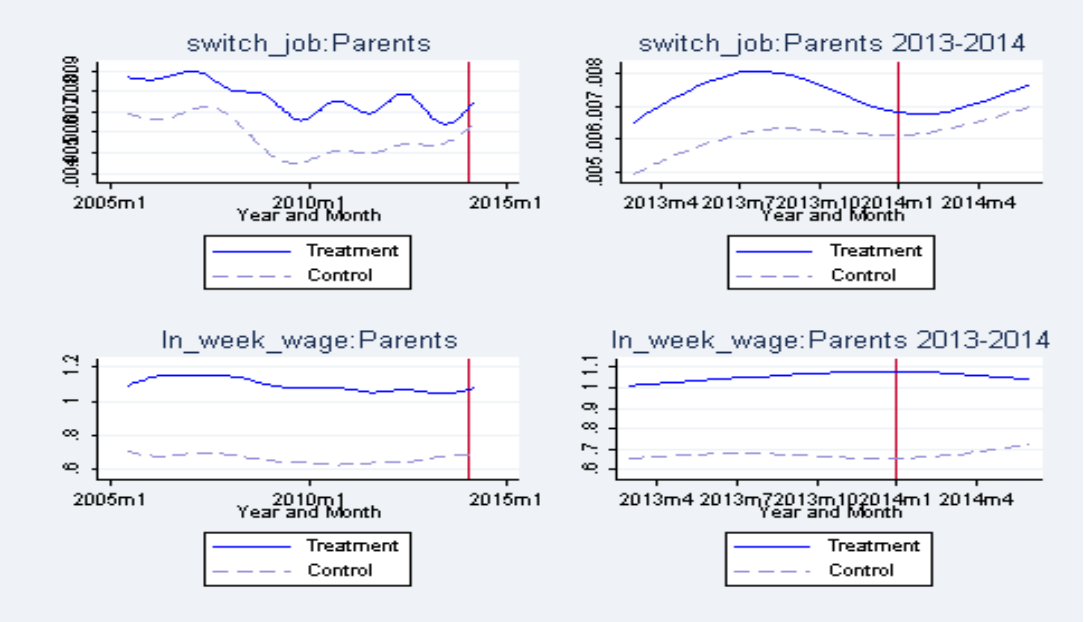
Note:
(1) Solid line represents treatment states (states expanding Medicaid coverage in 2014)
(2) Dashed line represents control states (states not expanding Medicaid coverage in 2014)
(3) Vertical solid line – Implementation of 2014 ACA Medicaid expansion in January 2014
(4) Less Educated is defined as High School Diploma or Less
(5) Graphs on the left column shows trends in outcome variables switch_job and ln_week_wage from 2005 till August 2014
(6) Graphs on the right column show trends in outcome variables switch_job and ln_week_wage from 2013-2014. It specifically looks at the trends observed in CPS Basic Monthly data between the year before implementation (January 2013) up till the latest month of CPS Basic data (August 2014).

Figure 1b: Trends in Outcome Variables among the Less-Educated, Childless Adults



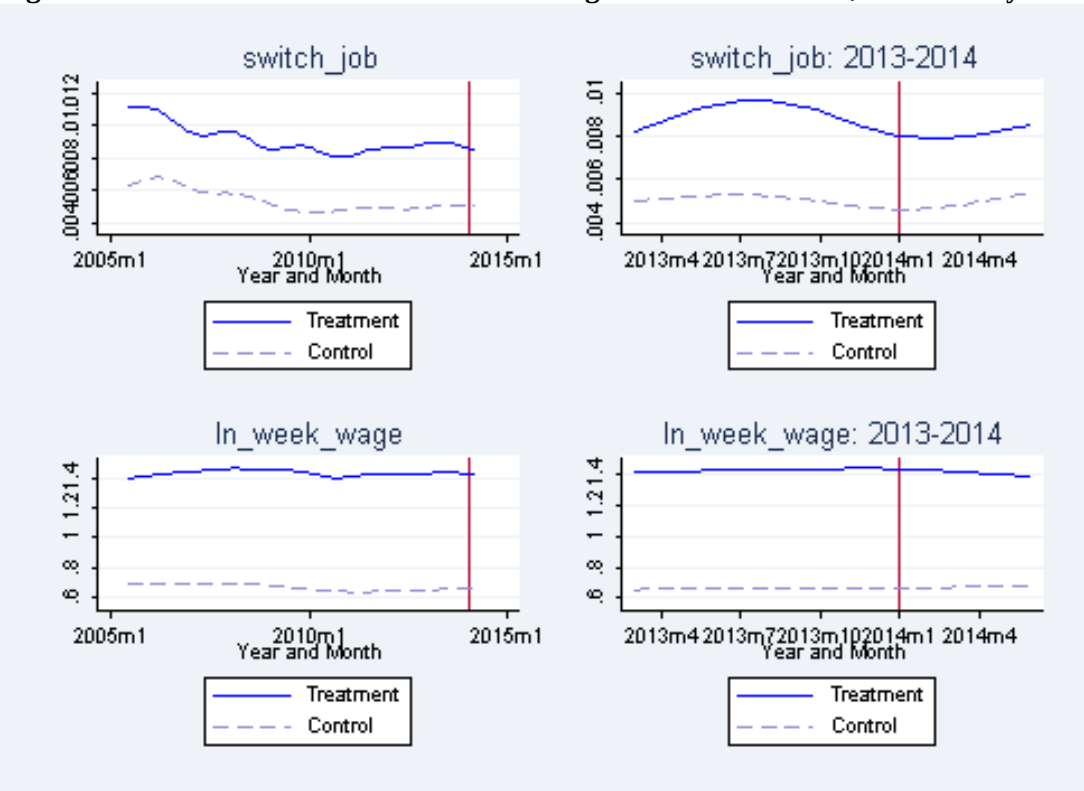
Note:
see Notes to Figure 1a

Figure 1c: Trends in Outcome Variables among the Less-Educated, Parents



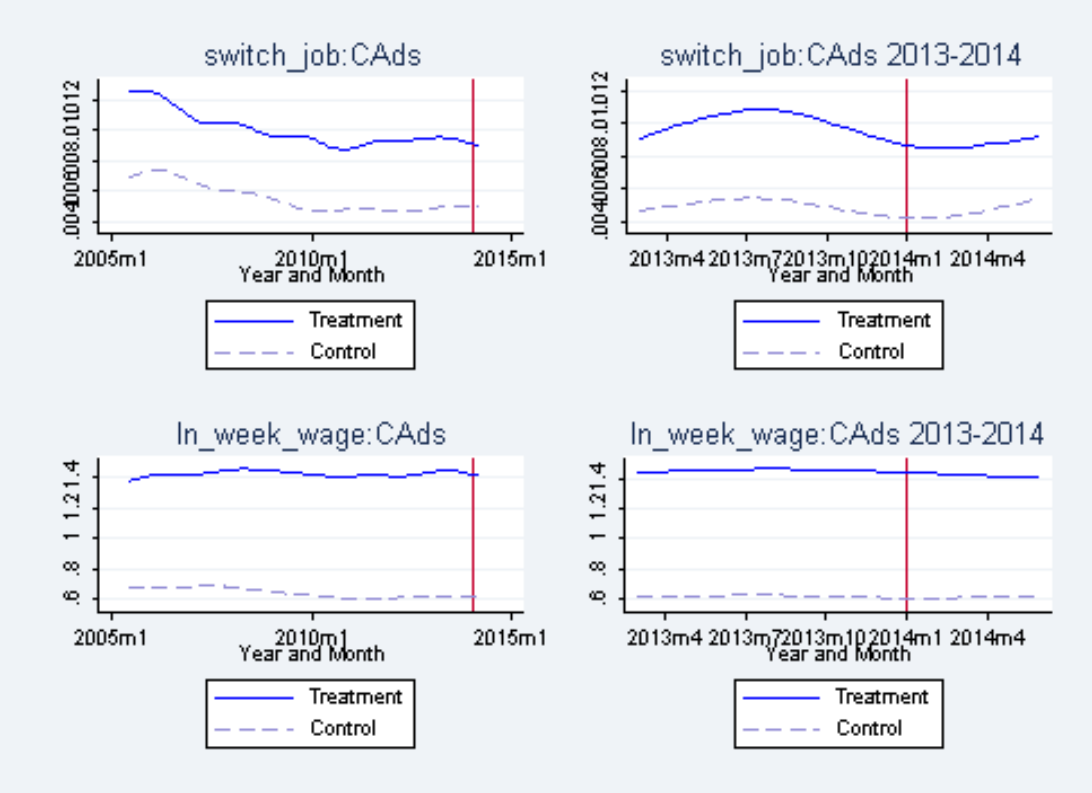
Note:
See Notes to Figure 1a

Figure 2a: Trends in Outcome Variables among the More-Educated, Non-Elderly Adults



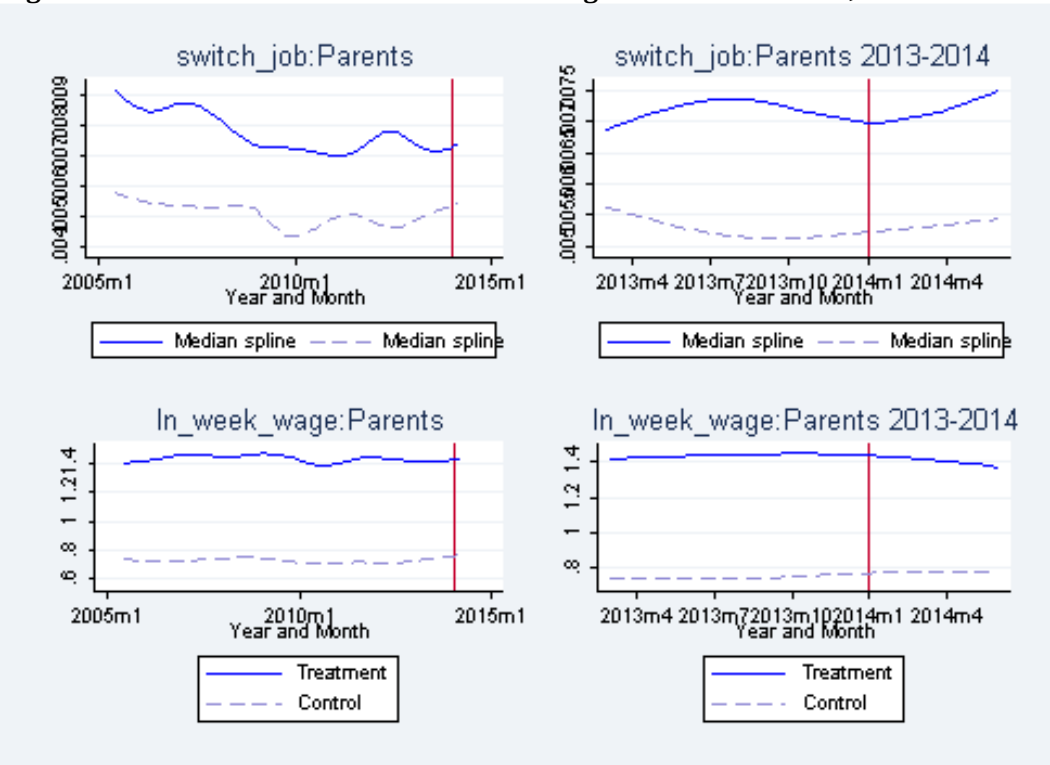
Note:
(1) Solid line represents treatment states (states expanding Medicaid coverage in 2014)
(2) Dashed line represents control states (states not expanding Medicaid coverage in 2014)
(3) Vertical solid line – Implementation of 2014 ACA Medicaid expansion in January 2014
(4) More Educated is defined as more than High School Diploma, Less Educated defined as High School Diploma or less
(5) Graphs on the left column shows trends in outcome variables switch_job and ln_week_wage from 2005 till August 2014
(6) Graphs on the right column shows trends in outcome variables switch_job and ln_week_wage from 2013-2014. It specifically looks at the trends observed in CPS Basic Monthly data between the year before implementation (January 2013) up till the latest month of CPS Basic data (August 2014).

Figure 2b: Trends in Outcome Variables among the More-Educated, Childless Adults



Note:
See Notes to Figure 2a

Figure 2c: Trends in Outcome Variables among the More-Educated, Parents



Note:
See Notes to Figure 2a

Table 3a. Effect of ACA Medicaid Expansion among Non-elderly Adults (HS Diploma or Less Education)

	Indicator: job switch	Indicator: Log of weekly wages
ACA Medicaid Expansion Effect	-0.0010	-0.0057
(2014 January-August)	(0.0001)	(0.0100)
ACA Medicaid Expansion Effect (19-34 year olds) (6)	-0.001	-0.008
(2014 January-August)	(0.0024)	(0.0174)
ACA Medicaid Expansion Effect (58-64 year olds) (7)	0.001	-0.004
(2014 January-August)	(0.0022)	(0.0229)
Baseline Average in Medicaid Expanding State	0.021	6.201
Baseline Average in Non-Medicaid Expanding State	0.023	6.169
Post 2014 Average in Medicaid Expanding State	0.022	6.269
Post 2014 Average in Non-Medicaid Expanding State	0.024	6.256
Number of Observations	1,455,563	506,842

Notes: (1) Cells of the table contain coefficients and standard errors in parentheses. Coefficients in the first row are from the interaction of a dummy variable for treatment group (less-educated, non-elderly individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (2) Data: CPS Basic Monthly data from January 2005 to August 2014. The population is 19-64 years old with education level of a high school diploma or less. (3) Dependent variables—column 1: indicator variable that equals 1 if individual switches jobs during the current month and 0 otherwise; and column 2: log of weekly wages. (4) Other regressors are age, gender, race/ethnicity, marital status, monthly state unemployment rate, year-specific fixed effects, month-specific fixed effects, state fixed effects, and state-specific linear trends. (5) Means of dependent variables are obtained for treatment and control groups before and after the ACA Medicaid expansion. (6) Coefficients in the third row are from the interaction of a dummy variable for treatment group (less-educated 19-34 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014) in the regressions that use a sub-population of less-educated 19-34 year old individuals. (7) Coefficients in the third row are from the interaction of a dummy variable for treatment group (less-educated 58-64 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014) in the regressions that use a sub-population of less-educated 58-64 year old individuals.

Table 3b. Effect of ACA Medicaid Expansion among Non-elderly Adults (More than HS Diploma)

	Indicator: job switch	Indicator: Log of weekly wages
ACA Medicaid Expansion Effect	-0.000265	0.006
(2014 January-August)	(0.0009)	(0.0090)
ACA Medicaid Expansion Effect (19-34 year olds) (6)	-0.001	0.002
(2014 January-August)	(0.0019)	(0.0128)
ACA Medicaid Expansion Effect (58-64 year olds) (7)	-0.001	-0.006
(2014 January-August)	(0.0021)	(0.0310)
Baseline Average in Medicaid Expanding State	0.021	6.614
Baseline Average in Non-Medicaid Expanding State	0.021	6.526
Post 2014 Average in Medicaid Expanding State	0.020	6.682
Post 2014 Average in Non-Medicaid Expanding State	0.020	6.607
Number of Observations	2,751,140	918,881

Notes: (1) Cells of the table contain coefficients and standard errors in parentheses. Coefficients in the first row are from the interaction of a dummy variable for treatment group (more-educated, non-elderly individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (2) Data: CPS Basic Monthly data from January 2005 to August 2014. The population is 19-64 years old with education level of more than a high-school diploma. (3) Dependent variables—column 1: indicator variable that equals 1 if individual switches jobs during the current month and 0 otherwise; and column 2: log of weekly wages. (4) Other regressors are age, gender, race/ethnicity, marital status, monthly state unemployment rate, year-specific fixed effects, month-specific fixed effects, state fixed effects, and state-specific linear trends. (5) Means of dependent variables are obtained for treatment and control groups before and after the ACA Medicaid expansion. (6) Coefficients in the first row are from the interaction of a dummy variable for treatment group (less-educated, 19-34 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (7) Coefficients in the first row are from the interaction of a dummy variable for treatment group (less-educated, 58-64 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014).

Table 4a. Effect of ACA Medicaid Expansion among Non-elderly Childless Adults (HS Diploma or Less Education)

	Indicator: job switch	Indicator: Log of weekly wages
ACA Medicaid Expansion Effect	0.001	-0.001
(2014 January-August)	(0.0015547)	(0.0097848)
Baseline Average in Medicaid Expanding State	0.020	6.231
Baseline Average in Non-Medicaid Expanding State	0.022	6.192
Post 2014 Average in Medicaid Expanding State	0.021	6.282
Post 2014 Average in Non-Medicaid Expanding State	0.026	6.281
Number of Observations	913,761	321,956

Notes: (1) Cells of the table contain coefficients and standard errors in parentheses. Coefficients in the first row are from the interaction of a dummy variable for treatment group (states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-May, 2014). (2) Data: CPS Basic Monthly data from January 2005 to August 2014. The population is non-elderly childless adults (19-64 years old) with education level of a high school diploma or less. (3) Dependent variables—column 1: indicator variable that equals 1 if individual switches jobs during the current month and 0 otherwise; and column 2: log of weekly wages. (4) Other regressors are age, gender, race/ethnicity, marital status, monthly state unemployment rate, year-specific fixed effects, month-specific fixed effects, state fixed effects, and state-specific linear trends. (5) Means of dependent variables are obtained for treatment and control groups before and after the ACA Medicaid expansion.

Table 4b. Effect of ACA Medicaid Expansion among Non-elderly Childless Adults (More than HS Diploma)

	Indicator: job switch	Indicator: Log of weekly wages
ACA Medicaid Expansion Effect	-0.00036	0.001
(2014 January-August)	(0.0012)	(0.0106)
Baseline Average in Medicaid Expanding State	0.017	6.724
Baseline Average in Non-Medicaid Expanding State	0.019	6.616
Post 2014 Average in Medicaid Expanding State	0.018	6.813
Post 2014 Average in Non-Medicaid Expanding State	0.019	6.705
Number of Observations	1,686,368	574,223

Notes: (1) Cells of the table contain coefficients and standard errors in parentheses. Coefficients in the first row are from the interaction of a dummy variable for treatment group (states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-May, 2014). (2) Data: CPS Basic Monthly data from January 2005 to August 2014. The population is non-elderly childless adults (19-64 years old) with education level of more than a high school diploma. (3) Dependent variables—column 1: indicator variable that equals 1 if individual switches jobs during the current month and 0 otherwise; and column 2: log of weekly wages. (4) Other regressors are age, gender, race/ethnicity, marital status, monthly state unemployment rate, year-specific fixed effects, month-specific fixed effects, state fixed effects, and state-specific linear trends. (5) Means of dependent variables are obtained for treatment and control groups before and after the ACA Medicaid expansion.

Table 5a. Effect of ACA Medicaid Expansion among Non-elderly Parents (HS Diploma or Less Education)

	Indicator: job switch		Indicator: Log of weekly wages
ACA Medicaid Expansion Effect	-0.004	*	-0.013
(2014 January-August)	(0.0024)		(0.0161)
Baseline Average in Medicaid Expanding State	0.022		6.184
Baseline Average in Non-Medicaid Expanding State	0.023		6.157
Post 2014 Average in Medicaid Expanding State	0.022		6.262
Post 2014 Average in Non-Medicaid Expanding State	0.024		6.242
Number of Observations	541,802		184,886

Notes: (1) Cells of the table contain coefficients and standard errors in parentheses. Coefficients in the first row are from the interaction of a dummy variable for treatment group (states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (2) Data: CPS Basic Monthly data from January 2005 to August 2014. The population is non-elderly parents (19-64 years old) with education level of a high school diploma or less. (3) Dependent variables—column 1: indicator variable that equals 1 if individual switches jobs during the current month and 0 otherwise; and column 2: log of weekly wages. (4) Other regressors are age, gender, race/ethnicity, marital status, monthly state unemployment rate, year-specific fixed effects, month-specific fixed effects, state fixed effects, and state-specific linear trends. (5) Means of dependent variables are obtained for treatment and control groups before and after the ACA Medicaid expansion.

Table 5b. Effect of ACA Medicaid Expansion among Non-elderly Parents (More than HS Diploma)

	Indicator: job switch	Indicator: Log of weekly wages
ACA Medicaid Expansion Effect	-0.0000517	0.012
(2014 January-August)	(0.0014466)	(0.0113537)
Baseline Average in Medicaid Expanding State	0.023	6.550
Baseline Average in Non-Medicaid Expanding State	0.023	6.470
Post 2014 Average in Medicaid Expanding State	0.022	6.610
Post 2014 Average in Non-Medicaid Expanding State	0.022	6.547
Number of Observations	1,064,772	344,658

Notes: (1) Cells of the table contain coefficients and standard errors in parentheses. Coefficients in the first row are from the interaction of a dummy variable for treatment group (more-educated, non-elderly individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (2) Data: CPS Basic Monthly data from January 2005 to August 2014. The population is non-elderly parents (19-64 years old) with education level of more than a high school diploma. (3) Dependent variables—column 1: indicator variable that equals 1 if individual is switching jobs during the current month and 0 otherwise; and column 2: log of weekly wages. (4) Other regressors are age, gender, race/ethnicity, marital status, monthly state unemployment rate, year-specific fixed effects, month-specific fixed effects, state fixed effects, and state-specific linear trends. (5) Means of dependent variables are obtained for treatment and control groups before and after the ACA Medicaid expansion. (6) Coefficients in the third row are from the interaction of a dummy variable for treatment group (more-educated 19-34 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (7) Coefficients in the third row are from the interaction of a dummy variable for treatment group (more-educated 58-64 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014)

Table 6. DDD- Effect of ACA Medicaid Expansion On the Difference Between Non-elderly Less Educated (HS Diploma or HS Drop Out) Adults Relative to Non-Elderly, More Educated (More than High School Diploma) Adults in Medicaid Expanding States Relative to Non-Medicaid Expanding States

	Indicator: job switch		Indicator: Log of weekly wages	
ACA Medicaid Expansion Effect	-0.003	**	-0.029	
(2014 January-August)	(0.0011)		(0.0188)	
ACA Medicaid Expansion Effect (19-34 year olds) (6)	-0.003		-0.041	*
(2014 January-August)	(0.0022)		(0.0228)	
ACA Medicaid Expansion Effect (58-64 year olds) (7)	-0.002		(0.017)	
(2014 January-August)	(0.0024)		(0.0358)	
Baseline Mean Difference (childless adults-parents): Medicaid Expanding States	0.000369		-0.413	
Baseline Mean Difference (childless adults-parents): Non-Medicaid Expanding States	0.0016779		-0.357	
Post Policy Mean Difference (childless adults-parents): Medicaid Expanding States	0.001389		-0.413	
Post Policy Mean Difference (childless adults-parents): Non-Medicaid Expanding States	0.0038985		-0.352	
Number of Observations	4,206,703		1,425,723	

Notes: (1) Cells of the table contain coefficients and standard errors in parentheses. Coefficients in the first row are from the interaction of a dummy variable for treatment group (less-educated non-elderly individual in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (2) Data: CPS Basic Monthly data from January 2005 to August 2014. The population is non-elderly adults (19-64 years old) with education level of more than a high school diploma. (3) Dependent variables—column 1: indicator variable that equals 1 if individual switches jobs during the current month and 0 otherwise; and column 2: log of weekly wages. (4) Other regressors are age, gender, race/ethnicity, marital status, monthly state unemployment rate, year-specific fixed effects, month-specific fixed effects, state fixed effects, and state-specific linear trends. (5) Means of dependent variables are obtained for treatment and control groups before and after the ACA Medicaid expansion. (6) Coefficients in the 3rd row are from the interaction of a dummy variable for treatment group (less-educated, 19-34 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014). (7) Coefficients in the 3rd row are from the interaction of a dummy variable for treatment group (less-educated 58-64 year old individuals in states that participated in the 2014 ACA Medicaid expansion) and a dummy variable for the period after the Medicaid expansion (January-August, 2014).

Appendix:

Appendix Table 1a: Test for Equality of Pre-Reform Trends among Non-elderly Adults (with HS Diploma or Less) Between Medicaid Expanding States and Non Medicaid Expanding States

	Indicator: job switch	Indicator: Log of weekly wages	
Interaction of time trend and a dummy variable for treatment group	-0.000002	-0.0002	*
	(0.00001)	(0.0001)	
Baseline Average in Medicaid Expanding State	0.657	6.201	
Baseline Average in Non-Medicaid Expanding State	0.662	6.169	
Number of observations	1,360,979	472,206	

Notes: (1) Data: CPS Basic Monthly data from January 2005 to December 2013, which is prior to the implementation of the ACA Medicaid expansion in Jan 2014. (2) Cells of the table contain: coefficients, and standard errors in parentheses. Coefficients are from the interaction of a dummy variable for treatment group and a linear measure for time trend (number of months since January 2005), which shows whether there was a different time trend for the control vs. the treatment group in the period prior to policy enactment. (3) Other regressors are a linear time trend, a dummy variable for the treatment group, and all other explanatory variables included in our main specification.

Appendix Table 1b: Test for Equality of Pre-Reform Trends among Non-elderly Childless Adults (with HS Diploma or Less) Between Medicaid Expanding States and Non Medicaid Expanding State

	Indicator: job switch	Indicator: Log of weekly wages
Interaction of time trend and a dummy variable for treatment group	-0.000006	-0.0001
	(0.00001)	(0.0001)
Baseline Average in Medicaid Expanding State	0.020	6.231
Baseline Average in Non-Medicaid Expanding State	0.022	6.192
Number of observations	852,921	299,375

Notes: (1) Data: CPS Basic Monthly data from January 2005 to December 2013, which is prior to the implementation of the ACA Medicaid expansion in Jan 2014. (2) Cells of the table contain: coefficients, and standard errors in parentheses. Coefficients are from the interaction of a dummy variable for treatment group and a linear measure for time trend (number of months since January 2005), which shows whether there was a different time trend for the control vs. the treatment group in the period prior to policy enactment. (3) Other regressors are a linear time trend, a dummy variable for the treatment group, and all other explanatory variables included in our main specification.

Appendix Table 1c: Test for Equality of Pre-Reform Trends among Non-elderly Parents (with HS Diploma or Less) Between Medicaid Expanding States and Non Medicaid Expanding State

	Indicator: job switch	Log of weekly wages	
Interaction of time trend and a dummy variable for treatment group	0.000004	-0.0004	**
	(0.00002)	(0.0002)	
Baseline Average in Medicaid Expanding State	0.022	6.184	
Baseline Average in Non-Medicaid Expanding State	0.023	6.157	
Number of observations	508,058	172,831	

Notes: (1) Data: CPS Basic Monthly data from January 2005 to December 2013, which is prior to the implementation of the ACA Medicaid expansion in Jan 2014. (2) Cells of the table contain: coefficients, and standard errors in parentheses. Coefficients are from the interaction of a dummy variable for treatment group and a linear measure for time trend (number of months since January 2005), which shows whether there was a different time trend for the control vs. the treatment group in the period prior to policy enactment. (3) Other regressors are a linear time trend, a dummy variable for the treatment group, and all other explanatory variables included in our main specification.

Appendix Table 2: Test for Equality of Pre-Reform Trends among Less Educated (High School Diploma or Less), Non-elderly Adults Relative to More Educated (More than High School Diploma) Non-elderly Adults, in States with and Without Medicaid Expansions

	Indicator: job switch		Indicator: Log of weekly wages	
Interaction of time trend and a dummy variable for treatment group	-0.00002	*	-0.0004	*
	(0.00001)		(0.0002)	
Baseline Mean Difference (less educated – more educated): Medicaid Expanding States	0.000369		-0.413	
Baseline Mean Difference (less educated – more educated): Non-Medicaid Expanding States	0.0016779		-0.357	
Number of observations	3,916,172		1,322,906	

Notes: (1) Data: CPS Basic Monthly data from January 2005 to December 2013, which is prior to the implementation of the ACA Medicaid expansion in Jan 2014. (2) Cells of the table contain: coefficients, and standard errors in parentheses. Coefficients are from the interaction of a dummy variable for treatment group and a linear measure for time trend (number of months since January 2005), which shows whether there was a different time trend for the control vs. the treatment group in the period prior to policy enactment. (3) Other regressors are a linear time trend, a dummy variable for the treatment group, and all other explanatory variables included in our main specification

The Effects of Eliminating the Work Search Requirement on Job Match Quality and Other Long-Term Employment Outcomes

Marta Lachowska,* Merve Meral, and Stephen A. Woodbury

This paper was prepared with funding from the U.S. Department of Labor. The views expressed are those of the authors and should not be attributed to the Federal Government or the Department of Labor.

Abstract

We exploit data from the 1986–87 Washington Alternative Work Search experiment (merged with nine years of follow-up administrative wage records) to estimate the causal effects of eliminating the unemployment insurance (UI) work search requirement (WSR) on duration of nonemployment, tenure with first post-claim employer, number of post-claim employers, long-term earnings, employment, and hours worked. For UI claimants as a whole, we find that eliminating the WSR had little influence, either positive or negative, on long-term post-claim outcomes. In contrast, for permanent job losers, we find strong evidence that eliminating the WSR had a negative effect on employment outcomes, resulting in a longer time to reemployment, lower earnings, and a shorter duration of tenure with first post-claim employer. For claimants who were not permanent job losers, eliminating the WSR resulted in more UI benefit payments and longer unemployment durations, but made no difference for their employment outcomes. We conclude that, in addition to reducing moral hazard associated with UI, the WSR is an important policy for improving the long-term employment outcomes of permanent job losers.

JEL classification: C21, C93, I38, J18, J24, J38, J64, J65, J68

Keywords: Unemployment insurance, random-assignment experiment, work search requirement, reemployment policy, long-term evaluation of public policy, administrative data

*Lachowska: W.E. Upjohn Institute for Employment Research, 300 South Westnedge Avenue, Kalamazoo, MI 49007 (marta@upjohn.org).

Meral: Department of Economics, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747-2300 (mcebi@umassd.edu).

Woodbury: Department of Economics, Michigan State University, 486 West Circle Drive, East Lansing, MI 48824 (woodbur2@msu.edu) and W.E. Upjohn Institute for Employment Research.

We are grateful to Randall Eberts, Wayne Gordon, Peter Mueser, Peter Orazem, Suzanne Simonetta, Jeffrey Smith, and Demetra Nightingale for helpful comments on an earlier version of this paper. We thank Rod Anderson for research assistance. Lachowska gratefully acknowledges support from the Department of Labor Scholars Grant.

List of Abbreviations

ER—Exception Reporting

JSA—Job Search Assistance

NWS—New Work Search

UI—Unemployment Insurance

WAWS—Washington Alternative Work Search experiment

WSR—Work Search Requirement

1 Introduction

The work search requirement (WSR) for unemployment insurance (UI) recipients has been a central part of UI in the United States since the system began in the 1930s. Typically, to be eligible for UI benefits, a claimant initially needs an adequate work history and must have lost her job through lack of work and no fault of her own. In addition, to remain eligible, the worker must be "able, available, and searching" for work—that is, must satisfy the *work search requirement*, or WSR.

Although the WSR aims to reduce the moral hazard associated with UI—that is, to counter the incentive to reduce job search effort and take longer to become reemployed—it may also pressure workers into accepting a relatively poor job match, leading to an unstable pattern of employment and lower long-term earnings.[1] Hence, eliminating the WSR could allow the claimants to search for a better job match and lead to improved employment outcomes—the *improved job match hypothesis*. Alternatively, eliminating the requirement could prolong duration of unemployment, making the claimant less attractive to employers and hence worsen employment outcomes—the *negative duration dependence hypothesis*.[2] Finally, eliminating the work search requirement could impose greater costs to the UI system, without any effect on employment outcomes—the *moral hazard hypothesis*.

Understanding the effects of the WSR on employment outcomes is of ongoing importance because in recent years most states have relaxed enforcement of the requirement by shifting toward taking claims over the phone or on-line (see O'Leary [2006] and Ebenstein and

[1] A UI claimant does not need to accept the first available job offer, but he or she is required to accept a job offer that satisfies the "suitable work" condition. In practice, claimants do not need to accept work that is not in line with their training and experience. The work search requirement could nevertheless pressure a claimant to accept a less attractive job offer that meets the suitable work condition instead of holding out for a better offer.
[2] See, e.g., Notowidigdo, Kroft, and Lange (2013) for recent evidence of scarring effects of long spells of unemployment.

Stange [2010]). Because telephone and on-line claiming in effect reduces the frequency of in-person contact between a claimant and the state workforce agency, it is important to know whether a more "hands-off" approach to the WSR has any beneficial effect on post-unemployment job match quality.

The aim of this paper is to examine the effects of eliminating the WSR on post-unemployment job match quality, proxied by employment tenure, and other long-term employment outcomes, such as duration of nonemployment, the number of post-claim employers, earnings, hours worked, and employment of UI claimants. To do this, we add nine years of quarterly follow-up wage records to the original data from the Washington Alternative Work Search (WAWS) experiment (Johnson and Klepinger 1991, 1994). In the WAWS experiment, all eligible UI claimants at the Tacoma Employment Service Center between July 1986 and August 1987 were randomly assigned to a control group, which imposed a standard WSR, or to an exception reporting (ER) treatment group, which effectively eliminated the WSR.

Claimants in the control group were told to contact at least three employers per week and be prepared to give evidence that they had done so in an *eligibility review interview*, usually conducted 13–15 weeks after the initial claim. Claimants in the ER treatment group were told (at the time of their initial claim) to actively seek work, but were also told that they would not be called in for an eligibility review interview, and that weekly UI benefits would be mailed unless they called the Tacoma Employment Service Center to report that they had stopped looking for work or had taken a job. As such, ER amounted to an "honor system" with no WSR (Johnson and Klepinger 1991, pp. 3–9).

When studying the short-term effects of ER, Johnson and Klepinger (1991, 1994) find that eliminating the WSR substantially increased benefits received, the duration of benefit

receipt, and the probability of exhausting benefits, but without affecting earnings or hours worked during the claim quarter or the benefit year. This combination of increased benefit receipt without any changes in earnings or hours suggests that ER led to increased abuse of the UI system. At the same time, however, ER also increased the probability that a worker returned to a former employer. Although this increased likelihood of return to a past employer suggests that relaxing the WSR may have been beneficial to at least some of the claimants (in that they reestablished a previous job match), Johnson and Klepinger find no evidence of improved short-term post-unemployment outcomes. On balance, then, Johnson and Klepinger's findings suggest that eliminating the WSR led to increased abuse of the UI system by claimants but did not lead to better employment outcomes.

Other studies of the WSR arrive at quite different conclusions from the WAWS experiment. For example, the evaluation of the 1994 Maryland UI Work-Search Demonstration (Klepinger, Johnson, and Joesch 2002) concluded that although a *relaxed* enforcement of WSR prolonged the duration of UI receipt, it also increased the probability of subsequent employment and led to higher earnings in the quarters following the experiment.[3] Poe-Yamagata et al. (2011) find that an increased emphasis on WSR under the 2005 Reemployment and Eligibility Assessment initiative decreased the duration of UI receipt and had a positive impact on reemployment probability in the short-run. Finally, Ashenfelter, Ashmore, and Deschênes (2005) find that reducing the enforcement of the WSR did not lead to increased abuse of the UI system by the claimants. Hence, the issue of whether a relaxed WSR leads to more abuse or has the

[3] The treatment resembling the WAWS' ER treatment in the Maryland experiment only relaxed some aspects of WSR. This treatment did not include automatic payments to the claimants. Instead, the claimants needed to inform the UI office on a weekly basis that they had not found work and were actively searching. This treatment group was, however, not required to report their employer contacts. In effect, the Maryland treatment relaxed some features of the WSR, but did not eliminate it all together.

positive effect of helping claimants obtain more stable and better paying post-unemployment jobs remains a matter of debate.

Studying the long-term effects of eliminating the WSR is related to the more general issue of how design of UI—e.g., the generosity and duration of benefits—affects earnings and employment. Thanks to the availability of high-quality microdata, this literature has expanded in the recent decades. Addison and Blackburn (2000) and Tatsiramos and van Ours (2014) review the literature on the relationship between UI benefit generosity and post-unemployment earnings. Both literature surveys conclude that the evidence has been mixed. For example, Ehrenberg and Oaxaca (1976), Burgess and Kingston (1976), McCall and Chi (2008), Caliendo, Tatsiramos, and Uhlendorff (2012), and Nekoei and Weber (2013) find a positive association between a more generous UI system and reemployment earnings, whereas Addison and Portugal (1989), Gregory and Jukes (2001), and Schmider, von Wachter, and Bender (2012) find a negative association. Finally, some research has not found any convincing relationship between reemployment earnings and either UI benefit generosity (Classen 1977; Belzil 2001) or longer potential duration of UI benefits (Lalive 2007; Card, Chetty, and Weber 2007).

A subset of this literature studies whether the design of UI has an impact on post-unemployment job match quality, measured by job or employment tenure. The conclusions have varied. Belzil (2001), Card, Chetty, and Weber (2007), and van Ours and Vodopivec (2008) find little or no relationship between UI generosity and subsequent tenure, whereas Centeno (2004), Centeno and Novo (2009), and Tatsiramos (2009) conclude that a more generous UI leads to a longer post-unemployment tenure.

Consequently, whether there is a link between various aspects of the UI system and post-unemployment job-market outcomes remains unclear. The controversy is due, in part, to the lack

of long-term post-unemployment data that can be matched to the kind of exogenous variation necessary to identify a causal effect. Because the WAWS experiment randomly assigned a group of new UI claimants to a treatment that effectively eliminated the WSR, in this paper we are able to study the causal effect of eliminating the WSR on long-run outcomes. By using nine years of post-experimental quarterly earnings records, merged to data from a random-assignment experiment, we are able to address two main questions: "How does elimination of the WSR affect the post-claim job match quality and long-term employment outcomes?" and "Does the effect vary by different groups of claimants?"

We address these questions by estimating regression models comparing the long-term outcomes of claimants assigned to the ER and control groups. We measure job match quality as the duration of tenure with the first post-claim employer and we measure other long-term employment outcomes along several dimensions: the duration of nonemployment, the number of post-unemployment employers, long-term earnings (and the volatility of those earnings), annual probability of employment, and hours worked in the nine years following the experiment.

Because it seems likely the WSR may have different effects on different groups of claimants, we estimate separate long-term effects for claimants who suffered permanent job loss, were temporarily laid off, quit for good cause, and were temporary or seasonal workers. We also examine how relaxing the WSR might affect long-term unemployed claimants; we do this by estimating the effects of ER separately for claimants with high and low probabilities of exhausting their UI benefits.

The paper has the following main findings. Although, for UI claimants as a whole, we find that the long-term employment outcomes of ER claimants were no different from outcomes of the comparison group, we find significant differences among various subgroups.

For permanent job losers, eliminating the WSR resulted in clearly worse employment outcomes: greater earnings losses in the year following job loss, a longer spell of nonemployment, and shorter tenure with the first post-claim employer. In contrast, eliminating the WSR had no impact on employment outcomes for workers who were not permanent job losers—those on a temporary layoff, quits, and contract or seasonal workers. That these claimants claimed more benefits for a longer period of time, but had employment prospects no different than workers in the control group, is consistent with the interpretation that they continued claiming benefits even after becoming reemployed.

The results for claimants who were not permanent job losers imply that the WSR plays an important role in mitigating claimant moral hazard: without the WSR, these claimants would draw more UI benefits, but would not ultimately have improved employment outcomes. The results also show that the WSR is an important policy for improving the welfare of permanent job losers, who in absence of the WSR would have worse employment outcomes. As permanent layoffs as a share of all layoffs have increased in the past 20 years (O'Leary, 2007), the findings of this paper are relevant to policymakers concerned with the current reemployment prospects of permanent job losers.

The rest of the paper is organized as follows. Section 2 briefly describes the design of the Washington experiment, describes the intention-to-treat effects, and includes a discussion of the effect of eliminating the WSR on returning to a former employer. Section 3 describes the methods for estimating the long-term effects and differences in long-term effects for various subgroups. Section 4 presents the results, and section 5 summarizes the findings and concludes. To keep the main discussion as direct as possible, we relegate a detailed description of the data

and details of how we created a long-term panel, as well as sample definitions, to a Data Appendix.

2 Exception Reporting and the Washington Alternative Work Search Experiment

The main purpose of the WAWS experiment was to test alternative means of reducing the duration of UI receipt and unemployment duration. To be eligible for UI in Washington, a claimant must have worked at least 680 hours in roughly the year before claiming UI, must have been laid off for lack of work and through no fault of her own, and must be "able, available, and searching" for work. This last criterion for UI eligibility is the *work search requirement* (WSR). In order to fulfill the WSR in Washington, the Employment Security Department personnel tell the claimants to contact at least three employers per week and to be prepared to give evidence that they have done so in an eligibility review interview, which may be conducted 13–15 weeks after the claimant files for benefits. For an eligibility review interview, a claimant reports to the public Employment Service for a one-hour group "interview" (or lecture) followed by (in some cases) a 15-minute individual interview during which employer contacts are checked.

The WAWS experiment tested the effects of eliminating this WSR by randomly assigning new UI claimants to a control group (subject to the standard WSR) and an ER treatment group. The latter were told (at the time of their initial claim) to actively seek work, but also that they would not be called in for an eligibility review interview (so they did not need to keep a record of job search contacts), and that weekly UI benefits would be mailed unless they called the Tacoma Employment Service Center to report they had stopped looking for work or had taken a job. In effect, ER amounted to an honor system with no WSR (Johnson and

Klepinger 1991, pp. 3–9). Random assignment occurred between July 1986 and August 1987 at the Tacoma Employment Service Center, based on the last digit of each claimant's Social Security number (see the Data Appendix for details).

2.1 Sample definition

Because the follow-up administrative wage records available to us begin in the first quarter of 1987, we do not have data on earnings, hours, and employer information for the first post-claim quarter for those who claimed in the third quarter of 1986 (that is, July, August, and September). Because of this data limitation, the sample we use is smaller than the sample studied by Johnson and Klepinger (1991, 1994); the Data Appendix provides details on how we define our analysis sample.

The experiment also tested a policy alternative called a "new work search" (NWS) policy, similar to the standard WSR except that selected claimants were called for an eligibility review interview earlier than usual (in week 6 after the claim rather than week 13–15 and at discretion of the UI office) and received a detailed job development plan (see Johnson and Klepinger [1991, p. 4]).[4] Since there is considerable variation between the states in the implementation of the eligibility review interview (see O'Leary [2006]), the NWS policy treatment could conceivably be a "standard" WSR in another state. As we document in Tables 3 below and in Table A1 in the Results Appendix, we argue that because the NWS policy differed little from the standard WSR in Washington at the time and because there is no evidence that

[4] The WAWS experiment also included an "intensive services" treatment, in which claimants were assigned to job search assistance (see Johnson and Klepinger [1991]). We study the long-term effects of this treatment in Cebi, Lachowska, and Woodbury (2014).

NWS policy affected outcomes, we can treat the NWS policy group as an alternative control group, and hence increase the sample size by pooling the NWS policy group with the controls.[5]

Table 1 offers a profile of how the different treatments worked in practice by showing proportions of the control, ER, and NWS policy groups that were called for an eligibility review interview and received various employment services. Two points are worth noting. First, almost none of the ER claimants were subjected to an eligibility review interview, consistent with the design of the treatment. ER claimants were also less likely to receive employment services, especially those requiring some initiative on the part of the claimant, such as assistance with a job development plan. The main services provided to ER claimants were job referral and placement, which are typically initiated by the Employment Service.

Second, Table 1 shows that when compared to the control group, the NWS policy group was more likely to receive an eligibility review interview and a job development plan, both likely due to the earlier scheduling of an eligibility review interview and the additional emphasis placed on a job development plan for claimants assigned to this group (see Johnson and Klepinger [1991, pp. 3–9]). Otherwise, the claimants assigned to the control and NWS groups received a similar mix of employment services (that is, job consultation, receipt of or referral to training, testing, support services, contacting an employer on the claimant's behalf, or any other contact with the Employment Service), suggesting that this treatment was effectively very similar to the standard WSR experienced by the controls.[6]

[5] In Table A1 in the Results Appendix, we show that there is no statistically significant difference in any of the short-term outcomes between the control and the NWS policy groups. In Tables A2–A9 in the Results Appendix, we show that our conclusions regarding the effect of ER on job-match quality and other long-term outcomes are unchanged if we limit the estimation sample to only include the ER and control groups (N = 3,145). Together, these findings strengthen our rationale for pooling the NWS policy group together with the control group.

[6] The differences between NWS policy group and the controls in the receipt of these six employment services were not statistically significant.

Since neither we nor Johnson and Klepinger (1991, 1994) find evidence that the NWS policy had a differential impact on outcomes when compared to the control group, we pool the control group together with the NWS policy group as a way to increase the size of our analysis sample. We refer to this larger, pooled control group as the *comparison group*.

2.2 Descriptive statistics

Table 2 displays various mean characteristics of the control, ER, and NWS policy groups, and the differences among them. The characteristics can be classified as

- demographic — sex, race, age, schooling, veteran status, marital and household status
- pre-claim — earnings and hours in the three prior years; industry and occupation before the claim; whether the individual had a prior UI claim
- claim-related — reason for job loss, whether the claimant had a recall date or was placed through a union hiring hall, UI benefits and claim type, and reservation wage

In general, the randomization protocol appears to have been successful, although there is evidence of nonrandomness between the controls and ER groups for some observables, for example, the distribution of age, schooling, industry, and reason for job loss across the groups. Also, relatively few ER group claimants were on standby or in a union that referred claimants to jobs. Johnson and Klepinger (1991, 1994) suggest that this difference is a matter of reporting rather than actual status: because claimants in the ER group did not need to submit continued claims for UI, the UI staff had no incentive to record the standby or union status of claimants in this group. A baseline survey completed by claimants (reported in Johnson and Klepinger [1994,

p. 704] but not available to us) supports the claim and shows no difference between the groups in the proportion on standby or placed by a union. Nonetheless, the measurable differences between the control and ER groups offer a rationale for regression-adjustment in comparing the groups.

Because the difference between control and ER groups could be due, in part, to using a smaller sample than Johnson and Klepinger (1991, 1994), in Table 2 we make additional comparisons of mean characteristics of the control group with the NWS policy group and of the ER group with a pooled sample of the control group and the NWS group (i.e., the comparison group). We note two things. First, randomization into the control and NWS groups appears to have been successful. Second, for only 3 characteristics out of 60 shown are the differences between claimants assigned to ER and the pooled control and NWS group with a p-value < 0.05.

2.3 Replication of Johnson and Klepinger's main results

Table 3 replicates the estimated effects of the ER treatment on various short-term outcomes considered by Johnson and Klepinger (1994). We group the outcome variables into two categories: 1) variables pertaining to UI receipt (total UI benefits paid, weeks of UI payments, and proportion that exhausted UI benefits); and 2) variables pertaining to short-term post-claim employment outcomes (proportion employed, hours worked, earnings, and proportion who returned to previous employer or industry). Each cell in the third and fourth columns from the left is a point estimate and a standard error from a separate regression. We will follow this convention throughout the paper.

Like Johnson and Klepinger (1994), we find that, on average, claimants in the ER group received more UI benefits (an additional $445 in Table 3), received benefits for an additional 3 weeks, and were more likely to exhaust their benefits (by about 11 percentage points) compared

with the comparison group. Also like Johnson and Klepinger, we find no statistically significant difference between the ER and comparison groups in hours worked or earnings in year 0 (the benefit year) or year 1 (the subsequent year).

On one hand, these findings suggest that eliminating the WSR may have led to abuse of the system by the claimants—the ER group received more UI benefits than the comparison group, but their earnings and work hours did not fall relative to the comparison group. It would seem that claimants in the ER group may have returned to work without informing the UI agency, and hence continued to receive benefits to which they were not entitled.

On the other hand, the estimates in Table 3 suggest that the ER group had a marginally lower probability of employment in the first post-claim quarter and in the year of the experiment. That the total earnings and hours of ER claimants in year 0 and 1 did not fall in spite of this lower probability of reemployment suggests that the ER claimants who did become reemployed could have worked at higher wages than the comparison group. This interpretation is consistent with the findings in Johnson and Klepinger (1994), who impute hourly wages using a Heckman selection-correction model and find that hourly wages increased for ER claimants (we do not attempt to impute hourly wages).[7] This potential hourly wage gain for ER claimants who were reemployed suggests they may have found better job matches. This interpretation is also consistent with the finding that ER claimants had almost a 3 percentage-point higher likelihood of returning to a former employer than the comparison group.

[7] Johnson and Klepinger (1994) find higher imputed hourly wages for ER claimants, but unlike us, they do not find a statistically significant decrease in the probability of reemployment. Our finding appears to be in part due to pooling together the NWS policy group and the control group. When comparing the claimants assigned to ER and the control group, the decrease in the probability of employment during the first year is negative, but not statistically different from zero. In Table A1, we show that the NWS policy group had a higher probability of reemployment in the first post-claim quarter (by about 0.5 percentage points) and in the year of the experiment (by about 0.8 percentage points) than the controls, but this gain is not statistically significant. Pooling the NWS policy group and the control group increases the average reemployment probability sufficiently to explain the statistically negative effect in Table 3.

Together, these findings suggest that eliminating the WSR may have improved the employment prospects of some claimants by allowing them more time to establish (or reestablish) a successful job match and earn higher wages. In section 4, we address this issue further by studying whether ER resulted in any long-term job match quality gains, and if so, for what type of claimant.

2.4 Post-claim employment outcomes

The administrative wage records allow us to follow each claimant's post-experiment employment, earnings, and hours for nine years. Because administrative wage records also include quarterly information about each claimant's employer account number (EAN), we construct post-claim employment outcomes not considered by Johnson and Klepinger (1991, 1994). First, for each claimant, we compute the number of unique employers (identified by EANs) we observe from the first quarter after the initial claim to the last follow-up quarter in which we can observe every claimant. We refer to this variable as *number of post-claim employers*. Figure 1 presents a histogram of the number of employers for the comparison group and the ER group. Table 5 shows the mean, median, and the standard deviation of this variable.

Second, we construct the variable *quarters of nonemployment* by computing the number of consecutive post-claim quarters in which a claimant is observed without covered earnings. This variable allows us to examine whether ER resulted in an increase in the time to reemployment beyond what we can infer from UI claims records that can only measure duration of *insured* unemployment. Figure 2 shows the distribution of this variable.

Third, we measure the volatility of post-claim earnings by standard deviation of earnings from year 0 to year 9. We refer to this variable as *standard deviation of post-claim earnings*.

Finally, we construct a proxy for post-claim job match quality. For each claimant, we compute the number of quarters in which a claimant is observed with earnings from the first post-claim employer. This variable ranges from 0, if no EAN is observed, to 40, if the claimant is with the same EAN throughout our window of observation. We refer to this employment tenure variable as *quarters with first post-claim employer*. Figure 3 shows the distribution of this variable.

3 Methods

In order to estimate the effect of ER on post-unemployment outcomes, we merge the WAWS experimental data on each claimant (derived from UI claims records, administrative wage records, and Employment Service records) with quarterly administrative records on the claimant's employment, earnings, and hours worked in the 40 quarters following the claim quarter (and the enrollment in the experiment).

The effect of assignment to the ER treatment group on outcomes can be obtained by pooling the comparison group (consisting of the control and NWS policy groups) and ER group and estimating linear models of the following form:

$$y_i = \alpha + \beta ER_i + X_i\gamma + u_i, \tag{1}$$

where y_i is an outcome for individual i in any of the years following enrollment in the experiment; ER_i is an indicator for assignment to the ER group (that is, the group not subject to the WSR); X_i includes all of the variables listed in Table 2, as well as the unemployment rate in the county where the claim was filed and indicators for the quarter the individual claimed benefits; and u_i denotes i's unobservable traits.

The identifying assumption is that assignment to treatment indicator is independent of any individual characteristics, including those unobserved by the researcher: $E(u|ER) = 0$. As

Johnson and Klepinger (1994) note, because the random assignment to control and ER treatment groups appears to have succeeded, this assumption is reasonable. In this case, the ordinary least squares (OLS) estimator of β is a consistent estimator of the intention-to-treat effect on outcome y. Including the demographic variables (X) reduces sampling error and controls for observable differences between treatment and control groups that may arise even under random assignment.

The outcomes (y) include the claimant's post-experiment employment, earnings, hours, quarters with first post-claim employer, number of post-claim employers, quarters of nonemployment, and standard deviation of earnings. By estimating a model for each of the nine years following enrollment in the experiment, we can trace out the path of long-term effect of assignment to the ER group on hours, earnings, and probability of employment. For the remaining outcomes—number of post-claim employers, quarters of nonemployment, standard deviation of earnings, and quarters with first post-claim employer—we also estimate linear models. Since the first three outcomes listed above are count variables, we have also estimated Poisson maximum-likelihood models. Our findings remain qualitatively unchanged.

Taken together, all these outcomes capture different, but not necessarily independent dimensions of the effect of assignment to ER. If, according to the improved job match hypothesis, eliminating the WSR prolonged the duration of unemployment, but had a beneficial effect on post-claim outcomes, we would expect the estimate of β to have a positive effect on post-claim hours, earnings, employment, and the number of quarters with the first post-claim employer. On the contrary, if eliminating the WSR only prolongs the unemployment spell, then, according to the negative duration dependence hypothesis, we would expect the estimate of β to have a negative effect on post-claim hours, earnings, employment, and the number of quarters with the first post-claim employer.

The effect of ER on the remaining outcomes—quarters of nonemployment, number of post-claim employers, and standard deviation of earnings—is more ambiguous and ought to be considered jointly with the estimated effect on other outcomes. For example, if ER did not have any effect on the level of post-claim earnings but at the same time had a negative effect on the volatility of post-claim earnings, it could be argued that ER had a beneficial effect, since, on average, claimants assigned to ER are earning just as much but experience less variability. Analogously, a longer duration of nonemployment and fewer post-claim employers should be interpreted jointly with the effect on post-claim earnings of ER claimants, since it is difficult to interpret the effect of ER on these outcomes in isolation.

3.1 Effect of ER by reason for job loss

In order to study whether the effects of eliminating the WSR are different for claimants on permanently laid off than for claimants who lost their jobs for other reasons, we estimate separate models by five mutually exclusive reasons for job loss: 1) quit for reasons satisfying the standard for "good cause," 2) lost job permanently, 3) temporary layoff, 4) contract ended/seasonal layoff, and 5) lost job for reasons unknown. The Data Appendix explains in detail how we created these indicators.

We estimate Equation (1) for each of the five reasons for job loss, where each model compares outcomes for claimants assigned to ER who lost their jobs due to a given reason to claimants in the control group who lost their job for the same reason. Since reason for job loss is pre-determined with respect to treatment assignment, the coefficient on the ER indicator yields an intention-to-treat effect of eliminating the WSR for a given reason-for-job-loss category of claimants.

3.2 Effect of ER by likelihood of benefit exhaustion

In order to study whether the long-term unemployed benefit from the elimination of the WSR, we study claimants with a high and low probability of exhausting benefits separately. In practice, we construct an ex ante probability of benefit exhaustion. First, using a probit, we estimate a likelihood of benefit exhaustion over the comparison group sample. To estimate the probit, we include all of the variables in Table 2, plus the unemployment rate in the county and month the claim was filed and quarter of claim in the conditioning set. Second, we assign the predicted likelihood values to all the claimants in the analysis sample.[8] We define a claimant as "high probability" if the claimant's ex ante probability of exhausting benefits is higher than the comparison group average, which equals 26.4 percent.[9] We define a claimant as "low probability" if the claimant has an ex ante probability that is lower than the comparison group average.

3.3 Threats to validity

Since WAWS is a random-assignment experiment, it has high internal validity. However, external validity might be compromised if the inferences and conclusions cannot be generalized from the population and setting in which they are studied to other populations and settings. We believe that external validity of the study is reasonably high, as the state of Washington is not an outlier with respect to the characteristics of its population. As Johnson and Klepinger (1994) note, the UI practices implemented in the state of Washington at the time of the demonstration

[8] This bears similarities to estimating a worker profiling score; see Berger et al. (1997) and Berger et al. (2000).
[9] In order to increase the number of observations and avoid colinearity problems, we estimate the likelihood model using the pooled NWS policy group and control group for all the quarters of the experiment. The mean of value of exhausted benefits is 26.4 percent in this sample, which is slightly higher than the mean value in Table 3, where it is 23.1 percent. Table A10 in the Results Appendix shows the estimated coefficients for the model predicting benefit exhaustion.

(that is, the standard WSR that the claimants assigned to the control group were subject to) did not deviate from the approach used in most other states at that time. It is also worthwhile to note that the average unemployment rate in Tacoma, the location of WAWS experiment, was at the time about 7.9 percent. Therefore, the estimated effects pertain to relatively slack labor market conditions, a setting that makes our findings of current interest.

Another concern regarding external validity is whether compliance with the experimental protocol is specific to a given demonstration. In the case of ER, the issue of noncompliance (opting out of treatment) is not really a concern because the ER treatment is in the form of information and instructions supplied to claimants when they file for benefits. That is, the treatment does not include a follow up, and hence the possibility of noncompliance as would be the case with a training program or job search assistance.

A potential threat to external validity is whether turning a temporary and local experimental program into a permanent and widespread policy might change the economic environment in such a way that the conclusions from the smaller-scale experiment cannot be generalized. For example, in the permanent absence of the WSR, more workers might be induced to enter the UI system, thus changing the composition of the pool of claimants from that studied in the original WAWS demonstration. This would reduce the external validity of the experiment.

Finally, we discuss attrition from our long-term panel and the reliability of our follow-up outcome measures in the Data Appendix.

4 Results

4.1 Baseline results

Table 4 reports the estimated long-term effect of assignment to ER on the probability of employment, hours worked, and total earnings in each of the nine years following enrollment in the WAWS experiment. In order for the treatment effects to be interpreted as deviations from the comparison group mean, we present the mean and standard deviation of the comparison group to the left of the β estimate. Except for the 2 percentage point lower probability of employment in the year of the experiment, ER did not have a statistically significant effect on employment in the other post-experimental years, nor did it have an effect on hours worked or earnings.

Table 5 reports the estimated effect of assignment to ER on the other post-claim long-term employment outcomes: the number of post-claim employers, quarters of nonemployment, the standard deviation of subsequent earnings, and our proxy for job match quality—quarters with the first post-claim employer. As in the previous table, we present the mean and standard deviation of the comparison group to the left of each estimated coefficient. Since Figures 1–3 imply that some of these variables have a long right-tail, we also present the comparison-group median.

On average, a claimant in the comparison group spent about two years with the first post-claim employer, but the median tenure equals only three quarters. Rounding down the mean, we see that the mean and the median number of post-claim employers in the 40 quarters following the experiment equals four. The median number of quarters of nonemployment equals one quarter, while the mean equals about 3.6 quarters.

Turning to the β coefficient, we see that the point estimates in Table 5 suggest that ER prolonged the duration of nonemployment but also increased tenure with first employer, reduced the number of post-claim employers, and reduced the volatility of earnings. However, all the point estimates in Table 5 are small, and no point estimate is statistically different from zero. In

sum, the results in Tables 4 and 5 suggest that for UI claimants as a whole, eliminating the WSR did not have a statistically significant effect on any employment outcome in the nine years following the experiment.

4.2 The effect of ER by reason for job loss

In order to see if the effect of ER on outcomes differs depending on reason for job loss, in Table 6, each row presents the estimated effect of ER on a selected outcome, by reason for job loss. Table 7 complements Table 6 by presenting the mean and standard deviation of each outcome for each reason-for-job-loss category for the claimants in the comparison group.

Turning to the effect of ER on UI receipt outcomes (total UI benefits paid, weeks of UI payments, and whether a claimant exhausted benefits), we see that the estimates in Table 6 are numerically similar to the estimates from Table 3.[10] For every reason for job loss category, the ER claimants received between about $410 and $510 more in total UI benefits, for about 3–4 weeks longer, and were about 10 percentage points more likely to exhaust benefits than claimants in the comparison group.

Caution must be exercised when comparing the results *across* the groups in Table 6, as the comparison group baseline average is different depending on reason for job loss; see the means of outcomes of the comparison group in Table 7.[11] Taking these differences into account, it turns out that, relative to the comparison group average, the increase in total UI benefits paid and weeks of UI payments is similar across the reason for job loss categories; however, the

[10] To save space, we present only benefit year outcomes and not both benefit year and first spell outcomes, as in Table 3.

[11] For example, claimants who are unemployed due to a permanent job loss are more likely to be female, white, college educated, and work more in finances and services compared to the entire sample of UI claimants in WAWS. They are also likely to have had a prior UI claim. Claimants temporarily laid off are on the other hand more likely to be male, younger, less likely to have a college degree, but more likely to work in construction or manufacturing. They are also less likely to have had a prior UI claim. The underlying descriptive statistics are available from the authors.

likelihood of benefit exhaustion for ER claimants on temporary layoff is strikingly 79 percent higher (that is, 0.108/0.136). The likelihood of exhausting benefits is 40–52 percent higher for ER claimants who became unemployed for other reasons.

Turning to the year 0 employment outcomes (in Table 6), we see that ER claimants who were permanently laid off had lower chances of employment, worked fewer hours, and had lower total earnings compared to comparison claimants laid off permanently. This decrease is, however, only transitory: by year 1, the outcomes for the ER claimants were statistically undistinguishable from the comparison group.[12] The temporary negative effect on employment outcomes during the year of the experiment is consistent with the ER claimants taking almost 1.5 quarters longer to find employment than the baseline average of 4.2 quarters or (see Table 7). It appears that the longer duration of insured unemployment resulted in a longer duration of nonemployment. In Table 6, we also see that ER claimants who were permanently laid off had a shorter tenure with their first post-claim employer by about 1.65 quarters. This suggests that the first job match of permanently laid off claimants assigned to ER was less successful than the first job match of permanently laid off claimants in the comparison group.

The effect of ER claimants on temporary layoff is very different. We see that the only statistically significant employment outcome effect is a decrease in the number of post-claim employers. We also see that ER claimants on temporary layoff had a 4.4 percentage point higher probability of returning to a previous employer, but this effect is not statistically significant (t-value is 1.42). Overall, the marginally improved probability of returning to a former employer and the reduction in job changing following ER did not lead to long-term gains in earnings or employment.

[12] Also in later years the employment outcome differences are not statistically different from zero; we do not show this in Table 6 to conserve space.

Interestingly, the largest group, the claimants who lost their jobs for reasons unknown to us, had higher total earnings in year 1, experienced a shorter duration on nonemployment by almost a quarter, and were more likely to return to their pre-claim employers by 4.7 percentage points. This is intriguing, as this is the only group in Table 6 for which there is a statistically significant effect on return to same employer. Initially, we expected that the increase in the probability of return to same employer reported in Table 3 would be explained by a higher probability of return by claimants placed on recall. However, as Table 6 shows, this effect is driven by the group whose reasons for unemployment are unknown to us.

Other than a higher probability of return to former industry for ER claimants who were seasonal or contract workers, for the claimants in the remaining category, claimants who quit, ER did not have a statistically different effect on any employment outcomes.

4.3 The effect of ER by likelihood of benefit exhaustion

In Table 8, we show the long-term effects of assignment to ER on the probability of employment, hours worked, and earnings during the nine years following enrollment in the experiment for claimants likely to exhaust their benefits, i.e., claimants whose predicted likelihood is higher than the comparison-group average. We see that in year 0, ER claimants had a 4.4 percentage point lower likelihood of reemployment than claimants in the comparison group. We also see a negative effect on employment in year 3 but not in the years before and after, which may question how much stock we can put on this finding. There is no statistically significant effect on any of the other employment outcomes, hours worked and earnings.

In Table 9, we present the results for claimants statistically unlikely to exhaust benefits. We see that the outcomes for these ER claimants were not statistically different from the outcomes of claimants assigned to the comparison group.

Finally, in Table 10 we show the effect of ER for claimants with both high and low probability of exhausting benefits on the remaining employment outcomes: number of post-claim employers, quarters of nonemployment, standard deviation of earnings, quarters with the first post-claim employer, and the likelihood to return to a former employer and former industry. Except for an increase in the probability to return to a former employer for ER claimants unlikely to exhaust benefits, in no remaining case is the effect of ER statistically different from zero.

5 Discussion and Summary

A longstanding concern about strict enforcement of the UI work search requirement (WSR) is that it may pressure unemployed job seekers to accept a job "too soon," reducing job match quality and long-term earnings. In addition to being undesirable for workers this could be detrimental to employers, many of whom value long-term relationships and are willing to pay higher wages to encourage long tenure; see Farber (1999).

The Washington Alternative Work Search experiment tested the effects of eliminating the WSR by randomly assigning new UI claimants to a control group and to an "exception reporting" (ER) honor system in which claimants were told to search actively for reemployment but were also told their benefits would be sent to them unless they told the UI agency that they had found a job or had stopped looking for work. By appending nine years of administrative wage records to the original data from the experiment, we are able to examine the long-term

effects of ER—that is, the effects on employment tenure, number of post-claim employers, employment, hours, and earnings.

In the short term, ER increased the duration of UI benefit receipt, benefits received, and the probability of exhausting benefits. Although it also increased the probability that a worker would return to a former employer, which could be a positive outcome, in the long-term, (that is, in the nine years following the experiment), ER had no effect on earnings, hours worked, or other employment outcomes. We also find no evidence of a statistically significant effect of ER on time to reemployment, post-claim employment tenure, number of post-claim employers, or volatility of earnings. Overall, then, ER increased claimant moral hazard and the costs to the UI system without observable gains for workers.

We also study the effects of ER by reason for unemployment, and find differences among different groups of claimants. First, eliminating the WSR was harmful in the short run for claimants who lost their job as a result of a permanent layoff, consistent with negative duration dependence. During the year of the experiment, these claimants experienced lower probability of reemployment, worked fewer hours, and had lower earnings. Moreover, in the long term, these claimants were reemployed about 1.4 quarters later then the comparison group and experienced shorter job tenure with their first post-claim employer by 1.65 quarters. Both of these effects are economically large and imply strongly that the WSR is a policy that benefits UI claimants who were permanently laid off.

Second, it appears that eliminating the WSR led to more abuse of the UI system by all groups of claimants who were not permanent job losers—claimants who quit, claimants on temporary layoff, or claimants who were contract or seasonal workers. For these claimants, ER led to more benefit payments, a longer spell of insured unemployment, and a higher likelihood of

exhausting benefits. However, the probability of reemployment, the number of hours worked, and earnings for these claimants were no different from those assigned to the comparison group (who were subject to the WSR). This implies that eliminating the WSR led to increased claimant moral hazard—UI benefits drawn were greater, but for the ER claimants who were not permanent job losers, the employment outcomes were no different than for claimants subjected to the standard WSR.

Overall, eliminating the WSR was costly to the UI system without convincingly improving employment outcomes for any claimant category considered. The clear conclusion for policy is that the WSR is an important tool for improving outcomes of permanent job losers and for reducing moral hazard associated with UI for other UI claimants.

References

Addison, John T., and P. Portugal. 1989. "Job Displacement, Relative Wage Changes, and Duration of Unemployment." *Journal of Labor Economics* 7(3): 281–302.

Addison, John T., and McKinley L. Blackburn. 2000. "The Effects of Unemployment Insurance on Postunemployment Earnings." *Labour Economics* 7: 21–53.

Ashenfelter, O., Ashmore, D., and O. Deschênes. 2005. "Do Unemployment Insurance Recipients Actively Seek Work? Evidence from Randomized Trials in Four U.S. States." *Journal of Econometrics* 125: 53–75.

Berger, M. C., D. A. Black, A. Chandra, and S. N. Allen. 1997. "Kentucky's Statistical Model of Worker Profiling for Unemployment Insurance." *Kentucky Journal of Economics and Business* 16: 1–18.

Berger, M. C.; Black, D. A. and J. A. Smith "Evaluating Profiling as a Means of Allocating Government Services," in Michael Lechner and Friedhelm Pfeiffer (Eds.), *Econometric evaluation of labour market policies*. Heidelberg: Physica-Verlag, 2000, pp. 59–84.

Belzil, Christian. 2001. "Unemployment Insurance and Subsequent Job Duration: Job Matching Versus Unobserved Heterogeneity." *Journal of Applied Econometrics* 16(5): 619–636.

Burgess, P. L., and J. L. Kingston. 1976. "The Impact of Unemployment Insurance Benefits on Reemployment Success." *Industrial and Labor Relations Review* 30(1): 25–31.

Card, D., R. Chetty, and A. Weber. 2007. "Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market." *Quarterly Journal of Economics* 122(4): 1511–1560.

Cebi, Merve, Marta Lachowska, and Stephen A. Woodbury. 2014. "Long-Term Effects of Job Search Assistance." Unpublished manuscript.

Centeno, Mario. 2004. “The Match Quality Gains for Unemployment Insurance.” *Journal of Human Resources* 39(Summer): 839–863.

Centeno, Mario, and Álvaro A. Novo. 2009. “Reemployment Wages and UI Liquidity Effect: A Regression Discontinuity Approach.” *Portuguese Economic Journal* 8 (4): 45–52.

Classen, Kathleen. 1977. “The Effect of Unemployment Insurance on the Duration of Unemployment and Subsequent Earnings.” *Industrial and Labor Relations Review* 30 (July 1977): 438–444.

Ebenstein, Avraham, and Kevin Stange. 2010. “Does Inconvenience Explain Low Take-Up? Evidence from Unemployment Insurance.” *Journal of Policy Analysis and Management* 29(1): 111–136.

Ehrenberg, R. G., and R. L. Oaxaca. 1976. “Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Gain.” *American Economic Review* 66(5):754–766.

Farber, Henry S., 1999. “Mobility and Stability: The Dynamics of Job Change in Labor Markets” in O. Ashenfelter & D. Card (eds), Handbook of Labor Economics, edition 1, volume 3, chapter 37, pages 2439–2483.

Gregory, M., and R. Jukes. 2001. “Unemployment and Subsequent Earnings: Estimating Scarring among British Men 1984–1994.” *The Economic Journal* 111: F607–F625.

Johnson, Terry R., and Daniel R. Klepinger. 1991. “Evaluation of the Impacts of the Washington Alternative Work Search Experiment.” Unemployment Insurance Occasional Paper 91-4. Washington, DC: U.S. Department of Labor, Employment and Training Administration.

———. 1994. “Experimental Evidence on Unemployment Insurance Work search Policies.” *Journal of Human Resources* 29 (Summer): 695–717.

Klepinger, Daniel R., Terry R. Johnson, and Jutta M. Joesch. 2002. “Effects of Unemployment Insurance Work-Search Requirements: The Maryland Experiment.” *Industrial and Labor Relations Review* 56(1): 3–32.

Lalive, R. 2007. “Unemployment Benefits, Unemployment Duration, and Post-Unemployment Jobs: A Regression Discontinuity Approach.” *American Economic Review* 97(2): 108–112.

McCall, B. and W. Chi. 2008. “Unemployment Insurance, Unemployment Durations and Re-employment Wages.” *Economics Letters* 99(1): 115–118.

Nekoei, A., and A. Weber. 2013. “Does Extending Unemployment Benefits Improve Job Quality.” Unpublished manuscript.

Notowidigdo, M., K. Kroft, and F. Lange. 2013. “Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment.” *Quarterly Journal of Economics* 128(3): 1123–1167.

O’Leary, Christopher J. 2006. “State UI Job Search Rules and Reemployment Services.” *Monthly Labor Review* 130(June): 27–37.

———. 2007. “Do Unemployment Insurance Beneficiaries Still Return to their Prior Employers?” Prepared for the 29 Annual Research Conference of the Association for Public Policy Analysis and Management.

Poe-Yamagata, E., J. Benus, N. Bill, H. Carrington, M. Michaelides, and T. Shen. 2011. “Impact of the Reemployment and Eligibility Assessment (REA) Initiative.” Department of Labor Employment and Training Administration Occasional Paper 2012-08.

Schmider, Johannes, Till von Wachter, and Stefan Bender. 2013. “The Causal Effect of Unemployment Duration on Wages: Evidence from Unemployment Insurance Extensions.” NBER Working Paper No. 19772. Cambridge, MA: National Bureau of Economic Research.

Tatsiramos, K. 2009. “Unemployment Insurance in Europe: Unemployment Duration and Subsequent Employment Stability.” *Journal of the European Economic Association* 7(6): 1225–1260.

Tatsiramos, K., and J. C. van Ours. 2014. “Labor Market Effects of Unemployment Insurance Design.”*Journal of Economic Surveys* 28(2): 284–311.

van Ours, Jan C., and Milan Vodopivec. 2008. “Does Reducing Unemployment Insurance Generosity Reduce Job Match Quality?” *Journal of Public Economics* 92: 684–695.

Results

Table 1
Eligibility Review Interviews and Employment Services Received by Control, Exception Reporting, and New Work Search Groups

Service[a]	(1) Control	(2) Exception Reporting	(3) New Work Search	(4) Difference between (1) and (2)*	(5) Difference between (1) and (3)*
Eligibility review interview	0.250	0.004	0.322	0.000	0.000
Employment services					
job referral/placement	0.185	0.155	0.160	0.027	0.102
job development plan	0.114	0.007	0.182	0.000	0.000
other employment service[b]	0.107	0.062	0.116	0.000	0.466
Sample size	1,539	1,606	1,073		

Source: Author's tabulations of the Washington Alternative Work Search experimental data, from UI claims records, administrative wage records, and Employment Service records. See the Data Appendix for details.

Notes: Universe consists of exception reporting, control and new work search groups during fall 1986, winter 1987, and spring 1987.

* *p* -value for test of difference of means.

a. A claimant may receive more than one category of services.

b. Job consultation, receipt of or referral to training, testing, support services, job development (contacting an employer on the claimant's behalf), or any other contact with the Employment Service.

Table 2
Characteristics of Control, New Work Search, and Exception Reporting Groups

Covariate	Control	Exception Reporting	Difference[1] (*p* -value)	Control	New Work Search	Difference[1] (*p* -value)	New Work Search and Control	Exception Reporting	Difference[1] (*p* -value)
Male	0.718	0.717	0.935	0.718	0.713	0.779	0.716	0.717	0.958
Race									
white	0.819	0.828	0.488	0.819	0.829	0.479	0.823	0.828	0.677
black	0.097	0.099	0.885	0.097	0.087	0.350	0.093	0.099	0.521
other	0.084	0.073	0.252	0.084	0.084	0.996	0.084	0.073	0.201
Age									
≤ 24	0.218	0.210	0.592	0.218	0.192	0.111	0.207	0.210	0.833
25-34	0.389	0.404	0.394	0.389	0.391	0.909	0.390	0.404	0.367
35-44	0.240	0.207	**0.029**	0.240	0.222	0.285	0.232	0.207	0.058
45-54	0.103	0.111	0.459	0.103	0.129	0.040	0.113	0.111	0.804
≥ 54	0.051	0.068	**0.042**	0.051	0.066	0.093	0.057	0.068	0.154
Schooling									
less than high school	0.159	0.123	**0.004**	0.159	0.148	0.471	0.154	0.123	**0.004**
high school	0.537	0.566	0.099	0.537	0.542	0.774	0.539	0.566	0.088
some college	0.225	0.240	0.303	0.225	0.242	0.298	0.232	0.240	0.535
college graduate	0.080	0.071	0.343	0.080	0.067	0.220	0.075	0.071	0.657
Veteran	0.196	0.190	0.654	0.196	0.215	0.235	0.204	0.190	0.264
Marital status/gender									
married male	0.270	0.264	0.721	0.270	0.242	0.116	0.258	0.264	0.688
married female	0.099	0.094	0.609	0.099	0.096	0.772	0.098	0.094	0.670
Household status									
no dependents	0.309	0.329	0.212	0.309	0.322	0.485	0.314	0.329	0.296
1 dependent	0.155	0.148	0.546	0.155	0.169	0.360	0.161	0.148	0.250
2 or more dependents	0.236	0.229	0.626	0.236	0.207	0.081	0.224	0.229	0.731
homeowner	0.286	0.285	0.934	0.286	0.253	0.067	0.273	0.285	0.399
Pre-claim earnings ($)									
1 year before	13,841	13,559	0.436	13,841	13,531	0.447	13,713	13,559	0.632
2 years before	11,900	11,571	0.417	11,900	11,639	0.563	11,793	11,571	0.538
3 years before	10,744	10,737	0.988	10,744	10,801	0.904	10,767	10,737	0.936
Pre-claim hours									

Covariate	Control	Exception Reporting	Difference[1] (*p* -value)	Control	New Work Search	Difference[1] (*p* -value)	New Work Search and Control	Exception Reporting	Difference[1] (*p* -value)
1 year before	1334	1313	0.376	1334	1286	0.073	1314	1313	0.938
2 years before	1101	1064	0.178	1101	1076	0.414	1091	1064	0.271
3 years before	946	931	0.599	946	964	0.583	954	931	0.382
Occupation									
professional	0.105	0.102	0.772	0.105	0.106	0.936	0.106	0.102	0.714
clerical	0.122	0.133	0.379	0.122	0.116	0.661	0.120	0.133	0.222
sales	0.058	0.059	0.933	0.058	0.050	0.407	0.055	0.059	0.605
service	0.101	0.101	0.988	0.101	0.123	0.073	0.110	0.101	0.357
agric., fishery, forestry	0.026	0.028	0.726	0.026	0.021	0.455	0.024	0.028	0.436
processing	0.038	0.033	0.420	0.038	0.035	0.698	0.037	0.033	0.481
machine trades	0.086	0.090	0.700	0.086	0.107	0.066	0.095	0.090	0.594
benchwork	0.046	0.048	0.811	0.046	0.049	0.700	0.047	0.048	0.944
structural work	0.266	0.265	0.910	0.266	0.274	0.667	0.270	0.265	0.728
miscellaneous	0.151	0.143	0.486	0.151	0.116	**0.011**	0.137	0.143	0.614
Industry									
agriculture	0.025	0.025	0.969	0.025	0.021	0.588	0.023	0.025	0.749
mining	0.001	0.001	0.539	0.001	0.001	0.785	0.001	0.001	0.590
construction	0.205	0.196	0.520	0.205	0.190	0.338	0.199	0.196	0.816
manufacturing	0.237	0.232	0.778	0.237	0.263	0.125	0.247	0.232	0.267
transportation, utilities	0.038	0.054	**0.028**	0.038	0.034	0.577	0.036	0.054	**0.005**
wholesale trade	0.070	0.060	0.237	0.070	0.048	**0.023**	0.061	0.060	0.845
retail trade	0.159	0.158	0.938	0.159	0.158	0.994	0.158	0.158	0.934
finance, ins., real estate	0.028	0.031	0.597	0.028	0.031	0.674	0.029	0.031	0.706
services	0.174	0.172	0.866	0.174	0.172	0.909	0.173	0.172	0.896
government	0.045	0.057	0.135	0.045	0.054	0.318	0.049	0.057	0.240
unclassified	0.018	0.014	0.390	0.018	0.027	0.129	0.022	0.014	0.083
Prior UI claim									
none	0.804	0.804	0.969	0.804	0.791	0.408	0.799	0.804	0.701
duration ≤ 15 weeks	0.104	0.100	0.688	0.104	0.106	0.852	0.105	0.100	0.584
duration > 15 weeks	0.092	0.097	0.638	0.092	0.103	0.353	0.096	0.097	0.964
Reason for job loss									
permanent layoff	0.172	0.153	0.149	0.172	0.157	0.291	0.166	0.153	0.280
temporary layoff with recall date	0.231	0.265	**0.027**	0.231	0.253	0.179	0.240	0.265	0.073
contract/seasonal	0.155	0.154	0.908	0.155	0.156	0.981	0.155	0.154	0.886

Covariate	Control	Exception Reporting	Difference[1] (*p* -value)	Control	New Work Search	Difference[1] (*p* -value)	New Work Search and Control	Exception Reporting	Difference[1] (*p* -value)
quit	0.172	0.000	0.697	0.172	0.167	0.719	0.170	0.000	0.533
Employer-attached/placed by union[2]	0.355	0.286	**0.000**	0.355	0.371	0.418	0.362	0.286	**0.000**
UI benefits/claim type									
weekly amount ($)	146	145	0.564	146	145	0.640	146	145	0.686
maximum amount ($)	3,868	3,830	0.529	3,868	3,849	0.776	3,860	3,830	0.576
potential duration	26.0	25.9	0.887	26.0	26.0	0.875	26.0	25.9	0.810
replacement rate (percent)[3]	61.7	61.6	0.841	61.7	61.4	0.733	61.6	61.6	0.961
combined wage claim[4]	0.049	0.044	0.492	0.049	0.045	0.635	0.047	0.044	0.597
ex-service member claim	0.034	0.035	0.868	0.034	0.034	0.923	0.034	0.035	0.890
federal employee claim	0.009	0.018	**0.031**	0.009	0.014	0.241	0.011	0.018	0.060
Reservation wage (hourly)									
≤ $5.00	0.190	0.181	0.479	0.190	0.175	0.325	0.184	0.181	0.770
$5.01–$7.00	0.151	0.164	0.294	0.151	0.142	0.519	0.147	0.164	0.129
$7.01–$10.00	0.138	0.161	0.065	0.138	0.157	0.180	0.145	0.161	0.165
$10.01–$20.00	0.143	0.130	0.272	0.143	0.138	0.717	0.141	0.130	0.296
> $20.00	0.110	0.106	0.719	0.110	0.117	0.581	0.113	0.106	0.491
Sample size	1,539	1,606		1,539	1,073		2,612	1,606	

Notes: Universe consists of exception reporting, control and new work search groups during fall 1986, winter 1987, and spring 1987.

Source: Author's tabulations of the Washington Alternative Work Search experimental data.

1. Bold denotes *p* -values for the test of mean differences between groups < .05.

2. Claimants were not required to search for work if they were on layoff with a set recall date or if they were placed through a union hiring hall.

3. The replacement rate is the weekly benefit amount as a percentage of average weekly earnings before the UI claim.

4. Combined wage claims use earnings from more than one state to calculate base period earnings.

Table 3

Effect of Exception Reporting on Selected Outcomes as Deviations from the Comparison Group (Control and New Work Search Group pooled)

Outcome	Comparison (Control and New Work Search) Mean (Std.Dev.)		Exception Reporting Coefficient (Std. error)	
UI Receipt Outcomes				
Benefit year				
Total UI benefits paid ($)	1,956	(1728)	451***	(47)
Weeks of UI payments	14.18	(10.5)	3.26***	(0.31)
Exhausted benefits (proportion)	0.231	(0.422)	0.114***	(0.014)
First spell				
Total benefits in first UI spell ($)	1,638	(1636)	445***	(48)
Weeks of first UI spell	13.58	(10.58)	3.43***	(0.33)
Employment Outcomes				
First quarter outcomes[a]				
Employed (proportion)	0.695	(0.46)	-0.032**	(0.014)
Hours worked	201.3	(210)	-7.3	(6.1)
Total Earnings ($)	2,285	(2676)	-84	(71)
Year 0 outcomes[a]				
Employed (proportion)	0.89	(0.313)	-0.020*	(0.010)
Hours worked	1016	(746)	-29.0	(21.7)
Total Earnings ($)	11,617	(10143)	-277	(253)
Year 1 outcomes[a]				
Employed (proportion)	0.843	(0.364)	-0.005	(0.011)
Hours worked	1134	(838)	0.9	(25.2)
Total Earnings ($)	13,122	(11210)	321	(305)
Other outcomes				
Returned to same employer (proportion)	0.322	(0.467)	0.029**	(0.013)
Returned to same industry (proportion)	0.441	(0.497)	0.019	(0.014)
Sample size			4,218	

Notes: Universe consists of exception reporting, control, and new work search groups during fall 1986, winter 1987, and spring 1987. Robust standard errors are in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression-adjusted differences control for all variables displayed in Table 2 plus the quarter in which the claim was filed and the unemployment rate in the county and month in which the claim was filed.

a. First quarter is the quarter following the claim quarter. Year 0 is defined as the sum of the first, second, third, and fourth quarter after the claim quarter. Year 1 is defined as the sum of the fifth, sixth, seventh, and eighth quarter after the claim quarter. Earnings are expressed in 1988:4 prices.

Table 4

Effect of Exception Reporting on Long-Term Employment, Hours, and Earnings as Deviations from the Comparison Group

Outcome	Proportion employed				Hours worked				Total Earnings ($)			
	Comparison group		Exception Reporting		Comparison group		Exception Reporting		Comparison group		Exception Reporting	
	Mean (Std. dev.)		Coeff.	(Std. error)	Mean (Std. dev.)		Coeff.	(Std. error)	Mean (Std. dev.)		Coeff.	(Std. error)
Year relative to claim												
Year 0	0.890	(0.313)	-0.020*	(0.010)	1,016	(746)	-29.0	(21.7)	11,617	(10,143)	-277	(253)
Year 1	0.843	(0.364)	-0.005	(0.011)	1,134	(838)	0.9	(25.2)	13,122	(11,210)	321	(305)
Year 2	0.810	(0.393)	-0.008	(0.012)	1,133	(867)	-13.2	(25.9)	13,828	(12,014)	-180	(329)
Year 3	0.776	(0.417)	-0.015	(0.013)	1,104	(883)	7.0	(26.8)	13,615	(12,204)	-114	(350)
Year 4	0.741	(0.438)	-0.011	(0.014)	1,038	(886)	4.6	(27.0)	13,199	(12,714)	-96	(364)
Year 5	0.713	(0.452)	-0.007	(0.014)	992	(897)	29.3	(27.5)	12,728	(12,752)	349	(368)
Year 6	0.681	(0.466)	0.005	(0.014)	942	(912)	27.0	(27.7)	12,001	(12,925)	304	(368)
Year 7	0.659	(0.474)	0.003	(0.014)	932	(922)	9.8	(28.0)	11,621	(12,674)	445	(374)
Year 8	0.637	(0.481)	-0.003	(0.015)	891	(915)	6.7	(28.1)	11,442	(12,918)	415	(387)
Year 9	0.624	(0.484)	-0.001	(0.015)	913	(943)	1.7	(28.6)	11,780	(13,532)	280	(401)
Sample size			4,218				4,218				4,218	

Notes: See annotations to Table 3.

Table 5
Effect of Exception Reporting on Selected Long-Term Employment Outcomes as Deviations from the Comparison Group

	Comparison group			Exception Reporting	
	Mean	Median	(Std. dev.)	Coeff.	(Std. error)
Outcome					
Number of post-claim employers	4.68	4	(3.78)	-0.13	(0.11)
Quarters of nonemployment	3.63	1	(8.34)	0.12	(0.25)
Std. deviation of post-claim earnings ($)	5,850	5,253	(3,965)	-70	(115)
Quarters with first post-claim employer	8.05	3	(10.85)	0.23	(0.33)
Sample size				4,218	

Notes: See annotations to Table 3. Number of post-claim employers is defined as the number of unique employers we observe from the first quarter after the claim until the last quarter. Quarters of nonemployment is defined as the number of consecutive post-claim quarters without covered earnings. Std. deviation of earnings is defined as the standard deviation of earnings in the nine years following the claim.

Table 6

Effects of Exception Reporting on Selected Outcomes by Reason for Job Loss as Deviations from the Comparison Group

	Reason for Job Loss				
Outcome	Quit	Permanent job loss	Temporary layoff	Contract/Seasonal layoff	Unknown
UI Receipt Outcomes					
Benefit year outcomes					
Total UI benefits paid ($)	456***	491***	416***	518***	485***
	(117)	(118)	(102)	(120)	(93)
Weeks of UI payments	3.99***	3.64***	2.93***	3.33***	3.18***
	(0.81)	(0.82)	(0.65)	(0.77)	(0.61)
Exhausted benefits (proportion)	0.117***	0.157***	0.108***	0.096***	0.123***
	(0.038)	(0.039)	(0.025)	(0.036)	(0.027)
Employment Outcomes					
Year 0 outcomes					
Employed (proportion)	-0.048	-0.072**	-0.011	-0.006	0.004
	(0.030)	(0.028)	(0.016)	(0.022)	(0.021)
Hours worked	-56.9	-105.2*	14.5	-34.5	-25.6
	(60.0)	(58.9)	(41.3)	(53.0)	(42.4)
Total Earnings ($)	-450	-1,701**	-419	-27	408
	(583)	(663)	(526)	(709)	(477)
Year 1 outcomes					
Employed (proportion)	-0.036	-0.050	0.004	0.007	0.018
	(0.031)	(0.031)	(0.020)	(0.028)	(0.023)
Hours worked	-45.4	-36.8	56.5	-11.2	16.7
	(65.0)	(65.4)	(51.1)	(63.5)	(50.4)
Total Earnings ($)	-499	-673	46	1,066	1,264**
	(699)	(778)	(626)	(822)	(595)
Other post-claim outcomes					
Number of post-claim employers	-0.14	0.08	-0.40*	0.18	-0.23
	(0.27)	(0.30)	(0.21)	(0.36)	(0.21)
Quarters of nonemployment	0.55	1.44**	0.29	0.22	-0.99*
	(0.76)	(0.71)	(0.39)	(0.59)	(0.51)
Std. deviation earnings ($)	-248	-7	-125	-351	327
	(276)	(293)	(235)	(311)	(229)
Quarters with first post-claim employer	-0.83	-1.65**	1.21	0.49	0.88
	(0.70)	(0.77)	(0.75)	(0.91)	(0.68)
Returned to same employer (proportion)	0.001	-0.017	0.044	0.058	0.047*
	(0.028)	(0.027)	(0.031)	(0.039)	(0.027)
Returned to same industry (proportion)	-0.027	-0.044	0.045	0.070*	0.038
	(0.034)	(0.035)	(0.030)	(0.041)	(0.028)
Sample size	729	679	1,052	653	1,105

Notes: See annotations to Table 3.

Table 7
Mean and Standard Deviation (in parentheses) of Outcomes of the Comparison Group by Reason for Job Loss

	Reason for Job Loss				
Outcome	Quit	Permanent job loss	Temporary layoff	Contract/Seasonal layoff	Unknown
UI Receipt Outcomes					
Benefit year outcomes					
Total UI benefits paid ($)	1,857	2,322	1,685	2,244	1,867
	(1,763)	(1,813)	(1,580)	(1,751)	(1,710)
Weeks of UI payments	14.2	16.9	12.1	15.4	13.7
	(11.2)	(11.2)	(11.2)	(11.2)	(11.2)
Exhausted benefits (proportion)	0.273	0.316	0.136	0.236	0.235
	(0.446)	(0.466)	(0.343)	(0.425)	(0.424)
Employment Outcomes					
Year 0 outcomes					
Employed (proportion)	0.849	0.868	0.944	0.924	0.862
	(0.358)	(0.338)	(0.23)	(0.266)	(0.345)
Hours worked	860	920	1,228	978	1,005
	(779)	(738)	(710)	(690)	(751)
Total Earnings ($)	8,083	9,651	15,300	13,437	10,721
	(8,514)	(9,586)	(10,954)	(10,698)	(9,071)
Year 1 outcomes					
Employed (proportion)	0.806	0.829	0.885	0.874	0.819
	(0.396)	(0.377)	(0.319)	(0.332)	(0.385)
Hours worked	1,005	1,117	1,293	1,099	1,103
	(848)	(819)	(847)	(764)	(857)
Total Earnings ($)	10,085	11,715	16,185	14,949	12,118
	(9,694)	(10,482)	(12,288)	(11,638)	(10,462)
Other post-claim outcomes					
Number of post-claim employers	4.59	4.52	4.51	5.62	4.43
	(3.5)	(3.69)	(3.5)	(4.4)	(3.77)
Quarters of nonemployment	4.60	4.20	2.21	2.85	4.38
	(9.42)	(9.15)	(5.49)	(6.99)	(9.63)
Std. deviation earnings ($)	5,229	5,640	6,212	6,668	5,576
	(3,754)	(4,153)	(3,842)	(4,161)	(3,873)
Quarters with first post-claim employer	6.39	7.40	9.31	8.13	8.32
	(9.33)	(10.7)	(11.52)	(10.77)	(11.12)
Returned to same employer (proportion)	0.151	0.127	0.52	0.347	0.359
	(0.358)	(0.333)	(0.5)	(0.477)	(0.48)
Returned to same industry (proportion)	0.286	0.273	0.616	0.48	0.463
	(0.452)	(0.446)	(0.487)	(0.5)	(0.499)
Sample size	729	679	1,052	653	1,105

Notes: Universe consists of control and new work search groups during fall 1986, winter 1987, and spring 1987.

Table 8

Effect of Exception Reporting on Long-Term Employment, Hours, and Earnings for Claimants with a Higher than Average Predicted Likelihood of Benefit Exhaustion

Outcome	Proportion employed				Hours worked				Total Earnings ($)			
	Comparison group		Exception Reporting		Comparison group		Exception Reporting		Comparison group		Exception Reporting	
	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)
Year relative to claim												
Year 0	0.834	(0.372)	-0.044**	(0.020)	805	(744)	-45.1	(37.7)	8838	(8807)	-452	(453)
Year 1	0.794	(0.405)	-0.013	(0.021)	964	(832)	6.3	(43.6)	10688	(10086)	314	(555)
Year 2	0.759	(0.428)	-0.020	(0.023)	956	(851)	5.1	(44.4)	11278	(11189)	-199	(588)
Year 3	0.729	(0.445)	-0.054**	(0.024)	942	(864)	18.7	(46.0)	11072	(11543)	29	(609)
Year 4	0.673	(0.469)	-0.022	(0.025)	869	(875)	36.7	(46.5)	10480	(11972)	184	(630)
Year 5	0.627	(0.484)	-0.013	(0.025)	818	(892)	45.0	(46.8)	9882	(11626)	720	(629)
Year 6	0.608	(0.489)	-0.013	(0.026)	772	(884)	44.0	(46.3)	9092	(11523)	861	(615)
Year 7	0.578	(0.494)	-0.006	(0.026)	755	(902)	37.6	(47.1)	8689	(11137)	884	(610)
Year 8	0.545	(0.498)	0.006	(0.026)	720	(885)	20.0	(46.6)	8549	(11353)	817	(627)
Year 9	0.539	(0.499)	0.004	(0.026)	757	(919)	-3.7	(48.1)	8826	(11712)	844	(665)
Sample size			1,543				1,543				1,543	

Notes: See annotations to Table 3.

The comparison group consist of claimants in the control and new work search groups with a higher than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

Table 9

Effect of Exception Reporting on Long-Term Employment, Hours, and Earnings for Claimants with a Lower than Average Predicted Likelihood of Benefit Exhaustion

Outcome	Proportion employed				Hours worked				Total Earnings ($)			
	Comparison group		Exception Reporting		Comparison group		Exception Reporting		Comparison group		Exception Reporting	
	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)
Year relative to claim												
Year 0	0.921	(0.27)	-0.007	(0.011)	1132	(721)	-27.0	(29.1)	13153	(10502)	-449	(416)
Year 1	0.87	(0.336)	-0.004	(0.014)	1228	(827)	-8.9	(33.4)	14467	(11571)	144	(472)
Year 2	0.838	(0.369)	-0.005	(0.015)	1231	(860)	-28.2	(34.3)	15237	(12224)	-288	(488)
Year 3	0.803	(0.398)	0.006	(0.016)	1194	(881)	0.5	(34.9)	15022	(12335)	-225	(498)
Year 4	0.779	(0.415)	-0.006	(0.017)	1131	(879)	-14.5	(34.7)	14702	(12865)	-292	(507)
Year 5	0.761	(0.427)	-0.006	(0.017)	1089	(885)	16.5	(35.7)	14302	(13074)	48	(516)
Year 6	0.722	(0.448)	0.012	(0.018)	1035	(914)	12.2	(36.3)	13609	(13372)	-67	(515)
Year 7	0.704	(0.457)	0.003	(0.018)	1030	(918)	-8.2	(36.6)	13243	(13175)	200	(530)
Year 8	0.687	(0.464)	-0.013	(0.019)	985	(918)	-5.5	(37.0)	13042	(13445)	181	(550)
Year 9	0.671	(0.47)	-0.009	(0.019)	1000	(945)	-2.0	(37.5)	13413	(14182)	-37	(562)
Sample size			2,675				2,675				2,675	

Notes: Same as above.

The comparison group consist of claimants in the control and new work search groups with a lower than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

Table 10

Effect of Exception Reporting on Long-Term Employment Outcomes by Varying Likelihood of Exhausting Benefits

Sample	Probability of exhausting benefits higher than average[a]				Probability of exhausting benefits lower than average[b]			
	Comparison group		Exception Reporting		Comparison group		Exception Reporting	
	Mean	(Std. dev.)	Coeff.	(Std. error)	Mean	(Std. dev.)	Coeff.	(Std. error)
Outcome								
Number of post-claim employers	4.67	(4.00)	-0.31	(0.20)	4.68	(3.64)	-0.09	(0.15)
Quarters of nonemployment	5.17	(10.46)	0.05	(0.52)	2.78	(6.75)	0.18	(0.28)
Std. deviation of post-claim earnings ($)	5,491	(4,139)	-70	(209)	6,049	(3,853)	-26	(155)
Quarters with first post-claim employer	6.82	(9.76)	-0.23	(0.50)	8.72	(11.35)	0.3	(0.46)
Returned to same employer (proportion)	0.215	(0.411)	0.001	(0.021)	0.381	(0.486)	0.038*	(0.020)
Returned to same industry (proportion)	0.328	(0.47)	-0.009	(0.024)	0.503	(0.5)	0.024	(0.020)
Sample size			1,543				2,675	

Notes: See annotations to Table 3.

a. The comparison group consist of claimants in the control and new work search groups with a higher than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

b.The comparison group consist of claimants in the control and new work search groups with a lower than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

Figure 1: Distribution of the number of post-claim employers

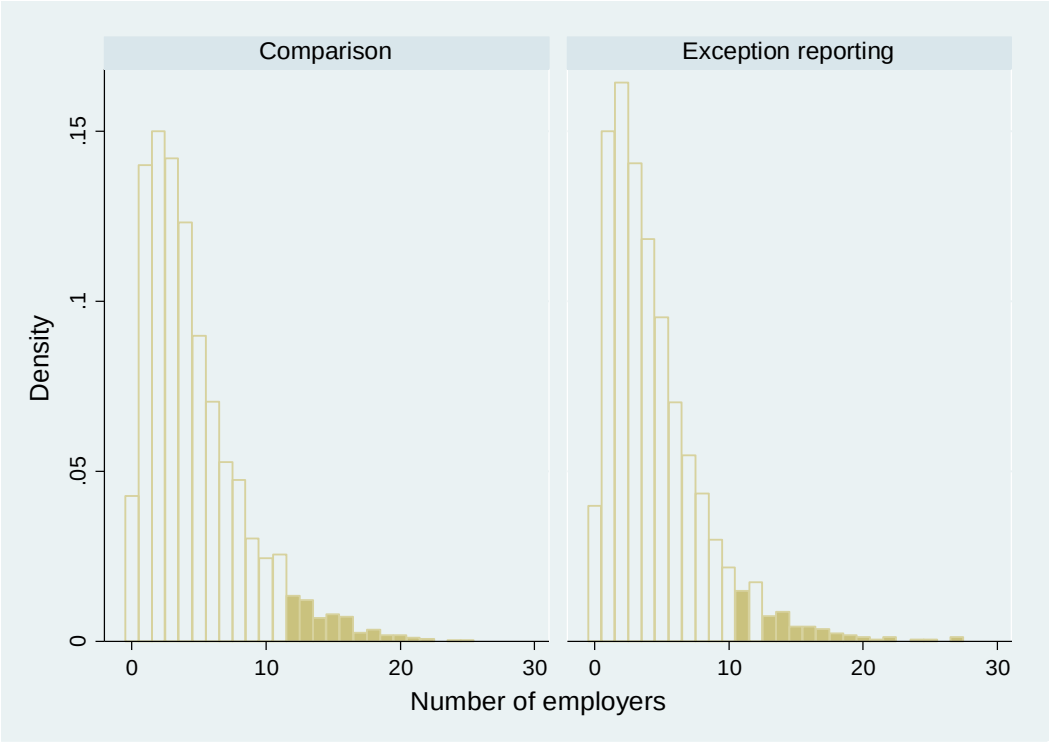


Figure 2: Distribution of quarters of nonemployment

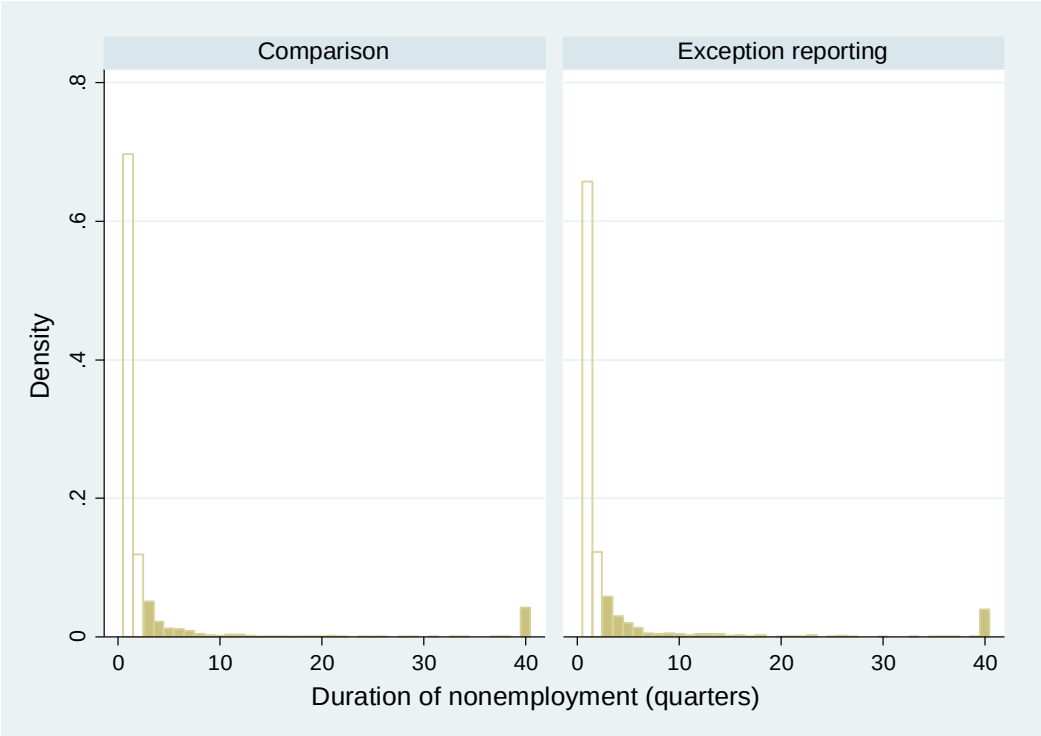
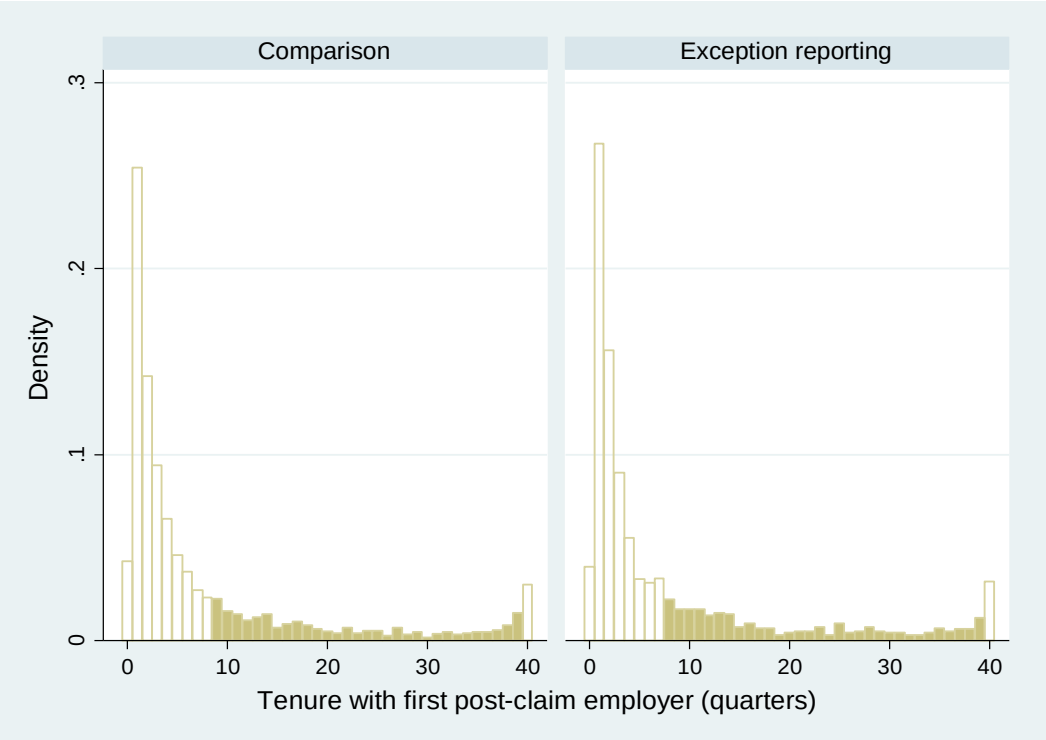


Figure 3: Distribution of the number of quarters with first post-claim employer



Results Appendix

Table A1

Effect of New Work Search on Selected Outcomes as Deviations from the Control Group

Outcome	Control group Mean (Std. dev.)		New Work Search Coefficient (Std.error)	
UI Receipt Outcomes				
Benefit year				
Total UI benefits paid ($)	$1,978	(1731)	-50	(60)
Weeks of UI payments	14.13	(10.40)	0.07	(0.40)
Exhausted benefits (proportion)	0.227	(0.419)	0.007	(0.016)
First spell				
Total benefits in first UI spell ($)	$1,654	(1642)	-34	(59)
Weeks of first UI spell	13.48	(10.46)	0.15	(0.41)
Employment Outcomes				
First quarter outcomes[a]				
Employed (proportion)	0.692	(0.462)	0.005	(0.018)
Hours worked	199.8	(212)	1.2	(7.9)
Total Earnings ($)	$2,285	(2,786)	-16	(88)
Year 0 outcomes[a]				
Employed (proportion)	0.888	(0.316)	0.008	(0.012)
Hours worked	1014	(751)	5.3	(27.3)
Total Earnings ($)	$11,701	(10,400)	-135	(314)
Year 1 outcomes[a]				
Employed (proportion)	0.847	(0.360)	-0.005	(0.014)
Hours worked	1138	(823)	-7.7	(31.7)
Total Earnings ($)	$13,304	(11,315)	-326	(373)
Other outcomes				
Returned to same employer (proportion)	0.316	(0.465)	0.008	(0.017)
Returned to same industry (proportion)	0.428	(0.495)	0.021	(0.018)
Sample size			2,612	

Notes: Universe consists of new work search and control groups during fall 1986, winter 1987, and spring 1987. Robust standard errors are in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression-adjusted differences control for all variables displayed in Table 2 plus the quarter in which the claim was filed and the unemployment rate in the county and month in which the claim was filed.

[a] First quarter is the quarter following the claim quarter. Year 0 is defined as the sum of the first, second, third, and fourth quarter after the claim quarter. Year 1 is defined as the sum of the fifth, sixth, seventh, and eighth quarter after the claim quarter. Earnings are expressed in 1988:4 prices.

Table A2

Effect of Exception Reporting on Selected Outcomes as Deviations from the Control Group

Outcome	Control group Mean (Std. dev.)		Exception Reporting Coefficient (Std.error)	
UI Receipt Outcomes				
Benefit year				
Total UI benefits paid ($)	$1,978	(1731)	429***	(54)
Weeks of UI payments	14.13	(10.40)	3.27***	(0.35)
Exhausted benefits (proportion)	0.227	(0.419)	0.116***	(0.015)
First spell				
Total benefits in first UI spell ($)	$1,654	(1642)	424***	(54)
Weeks of first UI spell	13.48	(10.46)	3.47***	(0.37)
Employment Outcomes				
First quarter outcomes[a]				
Employed (proportion)	0.692	(0.462)	-0.027*	(0.016)
Hours worked	199.8	(212)	-6.0	(7.0)
Total Earnings ($)	$2,285	(2,786)	-84	(82)
Year 0 outcomes[a]				
Employed (proportion)	0.888	(0.316)	-0.015	(0.011)
Hours worked	1014	(751)	-23.1	(24.7)
Total Earnings ($)	$11,701	(10,400)	-315	(290)
Year 1 outcomes[a]				
Employed (proportion)	0.847	(0.360)	-0.006	(0.013)
Hours worked	1138	(823)	-0.7	(28.4)
Total Earnings ($)	$13,304	(11,315)	186	(342)
Other outcomes				
Returned to same employer (proportion)	0.316	(0.465)	0.035**	(0.015)
Returned to same industry (proportion)	0.428	(0.495)	0.031*	(0.016)
Sample size			3,145	

Notes: Universe consists of exception reporting and control groups during fall 1986, winter 1987, and spring 1987. Robust standard errors are in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Regression-adjusted differences control for all variables displayed in Table 2 plus the quarter in which the claim was filed and the unemployment rate in the county and month in which the claim was filed.

[a] First quarter is the quarter following the claim quarter. Year 0 is defined as the sum of the first, second, third, and fourth quarter after the claim quarter. Year 1 is defined as the sum of the fifth, sixth, seventh, and eighth quarter after the claim quarter. Earnings are expressed in 1988:4 prices.

Table A3

Effect of Exception Reporting on Long-Term Employment, Hours, and Earnings as Deviations from the Control Group

Outcome	Proportion employed				Hours worked				Total Earnings ($)			
	Control group		Exception Reporting		Control group		Exception Reporting		Control group		Exception Reporting	
	Mean (Std. dev.)		Coeff.	(Std. error)	Mean (Std. dev.)		Coeff.	(Std. error)	Mean (Std. dev.)		Coeff.	(Std. error)
Year relative to claim												
Year 0	0.888	(0.316)	-0.015	(0.011)	1,014	(751)	-23.1	(24.7)	11,701	(10,400)	-315	(290)
Year 1	0.847	(0.36)	-0.006	(0.013)	1,138	(823)	-0.7	(28.4)	13,304	(11,315)	186	(342)
Year 2	0.817	(0.386)	-0.013	(0.014)	1,140	(858)	-19.8	(29.0)	14,154	(12,266)	-488	(372)
Year 3	0.781	(0.414)	-0.017	(0.015)	1,100	(881)	17.2	(30.1)	13,778	(12,416)	-209	(396)
Year 4	0.749	(0.434)	-0.014	(0.015)	1,044	(889)	8.9	(30.4)	13,503	(12,977)	-284	(413)
Year 5	0.728	(0.445)	-0.016	(0.015)	1,002	(893)	31.2	(30.7)	12,993	(12,832)	207	(414)
Year 6	0.686	(0.464)	0.006	(0.016)	953	(920)	22.3	(31.3)	12,204	(12,990)	196	(418)
Year 7	0.667	(0.471)	0	(0.016)	960	(925)	-7.5	(31.5)	12,075	(12,941)	136	(424)
Year 8	0.643	(0.479)	-0.005	(0.016)	904	(916)	5.2	(31.6)	11,800	(13,127)	220	(438)
Year 9	0.63	(0.483)	0	(0.017)	923	(943)	4.9	(32.3)	12,115	(13,817)	127	(456)
Sample size			3,145				3,145				3,145	

Notes: See annotations to Table A2.

Table A4

Effect of Exception Reporting on Selected Long-Term Employment Outcomes as Deviations from the Control Group

	Control group			Exception Reporting	
	Mean	Median	(Std. dev.)	Coeff.	(Std. error)
Outcome					
Number of post-claim employers	4.73	4	(3.79)	-0.14	(0.13)
Quarters of nonemployment	3.60	1	(8.24)	0.08	(0.28)
Std. deviation of post-claim earnings ($)	6,026	5,407	(4,074)	-214*	(130)
Quarters with first post-claim employer	7.79	3	(10.6)	0.40	(0.37)
Sample size				3,145	

Notes: See annotations to Table A2. Number of post-claim employers is defined as the number of unique employers we observe from the first quarter after the claim until the last quarter. Quarters of nonemployment is defined as the number of consecutive post-claim quarters without covered earnings. Std. deviation of earnings is defined as the standard deviation of earnings in the nine years following the claim.

Table A5

Effects of Exception Reporting on Selected Outcomes by Reason for Job Loss as Deviations from the Control Group

	Reason for Job Loss				
Outcome	Quit	Permanent job loss	Temporary layoff	Contract/ Seasonal layoff	Unknown
UI Receipt Outcomes					
Benefit year outcomes					
Total UI benefits paid ($)	541***	444***	307***	472***	510***
	(133)	(134)	(116)	(142)	(104)
Weeks of UI payments	4.66***	3.46***	2.14***	3.35***	3.38***
	(0.92)	(0.92)	(0.74)	(0.91)	(0.68)
Exhausted benefits (proportion)	0.142***	0.156***	0.086***	0.088**	0.134***
	(0.042)	(0.044)	(0.028)	(0.042)	(0.030)
Employment Outcomes					
Year 0 outcomes					
Employed (proportion)	-0.036	-0.058*	-0.002	-0.006	0.004
	(0.034)	(0.031)	(0.018)	(0.024)	(0.024)
Hours worked	-91.7	-94.0	58.0	-9.6	-14.7
	(70.1)	(68.0)	(46.9)	(58.6)	(48.7)
Total Earnings ($)	-597	-1,631**	-17	91	123
	(673)	(758)	(591)	(875)	(545)
Year 1 outcomes					
Employed (proportion)	-0.028	-0.059*	0.016	0.001	0.005
	(0.037)	(0.034)	(0.024)	(0.032)	(0.025)
Hours worked	-36.6	-36.3	61.3	-29.9	7.1
	(73.8)	(72.4)	(58.9)	(70.7)	(57.7)
Total Earnings ($)	-569	-655	271	635	935
	(827)	(872)	(706)	(937)	(675)
Other post-claim outcomes					
Number of post-claim employers	-0.07	0.07	-0.32	-0.19	-0.15
	(0.30)	(0.34)	(0.25)	(0.42)	(0.24)
Quarters of nonemployment	0.11	1.04	0.36	0.32	-0.69
	(0.89)	(0.78)	(0.39)	(0.61)	(0.58)
Std. deviation earnings ($)	-217	-135	-13	-644*	53
	(333)	(320)	(268)	(353)	(254)
Quarters with first post-claim employer	-0.09	-1.51*	1.33	0.58	0.99
	(0.74)	(0.87)	(0.87)	(1.00)	(0.79)
Returned to same employer (proportion)	0.021	-0.013	0.056	0.050	0.062**
	(0.033)	(0.030)	(0.036)	(0.045)	(0.032)
Returned to same industry (proportion)	0.017	-0.013	0.058*	0.081*	0.035
	(0.038)	(0.039)	(0.035)	(0.048)	(0.032)
Sample size	550	511	780	486	818

Notes: See annotations to Table A2.

Table A6

Mean and Standard Deviation (in parentheses) of Outcomes of the Control Group by Reason for Job Loss

Outcome	Reason for Job Loss				
	Quit	Permanent job loss	Temporary layoff	Contract/ Seasonal layoff	Unknown
UI Receipt Outcomes					
Benefit year outcomes					
Total UI benefits paid ($)	1,794	2,376	1,767	2,283	1,846
	(1,734)	(1,822)	(1,592)	(1,853)	(1,649)
Weeks of UI payments	13.4	16.9	12.6	15.3	13.5
	(10.8)	(10.7)	(9.7)	(10.)	(10.4)
Exhausted benefits (proportion)	0.253	0.313	0.149	0.226	0.224
	(0.435)	(0.465)	(0.357)	(0.419)	(0.417)
Employment Outcomes					
Year 0 outcomes					
Employed (proportion)	0.834	0.868	0.941	0.925	0.867
	(0.373)	(0.339)	(0.236)	(0.264)	(0.339)
Hours worked	889	923	1,211	969	1,008
	(826)	(743)	(706)	(685)	(749)
Total Earnings ($)	8,180	9,846	15,223	13,571	11,043
	(8,783)	(9,779)	(11,568)	(10,993)	(9,192)
Year 1 outcomes					
Employed (proportion)	0.796	0.845	0.882	0.874	0.834
	(0.404)	(0.362)	(0.323)	(0.332)	(0.373)
Hours worked	979	1,120	1,309	1,115	1,118
	(835)	(793)	(834)	(752)	(842)
Total Earnings ($)	10,032	11,924	16,215	15,567	12,480
	(9,757)	(10,566)	(12,654)	(11,881)	(10,329)
Other post-claim outcomes					
Number of post-claim employers	4.63	4.63	4.50	5.93	4.36
	(3.57)	(3.6)	(3.53)	(4.46)	(3.74)
Quarters of nonemployment	5.07	4.19	2.05	2.78	4.07
	(10.24)	(9.09)	(4.69)	(6.75)	(9.12)
Std. deviation earnings ($)	5,217	5,985	6,228	7,067	5,794
	(3,920)	(4,451)	(3,861)	(4,278)	(3,849)
Quarters with first post-claim employer	5.84	7.06	9.30	7.77	8.24
	(9.09)	(10.24)	(11.62)	(10.16)	(10.89)
Returned to same employer (proportion)	0.136	0.125	0.51	0.347	0.369
	(0.343)	(0.331)	(0.501)	(0.477)	(0.483)
Returned to same industry (proportion)	0.26	0.242	0.606	0.456	0.487
	(0.44)	(0.429)	(0.489)	(0.499)	(0.5)
Sample size	550	511	780	486	818

Notes: Universe consists of the control group during fall 1986, winter 1987, and spring 1987.

Table A7

Effects of Exception Reporting as Deviations from the Control Group on Long-Term Employment, Hours, and Earnings for Claimants with a Higher than Average Predicted Likelihood of Benefit Exhaustion

Outcome	Proportion employed				Hours worked				Total Earnings ($)			
	Comparison group		Exception Reporting		Comparison group		Exception Reporting		Comparison group		Exception Reporting	
	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)
Year relative to claim												
Year 0	0.828	(0.378)	-0.038	(0.023)	784.9	(754)	-25.9	(43.1)	8723	(8684)	-344	(509)
Year 1	0.788	(0.409)	-0.008	(0.024)	958	(830)	12.7	(49.0)	10636	(10077)	369	(619)
Year 2	0.758	(0.429)	-0.019	(0.025)	945.2	(837)	16.2	(49.6)	11432	(11487)	-355	(671)
Year 3	0.722	(0.449)	-0.046*	(0.027)	914.1	(876)	46.8	(52.0)	11110	(12146)	-23	(704)
Year 4	0.68	(0.467)	-0.030	(0.028)	855.6	(875)	49.5	(52.2)	10524	(11958)	133	(709)
Year 5	0.642	(0.48)	-0.028	(0.028)	825.5	(894)	36.1	(52.7)	10058	(11560)	531	(702)
Year 6	0.599	(0.491)	-0.004	(0.029)	763.5	(898)	51	(52.6)	8903	(11460)	1,032	(688)
Year 7	0.577	(0.495)	-0.004	(0.029)	759.7	(904)	32.8	(53.1)	8706	(11176)	857	(682)
Year 8	0.537	(0.499)	0.013	(0.029)	701.5	(878)	38	(52.2)	8421	(11168)	929	(694)
Year 9	0.537	(0.499)	0.006	(0.029)	766.2	(929)	-13.2	(54.4)	8853	(11680)	820	(736)
Sample size			1,166				1,166				1,166	

Notes: See annotations to Table A2.

The control group consists of claimants in the control and new work search groups with a higher than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

Table A8

Effecst of Exception Reporting as Deviations from the Control Group on Long-Term Employment, Hours, and Earnings for Claimants with a Lower than Average Predicted Likelihood of Benefit Exhaustion

Outcome	Proportion employed				Hours worked				Total Earnings ($)			
	Comparison group		Exception Reporting		Comparison group		Exception Reporting		Comparison group		Exception Reporting	
	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)	Mean (Std.dev.)		Coeff.	(Std. error)
Year relative to claim												
Year 0	0.921	(0.27)	-0.006	(0.012)	1142	(718)	-34.8	(32.6)	13371	(10902)	-636	(478)
Year 1	0.879	(0.326)	-0.013	(0.015)	1239	(802)	-17.5	(36.8)	14800	(11695)	-149	(530)
Year 2	0.851	(0.356)	-0.017	(0.016)	1249	(851)	-43.9	(38.3)	15681	(12430)	-692	(553)
Year 3	0.814	(0.389)	-0.005	(0.018)	1204	(868)	-5.7	(38.9)	15275	(12319)	-422	(556)
Year 4	0.787	(0.41)	-0.013	(0.019)	1149	(879)	-29.2	(39.0)	15173	(13230)	-712	(578)
Year 5	0.776	(0.417)	-0.02	(0.019)	1101	(877)	8	(39.8)	14639	(13217)	-225	(583)
Year 6	0.734	(0.442)	0.001	(0.020)	1059	(916)	-8	(40.9)	14055	(13428)	-463	(584)
Year 7	0.718	(0.45)	-0.01	(0.020)	1073	(919)	-47.9	(41.1)	13964	(13474)	-477	(600)
Year 8	0.703	(0.457)	-0.027	(0.021)	1018	(917)	-34.8	(41.5)	13695	(13754)	-426	(621)
Year 9	0.682	(0.466)	-0.018	(0.021)	1010	(940)	-9.7	(42.1)	13944	(14572)	-532	(642)
Sample size			1,979				1,979				1,979	

Notes: Same as above.

The control group consists of claimants in the control and new work search groups with a lower than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

Table A9

Effects of Exception Reporting as Deviations from the Control Group on Long-Term Employment Outcomes by Varying Likelihood of Exhausting Benefits

Sample	Probability of exhausting benefits higher than average[a]				Probability of exhausting benefits lower than average[b]			
	Control group		Exception Reporting		Control group		Exception Reporting	
	Mean	(Std. dev.)	Coeff.	(Std. error)	Mean	(Std. dev.)	Coeff.	(Std. error)
Outcome								
Number of post-claim employers	4.81	(4.03)	-0.44*	(0.23)	4.69	(3.66)	-0.1	(0.17)
Quarters of nonemployment	5.20	(10.45)	0.01	(0.59)	2.70	(6.52)	0.24	(0.30)
Std. deviation of post-claim earnings ($)	5,676	(4,317)	-254	(243)	6,221	(3,920)	-188	(175)
Quarters with first post-claim employer	6.32	(9.3)	0.25	(0.55)	8.62	(11.19)	0.43	(0.51)
Returned to same employer (proportion)	0.204	(0.404)	0.012	(0.024)	0.378	(0.485)	0.043*	(0.022)
Returned to same industry (proportion)	0.298	(0.458)	0.02	(0.027)	0.501	(0.5)	0.028	(0.022)
Sample size			1,166				1,979	

Notes: See annotations to Table A2.

a. The comparison group consist of claimants in the control group with a higher than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

b.The comparison group consist of claimants in the control group with a lower than average probability of exhausting benefits. The average probability of exhausting benefits equals 0.264.

Table A10

Estimated probability of exhausting benefits (probit coefficients). Dependent variable: exhausted benefits

Covariates	Probability of exhausting benefits
Mean of dependent variable	0.264
Male	0.087
	(0.064)
Race/Etnicity	
Black non-Hispanic	0.134*
	(0.070)
Hispanic/Am.Indian/Asian/other	0.137*
	(0.076)
Age	
25-34	0.184***
	(0.061)
35-44	0.356***
	(0.069)
45-54	0.545***
	(0.083)
≥ 54	0.620***
	(0.101)
Schooling	
high school	-0.101
	(0.062)
some college	-0.156**
	(0.071)
college graduate	-0.380***
	(0.098)
Pre-claim earnings ($)	
Earnings 1 year before claim	0.000
	(0.000)
Earnings 2 years before claim	0.000
	(0.000)
Earnings 3 years before claim	-0.000*
	(0.000)
Pre-claim hours	
Hours 1 year before claim	-0.000**
	(0.000)
Hours 2 years before claim	-0.000
	(0.000)
Hours 3 years before claim	0.000
	(0.000)
Occupation dummies?	Yes

Covariates	Probability of exhausting benefits
Industry dummies?	Yes
Reason for job loss	
permanent layoff	0.189***
	(0.064)
temporary layoff with recall date	-0.238***
	(0.068)
contract/seasonal	0.030
	(0.072)
Employer-attached/placed by union	-0.348***
	(0.056)
Prior UI claim	
duration ≤ 15 weeks	-0.418***
	(0.086)
duration > 15 weeks	0.155**
	(0.075)
UI benefits/claim type	
weekly amount ($)	0.008***
	(0.002)
maximum amount ($)	-0.000*
	(0.000)
potential duration	-0.011
	(0.014)
replacement rate (percent)	0.002
	(0.002)
combined wage claim	-0.098
	(0.097)
ex-service member claim	-0.053
	(0.180)
federal employee claim	0.353**
	(0.164)
Veteran	-0.093*
	(0.053)
Reservation wage (hourly)	
≤ $5.00	0.030
	(0.076)
$5.01–$7.00	0.180**
	(0.078)
$7.01–$10.00	0.114
	(0.078)
$10.01–$20.00	0.040
	(0.091)
> $20.00	0.184**
	(0.088)
Marital status/gender	
married male	-0.234***
	(0.067)

Covariates	Probability of exhausting benefits
married female	0.177**
	(0.085)
Household status	
1 dependent	-0.033
	(0.064)
2 or more dependents	0.002
	(0.063)
homeowner	-0.059
	(0.061)
unemployment rate in the area	-0.143**
	(0.062)
Quarter of claim	
Summer 1986	0.333***
	(0.076)
Fall 1986	0.173**
	(0.087)
Winter 1987	0.192
	(0.152)
Spring 1987	0.061
	(0.092)
Constant	-0.079
	(0.599)
Sample size	4,811

Notes: Standard errors are in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Data Appendix

This appendix describes the original data from the Washington Alternative Work Search (WAWS) experiment (Johnson and Klepinger 1991, 1994) and how we constructed the long-term panel used to examine the long-term effects of the WAWS treatments on labor market outcomes (sections 1 and 2). We also provide a discussion of the extent and severity of attrition from the long-term sample (section 3).

1. Description of the Experimental Data

The data we use to replicate and extend the short-term findings of the Washington experiment were provided to the W.E. Upjohn Institute by the Washington State Employment Security Department (ESD) in June 1989. According to ESD, they are the same data provided to Terry Johnson and Daniel Klepinger, who performed the evaluation at Battelle Memorial Institute for the U.S. Department of Labor (Johnson and Klepinger 1991, 1994); however, they do not include responses to the survey referred to by Johnson and Klepinger (1991). We refer to these as "the experimental data."

The experimental data combine data from three sources:

- UI claims records — the records maintained by the ESD to track each worker's benefit eligibility and the timing and amount of benefits paid in each week of the benefit year
- Administrative wage records — the quarterly records provided by each covered employer in the state to the ESD to determine the employer's UI payroll tax liability and to track the wages and hours of each employee (which in turn are used to determine UI eligibility and the weekly benefit amount)

- Employment Service records — the records maintained by the state Employment Service on each job seeker, which include a range of personal characteristics of claimants that are inessential to running the UI system but useful in referring and placing workers in jobs

For example, data on UI benefits, including weeks of benefits paid and whether a worker exhausted benefits, are from UI claims records; quarterly earnings, hours, and data on industry of employment and whether a worker returned to the pre-UI employer are from administrative wage records; and individual characteristics (including occupation) and whether a worker was on standby are from the Employment Service records.

Details of the assignment protocol

The experimental data do not include an explicit treatment indicator, so we created one using the last digit of each claimant's pseudo-Social Security number and the date of his or her enrollment in the experiment. (The last number of the pseudo-SSN was not altered from the last number of the true SSN.) Initially, the Tacoma UI office assigned claimants with SSN ending in 0, 1, or 2 to the Exception Reporting group; 3, 4, or 5 to the control group; 6 or 7 to the New Work Search (NWS) policy group; and 8 or 9 to the Job Search Assistance (JSA) group (Johnson and Klepinger refer to this as the Intensive Services group). However, starting with the week of May 10, 1987, ESD stopped enrolling claimants in the Exception Reporting (ER) treatment because it was clear that claimants assigned to this group were claiming more benefits and experiencing longer spells of insured unemployment than the control group. Thereafter, the Tacoma office assigned claimants with SSN ending in 0, 1, or 2 to JSA. Accordingly, we assigned claimants with SSN = 0, 1, or 2 and benefit year starting in week 27 of 1986 through week 18 of 1987 (inclusive) to the ER group, and claimants with SSN = 0, 1, or 2 and benefit

year starting in week 19 of 1987 through week 35 of 1987 (inclusive) to the JSA treatment group (as well as all claimants with SSN = 8 or 9).

Because the assignment procedure changed in May 1987, the number of claimants in each of the treatment groups does not reflect the 30-30-20-20 proportions that would be suggested by the original assignment protocol. Rather, 30 percent of all assigned claimants are controls, 23 percent (rather than 30 percent) are in the ER treatment, 21 percent are in the New Work Search Policy treatment, and 27 percent (rather than 20 percent) are in the JSA treatment.[1]

Replicating Johnson and Klepinger's sample

Claimants who failed the separation requirement for UI need to be dropped from the sample, but the experimental data do not include a binary indicator of whether the claimant satisfied that requirement. The data do include two "separation issue" variables, one for the total number of separation issues raised during the benefit year, the other a set of weekly variables indicating whether a separation issue was raised in that week. The two are inconsistent, and our solution was to drop all workers with at least one separation issue during the benefit year and zero benefits paid, on the assumption that they were ineligible due to a separation issue.

For our project studying the long-term effects of assignment to intensive services (JSA), we also dropped four claimants who were assigned to the control group but who received intensive services. Presumably, these claimants happened to be at the Tacoma Employment Service Center when an intensive services workshop was conducted and were either directed to attend (mistakenly) or attended voluntarily. This "cross-over" of four claimants from the control group to the JSA treatment group represents an assignment error rate of about 0.15 percent (2,597 controls received at least one week of benefits and did not attend the intensive services

1. These proportions are with respect to the full sample of 9,607, which is the number of benefit years in the experimental data before dropping any claimants who never received a benefit check.

workshop). We do not drop these claimants from the study of the effects of ER on long-term outcomes.

The sample we use to study the effects of ER on long-term outcomes is smaller than the sample used by Johnson and Klepinger (1991, 1994). This is for two reasons. First, note that the ER treatment consisted of four experimental cohorts, with the first cohort assigned in week 27 of 1986:III and the last cohort assigned in week 18 of 1987:II. In order to make sure that we observe the ER, control, and new work search policy group over the same time period, we only keep claimants who claimed between week 27 of 1986 (inclusive) and week 18 of 1987 (inclusive). Second, because the follow-up administrative wage records available to us begin in 1987:I, for the first experimental cohort we do not have follow-up records for the first post-claim quarter (denoted Q(t+1)), 1986:IV; see the table below.

Structure of the long-term panel

Experimental cohort	Claim Quarter, Q(t)	Q(t+1)	Q(t+2)	Q(t+3)		Q(t+42)
1	1986:Q3	N/A	1987:Q1	1987:Q2	…	1997:Q1
2	1986:Q4	1987:Q1	1987:Q2	1987:Q3	…	1997:Q2
3	1987:Q1	1987:Q2	1987:Q3	1987:Q4	…	1997:Q3
4	1987:Q2	1987:Q3	1987:Q4	1988:Q1	…	1997:Q4

Because we are missing administrative records for the first post-claim quarter for experimental cohort 1 (see the table), we *exclude* all the claimants from this cohort. Hence, our final analysis sample consists of claimants assigned to ER, control, and NWS group during who claimed between 1986:IV (specifically, claimants who claimed beginning with week 40 of 1986) and 1987:II (specifically, claimants who claimed in week 18 of 1987).

Because the administrative wage records all end in 1997:IV and the last quarter the ER treatment was conducted in 1987:II, we have at least 42 quarters of post-claim quarter observations; see the table above. In order to have nine complete years of follow-up administrative wage records for all claimants in the analysis sample, we decided to study long-term outcomes for 40 quarters after the claim quarter.

Although the sample we use is not identical to that used by Johnson and Klepinger (1991, 1994), both our short-term treatment effects and theirs are essentially similar, as can be seen by comparing Tables 2 and 3 in our ER paper with Tables 1, 2, and 3 in Johnson and Klepinger (1994).

Claimants with two benefit years

Each observation in the experimental data represents a UI claimant's "benefit year"—that is, the year following the initial UI claim, during which the claimant could receive UI benefits. For 9,207 initial UI claimants, one observation (that is, one benefit year) appears in the experimental data; however, for 200 claimants, two observations appear. These latter claimants each started a benefit year in the early months of the WAWS experiment (in July or August 1986). This benefit year lapsed a year later, and these 200 claimants established a new benefit year in July or August 1997, while the Washington experiment was still in progress. As a result, they were again assigned to a treatment or to the control group and are included in our sample.

Benefit duration

Another key variable we constructed is the *duration of the first UI claim spell*. The experimental data include a variable for the amount of UI benefit paid in each of the 52 weeks of the benefit year, so we followed each claimant's benefit payments from the time of the initial claim until the series lapsed for at least four weeks. This follows the procedure used by

Spiegelman, O'Leary and Kline (1992, 1995) and Corson, Decker, Dunstan, and Kerachsky (1992) in a similar setting, although appears to differ from Johnson and Klepinger (1991, 1994), who ended the first spell with a lapse of one week. Inspection of the data shows that many claimants with a lapse of three or fewer weeks had a subsequent claim series of at least another four weeks, which suggested a single spell of insured unemployment briefly interrupted by time out of the labor force, rather than reemployment. We have checked the sensitivity of the findings and hazards we report to different definitions of a first spell (a one-week lapse, two-weeks, and so on) and obtain similar findings in each case.

We counted claimants who were eligible for benefits, but who never received a payment, as having a one-week UI claim spell (corresponding to the waiting week). There are 726 such claimants. Another 1,830 claimants did not receive benefits within two weeks of their initial claim, but did receive benefits later in the benefit year. For these, we started the first UI spell in the week before the first payment (to account for the waiting week) and again followed the benefit payments until the series lapsed for at least four weeks.

For 7,051 claimants (73.4 percent of those in the experiment), the first spell of UI accounts for all benefits paid in the benefit year. The rest received additional benefits later in the benefit year. We refer to the total number of weeks in which a claimant received benefits as the "compressed UI spell." Although useful as a measure of the total effect of a treatment, a compressed spell does not correspond to a true duration because it combines spells of benefit receipt separated by at least four weeks. Accordingly, the hazard functions we display (and our interpretation of the treatment effects on search behavior) focus on the duration of the first claim spell.

Reason for job loss

Reason for job loss is another key claim characteristic. To construct this variable we create five mutually exclusive categories by using "unemployed due to lack of work" and "reason for lack of work" indicators from the Employment Service records. We define "unemployed due to permanent job loss" if a claimant was either laid off permanently because of a plant or company closure; "unemployed due to temporary layoff" if a claimant was laid off temporarily either with a known recall date or without a recall date; "unemployed due to contract completion or seasonal layoff" if a claimant was laid off because of job or contract completion or because of a seasonal layoff; "unemployment not due to job loss," which we call "quit" for short, if the reason for lack of work was missing and the claimant was unemployed not due to lack of work. Note that, typically, if a worker voluntarily quit his or her job, he or she is not eligible for UI benefits. Such a worker might still receive benefits, if the reason for the quit meets the "standard for good cause," such as showing that the claimants left due to a hostile work environment (e.g., because of discrimination or sexual harassment). Finally, we define "reason for unemployment unknown" if there is no information regarding if the claimant was unemployed due to lack of work or if the reason for lack of work is missing. This last group is the largest of the groups.

2. Constructing a Long-Term Panel

To construct a long-term panel, we appended additional administrative wage records to the experimental data described above. In Washington, wage records include the following for each worker in each quarter:

- a worker identifier (pseudo-Social Security number)
- the year and quarter

- the pseudo-employer account number and earnings received from that employer in that quarter (for each employer from whom the worker received earnings), and
- hours worked in the quarter (again, for each employer)

Coverage of the UI system is nearly universal (self-employed workers are the only significant group of "above-ground" workers who are not covered), and any UI-covered worker who has earnings in a given quarter from an employer in the state appears in the wage records. As a result, wage records can be used to construct an earnings history of most workers who were in the WAWS experiment. For this study, we had available the population of Washington administrative wage records for quarters 1987:I through 1997:IV inclusive, so we can observe up to nine years of earnings following the Washington experiment by matching workers in the experiment with their wage records.

How reliable are earnings histories constructed from wage records likely to be? Because wage records are central to financing and administering UI, most states randomly audit employer wage reports. Analyses of these audits by Blakemore et al. (1996) and Burgess, Blakemore, and Low (1998) suggest that small employers and employers with high turnover tend to underreport their workers' earnings, raising questions about the value of wage records for research. However, validation studies comparing wage records (whose source is the employer) with survey data (whose source is the worker) suggest that the reliability of wage records is similar to that of surveys. Kornfeld and Bloom (1999) performed a landmark study comparing UI wage records with survey data in a 12-state sample of over 12,000 low-wage workers who participated in the National JTPA Study. They concluded that, except for young males with past arrests, "UI wage records provide a valid alternative to surveys" for the purpose of evaluating employment and earnings outcomes of training programs (Kornfeld and Bloom 1999, p. 171). Wallace and

Haveman's (2007) validation study focused on welfare recipients in Wisconsin and found that, despite discrepancies, wage records and survey data gave similar results on employment and earnings outcomes. Wallace and Haveman conclude that, given their availability, low cost, and similarity across states, UI wage records are preferable to surveys for monitoring labor market outcomes of low-wage workers.

Table 1 shows the results of matching workers enrolled in the Washington experiment with the Washington wage records. The table shows the number of matches for the all claimants in the experiment (the "Total" column) and each treatment group (including controls) by year.[2] A "match" occurs when a claimant is observed with earnings in the Washington administrative wage records in a given year. The table also shows "match rates" (in parentheses) defined as the proportion of claimants initially enrolled in the experiment (or each treatment) observed with positive earnings (or "matched") in a given year. In the table, "year 0" refers to the claim year, defined as the quarter in which the initial UI claim was filed and the three following quarters.[3]

The Washington experiment enrolled new claimants between July 1986 and August 1987, so it should be possible to match most of the enrolled claimants with their 1987 wage records. (Claimants would not have a 1987 match if they claimed benefits in 1986 or early 1987 and never found reemployment, withdrew from the labor force, or for a few other reasons — see the following section.) As Table 1 shows, 86.5 percent of claimants enrolled in the experiment could be matched with wage records at some time during the claim year. This match rate falls unevenly to 62.5 percent in year 9, or at an average of just under 3 percentage points per year. Specifically,

2. The "Total" column gives the sum of the full control, Exception Reporting, New Work Search, and JSA groups. The full control group is used with the Exception Reporting and New Work Search treatments. The restricted control group, which drops claimants who never received a benefit, is used with the JSA treatment because only JSA assignees who received a first benefit payment received a JSA call-in notice.
3. This is the definition we use in all long-term analyses. It differs from the definition we use in the short-term analyses, where "claim year" refers to the quarter in which the initial UI claim was filed and the four following quarters.

the match rate falls by about 4 percentage points between years 0 and 1, then by 3 percentage points per year until year 5, after which it falls somewhat more slowly.

3. Attrition from the Long-term Sample

The match rates shown in Table 1 decline over time for two reasons. First, workers may remain unemployed or drop out of the labor force after claiming UI, so they will have no earnings and none will appear in the wage records. Second, workers could become self-employed, leave the formal labor force for the underground economy, or leave Washington State and find employment elsewhere.[4] In these latter cases, a worker will have earnings, but those earnings will not be recorded in the UI wage records of Washington State. (Self-employed workers are not covered by UI, earnings in the underground economy are not reported, and out-of-state earnings will be picked up in the wage records of another state.)

In the first case (continued unemployment or departure from the labor force) wage records give a correct picture of the individual's labor market status. In the second (movement to self-employment, the underground economy, or out of Washington), we have a form of sample attrition. There is no way of distinguishing between the two cases — if an individual has covered earnings in Washington, they appear (or should) in administrative wage records; otherwise, we observe a missing value for the individual in a given quarter. (In wage records, there is no difference between zero earnings and missing earnings.)

Attrition of participants from a long-term panel poses a threat to the validity of an experimental study if the subjects who leave the sample differ systematically and in unobserved ways from those who remain (see for example, the discussion and references in Murnane and

4. The WAWS experimental design attempted to lessen the problem of losing workers who move to another state by excluding interstate claims; however, this by no means eliminates the possibility.

Willett 2011). Sample attrition is usually considered as a problem of nonresponse: experiments conducted with household surveys depend on both a survey center's ability to find participants and on participants' cooperation over a long period of time (Hausman and Wise 1979, McFadden 1985). Administrative data have an advantage over survey data because they are potentially less vulnerable to attrition: anyone who receives covered earnings in given state in a given quarter should be observed in UI administrative wage records, regardless of whether that individual can be found or is willing to cooperate.[5]

Nevertheless, earnings histories constructed from UI administrative data could be subject to attrition that would bias experimental estimators. In the case of the WAWS experiment, if a treatment increased the long-term probability of becoming self-employed, moving to the underground economy, or taking a job in another state, then sample attrition from the treatment group would be greater than from the control group, the earnings of those assigned to the treatment would be understated in Washington State administrative wage records, and the estimated treatment effects estimator would be downward-biased.

To investigate the extent to which sample attrition could be a threat to the validity of the estimators, we perform a difference-in-differences analysis of the characteristics of successive groups of claimants who are observed with earnings. We first calculate the claim year (year 0) differences in average characteristics between a given treatment group (with earnings) and the control group (also with earnings), then do the same for each subsequent year. We then calculate the differences in these differences between the claim year and each subsequent year (along with the standard error of each difference in differences).

5. Also, with survey data, attrition is typically an absorbing state — once a subject leaves the sample, he or she does not return. This is not the case with administrative data, where a subject may have no earnings for one or several quarters but then return to work and appear again in the administrative records.

Specifically, for a given characteristic *x*, the difference in differences (or *change* in the differences) between the treatment group (*T*) and the control group (*C*), between year *t* (a subsequent year) and year 0 (the claim year), can be written:

$$(\overline{x}_{t,T} - \overline{x}_{t,C}) - (\overline{x}_{0,T} - \overline{x}_{0,C}) \tag{1}$$

If the differences in differences are insubstantial for most characteristics, then we would infer that the observable characteristics of claimants who are leaving the treatment and control samples over time (that is, no longer observed with earnings) are similar. It does not necessarily follow that the *unobservable* characteristics of claimants leaving the two groups over time are also similar; however, given that we find the selection on observables assumption to be reasonable based on pre-treatment outcome tests, it seems plausible that unobservables and observables are correlated in these samples. If so, then finding that the observable characteristics of claimants who leave the treatment and controls groups over time are similar would suggest that selective attrition is not a threat to the validity of the estimators we use.

Tables 2, 3, and 4 display the findings of such a difference-in-differences analysis. Consider first Table 2, which pertains to attrition from the JSA treatment. Columns 1 and 2 show sample means for the control and JSA groups in the claim year (year 0), and column 3 shows the difference for each characteristic in the claim year. Columns 4, 5, and 6 show the difference (or change) between this claim-year difference and the difference in years 3, 6, and 9. (We have also computed differences in differences for years 1, 2, 4, 5, 7, and 8, but for the sake of economy we do not report them. As discussed below, and their implications are the same as those for the years shown.)

For example, the claim-year difference in the proportion of the control and JSA groups who were male was about 2 percentage points (0.0192, *p*-value = 0.17) (see column 1, Table 2).

By year 3, this difference had increased by 0.0001 (that is, to 0.0193, *p*-value = 1.00), by year 6 it had increased by 0.0023 (*p*-value = 0.91), and by year 9 it had fallen by 0.0037 (*p*-value = 0.86). None of these differences in differences is significant, statistically or in practical terms (the absolute values of the standard errors of these differences in differences are at least 5 times the point estimates).

For all but five variables shown in Table 2, the claim-year difference between the control and JSA groups had a *p*-value greater than 0.10; that is, for all but five characteristics, there was no initial difference between the control and JSA groups. Further, in all these cases, there was no subsequent change over time in the difference between the control and JSA groups — that is, the differences in differences over the following 9 years all have *p*-values greater than 0.10.

For five variables, the claim-year difference between the control and JSA groups has a *p*-value of 0.10 or less: claimants from households with 1 dependent, who formerly worked in wholesale trade, who formerly worked in government, who were ex-service members, and whose reservation wage was between $10 and $20. For each of these characteristics, the initial difference between the control and JSA groups persists throughout the 9-year follow-up period — the estimated differences in differences all have *p*-values greater than 0.10.

Overall, the inference is that, if there was no initial difference between the control and JSA groups in a given characteristic, none appeared over time. And if there was initially a difference in a characteristic between the control and JSA groups, that difference did not change over time. Assignment to treatment appears to have had no impact on the characteristics of the JSA claimants compared with those of the controls.

Consider now Table 3 (comparing controls with Exception Reporting claimants) and Table 4 (comparing controls with New Work Search claimants). Columns 3, 4, and 5 of those

tables show only two differences in differences with a *p*-value of 0.10 or less: for both the Exception Reporting and New Work Search groups, the year 9 differences in differences for the proportion of females aged 45–54 have *p*-values between 0.08 and 0.09. That is, by year 9, the difference between the control and Exception Reporting groups in the proportion of females aged 45–54 had increased by 1.4 percentage point (*p*-value = 0.09); a similar finding holds for women aged 45–54 in the control and New Work Search groups (an increase of 1.6 percentage points by year 9, *p*-value = 0.08).

As mentioned above, in addition to the differences in differences for years 3, 6, and 9 shown in Tables 2, 3, and 4, we have computed corresponding differences in differences for years 1, 2, 4, 5, 7, and 8. None of these estimates has a *p*-value of 0.10 or less.

We conclude that the observable characteristics of claimants who left the control group over time and those who left each of the treatment groups over time are essentially the same. Again, although this does not dispose of the possibility that those who left the treatment and control groups differed in unobservable ways, our finding that selection on observables is a reasonable assumption suggests that sample attrition probably does not threaten the validity of the estimators we use. Given that we are unable to detect any changes over time in the observable characteristics of controls compared with each of the treatment groups, it seems unlikely that sample attrition of any consequence is taking place.

References

Corson, Walter, Paul Decker, Shari Dunstan, and Stuart Kerachsky. 1992. "Pennsylvania Reemployment Bonus Demonstration" Unemployment Insurance Occasional Paper 92-1. Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Hausman, Jerry A. and David A. Wise. 1979. "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment." *Econometrica* 47: 455-473.

Johnson, Terry R. and Daniel R. Klepinger. 1991. "Evaluation of the Impacts of the Washington Alternative Work Search Experiment." Unemployment Insurance Occasional Paper 91-4. Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Johnson, Terry R. and Daniel R. Klepinger. 1994."Experimental Evidence on Unemployment Insurance Work search Policies." *Journal of Human Resources* 29 (Summer): 695–717.

Kornfeld, Robert and Howard S. Bloom. 1999. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics* 17 (January): 168–197.

McFadden, Daniel L. 1985. "Comment on Hausman and Wise." In *Social Experimentation*, edited by Jerry A. Hausman and David A. Wise. Chicago: University of Chicago Press. Pp. 214-218.

Murnane, Richard J. and John B. Willett. 2011. *Methods Matter*. New York: Oxford University Press.

Spiegelman, Robert G., Christopher J. O'Leary, and Kenneth Kline. 1992. "The Washington Reemployment Bonus Experiment: Final Report." Unemployment Insurance Occasional Paper 92-6. Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Spiegelman, Robert G., Christopher J. O'Leary, and Kenneth Kline. 1995. "Do Bonus Offers Shorten Unemployment Insurance Spells?: Results from the Washington Reemployment Bonus Experiment." *Journal of Policy Analysis and Management* 14 (Spring): 245–269.

Wallace, Geoffrey L. and Robert Haveman. 2007. "The Implications of Differences Between Employer and Worker Employment/Earnings Reports for Policy Evaluation." *Journal of Policy Analysis and Management* 26: 737–753.

Table 1

Matches[1] and match rates[2] (in parentheses) between population wage records and claimants enrolled in the Washington Alternative Work Search experiment, by year and treatment group

	Total[3]	Full control group[4]	Exception Reporting	New Work Search	Restricted control group[4]	JSA
Year 0	8,092	2,447	1,906	1,728	2,227	2,011
	(0.865)	(0.859)	(0.848)	(0.880)	(0.858)	(0.878)
Year 1	7,720	2,350	1,849	1,621	2,145	1,900
	(0.826)	(0.825)	(0.823)	(0.826)	(0.826)	(0.829)
Year 2	7,438	2,266	1,769	1,552	2,062	1,851
	(0.796)	(0.796)	(0.787)	(0.791)	(0.794)	(0.808)
Year 3	7,183	2,195	1,692	1,498	1,999	1,798
	(0.768)	(0.771)	(0.753)	(0.763)	(0.770)	(0.785)
Year 4	6,888	2,123	1,625	1,417	1,927	1,723
	(0.737)	(0.745)	(0.723)	(0.722)	(0.742)	(0.752)
Year 5	6,590	2,028	1,570	1,353	1,835	1,639
	(0.705)	(0.712)	(0.698)	(0.689)	(0.707)	(0.715)
Year 6	6,357	1,930	1,522	1,320	1,749	1,585
	(0.680)	(0.678)	(0.677)	(0.672)	(0.673)	(0.692)
Year 7	6,146	1,873	1,458	1,270	1,701	1,545
	(0.657)	(0.658)	(0.649)	(0.647)	(0.655)	(0.674)
Year 8	5,942	1,825	1,413	1,219	1,658	1,485
	(0.636)	(0.641)	(0.629)	(0.621)	(0.638)	(0.648)
Year 9	5,848	1,785	1,388	1,201	1,629	1,474
	(0.625)	(0.627)	(0.617)	(0.612)	(0.627)	(0.643)
Number enrolled	9,350	2,848	2,248	1,963	2,597	2,291

Source: Tabulated from the Washington Alternative Work Search experimental data base and population wage records from the Washington State Department of Employment Security.

1. A "match" occurs for a claimant in a given year if positive earnings were found in at least one quarter of the year for the claimant.

2. The match rate is the proportion of claimants initially enrolled in the experiment (or in a treatment) who were observed with earnings in a given year.

3. Sum of the full control, Exception Reporting, New Work Search, and JSA groups.

4. The full control group is used with the Exception Reporting and New Work Search treatments. The restricted control group, which drops claimants who never received a benefit, is used with the JSA treatment because only JSA assignees who received a first benefit payment received a JSA call-in notice.

Table 2
Characteristics of controls and JSA claimants with earnings in claim year, and differences in differences between controls and JSA claimants (subsequent years against claim year)
(standard errors in parentheses)

	Claim year			Difference in differences between control and JSA groups (year *t* – claim year)		
Covariate	Control mean (1)	JSA mean (2)	Difference (treatment – control) (3)	Year 3 (4)	Year 6 (5)	Year 9 (6)
Male	0.6974	0.7166	0.0192	0.0001	0.0023	-0.0037
	(0.0097)	(0.0101)	(0.0140)	(0.0203)	(0.0211)	(0.0216)
Age (years)						
≤ 24	0.2021	0.1949	-0.0071	0.0015	-0.0011	0.0071
	(0.0085)	(0.0088)	(0.0123)	(0.0179)	(0.0187)	(0.0192)
25-34	0.3767	0.3894	0.0126	-0.0057	-0.0121	-0.0263
	(0.0103)	(0.0109)	(0.0150)	(0.0218)	(0.0226)	(0.0231)
35-44	0.2506	0.2496	-0.0009	0.0100	0.0150	0.0207
	(0.0092)	(0.0097)	(0.0133)	(0.0195)	(0.0202)	(0.0208)
45-54	0.1159	0.1094	-0.0065	-0.0040	0.0050	0.0062
	(0.0068)	(0.0070)	(0.0097)	(0.0142)	(0.0145)	(0.0143)
≥ 54	0.0548	0.0567	0.0019	-0.0019	-0.0068	-0.0078
	(0.0048)	(0.0052)	(0.0071)	(0.0098)	(0.0088)	(0.0086)
Gender-age interactions						
male ≤ 24	0.1352	0.1348	-0.0004	0.0044	0.0008	0.0042
	(0.0072)	(0.0076)	(0.0105)	(0.0154)	(0.0161)	(0.0165)
male 25-34	0.2762	0.2904	0.0142	-0.0064	-0.0095	-0.0216
	(0.0095)	(0.0101)	(0.0139)	(0.0202)	(0.0211)	(0.0215)
male 35-44	0.1652	0.1735	0.0083	0.0039	0.0112	0.0146
	(0.0079)	(0.0084)	(0.0115)	(0.0168)	(0.0175)	(0.0180)
male 45-54	0.0790	0.0771	-0.0020	-0.0053	0.0002	0.0008
	(0.0057)	(0.0059)	(0.0083)	(0.0121)	(0.0122)	(0.0121)
male ≥ 54	0.0418	0.0408	-0.0010	0.0034	-0.0004	-0.0017
	(0.0042)	(0.0044)	(0.0061)	(0.0084)	(0.0075)	(0.0073)
female ≤ 24	0.0669	0.0602	-0.0067	-0.0029	-0.0019	0.0030
	(0.0053)	(0.0053)	(0.0075)	(0.0109)	(0.0113)	(0.0116)
female 25-34	0.1006	0.0990	-0.0016	0.0007	-0.0026	-0.0046
	(0.0064)	(0.0067)	(0.0092)	(0.0133)	(0.0140)	(0.0144)
female 35-44	0.0853	0.0761	-0.0092	0.0061	0.0038	0.0061
	(0.0059)	(0.0059)	(0.0084)	(0.0123)	(0.0128)	(0.0133)
female 45-54	0.0368	0.0323	-0.0045	0.0013	0.0048	0.0054
	(0.0040)	(0.0039)	(0.0056)	(0.0082)	(0.0084)	(0.0082)
female ≥ 54	0.0130	0.0159	0.0029	-0.0053	-0.0064	-0.0061
	(0.0024)	(0.0028)	(0.0037)	(0.0051)	(0.0047)	(0.0047)
Ethnicity						
white	0.8244	0.8424	0.0179	-0.0105	-0.0036	0.0026
	(0.0081)	(0.0081)	(0.0114)	(0.0165)	(0.0171)	(0.0173)
black	0.0997	0.0880	-0.0117	0.0065	0.0020	0.0065
	(0.0063)	(0.0063)	(0.0090)	(0.0128)	(0.0132)	(0.0134)
other	0.0759	0.0696	-0.0063	0.0040	0.0015	-0.0091

Covariate	Claim year: Control mean (1)	Claim year: JSA mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and JSA groups (year *t* – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
	(0.0056)	(0.0057)	(0.0080)	(0.0116)	(0.0121)	(0.0122)
Schooling (years)						
less than high school	0.1383	0.1273	-0.0110	0.0004	-0.0016	-0.0043
	(0.0073)	(0.0074)	(0.0104)	(0.0151)	(0.0154)	(0.0156)
high school	0.5384	0.5525	0.0141	-0.0004	0.0039	-0.0080
	(0.0106)	(0.0111)	(0.0153)	(0.0223)	(0.0231)	(0.0235)
some college	0.2434	0.2387	-0.0047	0.0018	-0.0011	0.0117
	(0.0091)	(0.0095)	(0.0132)	(0.0191)	(0.0199)	(0.0203)
college graduate	0.0799	0.0816	0.0016	-0.0017	-0.0012	0.0006
	(0.0057)	(0.0061)	(0.0084)	(0.0123)	(0.0128)	(0.0131)
Marital status/gender						
married male	0.2690	0.2636	-0.0054	-0.0084	-0.0138	-0.0121
	(0.0094)	(0.0098)	(0.0136)	(0.0198)	(0.0203)	(0.0208)
married female	0.0961	0.0865	-0.0096	0.0072	0.0107	0.0113
	(0.0062)	(0.0063)	(0.0089)	(0.0129)	(0.0133)	(0.0135)
Household status						
no dependents	0.3044	0.3068	0.0024	0.0045	0.0078	0.0170
	(0.0098)	(0.0103)	(0.0142)	(0.0206)	(0.0212)	(0.0216)
1 dependent	0.1504	0.1318	*-0.0187*	-0.0005	-0.0004	0.0055
	(0.0076)	(0.0075)	(0.0107)	(0.0156)	(0.0161)	(0.0166)
2 or more dependents	0.2393	0.2392	-0.0002	-0.0028	0.0025	-0.0098
	(0.0090)	(0.0095)	(0.0131)	(0.0191)	(0.0199)	(0.0203)
homeowner	0.2928	0.2740	-0.0188	0.0010	0.0019	0.0060
	(0.0096)	(0.0099)	(0.0139)	(0.0201)	(0.0208)	(0.0212)
Veteran	0.2411	0.2307	-0.0104	-0.0075	-0.0019	-0.0131
	(0.0091)	(0.0094)	(0.0131)	(0.0190)	(0.0196)	(0.0200)
Union/standby[a]	0.3992	0.4013	0.0021	0.0025	0.0015	-0.0109
	(0.0104)	(0.0109)	(0.0151)	(0.0219)	(0.0227)	(0.0231)
Occupation						
professional	0.0754	0.0885	0.0131	-0.0020	-0.0076	0.0020
	(0.0056)	(0.0063)	(0.0085)	(0.0124)	(0.0127)	(0.0129)
clerical	0.1392	0.1283	-0.0109	0.0008	-0.0016	0.0078
	(0.0073)	(0.0075)	(0.0105)	(0.0152)	(0.0158)	(0.0164)
sales	0.0548	0.0517	-0.0031	0.0025	0.0048	-0.0001
	(0.0048)	(0.0049)	(0.0069)	(0.0101)	(0.0105)	(0.0106)
service	0.1006	0.1109	0.0103	0.0071	0.0129	0.0126
	(0.0064)	(0.0070)	(0.0095)	(0.0138)	(0.0143)	(0.0144)
agric., fishery, forestry	0.0278	0.0239	-0.0040	0.0013	-0.0001	-0.0002
	(0.0035)	(0.0034)	(0.0049)	(0.0070)	(0.0073)	(0.0073)
processing	0.0382	0.0338	-0.0044	-0.0047	-0.0039	-0.0054
	(0.0041)	(0.0040)	(0.0057)	(0.0085)	(0.0088)	(0.0089)
machine trades	0.0786	0.0746	-0.0040	-0.0015	-0.0005	-0.0108
	(0.0057)	(0.0059)	(0.0082)	(0.0119)	(0.0123)	(0.0126)
benchwork	0.0557	0.0582	0.0025	-0.0004	-0.0014	-0.0043
	(0.0049)	(0.0052)	(0.0071)	(0.0103)	(0.0106)	(0.0108)

Table 2

Covariate	Claim year: Control mean (1)	Claim year: JSA mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and JSA groups (year t – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
structural work	0.2717	0.2740	0.0023	0.0019	0.0072	0.0023
	(0.0094)	(0.0099)	(0.0137)	(0.0200)	(0.0206)	(0.0211)
miscellaneous	0.1513	0.1437	-0.0076	-0.0045	-0.0069	-0.0005
	(0.0076)	(0.0078)	(0.0109)	(0.0158)	(0.0165)	(0.0169)
Industry						
agriculture	0.0211	0.0224	0.0013	-0.0011	-0.0017	-0.0031
	(0.0030)	(0.0033)	(0.0045)	(0.0065)	(0.0066)	(0.0067)
mining	0.0022	0.0005	-0.0017	-0.0003	-0.0005	-0.0007
	(0.0010)	(0.0005)	(0.0011)	(0.0015)	(0.0016)	(0.0017)
construction	0.2142	0.2014	-0.0128	-0.0088	-0.0021	-0.0128
	(0.0087)	(0.0089)	(0.0125)	(0.0181)	(0.0187)	(0.0191)
manufacturing	0.2335	0.2302	-0.0033	-0.0062	-0.0138	-0.0066
	(0.0090)	(0.0094)	(0.0130)	(0.0189)	(0.0195)	(0.0199)
transportation, utilities	0.0476	0.0532	0.0056	-0.0006	0.0013	-0.0009
	(0.0045)	(0.0050)	(0.0067)	(0.0099)	(0.0102)	(0.0106)
wholesale trade	0.0606	0.0487	*-0.0119*	0.0026	0.0001	0.0031
	(0.0051)	(0.0048)	(0.0070)	(0.0103)	(0.0109)	(0.0110)
retail trade	0.1576	0.1492	-0.0084	0.0027	0.0007	0.0050
	(0.0077)	(0.0079)	(0.0111)	(0.0162)	(0.0167)	(0.0170)
finance, ins., real estate	0.0269	0.0323	0.0054	-0.0004	0.0005	0.0034
	(0.0034)	(0.0039)	(0.0052)	(0.0075)	(0.0079)	(0.0080)
services	0.1720	0.1830	0.0110	0.0074	0.0100	0.0072
	(0.0080)	(0.0086)	(0.0118)	(0.0171)	(0.0177)	(0.0181)
government	0.0471	0.0607	*0.0135*	0.0021	0.0057	0.0022
	(0.0045)	(0.0053)	(0.0070)	(0.0101)	(0.0105)	(0.0108)
unclassified	0.0171	0.0184	0.0013	0.0024	-0.0001	0.0032
	(0.0027)	(0.0030)	(0.0041)	(0.0061)	(0.0062)	(0.0064)
Reason for job loss						
permanent layoff	0.1652	0.1507	-0.0146	0.0003	0.0072	0.0071
	(0.0079)	(0.0080)	(0.0112)	(0.0163)	(0.0169)	(0.0172)
temporary layoff	0.2236	0.2372	0.0136	0.0046	0.0081	-0.0039
	(0.0088)	(0.0095)	(0.0130)	(0.0188)	(0.0195)	(0.0199)
contract/seasonal	0.1702	0.1546	-0.0155	-0.0113	-0.0154	-0.0028
	(0.0080)	(0.0081)	(0.0113)	(0.0164)	(0.0169)	(0.0173)
quit	0.1608	0.1726	0.0118	0.0095	0.0105	0.0131
	(0.0078)	(0.0084)	(0.0115)	(0.0168)	(0.0174)	(0.0178)
missing	0.2802	0.2849	0.0047	-0.0031	-0.0104	-0.0136
	(0.0095)	(0.0101)	(0.0139)	(0.0202)	(0.0210)	(0.0214)
Prior claim						
none	0.7912	0.7892	-0.0020	0.0053	0.0004	0.0128
	(0.0086)	(0.0091)	(0.0125)	(0.0182)	(0.0188)	(0.0191)
duration ≤ 15 weeks	0.1078	0.0970	-0.0108	-0.0016	0.0031	0.0009
	(0.0066)	(0.0066)	(0.0093)	(0.0135)	(0.0141)	(0.0143)
duration > 15 weeks	0.1010	0.1139	0.0128	-0.0037	-0.0035	-0.0137
	(0.0064)	(0.0071)	(0.0095)	(0.0139)	(0.0141)	(0.0145)

Table 2

Covariate	Claim year: Control mean (1)	Claim year: JSA mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and JSA groups (year *t* – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
UI benefits/claim type						
weekly amount ($)	148.77	148.46	-0.31	-1.17	-1.74	-0.13
	(1.05)	(1.10)	(1.52)	(2.20)	(2.28)	(2.31)
maximum amount ($)	3,956.8	3,920.7	-36.1	-58.5	-74.1	-1.9
	(35.9)	(37.4)	(51.8)	(75.2)	(77.9)	(79.3)
potential benefit duration (weeks)	26.172	26.059	-0.113	-0.175	-0.227	-0.001
	(0.102)	(0.112)	(0.151)	(0.219)	(0.227)	(0.232)
replacement rate (percent)[b]	61.052	61.425	0.373	-0.129	0.586	-0.377
	(0.459)	(0.486)	(0.668)	(0.970)	(1.006)	(1.024)
combined wage claim[c]	0.0431	0.0502	0.0071	-0.0031	-0.0042	0.0012
	(0.0043)	(0.0049)	(0.0065)	(0.0093)	(0.0095)	(0.0099)
ex-service member claim	0.0265	0.0368	*0.0103*	0.0033	0.0058	0.0017
	(0.0034)	(0.0042)	(0.0054)	(0.0077)	(0.0080)	(0.0082)
federal employee claim	0.0184	0.0139	-0.0045	0.0014	0.0041	0.0059
	(0.0028)	(0.0026)	(0.0039)	(0.0056)	(0.0057)	(0.0058)
Reservation wage (hourly)						
≤ $5.00	0.1715	0.1626	-0.0089	0.0076	0.0071	0.0072
	(0.0080)	(0.0082)	(0.0115)	(0.0166)	(0.0171)	(0.0174)
$5.01–$7.00	0.1392	0.1546	0.0154	-0.0032	0.0027	0.0112
	(0.0073)	(0.0081)	(0.0109)	(0.0159)	(0.0165)	(0.0170)
$7.01–$10.00	0.1581	0.1621	0.0040	-0.0085	-0.0044	-0.0075
	(0.0077)	(0.0082)	(0.0113)	(0.0165)	(0.0172)	(0.0175)
$10.01–$20.00	0.1464	0.1283	*-0.0181*	-0.0029	-0.0116	-0.0122
	(0.0075)	(0.0075)	(0.0106)	(0.0153)	(0.0158)	(0.0162)
> $20.00	0.0925	0.0816	-0.0109	0.0043	0.0072	0.0068
	(0.0061)	(0.0061)	(0.0087)	(0.0126)	(0.0129)	(0.0131)
Pre-claim earnings ($)						
3 years before	10,870	10,610	-260	61	143	322
	(244)	(261)	(358)	(518)	(537)	(554)
2 years before	12,097	11,951	-146	-199	-151	1
	(238)	(248)	(344)	(500)	(518)	(528)
1 year before	14,241	14,095	-146	-155	-212	242
	(208)	(222)	(304)	(444)	(460)	(469)
Pre-claim hours						
3 years before	950.1	920.4	-29.7	8.6	3.9	14.4
	(16.9)	(17.4)	(24.3)	(35.3)	(36.5)	(37.2)
2 years before	1,101.9	1,089.2	-12.7	-5.7	-8.3	9.9
	(15.9)	(16.6)	(23.0)	(33.5)	(34.7)	(35.4)
1 year before	1,345.0	1,322.0	-22.9	17.5	9.8	30.1
	(13.7)	(15.1)	(20.4)	(29.6)	(30.7)	(31.3)
Pre-claim earnings variability						
annual (CV)	0.6418	0.6393	-0.0025	-0.0027	0.0102	0.0044
	(0.0110)	(0.0115)	(0.0159)	(0.0230)	(0.0237)	(0.0242)
seasonal (CV)	0.5709	0.5837	0.0128	-0.0077	0.0027	-0.0066
	(0.0080)	(0.0084)	(0.0116)	(0.0169)	(0.0174)	(0.0177)

Table 2

Covariate	Claim year: Control mean (1)	Claim year: JSA mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and JSA groups (year *t* – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
Quarters earnings >0	8.8572	8.7126	-0.1446	-0.0332	-0.0816	-0.0107
in 12 pre-claim quarters	(0.0698)	(0.0750)	(0.1025)	(0.1482)	(0.1529)	(0.1555)
< 6	0.1886	0.1890	0.0004	0.0045	0.0063	-0.0008
	(0.0083)	(0.0087)	(0.0120)	(0.0175)	(0.0179)	(0.0182)
6–8	0.1796	0.1934	0.0138	-0.0007	0.0031	0.0076
	(0.0081)	(0.0088)	(0.0120)	(0.0174)	(0.0180)	(0.0184)
> 8	0.6318	0.6176	-0.0142	-0.0038	-0.0095	-0.0068
	(0.0102)	(0.0108)	(0.0149)	(0.0216)	(0.0224)	(0.0228)
Pre-claim drop in earnings						
between pre-claim year 3	0.3426	0.3317	-0.0109	0.0059	0.0030	0.0043
and pre-clam year 2	(0.0101)	(0.0105)	(0.0145)	(0.0212)	(0.0219)	(0.0223)
between pre-claim year 2	0.3345	0.3158	-0.0188	0.0001	0.0030	-0.0007
and pre-clam year 1	(0.0100)	(0.0104)	(0.0144)	(0.0210)	(0.0217)	(0.0222)
Claimants with earnings >0						
Control group[d]	2,227	n/a	2,227	1,999	1,749	1,629
JSA group	n/a	2,011	2,011	1,798	1,585	1,474

Source: See Table 1.

Bold indicates *p* -value < .05 for test of difference between control and JSA groups (column 3).

Italic indicates *p* -value < .10 for test of difference between control and JSA groups (column 3).

a. Claimants were not required to search for work or participate in JSA if their job placement was handled by a union, or if they were on layoff with a set recall date.

b. Weekly benefit amount as a percentage of pre-claim average weekly earnings.

c. Combined wage claims use earnings from more than one state to calculate base period earnings.

d. The size of the control group in this table differs from size reported in Tables A-1, A-3, and A-4 because we drop claimants who never received benefits from the JSA group.

Table 2

Table 3
Characteristics of controls and Exception Reporting (ER) claimants with earnings in claim year, and differences in differences between controls and ER claimants (subsequent years against claim year) (standard errors in parentheses)

	Claim year			Difference in differences between control and ER groups (year *t* – claim year)		
Covariate	Control mean (1)	ER mean (2)	Difference (treatment – control) (3)	Year 3 (4)	Year 6 (5)	Year 9 (6)
Male	0.6935	0.7282	**0.0347**	-0.0139	-0.0190	-0.0170
	(0.0093)	(0.0102)	(0.0138)	(0.0202)	(0.0209)	(0.0214)
Age (years)						
≤ 24	0.2145	0.2078	-0.0068	-0.0042	-0.0058	0.0077
	(0.0083)	(0.0093)	(0.0125)	(0.0181)	(0.0188)	(0.0194)
25-34	0.3723	0.4145	**0.0422**	-0.0066	-0.0200	-0.0217
	(0.0098)	(0.0113)	(0.0149)	(0.0218)	(0.0225)	(0.0231)
35-44	0.2489	0.2183	**-0.0306**	0.0097	0.0128	0.0063
	(0.0087)	(0.0095)	(0.0129)	(0.0189)	(0.0195)	(0.0201)
45-54	0.1136	0.1097	-0.0040	-0.0029	0.0043	0.0091
	(0.0064)	(0.0072)	(0.0096)	(0.0141)	(0.0143)	(0.0143)
≥ 54	0.0507	0.0498	-0.0008	0.0040	0.0087	-0.0015
	(0.0044)	(0.0050)	(0.0067)	(0.0094)	(0.0088)	(0.0082)
Gender-age interactions						
male ≤ 24	0.1438	0.1401	-0.0038	0.0002	-0.0018	0.0063
	(0.0071)	(0.0080)	(0.0107)	(0.0155)	(0.0162)	(0.0167)
male 25-34	0.2726	0.3164	**0.0438**	-0.0125	-0.0228	-0.0226
	(0.0090)	(0.0107)	(0.0140)	(0.0203)	(0.0210)	(0.0217)
male 35-44	0.1626	0.1480	-0.0147	0.0020	0.0026	0.0033
	(0.0075)	(0.0081)	(0.0110)	(0.0161)	(0.0166)	(0.0171)
male 45-54	0.0756	0.0850	0.0094	-0.0085	-0.0039	-0.0047
	(0.0053)	(0.0064)	(0.0083)	(0.0121)	(0.0123)	(0.0122)
male ≥ 54	0.0388	0.0388	0.0000	0.0048	0.0069	0.0006
	(0.0039)	(0.0044)	(0.0059)	(0.0082)	(0.0076)	(0.0071)
female ≤ 24	0.0707	0.0677	-0.0030	-0.0044	-0.0040	0.0014
	(0.0052)	(0.0058)	(0.0077)	(0.0112)	(0.0116)	(0.0119)
female 25-34	0.0997	0.0981	-0.0016	0.0059	0.0027	0.0009
	(0.0061)	(0.0068)	(0.0091)	(0.0133)	(0.0139)	(0.0143)
female 35-44	0.0862	0.0703	*-0.0159*	0.0076	0.0102	0.0031
	(0.0057)	(0.0059)	(0.0082)	(0.0121)	(0.0126)	(0.0129)
female 45-54	0.0380	0.0247	**-0.0133**	0.0055	0.0082	*0.0138*
	(0.0039)	(0.0036)	(0.0053)	(0.0078)	(0.0080)	(0.0080)
female ≥ 54	0.0119	0.0110	-0.0008	-0.0008	0.0017	-0.0021
	(0.0022)	(0.0024)	(0.0032)	(0.0047)	(0.0046)	(0.0042)
Ethnicity						
white	0.8255	0.8221	-0.0034	-0.0115	0.0039	-0.0088
	(0.0077)	(0.0088)	(0.0116)	(0.0168)	(0.0172)	(0.0178)
black	0.0993	0.1070	0.0077	0.0086	-0.0096	0.0076
	(0.0060)	(0.0071)	(0.0093)	(0.0134)	(0.0135)	(0.0140)
other	0.0752	0.0708	-0.0044	0.0028	0.0057	0.0013

Covariate	Claim year: Control mean (1)	Claim year: ER mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and ER groups (year t – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
	(0.0053)	(0.0059)	(0.0079)	(0.0115)	(0.0120)	(0.0123)
Schooling (years)						
less than high school	0.1406	0.1196	**-0.0210**	0.0007	-0.0012	0.0018
	(0.0070)	(0.0074)	(0.0102)	(0.0148)	(0.0151)	(0.0154)
high school	0.5386	0.5525	0.0138	0.0007	0.0059	0.0014
	(0.0101)	(0.0114)	(0.0152)	(0.0222)	(0.0229)	(0.0234)
some college	0.2395	0.2508	0.0113	-0.0030	-0.0038	-0.0074
	(0.0086)	(0.0099)	(0.0132)	(0.0192)	(0.0198)	(0.0202)
college graduate	0.0813	0.0771	-0.0042	0.0016	-0.0009	0.0042
	(0.0055)	(0.0061)	(0.0082)	(0.0122)	(0.0125)	(0.0130)
Marital status/gender						
married male	0.2628	0.2712	0.0085	-0.0166	-0.0154	-0.0253
	(0.0089)	(0.0102)	(0.0135)	(0.0196)	(0.0202)	(0.0206)
married female	0.0960	0.0771	**-0.0189**	0.0040	0.0093	0.0126
	(0.0060)	(0.0061)	(0.0085)	(0.0123)	(0.0128)	(0.0130)
Household status						
no dependents	0.3102	0.3106	0.0004	0.0064	0.0176	0.0141
	(0.0094)	(0.0106)	(0.0141)	(0.0206)	(0.0212)	(0.0216)
1 dependent	0.1524	0.1516	-0.0008	0.0001	0.0039	0.0005
	(0.0073)	(0.0082)	(0.0110)	(0.0160)	(0.0166)	(0.0170)
2 or more dependents	0.2321	0.2356	0.0035	-0.0024	-0.0087	-0.0124
	(0.0085)	(0.0097)	(0.0129)	(0.0189)	(0.0194)	(0.0200)
homeowner	0.2857	0.2796	-0.0060	0.0098	0.0212	0.0071
	(0.0091)	(0.0103)	(0.0138)	(0.0200)	(0.0207)	(0.0211)
Veteran	0.2342	0.2324	-0.0017	-0.0072	-0.0018	-0.0205
	(0.0086)	(0.0097)	(0.0129)	(0.0188)	(0.0193)	(0.0197)
Union/standby[a]	0.3727	0.3274	**-0.0453**	0.0105	0.0149	0.0130
	(0.0098)	(0.0108)	(0.0145)	(0.0212)	(0.0218)	(0.0224)
Occupation						
professional	0.0732	0.0735	0.0003	0.0032	0.0016	0.0091
	(0.0053)	(0.0060)	(0.0080)	(0.0118)	(0.0121)	(0.0124)
clerical	0.1381	0.1343	-0.0038	-0.0007	0.0001	-0.0011
	(0.0070)	(0.0078)	(0.0105)	(0.0153)	(0.0159)	(0.0163)
sales	0.0552	0.0567	0.0015	-0.0024	0.0011	-0.0024
	(0.0046)	(0.0053)	(0.0070)	(0.0102)	(0.0105)	(0.0107)
service	0.1103	0.0855	**-0.0248**	0.0061	0.0074	0.0101
	(0.0063)	(0.0064)	(0.0090)	(0.0130)	(0.0135)	(0.0135)
agric., fishery, forestry	0.0262	0.0231	-0.0031	0.0002	-0.0008	0.0040
	(0.0032)	(0.0034)	(0.0047)	(0.0067)	(0.0070)	(0.0072)
processing	0.0364	0.0357	-0.0007	-0.0003	-0.0011	-0.0065
	(0.0038)	(0.0042)	(0.0057)	(0.0085)	(0.0087)	(0.0088)
machine trades	0.0826	0.0892	0.0066	0.0011	-0.0049	-0.0044
	(0.0056)	(0.0065)	(0.0086)	(0.0126)	(0.0128)	(0.0134)
benchwork	0.0568	0.0593	0.0025	0.0040	0.0015	-0.0012
	(0.0047)	(0.0054)	(0.0072)	(0.0104)	(0.0107)	(0.0109)

Table 3

Covariate	Claim year: Control mean (1)	Claim year: ER mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and ER groups (year t – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
structural work	0.2628	0.2907	**0.0279**	-0.0127	0.0064	-0.0043
	(0.0089)	(0.0104)	(0.0137)	(0.0199)	(0.0205)	(0.0211)
miscellaneous	0.1512	0.1459	-0.0054	0.0017	-0.0094	-0.0027
	(0.0072)	(0.0081)	(0.0109)	(0.0158)	(0.0164)	(0.0168)
Industry						
agriculture	0.0204	0.0220	0.0016	-0.0021	-0.0022	-0.0030
	(0.0029)	(0.0034)	(0.0044)	(0.0063)	(0.0065)	(0.0066)
mining	0.0020	0.0005	-0.0015	0.0003	0.0001	-0.0000
	(0.0009)	(0.0005)	(0.0011)	(0.0015)	(0.0016)	(0.0017)
construction	0.2084	0.2030	-0.0054	-0.0062	0.0100	-0.0008
	(0.0082)	(0.0092)	(0.0123)	(0.0179)	(0.0186)	(0.0190)
manufacturing	0.2272	0.2450	0.0178	0.0036	0.0001	0.0033
	(0.0085)	(0.0099)	(0.0130)	(0.0190)	(0.0196)	(0.0201)
transportation, utilities	0.0466	0.0540	0.0075	-0.0070	-0.0073	-0.0067
	(0.0043)	(0.0052)	(0.0067)	(0.0097)	(0.0099)	(0.0103)
wholesale trade	0.0597	0.0588	-0.0009	-0.0013	-0.0072	-0.0030
	(0.0048)	(0.0054)	(0.0072)	(0.0105)	(0.0110)	(0.0113)
retail trade	0.1655	0.1558	-0.0097	0.0007	0.0003	-0.0004
	(0.0075)	(0.0083)	(0.0112)	(0.0163)	(0.0169)	(0.0172)
finance, ins., real estate	0.0262	0.0294	0.0032	-0.0012	-0.0015	0.0002
	(0.0032)	(0.0039)	(0.0050)	(0.0072)	(0.0075)	(0.0077)
services	0.1778	0.1626	-0.0151	0.0156	0.0076	0.0130
	(0.0077)	(0.0085)	(0.0115)	(0.0168)	(0.0172)	(0.0177)
government	0.0499	0.0504	0.0005	-0.0024	-0.0018	-0.0023
	(0.0044)	(0.0050)	(0.0067)	(0.0096)	(0.0098)	(0.0102)
unclassified	0.0163	0.0184	0.0020	0.0002	0.0019	-0.0003
	(0.0026)	(0.0031)	(0.0040)	(0.0059)	(0.0062)	(0.0061)
Reason for job loss						
permanent layoff	0.1631	0.1511	-0.0120	-0.0009	0.0030	0.0178
	(0.0075)	(0.0082)	(0.0111)	(0.0162)	(0.0166)	(0.0172)
temporary layoff	0.2182	0.2513	**0.0331**	0.0063	0.0116	-0.0031
	(0.0084)	(0.0099)	(0.0130)	(0.0189)	(0.0195)	(0.0199)
contract/seasonal	0.1663	0.1584	-0.0079	0.0001	-0.0034	-0.0046
	(0.0075)	(0.0084)	(0.0113)	(0.0164)	(0.0169)	(0.0172)
quit	0.1716	0.1689	-0.0027	0.0031	0.0101	0.0011
	(0.0076)	(0.0086)	(0.0115)	(0.0168)	(0.0173)	(0.0177)
missing	0.2808	0.2702	-0.0106	-0.0085	-0.0214	-0.0112
	(0.0091)	(0.0102)	(0.0136)	(0.0198)	(0.0205)	(0.0210)
Prior claim						
none	0.8002	0.7943	-0.0058	0.0021	-0.0011	0.0034
	(0.0081)	(0.0093)	(0.0123)	(0.0179)	(0.0184)	(0.0189)
duration ≤ 15 weeks	0.1034	0.0986	-0.0048	0.0024	-0.0019	-0.0041
	(0.0062)	(0.0068)	(0.0092)	(0.0134)	(0.0138)	(0.0140)
duration > 15 weeks	0.0964	0.1070	0.0106	-0.0045	0.0030	0.0007
	(0.0060)	(0.0071)	(0.0093)	(0.0135)	(0.0138)	(0.0143)

Table 3

Covariate	Claim year: Control mean (1)	Claim year: ER mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and ER groups (year *t* – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
UI benefits/claim type						
weekly amount ($)	146.77	148.28	1.52	-0.80	-1.32	-1.63
	(1.03)	(1.13)	(1.52)	(2.21)	(2.28)	(2.33)
maximum amount ($)	3,888.1	3,943.0	54.9	-6.7	-37.2	-24.9
	(34.9)	(38.3)	(51.8)	(75.4)	(77.7)	(79.6)
potential benefit duration (weeks)	26.023	26.203	0.180	0.072	-0.034	0.097
	(0.100)	(0.112)	(0.150)	(0.215)	(0.221)	(0.228)
replacement rate (percent)[b]	61.197	60.188	-1.008	-0.369	0.132	-0.063
	(0.440)	(0.492)	(0.660)	(0.960)	(0.987)	(1.013)
combined wage claim[c]	0.0454	0.0414	-0.0039	0.0022	0.0028	0.0064
	(0.0042)	(0.0046)	(0.0062)	(0.0089)	(0.0093)	(0.0095)
ex-service member claim	0.0298	0.0299	0.0001	-0.0045	-0.0004	-0.0021
	(0.0034)	(0.0039)	(0.0052)	(0.0072)	(0.0073)	(0.0076)
federal employee claim	0.0168	0.0152	-0.0015	-0.0009	0.0006	0.0043
	(0.0026)	(0.0028)	(0.0038)	(0.0054)	(0.0054)	(0.0057)
Reservation wage (hourly)						
≤ $5.00	0.1814	0.1626	-0.0188	0.0104	0.0088	0.0105
	(0.0078)	(0.0085)	(0.0115)	(0.0167)	(0.0172)	(0.0175)
$5.01–$7.00	0.1373	0.1553	*0.0180*	-0.0043	0.0060	-0.0049
	(0.0070)	(0.0083)	(0.0108)	(0.0159)	(0.0164)	(0.0168)
$7.01–$10.00	0.1532	0.1668	0.0136	-0.0022	-0.0047	-0.0101
	(0.0073)	(0.0085)	(0.0112)	(0.0164)	(0.0170)	(0.0173)
$10.01–$20.00	0.1406	0.1443	0.0037	-0.0023	-0.0027	-0.0109
	(0.0070)	(0.0081)	(0.0107)	(0.0155)	(0.0160)	(0.0164)
> $20.00	0.0919	0.1049	0.0130	0.0042	0.0124	0.0179
	(0.0058)	(0.0070)	(0.0091)	(0.0133)	(0.0137)	(0.0141)
Pre-claim earnings ($)						
3 years before	10,669	11,160	491	312	77	112
	(232)	(271)	(357)	(519)	(531)	(546)
2 years before	11,919	12,188	269	43	-157	-111
	(228)	(260)	(345)	(504)	(517)	(533)
1 year before	14,073	14,204	131	97	-180	-37
	(200)	(230)	(305)	(448)	(457)	(469)
Pre-claim hours						
3 years before	939.1	950.0	11.0	18.6	0.4	0.3
	(16.1)	(18.1)	(24.3)	(35.3)	(36.3)	(37.2)
2 years before	1,095.2	1,095.4	0.2	4.9	-8.4	-9.7
	(15.2)	(17.2)	(23.0)	(33.4)	(34.5)	(35.3)
1 year before	1,340.7	1,350.3	9.6	13.8	4.7	8.9
	(13.2)	(15.3)	(20.2)	(29.2)	(30.1)	(30.9)
Pre-claim earnings variability						
annual (CV)	0.6470	0.6304	-0.0166	-0.0079	0.0037	0.0055
	(0.0106)	(0.0120)	(0.0160)	(0.0231)	(0.0237)	(0.0243)
seasonal (CV)	0.5729	0.5671	-0.0057	-0.0083	0.0018	0.0066
	(0.0077)	(0.0087)	(0.0116)	(0.0169)	(0.0174)	(0.0179)

Table 3

Covariate	Claim year: Control mean (1)	Claim year: ER mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and ER groups (year t – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
Quarters earnings >0	8.8038	8.7844	-0.0195	0.0687	-0.0390	-0.0067
in 12 pre-claim quarters	(0.0670)	(0.0771)	(0.1021)	(0.1474)	(0.1517)	(0.1546)
< 6	0.1913	0.2020	0.0107	-0.0096	-0.0040	-0.0039
	(0.0080)	(0.0092)	(0.0122)	(0.0175)	(0.0179)	(0.0184)
6–8	0.1843	0.1742	-0.0101	0.0012	0.0075	0.0157
	(0.0078)	(0.0087)	(0.0117)	(0.0170)	(0.0177)	(0.0181)
> 8	0.6244	0.6238	-0.0006	0.0084	-0.0036	-0.0118
	(0.0098)	(0.0111)	(0.0148)	(0.0215)	(0.0222)	(0.0227)
Pre-claim drop in earnings						
between pre-claim year 3	0.3367	0.3410	0.0043	0.0114	0.0089	0.0147
and pre-clam year 2	(0.0096)	(0.0109)	(0.0145)	(0.0211)	(0.0218)	(0.0223)
between pre-claim year 2	0.3327	0.3358	0.0031	0.0051	0.0021	-0.0002
and pre-clam year 1	(0.0095)	(0.0108)	(0.0144)	(0.0210)	(0.0217)	(0.0223)
Claimants with earnings >0						
Control group	2,447	n/a	2,447	2,195	1,930	1,785
ER group	n/a	1,906	1,906	1,692	1,522	1,388

Source: See Table 1.

Bold indicates p-value < .05 for test of difference between control and ER groups (column 3).

Italic indicates p-value < .10 for test of difference between control and ER groups (column 3), or for test of difference in differences between control and ER groups (year t – claim year) (column 6).

a. Johnson and Klepinger (1991, pp. 17-18) note that the difference between controls and ER claimants in union/standby is a reporting issue;

ER claimants did not need to submit continued claims, so there was no reason to record their union/standby status.

b. Weekly benefit amount as a percentage of pre-claim average weekly earnings.

c. Combined wage claims use earnings from more than one state to calculate base period earnings.

Table 3

Table 4
Characteristics of controls and New Work Search (NWS) claimants with earnings in claim year, and differences in differences between controls and NWS claimants (subsequent years against claim year) (standard errors in parentheses)

	Claim year			Difference in differences between control and NWS groups (year *t* – claim year)		
Covariate	Control mean (1)	NWS mean (2)	Difference (treatment – control) (3)	Year 3 (4)	Year 6 (5)	Year 9 (6)
Male	0.6935	0.7089	0.0154	-0.0086	-0.0119	-0.0176
	(0.0093)	(0.0109)	(0.0144)	(0.0210)	(0.0219)	(0.0224)
Age (years)						
≤ 24	0.2145	0.2072	-0.0074	0.0006	0.0005	-0.0027
	(0.0083)	(0.0098)	(0.0128)	(0.0187)	(0.0195)	(0.0200)
25-34	0.3723	0.3953	0.0230	-0.0027	-0.0133	-0.0050
	(0.0098)	(0.0118)	(0.0153)	(0.0224)	(0.0233)	(0.0239)
35-44	0.2489	0.2292	-0.0197	-0.0030	0.0063	-0.0007
	(0.0087)	(0.0101)	(0.0134)	(0.0195)	(0.0203)	(0.0209)
45-54	0.1136	0.1128	-0.0008	0.0029	0.0054	0.0146
	(0.0064)	(0.0076)	(0.0100)	(0.0147)	(0.0150)	(0.0151)
≥ 54	0.0507	0.0556	0.0049	0.0021	0.0011	-0.0062
	(0.0044)	(0.0055)	(0.0071)	(0.0100)	(0.0092)	(0.0087)
Gender-age interactions						
male ≤ 24	0.1438	0.1453	0.0014	-0.0029	-0.0052	-0.0081
	(0.0071)	(0.0085)	(0.0111)	(0.0161)	(0.0168)	(0.0173)
male 25-34	0.2726	0.2870	0.0145	-0.0049	-0.0100	-0.0072
	(0.0090)	(0.0109)	(0.0141)	(0.0207)	(0.0216)	(0.0223)
male 35-44	0.1626	0.1534	-0.0093	-0.0027	0.0070	0.0050
	(0.0075)	(0.0087)	(0.0114)	(0.0167)	(0.0174)	(0.0179)
male 45-54	0.0756	0.0799	0.0043	-0.0028	-0.0043	-0.0012
	(0.0053)	(0.0065)	(0.0084)	(0.0125)	(0.0125)	(0.0125)
male ≥ 54	0.0388	0.0434	0.0046	0.0047	0.0006	-0.0061
	(0.0039)	(0.0049)	(0.0063)	(0.0088)	(0.0080)	(0.0074)
female ≤ 24	0.0707	0.0619	-0.0088	0.0036	0.0057	0.0054
	(0.0052)	(0.0058)	(0.0078)	(0.0115)	(0.0120)	(0.0122)
female 25-34	0.0997	0.1082	0.0085	0.0022	-0.0033	0.0022
	(0.0061)	(0.0075)	(0.0096)	(0.0140)	(0.0147)	(0.0152)
female 35-44	0.0862	0.0758	-0.0104	-0.0003	-0.0007	-0.0057
	(0.0057)	(0.0064)	(0.0085)	(0.0125)	(0.0130)	(0.0134)
female 45-54	0.0380	0.0330	-0.0050	0.0057	0.0097	*0.0159*
	(0.0039)	(0.0043)	(0.0058)	(0.0086)	(0.0089)	(0.0090)
female ≥ 54	0.0119	0.0122	0.0003	-0.0026	0.0005	-0.0001
	(0.0022)	(0.0026)	(0.0034)	(0.0049)	(0.0048)	(0.0047)
Ethnicity						
white	0.8255	0.8264	0.0009	-0.0078	0.0002	0.0020
	(0.0077)	(0.0091)	(0.0119)	(0.0172)	(0.0178)	(0.0182)
black	0.0993	0.0926	-0.0067	0.0100	0.0019	0.0054
	(0.0060)	(0.0070)	(0.0092)	(0.0133)	(0.0137)	(0.0139)
other	0.0752	0.0810	0.0058	-0.0022	-0.0021	-0.0074

Covariate	Claim year: Control mean (1)	Claim year: NWS mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and NWS groups (year t – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
	(0.0053)	(0.0066)	(0.0085)	(0.0122)	(0.0127)	(0.0130)
Schooling (years)						
less than high school	0.1406	0.1348	-0.0057	-0.0008	-0.0085	-0.0113
	(0.0070)	(0.0082)	(0.0108)	(0.0157)	(0.0160)	(0.0162)
high school	0.5386	0.5347	-0.0039	0.0066	0.0035	-0.0094
	(0.0101)	(0.0120)	(0.0157)	(0.0229)	(0.0237)	(0.0243)
some college	0.2395	0.2494	0.0099	-0.0002	0.0076	0.0178
	(0.0086)	(0.0104)	(0.0135)	(0.0198)	(0.0206)	(0.0212)
college graduate	0.0813	0.0810	-0.0003	-0.0057	-0.0026	0.0029
	(0.0055)	(0.0066)	(0.0086)	(0.0125)	(0.0130)	(0.0136)
Marital status/gender						
married male	0.2628	0.2483	-0.0145	-0.0041	-0.0001	-0.0093
	(0.0089)	(0.0104)	(0.0137)	(0.0200)	(0.0207)	(0.0211)
married female	0.0960	0.0880	-0.0081	0.0111	0.0212	0.0151
	(0.0060)	(0.0068)	(0.0091)	(0.0133)	(0.0139)	(0.0140)
Household status						
no dependents	0.3102	0.2975	-0.0127	0.0039	0.0056	0.0101
	(0.0094)	(0.0110)	(0.0144)	(0.0211)	(0.0217)	(0.0222)
1 dependent	0.1524	0.1667	0.0142	0.0007	0.0046	-0.0004
	(0.0073)	(0.0090)	(0.0115)	(0.0169)	(0.0176)	(0.0180)
2 or more dependents	0.2321	0.2187	-0.0134	-0.0003	0.0024	-0.0003
	(0.0085)	(0.0099)	(0.0131)	(0.0192)	(0.0199)	(0.0206)
homeowner	0.2857	0.2541	**-0.0316**	0.0095	0.0214	0.0172
	(0.0091)	(0.0105)	(0.0139)	(0.0203)	(0.0211)	(0.0215)
Veteran	0.2342	0.2465	0.0124	-0.0125	-0.0045	-0.0206
	(0.0086)	(0.0104)	(0.0134)	(0.0196)	(0.0202)	(0.0207)
Union/standby	0.3727	0.3767	0.0040	0.0037	0.0111	-0.0016
	(0.0098)	(0.0117)	(0.0152)	(0.0222)	(0.0230)	(0.0236)
Occupation						
professional	0.0732	0.0666	-0.0066	-0.0048	-0.0001	0.0085
	(0.0053)	(0.0060)	(0.0080)	(0.0117)	(0.0122)	(0.0125)
clerical	0.1381	0.1273	-0.0108	0.0038	0.0034	-0.0033
	(0.0070)	(0.0080)	(0.0106)	(0.0156)	(0.0163)	(0.0166)
sales	0.0552	0.0503	-0.0048	0.0009	0.0002	0.0024
	(0.0046)	(0.0053)	(0.0070)	(0.0102)	(0.0105)	(0.0109)
service	0.1103	0.1227	0.0123	0.0013	0.0032	0.0129
	(0.0063)	(0.0079)	(0.0101)	(0.0146)	(0.0152)	(0.0154)
agric., fishery, forestry	0.0262	0.0231	-0.0030	0.0016	0.0001	0.0011
	(0.0032)	(0.0036)	(0.0048)	(0.0070)	(0.0072)	(0.0074)
processing	0.0364	0.0347	-0.0016	-0.0055	-0.0072	-0.0037
	(0.0038)	(0.0044)	(0.0058)	(0.0085)	(0.0087)	(0.0091)
machine trades	0.0826	0.0874	0.0048	0.0028	0.0014	-0.0026
	(0.0056)	(0.0068)	(0.0088)	(0.0129)	(0.0133)	(0.0138)
benchwork	0.0568	0.0625	0.0057	0.0046	0.0063	0.0043
	(0.0047)	(0.0058)	(0.0075)	(0.0109)	(0.0114)	(0.0116)

Table 4

Covariate	Claim year: Control mean (1)	Claim year: NWS mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and NWS groups (year t – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
structural work	0.2628	0.2812	0.0185	-0.0092	0.0011	-0.0234
	(0.0089)	(0.0108)	(0.0140)	(0.0204)	(0.0211)	(0.0216)
miscellaneous	0.1512	0.1343	-0.0169	0.0057	-0.0062	0.0061
	(0.0072)	(0.0082)	(0.0109)	(0.0161)	(0.0166)	(0.0172)
Industry						
agriculture	0.0204	0.0197	-0.0008	0.0021	0.0008	0.0014
	(0.0029)	(0.0033)	(0.0044)	(0.0065)	(0.0066)	(0.0069)
mining	0.0020	0.0023	0.0003	0.0006	0.0007	-0.0000
	(0.0009)	(0.0012)	(0.0015)	(0.0022)	(0.0024)	(0.0023)
construction	0.2084	0.1840	**-0.0244**	0.0004	0.0018	-0.0108
	(0.0082)	(0.0093)	(0.0124)	(0.0182)	(0.0187)	(0.0192)
manufacturing	0.2272	0.2517	*0.0245*	-0.0038	-0.0011	0.0012
	(0.0085)	(0.0104)	(0.0134)	(0.0197)	(0.0204)	(0.0209)
transportation, utilities	0.0466	0.0428	-0.0038	-0.0020	-0.0017	-0.0025
	(0.0043)	(0.0049)	(0.0065)	(0.0095)	(0.0098)	(0.0102)
wholesale trade	0.0597	0.0480	*-0.0116*	0.0062	0.0013	0.0038
	(0.0048)	(0.0051)	(0.0070)	(0.0105)	(0.0111)	(0.0113)
retail trade	0.1655	0.1615	-0.0041	-0.0042	-0.0080	0.0003
	(0.0075)	(0.0089)	(0.0116)	(0.0169)	(0.0175)	(0.0180)
finance, ins., real estate	0.0262	0.0301	0.0039	0.0042	0.0027	-0.0011
	(0.0032)	(0.0041)	(0.0052)	(0.0077)	(0.0080)	(0.0079)
services	0.1778	0.1875	0.0097	0.0018	0.0051	0.0128
	(0.0077)	(0.0094)	(0.0122)	(0.0177)	(0.0184)	(0.0189)
government	0.0499	0.0538	0.0040	-0.0018	-0.0006	-0.0066
	(0.0044)	(0.0054)	(0.0070)	(0.0101)	(0.0104)	(0.0106)
unclassified	0.0163	0.0185	0.0022	-0.0035	-0.0011	0.0016
	(0.0026)	(0.0032)	(0.0041)	(0.0059)	(0.0063)	(0.0065)
Reason for job loss						
permanent layoff	0.1631	0.1534	-0.0097	-0.0043	0.0026	0.0005
	(0.0075)	(0.0087)	(0.0114)	(0.0167)	(0.0173)	(0.0177)
temporary layoff	0.2182	0.2216	0.0034	0.0037	0.0031	-0.0026
	(0.0084)	(0.0100)	(0.0130)	(0.0190)	(0.0196)	(0.0201)
contract/seasonal	0.1663	0.1551	-0.0112	-0.0051	-0.0066	-0.0017
	(0.0075)	(0.0087)	(0.0115)	(0.0168)	(0.0173)	(0.0178)
quit	0.1716	0.1771	0.0054	0.0099	0.0079	0.0102
	(0.0076)	(0.0092)	(0.0119)	(0.0176)	(0.0181)	(0.0187)
missing	0.2808	0.2928	0.0121	-0.0042	-0.0071	-0.0064
	(0.0091)	(0.0110)	(0.0142)	(0.0208)	(0.0216)	(0.0222)
Prior claim						
none	0.8002	0.7882	-0.0120	0.0004	-0.0035	-0.0045
	(0.0081)	(0.0098)	(0.0127)	(0.0186)	(0.0193)	(0.0198)
duration ≤ 15 weeks	0.1034	0.1007	-0.0027	0.0028	-0.0002	0.0093
	(0.0062)	(0.0072)	(0.0095)	(0.0139)	(0.0144)	(0.0149)
duration > 15 weeks	0.0964	0.1111	0.0147	-0.0031	0.0037	-0.0047
	(0.0060)	(0.0076)	(0.0096)	(0.0141)	(0.0145)	(0.0149)

Table 4

Covariate	Claim year: Control mean (1)	Claim year: NWS mean (2)	Claim year: Difference (treatment – control) (3)	Difference in differences between control and NWS groups (year *t* – claim year): Year 3 (4)	Year 6 (5)	Year 9 (6)
UI benefits/claim type						
weekly amount ($)	146.77	145.82	-0.95	-0.71	-0.37	-0.97
	(1.03)	(1.22)	(1.59)	(2.32)	(2.40)	(2.46)
maximum amount ($)	3,888.1	3,839.1	-49.0	-23.6	-19.8	-13.7
	(34.9)	(41.5)	(54.2)	(79.0)	(81.7)	(84.0)
potential benefit duration (weeks)	26.023	25.840	-0.183	-0.009	-0.049	0.059
	(0.100)	(0.124)	(0.159)	(0.229)	(0.236)	(0.243)
replacement rate (percent)[a]	61.197	62.020	0.823	0.155	-0.144	-0.547
	(0.440)	(0.527)	(0.687)	(1.005)	(1.035)	(1.060)
combined wage claim[b]	0.0454	0.0503	0.0050	-0.0008	-0.0043	-0.0081
	(0.0042)	(0.0053)	(0.0067)	(0.0096)	(0.0099)	(0.0099)
ex-service member claim	0.0298	0.0376	0.0078	-0.0026	0.0030	-0.0019
	(0.0034)	(0.0046)	(0.0057)	(0.0081)	(0.0083)	(0.0085)
federal employee claim	0.0168	0.0145	-0.0023	0.0015	0.0002	0.0033
	(0.0026)	(0.0029)	(0.0039)	(0.0056)	(0.0055)	(0.0057)
Reservation wage (hourly)						
≤ $5.00	0.1814	0.1748	-0.0067	0.0089	-0.0045	0.0015
	(0.0078)	(0.0091)	(0.0120)	(0.0175)	(0.0179)	(0.0183)
$5.01–$7.00	0.1373	0.1325	-0.0048	-0.0032	0.0051	0.0015
	(0.0070)	(0.0082)	(0.0107)	(0.0158)	(0.0163)	(0.0169)
$7.01–$10.00	0.1532	0.1649	0.0117	-0.0098	-0.0041	-0.0059
	(0.0073)	(0.0089)	(0.0115)	(0.0168)	(0.0176)	(0.0180)
$10.01–$20.00	0.1406	0.1279	-0.0127	0.0097	0.0096	0.0039
	(0.0070)	(0.0080)	(0.0107)	(0.0157)	(0.0163)	(0.0168)
> $20.00	0.0919	0.0966	0.0047	-0.0037	0.0005	0.0065
	(0.0058)	(0.0071)	(0.0092)	(0.0133)	(0.0137)	(0.0141)
Pre-claim earnings ($)						
3 years before	10,669	10,558	-111	40	85	106
	(232)	(277)	(362)	(525)	(542)	(556)
2 years before	11,919	11,576	-343	-201	45	-166
	(228)	(263)	(348)	(507)	(526)	(537)
1 year before	14,073	13,625	-448	-80	144	132
	(200)	(235)	(309)	(453)	(470)	(479)
Pre-claim hours						
3 years before	939.1	939.4	0.4	2.9	2.9	13.2
	(16.1)	(19.3)	(25.2)	(36.7)	(37.9)	(38.8)
2 years before	1,095.2	1,079.1	-16.1	-11.0	4.8	-4.7
	(15.2)	(18.1)	(23.6)	(34.6)	(35.7)	(36.7)
1 year before	1,340.7	1,299.2	**-41.5**	0.3	19.1	16.0
	(13.2)	(16.1)	(20.8)	(30.4)	(31.4)	(32.2)
Pre-claim earnings variability						
annual (CV)	0.6470	0.6496	0.0026	-0.0070	-0.0116	-0.0100
	(0.0106)	(0.0126)	(0.0165)	(0.0240)	(0.0246)	(0.0253)
seasonal (CV)	0.5729	0.5812	0.0083	0.0016	-0.0025	-0.0063
	(0.0077)	(0.0092)	(0.0120)	(0.0176)	(0.0181)	(0.0186)

Table 4

	Claim year			Difference in differences between control and NWS groups (year t – claim year)		
Covariate	Control mean (1)	NWS mean (2)	Difference (treatment – control) (3)	Year 3 (4)	Year 6 (5)	Year 9 (6)
Quarters earnings >0	8.8038	8.6973	-0.1065	0.0067	0.0706	0.0520
in 12 pre-claim quarters	(0.0670)	(0.0813)	(0.1053)	(0.1529)	(0.1573)	(0.1618)
< 6	0.1913	0.2083	0.0171	-0.0069	-0.0087	-0.0015
	(0.0080)	(0.0098)	(0.0126)	(0.0183)	(0.0187)	(0.0192)
6–8	0.1843	0.1777	-0.0066	0.0042	-0.0023	-0.0041
	(0.0078)	(0.0092)	(0.0121)	(0.0177)	(0.0183)	(0.0186)
> 8	0.6244	0.6140	-0.0104	0.0027	0.0110	0.0056
	(0.0098)	(0.0117)	(0.0153)	(0.0223)	(0.0230)	(0.0235)
Pre-claim drop in earnings						
between pre-claim year 3	0.3367	0.3437	0.0070	0.0023	-0.0078	0.0094
and pre-clam year 2	(0.0096)	(0.0114)	(0.0149)	(0.0218)	(0.0226)	(0.0232)
between pre-claim year 2	0.3327	0.3362	0.0036	-0.0016	-0.0034	-0.0150
and pre-clam year 1	(0.0095)	(0.0114)	(0.0148)	(0.0217)	(0.0224)	(0.0230)
Claimants with earnings >0						
Control group	2,447	n/a	2,447	2,195	1,930	1,785
NWS group	n/a	1,728	1,728	1,498	1,320	1,201

Source: See Table 1.

Bold indicates p-value < .05 for test of difference between control and NWS groups (column 3).

Italic indicates p-value < .10 for test of difference between control and NWS groups (column 3), or for test of difference in differences between control and NWS groups (year t – claim year) (column 6).

a. Weekly benefit amount as a percentage of pre-claim average weekly earnings.

b. Combined wage claims use earnings from more than one state to calculate base period earnings.

Table 4

Filing Not Found:
Which Injuries Go Unreported to Worker Protection Agencies, and Why?

Alison Morantz
Stanford Law School

Abstract:

The underreporting of occupational injuries and illnesses to worker protection agencies has become a topic of great concern to researchers and policymakers. Although numerous studies have quantified the prevalence of the phenomenon, which specific *types* of injuries and establishments are most susceptible to underreporting is poorly understood. As a consequence, regulators have very little capacity to "red flag" employers that are likely to underreport the most injuries. This study begins to fill this gap in existing literature in four interrelated ways. First, I develop a simple theoretical model of the relationship between regulatory intensity, injury type, and underreporting. The model yields a number of concrete predictions about how the frequency of injuries, and mix of injury types, will respond to changes in the frequency and/or stringency of audits. Secondly, I propose a scheme for classifying different types of injuries by their relative "detectability." Third, using a dataset comprised of granular audit data obtained from the Occupational Safety and Health Administration (OSHA) and the Mine Safety and Health Administration (MSHA), I test the model's predictions regarding which types of injuries will be underreported the most across regimes and over time. Finally, I explore whether any observable, establishment-level covariates – such as the percentage of injuries contained in regulatory filings that are highly detectable – could be used, in a manner akin to the IRS, to identify likely violators. Overall, the results provide considerable grounds for optimism that mining injury data in this fashion could provide useful insights. Not only do my findings bear out most predictions of the model, but they also suggest that empirical algorithms could be devised, based exclusively on observable firm- and injury-level characteristics, to help labor regulators identify employers that hide workplace injuries.

This paper was prepared with funding from the U.S. Department of Labor. The views expressed are those of the authors and should not be attributed to the Federal Government or the Department of Labor.

I. Introduction[1]

In recent years, the underreporting of occupational injuries and illnesses to federal agencies has become an issue of great concern to Department of Labor (DOL) officials, Congress, and labor scholars alike. A growing body of scholarship suggests that injury and illness statistics recorded by employers at the behest of Department of Labor agencies, such as the Bureau of Labor Statistics (BLS), the Occupational Safety and Health Administration (OSHA), and the Mine Safety and Health Administration (MSHA), significantly undercount the true rates of occupational injury and illness.[2] Some researchers have estimated that as many as 70 percent of all injuries go unreported.[3] The underreporting phenomenon is not confined to minor injuries: one recent study suggests that at least one quarter of all amputations are not reported to OSHA.[4]

The deficiencies in the current injury and illness surveillance system compromise regulators' capacity to keep workers safe in several critical ways. First, customized injury prevention programs

---

[1] This project was funded by a grant from the Department of Labor's DOL Scholars Program. I could not have completed the project without the skilled research assistance of Austin Alleman, Rajlakshmi De, Kathleen Choi, and Alex Weiss. I am also deeply indebted to my colleague, Dr. Mark Cullen, Chief of Stanford University's Division of General Internal Medicine, for his extraordinary generosity in helping me devise a medically sound typology for identifying "hard-to-attribute" injuries.

[2] See, e.g., William J. Wiatrowski, "Examining the Completeness of Occupational Injury and Illness Data: An Update on Current Research," *Monthly Labor Review* (June 2014); Xiuwen S. Dong, et al., "Injury Underreporting Among Small Establishments in the Construction Industry," *American Journal of Industrial Medicine* 54 (2011): 339-349; Tahira M. Probst, Ty L. Brubaker, and Anthony Barsotti, "Organizational Injury Rate Underreporting: The Moderating Effect of Organizational Safety Climate," *Journal of Applied Psychology* 93 (2008): 1147-54; Monica Galizzi, et al., "Injured Workers' Underreporting in the Health Care Industry: An Analysis Using Quantitative, Qualitative, and Observational Data," *Industrial Relations* 49 (2010): 22-43; S. A. McCurdy, et al., "Reporting of Occupational Injury and Illness in the Semiconductor Manufacturing Industry," *American Journal of Public Health* 81 (1991): 85-89; John W. Ruser, "Examining Evidence on Whether BLS Undercounts Workplace Injuries and Illnesses," *Monthly Labor Review* (August 2008): 20-32; Alison Morantz, "Coal Mine Safety: Do Unions Make a Difference?" *Industrial and Labor Relations Review* 66 (2013): 88-116; Leslie Boden, and Alexander Ozonoff, "Capture-Recapture Estimates of Nonfatal Workplace Injuries and Illnesses," *Annals of Epidemiology* 18 (2008): 500-506; Leigh J. Paul, James P. Marcin, and Ted R. Miller, "An Estimate of the US Government's Undercount of Nonfatal Occupational Injuries," *Journal of Occupational and Environmental Medicine* 46 (2004): 10-18; K. D. Rosenman, et al., "How Much Work-Related Injury and Illness Is Missed By the Current National Surveillance System?" *Journal of Environmental Medicine* 48 (2006): 357-365.

[3] See Ruser (2008).

[4] See Wiatrowski (2014). As part of new severe injury and illness reporting requirements that went into effect in January of 2015, OSHA began requiring all employers under its jurisdiction to notify the agency within 8 hours of any work-related fatality, and to notify it within 24 hours of any hospitalization, amputation or enucleation. For a summary, see http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=NEWS_RELEASES&p_id=26673.

designed to assist workers in hazardous occupations will be of limited value if the data paint a distorted portrait of the distribution of injuries across industries and tasks.[5] Secondly, underreporting undermines the experience rating system upon which workers' compensation is based, whereby the most dangerous employers pay the highest insurance premiums. If firms can effectively shield workplace injuries from scrutiny – so that many injuries are neither reported to DOL nor trigger the filing of workers' compensation claims – the system will fail to induce the riskiest employers to invest more in accident prevention.[6] Finally, if firms that underreport injuries disproportionately violate other labor laws, the incapacity to identify the worst violators will hamper any coordinated regulatory strategy to protect workers' rights.[7]

Awareness of the underreporting problem has grown considerably since the turn of the millennium. In the mid-2000s, Congress held hearings on the issue and allocated funds to OSHA, BLS, and the National Institute of Occupational Safety and Health (NIOSH) for the express purpose of "follow[ing] up and expand[ing] on the previous research so as to understand the nature and magnitude of any undercount [of occupational injuries and illnesses] and attempt to identify solutions."[8] Yet to date, scholars still know relatively little about the prevalence and distribution of underreporting behavior across the U.S. economy. Most prior literature has compared employer-reported injury data to state workers' compensation filings to determine how many injuries are omitted from each data source. Much prior research also confines the scope of inquiry to a single

[5] William J. Wiatrowski, "Using Workplace Safety and Health Data for Injury Prevention," *Monthly Labor Review* (October 2013).

[6] See Boris Kralj, "Employer Responses to Workers' Compensation Insurance Experience Rating," *Industrial Relations* 49 (Winter 1994): 41-61; Sidney A. Shapiro, "Occupational Safety and Health Regulation," *Encyclopedia of Law and Economics, Cheltenham, Edward Elgar* 5540 (2000): 596-625.

[7] "[O]rganizations that have a lower commitment to safety are more likely to inadvertently or otherwise skew their injury data such that they appear to have similar safety outcomes as organizations with a positive safety climate." See Probst, Brubaker, and Barsotti (2008).

[8] See Wiatrowski (2014): 1-3.

industry or injury. Although often yielding valuable insights, these approaches sometimes yield biased estimates of the prevalence of underreporting to federal agencies, while failing to illuminate which *types* of injuries and workplaces are most vulnerable to underreporting.[9]

A few scholars have generated, and occasionally tried to test, hypotheses about which types of injuries are most likely to be underreported. As early as 1982, a study by the National Research Council conjectured that a subset of mining injuries that it labeled "intermediate" was less prone to underreporting than total injuries, although this assumption has never been empirically verified.[10] Several studies of particular industries[11] or discrete regulatory contexts[12] seem to bear out the notion that certain injuries – such as musculoskeletal injuries[13] and injuries occurring in smaller firms[14] and certain industrial sectors[15] – are highly vulnerable to underreporting. These findings are reminiscent of a parallel body of literature in the workers' compensation arena suggesting that some injuries,

---

[9] For a general, detailed discussion of these biases, see Wiatrowski (2014): 1-5. A recent (unpublished) report prepared by Eastern Research Group, and presented to the U.S. Department of Labor, also illustrates the pitfalls of relying exclusively on comparisons of workers' compensation claims to injuries reported to regulatory agencies. The report tabulated differences in "match rates" across workers' compensation records and MSHA Part 50 data by type of injury across two jurisdictions. See Eastern Research Group, *Final Report (Revised): Evaluation of the Accuracy and Completeness of Nonfatal Injury and Illness Reporting in the Mining Industry* (Lexington, MA: Eastern Research Group, June 11, 2013): 24-25, 34-36, http://www.dol.gov/asp/evaluation/reports/MSHA-Part50-Underreporting.pdf. Since the workers' compensation records used to conduct the matches were highly incomplete and themselves prone to underreporting, and there was no one-to-one correspondence between the categories used across the two datasets, no credible inferences could be drawn regarding whether certain types of injuries were more likely to be underreported than others. See also the report released by the U.S. Department of Labor, Office of the Inspector General (OIG), *MSHA Has Taken Steps to Detect and Deter Underreporting of Accidents and Occupational Injuries and Illnesses, But More Action Is Still Needed* (March 31, 2014), http://www.oig.dol.gov/public/reports/oa/2014/05-14-001-06-001.pdf.

[10] National Research Council, *Toward Safer Underground Coal Mines* (Washington, D.C.: National Academy Press, 1982).

[11] See Alison Morantz and Alexandre Mas, "Does Post-Accident Drug Testing Reduce Injuries? Evidence from a Large Retail Chain," *American Law and Economics Review* 10 (2008): 246-302, and Morantz (2013).

[12] See, e.g., Rosenman et al. (2006), which generates capture-recapture estimates of injuries reported in Michigan to multiple surveillance system, including the Bureau of Labor Statistics.

[13] See, e.g., Christine Daniels, and Peter Marlow, "Literature Review on the Reporting of Workplace Injury Trends," *Health and Safety Laboratory* 36 (2005); Wiatrowski (2014).

[14] See, e.g., Katherine L. Hunting, and James L. Weeks, "Transport Injuries in Small Coal Mines: An Exploratory Analysis," *American Journal of Industrial Medicine* 23 (1993): 391-406; Dong et al. (2011); Daniels and Marlow (2005); A. Oleinick, J. V. Gluck, and K. E. Guire, "Establishment Size and Risk of Occupational Injury," *American Journal of Industrial Medicine* 28 (1995): 1-21.

[15] See, e.g., Daniels and Marlow (2005).

especially hard-to-diagnose ones, are unusually prone to moral hazard effects.[16] Since approximately 2009, the Bureau of Labor Statistics has funded several independent, state-level studies that compare BLS non-fatal injury records with state workers' compensation data, match multiple-source data for cases involving work-related amputations and carpal tunnel syndrome, and interview employers about recordkeeping practices. Some of these studies have begun bearing fruit.[17] Finally, a few scholars in the field of occupational medicine have explored which injuries are most likely to go untreated because they are never reported to health care providers.[18] Yet no comprehensive mapping of the relationship between injury type and underreporting – or the practical implications of such relationships for the enforcement of occupational safety and health laws – has been attempted in prior work.

The paucity of knowledge regarding which types of injuries are most likely to evade detection limits the capacity of worker protection agencies to combat underreporting. The injury logs submitted to DOL contain granular detail on the frequency, type, and cause of occupational injuries and illnesses that occur each calendar year (or quarter). Yet like individuals that file income tax returns, employers have strong incentives to hide information they are statutorily obliged to report. In theory, DOL's worker protection agencies could emulate the Internal Revenue Service's fraud detection division and use statistical algorithms to target the worst violators. Just as an IRS official

---

[16] See, e.g., Georges Dionne and Pierre St-Michel, "Workers' Compensation and Moral Hazard," *The Review of Economics and Statistics* 73 (May 1991): 236-244.

[17] See, e.g., Wiatrowski (2014).

[18] For example, see K. D. Rosenman, et al., "Why Most Workers with Occupational Repetitive Trauma Do Not File for Workers' Compensation," *Journal of Occupational and Environmental Medicine* 42 (2000): 25–34; Laura Welch, and Katherine Hunting, "Injury Surveillance in Construction: What is an 'Injury', Anyway?" *American Journal of Industrial Medicine* 44 (2003): 191-196; David L. Parker, et al., "Characteristics of Adolescent Work Injuries Reported to the Minnesota Department of Labor and Industry," *American Journal of Public Health* 84 (1994): 606-611; Kris Siddharthan, et al., "Under-Reporting of Work-Related Musculoskeletal Disorders in the Veterans Administration," *International Journal of Health Care Quality Assurance* 19 (2006): 463-476; Oleinick, Gluck, and Guire (1995); Lenore S. Azaroff, Charles Levenstein, and David H. Wegman, "Occupational Injury and Illness Surveillance: Conceptual Filters Explain Underreporting," *American Journal of Public Health* 92 (2002): 1421-1429; Z. Joyce Fan, et al., "Underreporting of Work-Related Injury or Illness to Workers' Compensation: Individual and Industry Factors," *Journal of Occupational and Environmental Medicine* 48 (2006): 914-922.

might compare self-employment income or restaurant tips to state medians to help ferret out tax evasion, a DOL official could scrutinize the composition of injury logs for telltale signs of underreporting. Yet to date, no concerted attempt has been made to devise statistical algorithms that could help OSHA or MSHA channel regulatory resources toward the most likely offenders.

The present study opens up this line of inquiry in several interrelated ways. First, I develop a simple theoretical model of the relationship between regulatory intensity, injury type, and underreporting behavior. Starting with the simple observation that injuries differ in their relative chances of detection, the model presumes that the likelihood of underreporting will decline monotonically as detectability increases. It also yields a number of concrete predictions about how the frequency of injuries and mix of injury types will respond to changes in the frequency and/or stringency of audits. For example, the more intense the regulatory regime, the less detectable the type of injury that will best predict underreporting. In contrast to much existing literature, the model also implies that *ceteris paribus*, an increase in the frequency of audits and in their relative stringency can have very different effects. Whereas an increase in inspection frequency can have a similar deterrent effect on all types of injuries, a change in the thoroughness of inspections will have a disproportionate effect on the reporting of less-detectable injuries.

The paper's second important contribution is to propose a scheme, grounded in occupational medicine, for categorizing non-fatal injuries by their relative "detectability." I classify all occupational injuries into four types using three criteria: whether they are severe, whether they are traumatic, and whether they are easily attributed to activities performed on the job.

Third, using a dataset comprised of granular audit data from OSHA and MSHA, I test the model's predictions regarding which types of injuries will be underreported the most across regimes and over time. This task is not straightforward for two reasons. First, since the number of reporting

violations detected is largely a function of the stringency of audits performed, "audit stringency bias" can confound my empirical estimates of the magnitude of underreporting. Secondly, the prevalence of behavioral incentive programs that penalize workers for reporting injuries (or reward them for not doing so) can cause "stickiness" in firms' responses to fluctuations in audit stringency. Both factors complicate my identification strategy. Nevertheless, my empirical findings broadly bear out the predictions of the model.

Finally, I explore whether any establishment-level covariates – such as the percentage of reported injuries that are highly detectable – could be used as red flags, in a manner akin to the IRS, to target the most likely violators. Several of the red flags explored do turn out to have significant predictive value. For example, a reported percentage of *severe* injuries that places an employer in the top quartile of the sample is a significant predictor of the frequency of underreported injuries in the OSHA environment, and a reported percentage of *easy-to-attribute* injuries that places the employer in the top quartile of the sample is a significant predictor of the frequency *and* percentage of underreported injuries in the MSHA environment. On the basis of these findings, I suggest that DOL's worker protection agencies would do well to emulate their peers at the IRS by developing new data-mining techniques to combat underreporting. Although the likelihood that firms will behave strategically in a more dynamic, iterative enforcement environment complicates the task of translating theory into practice, I suggest that this challenge is probably surmountable.

The next section, Section Two, describes the origin and construction of the datasets analyzed. Section Three presents a simple, non-mathematical model of the relationship between injury detectability, inspection intensity, and underreporting. Section Four describes my empirical methodology. Section Five presents the study's key findings. Section Six considers several practical challenges involved in policy implementation, and Section Seven concludes.

II. Description of Data

Each year, MSHA and OSHA collect self-reported injury and illness data from tens of thousands of private establishments. MSHA receives Part 50 data on a quarterly basis from every mine nationwide, and OSHA receives similar information from a sample of about 80,000 establishments in high-risk industries selected for inclusion in its annual OSHA Data Initiative (ODI) survey.

The OSHA data used throughout the analysis were obtained from two auditing programs that jointly encompass the years 1997 through 2012. From October 1997 through February 2009, OSHA audited injury and illness records from high-hazard industries through its Audit and Verification Program - Occupational Injury and Illness Records ("Recordkeeping" or "RK") program. Since each audit typically examined injury logs from two years prior, the RK audits collectively examined injury logs from 1996-2006. In 2009, OSHA suspended the RK program and instead began conducting audits under the auspices of its short-lived Injury and Illness Recordkeeping National Emphasis Program ("NEP"). The NEP program, which ran from September 2009 through February 2012, audited injury records from 2007 through 2009. (Whereas each RK audit examined just one year of an establishment's injury and illness history, NEP audits spanned two calendar years.) The average annual frequency of audits declined slightly under the NEP program.[19] By combining information from these two OSHA programs, I created a dataset encompassing injury logs from the years 1996-2009.

The RK and NEP programs were substantively similar in many regards. Both required auditors to ascertain whether the employer's log of total injury and illness counts, employment numbers, and hours worked for the audited year matched the summary data submitted to OSHA; to

[19] Whereas the number of audits conducted from 1996-2006, under the auspices of the RK auditing program, exceeded 225 in every year except 1996, the average number of audits conducted annually under the NEP program (2007-2009) was 200. (Calculations are based on an analysis of data received from Dave Schmidt, OSHA.)

check the accuracy and completeness of records for a random sample of employees drawn from the pertinent employee roster; and to interview the establishment's designated record-keeper to "assess [the] recordkeeper's knowledge of the OSHA injury/illness recordkeeping requirements and to determine whether recordkeeping problems exist."[20]

Yet the programs also differed in critical respects. First, the selection criteria used to determine which establishments were audited varied somewhat across the two programs. While both programs generally focused on firms in high-hazard industries, the NEP program added an additional selection criterion that compared each firm's "DART" rate (the rate of injuries and illnesses resulting in days away from work, restricted work, or transfer to another position) to the mean DART rate among high-hazard industries.[21] Secondly, unlike their RK counterparts, NEP auditors were obliged to interview managers to determine the existence of "any incentive or disciplinary programs that may influence recordkeeping."[22] Third, although employee interviews were optional under the RK program, they were mandatory under the NEP program, and the NEP auditor was further instructed to "focus interviews on employees likely to be injured or ill."[23] Finally, if an NEP auditor found any evidence of underreporting, (s)he had the right to "expand the records inspection beyond the

[20] See, e.g., U.S. Department of Labor, Occupational Safety and Health Administration, *Audit and Verification Program of Occupational Injury and Illness Records,* Directive Number: CPL 02-00-138, January 12, 2006, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=3329; U.S. Department of Labor, Occupational Safety and Health Administration, *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*, Directive Number: 09-08 (CPL 02), September 30, 2009, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_09-08.pdf; and U.S. Department of Labor, Occupational Health and Safety Administration, *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*, Directive Number: 10-07 (CPL 02), September 28, 2010, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_10-07.pdf.

[21] Ibid. For example, while the RK's sample frame included *all* establishments above a given size cutoff that were part of the OSHA Data Initiative, the NEP program used a firm's reported Days Away and Restriction or Transfer (DART) rate as an additional selection criterion. The minimum size threshold in effect under the RK program was also lowered twice: once in late 1998 (from 60 to 50 employees) and again in December of 1999 (from 50 to 40 employees).

[22] Ibid.

[23] Ibid.

sampled employees," whereas RK auditors had no such authority.[24] In short, the NEP audits inspections were more stringent and comprehensive than RK audits. The data confirm, as one would expect, that reliance on employee interviews increased dramatically under the NEP program.[25]

Changes undertaken partway through the implementation of the NEP program, however, created sharp discontinuities in the mix of firms audited. During the first year of the program (Sept. 2009-Sept. 2010), which audited injury logs from 2007,[26] the agency targeted firms whose reported injury rates fell *below* the mean for high-hazard industries, 4.2 per 100 full-time employees. In the final 17 months of the program (Sept. 2010 – Feb. 2012), however, which audited injury logs from 2008 and 2009, the agency took the opposite approach of targeting firms whose reported DART rates fell *above* 4.2. As a consequence, injury rates from before and after January 2008 cannot be meaningfully compared.[27]

Information gleaned from OSHA audits is collected and stored at the audit level, making it very well suited to a study of this type. For each audit conducted, the dataset records information on both reported and unreported injuries. By examining the complete "injury set" for a given year, one

---

[24] Ibid.

[25] An average of 0.24 employee interviews per audit were conducted under the RK program, while the comparable figure for NEP audits was 3.19.

[26] In addition to the 2007 logs, injury logs from 2008 were also audited from February through September of 2010. See US DOL, OSHA, Directive Number: 10-07 (CPL 02) (2010); U.S. Department of Labor, Occupational Safety and Health Administration, *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*, Directive Number: 10-02 (CPL 02), February 19, 2010, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_10-02.pdf.

[27] In addition to the dramatic change in selection criteria after the 2007 injury logs were audited, firms evidently were given considerable *de facto* leeway to "correct" injury logs from prior years. Upon arriving at a workplace, each OSHA inspector was instructed to re-calculate the DART rate using the OSHA 300 Log for 2007 and "[i]f the re-calculated 2007 DART rate [fell] above the cutoff rate of this NEP (i.e., >4.2) the records inspection [would] not be conducted." See US DOL, OSHA, Directive Number: 09-08 (CPL 02) (2009), and US DOL, OSHA, Directive Number: 10-02 (CPL 02) (2010). In other words, if a firm responded to the announcement of the NEP program by retroactively recording more injuries on its 2007 logs, the audit would not be conducted and the firm would have been dropped from the dataset. A phone conversation with an employee of OSHA's Directorate of Enforcement Programs, September 4, 2014, also confirmed that this sometimes might have taken place. Unfortunately, I cannot discern from the data how frequently, if ever, such injury "backfilling" occurred. If it did occur, and firms with the most unreported injuries were especially likely to "backfill" their injury logs during the NEP program, then the figures presented here would probably *underestimate* the total number of injuries from 2007 that went unreported.

can ascertain how many (and what proportion of) injuries the audited firm failed to report. Moreover, for each audited establishment, the dataset records state, union status, total workers employed, and four-digit SIC code.[28] Importantly, however, the injury sets do *not* contain unique firm-level identifiers, nor do they contain the name or address of audited establishments. Therefore, audit results cannot be linked to other historical data on the same establishment, such as violations cited during prior OSHA inspections. Nor is it possible to determine if any firms were audited more than once.[29]

The MSHA dataset, encompassing the years 1992-2012, contains information gleaned from three different audit-like activities.[30] First, "compliance checks" performed during ordinary inspections include at least a cursory review of injury and illness records, which sometimes leads to the discovery of unreported injuries or illnesses. The second type, "Part 50" audits, have been undertaken since 1979 to assess mine operators' adherence with their statutory obligation to report all injuries and illnesses on a quarterly basis.[31] Thirdly, Potential Pattern of Violation ("PPOV") audits have been conducted since 2007 as part of a broader effort to single out the most dangerous mines for enhanced regulatory scrutiny. Mines that qualify for PPOV status on every criterion except their reportedly low rate of injury severity (total lost workdays X 200,000 / total hours worked) are targeted for PPOV audits on the theory that their numbers look "too good to be true" and thus could

[28] An SIC manual containing a list of these codes can be found on the OSHA website, https://www.osha.gov/pls/imis/sic_manual.html.

[29] One can obtain an upper bound for the prevalence of repeat audits by analyzing combinations of two establishment-specific fields: state and four-digit Standard Industrial Classification (SIC) code. If a given state and SIC code combination appears exclusively in a single audit-year, one can rest assured that the establishment in question was only audited once. Of the 2,019 unique combinations of state and four-digit SIC that appear in my data, 74.9% appear exclusively in a single audit-year. Therefore, no more than a quarter of the OSHA data utilized for the study could involve establishments that were audited repeatedly.

[30] Because of the small number of observations and somewhat ad hoc nature of its construction, the MSHA dataset does not include data from every year in this range. The final version includes injuries from the years 1992, 1994, 1996, 1997, and 2001-2012. Before 2001, however, the data only includes one mine for each year.

[31] According to Beth Nettles at MSHA, Part 50 audits dating back to this year appear in the MSHA database.

be inaccurate. When carrying out a PPOV audit, MSHA officials examine the most recent year (i.e. four quarters) of injury data.[32]

The scope and rigor of these three types of MSHA audits vary substantially. At one extreme are the PPOV audits, which are conducted by specially trained auditors and/or Education and Field Safety (EFS) staff, and scrutinize a broad array of data sources including medical records, medical claims forms, drug screening documents, employee interviews, and Part 50 information filed by mine operators. Since the program's inception, most (83%) of the mines subjected to PPOV audits have been underground mines.[33] At the opposite extreme are compliance checks conducted by ordinary MSHA inspectors, who simply compare data from injury logs (1000-1 and 7000-2 forms) and timesheets to the Part 50 information filed with MSHA's Denver office. Part 50 audits lie somewhere between these two extremes: although auditors are specifically tasked with finding unreported injuries, and scrutinize some of the same data sources examined in PPOV audits, they need not, and rarely do, examine them all.[34]

Unlike OSHA, MSHA does not store audit data at the injury level. For a given audited establishment, therefore, there is no straightforward way to compare injuries that were reported to those that were not. The only way to do so is to compare injuries that were *timely* reported in Part 50 filings with injuries that were *untimely* reported after the close of the fiscal (reporting) year. MSHA personnel indicated that injuries reported after the close of the fiscal year typically were reported at the behest of regulators following a compliance check or formal audit.[35] For the purposes of the

---

[32] Correspondence with Jay Mattos, MSHA, September 26, 2014.

[33] See Pattern of Violations Fact Sheet, http://www.msha.gov/POV/POVsinglesource.asp (last visited 9/26/2014) and "Mines Having Received POV or PPOV Notifications," available at http://www.msha.gov/POV/povmines.asp (last visited 9/26/2014).

[34] These institutional details are summarized in Eastern Research Group, *Final Report (Revised)* (2013).

[35] Telephone conference with Beth Nettles, MSHA, September 23, 2013.

empirical analysis, therefore, I compare two groups of injuries: those that were reported *after* the close of the fiscal year, and those that were reported by the same mine(s) in the same year(s) that the late-reported injuries occurred. For example, if a mine reported in 2013 an injury that took place in 2010, that injury would be compared to all injuries that the same mine timely reported in 2010. Because of these limitations, it is impossible to determine which respective proportions of late-reported injuries came to light as a result of PPOV audits, Part 50 audits, and compliance checks. Nor can I ascertain with certainty what number (or proportion) of injuries a given firm underreported in a given calendar year. Although mindful of these limitations, I create a "quasi injury set" for each mine that reported an injury after the close of the fiscal year. The quasi injury set contains both timely- and untimely-reported injuries from the year in question. Despite the likelihood of some measurement error, this methodology enables me to observe detailed injury information from all mines in the sample.

The substantive content of the OSHA and MSHA databases is largely similar. Both contain information on the source of injury, degree of injury, nature of injury, and affected body part. Both also enable me to observe the number of employees, state, year, and industry code. Yet there are a few salient differences. Although union status is available for all establishments audited by OSHA, in the mining sector union status is only available for coal mines.[36] Finally, the MSHA data contain information on "canvass code" (an industry coding scheme which the agency deems more useful than NAICS code), and on whether the mine operates at the surface or underground.

Both datasets are also susceptible to selection bias. The injury-reporting behavior of firms that are singled out for PPOV audits (in the case of MSHA) or NEP audits (in the case of OSHA) may differ from the behavior of firms that are not so targeted. Moreover, in the case of MSHA, firms that

[36] MSHA does not record data on union status, but the Department of Energy's Energy Information Administration (EIA) does record union status for coal mines on an annual basis. Using unique mine-level identifiers, I was able to append the union status field to the MSHA dataset.

underreported *no* injuries, or whose unreported injuries were never brought to light, are excluded entirely from the sample. Although the likelihood of selection bias (especially in the MSHA sample) may compromise the validity of out-of-sample predictions, the data still enable me to test several important predictions of the model, and may serve as useful bases for comparison in future studies that apply similar methodologies to a sample of randomly-selected establishments.

III. Injury Detectability, Regulatory Intensity, and Underreporting: A Simple Model

Although individuals may occasionally deviate from rational decision making due to cognitive biases or other behavioral limitations, most economic models presume that for-profit corporations operating in competitive markets engage in profit-maximizing behavior.[37] In the workplace safety context, profit maximization implies that a firm will compare the anticipated costs and benefits of abating a particular workplace hazard with the costs and benefits of allowing the hazard to persist. *Ceteris paribus*, an increase in the frequency or rigor of regulatory scrutiny raises the odds that a given hazard will be detected and penalized, thereby inducing rational firms to invest more in abatement. A sizable body of empirical literature has borne out these simple theoretical predictions.[38]

Parallel logic applies to the *reporting* of occupational injuries, which although mandated by statute (or regulation) is justifiably perceived as costly to employers. For example, reporting an

[37] See, e.g., Fritz Machlup, "Theories of the Firm: Marginalist, Behavioral, Managerial," *The American Economic Review* 57 (1967): 1-33; Jean-Jacques Laffont and David Martimort, *The Theory of Incentives: The Principal-Agent Model* (Princeton, NJ: Princeton University Press, 2002). Principal-agent theory – that is, the potential divergence between the incentives of a corporate entity *per se* and the incentives of the individual executives who make decisions on its behalf – is the leading paradigm used to explain instances in which corporations *do* exhibit apparent deviations from rationality. The pressure of market competition can play a significant role in mitigating such market imperfections.

[38] See, e.g., S. Lewis-Beck, and John R. Alford, "Can Government Regulate Safety? The Coal Mine Example," *The American Political Science Review* 74 (1980): 745-756; Wayne B. Gray, and John T. Scholz, "Does Regulatory Enforcement Work? A Panel Analysis of OSHA Enforcement," *Law and Society Review* 27 (1993): 177-214; David Weil, "If OSHA Is So Bad, Why Is Compliance So Good?" *The Rand Journal of Economics* 27 (1996): 618-640; Leon S. Robertson, and J. Philip Keeve, "Worker Injuries: The Effects of Workers' Compensation and OSHA Inspections," *Journal of Health Politics, Policy and Law* 8 (1983): 581-597.

injury makes it far more likely that the injury will be processed through the workers' compensation system, which (for all but the smallest employers) increases experience-rated premiums.[39] Full reporting of injuries may also harm the firms' eligibility for lucrative contracts[40], undermine workplace morale[41], place upward pressure on wages[42], and tarnish the firm's reputation.[43] Firms may also justifiably believe that reporting each and every injury will subject them to greater regulatory scrutiny, especially if underreporting is commonplace among their competitors.[44] For all these reasons, firms have strong incentives *not* to report all occupational injuries to worker protection agencies.

---

[39] Since it is well understood that experience rating itself encourages firms to underreport injuries, some researchers have even questioned whether the system has played an important role in lowering overall injury rates. See M. Harcourt, H. Lam, and S. Harcourt, "The Impact of Workers' Compensation Experience-Rating on Discriminatory Hiring Practices," *Journal of Economic Issues* 41 (2007): 681-99.

[40] A GAO report found, "Many employers did not report workplace injuries and illnesses for fear of increasing their workers' compensation costs or hurting their chances of winning contracts." Stephen Greenhouse, "Work-Related Injuries Underreported," *The New York Times,* November 16, 2009, http://www.nytimes.com/2009/11/17/us/17osha.html. Similarly, a Lockton Companies' report states that OSHA Incidence Rates are often used when deciding to hire a contractor. See Steven Polich, "Do Not Underestimate the Importance of OSHA Incidence Rates," *Lockton Companies,* December 2012, http://www.lockton.com/Resource_/PageResource/MKT/Polich_OSHA%20incidence%20rates_Dec%2012%20update.pdf. A Business Roundtable report likewise recommends that OSHA incidence rates be considered when selecting contractors for construction projects. See "Improving Construction Safety Performance," Report A-3, *The Business Roundtable* (January 1982).

[41] See Julian Barling, E. Kevin Kelloway, and Roderick D. Iverson, "Accidental Outcomes: Attitudinal Consequences of Workplace Injuries," *Journal of Occupational Health Psychology* 8 (2003): 74-85.

[42] See, e.g., Thomas J. Kniesner, and John D. Leeth, "Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States," *Journal of Risk and Uncertainty* 4 (1991): 75-90.

[43] For example, one popular business magazine advised its clientele, "Injuries tarnish a company's reputation and erase years of marketing gains." See Phil La Duke, "What Every Entrepreneur Should Know about Worker Safety," *Entrepreneur*, May 30, 2014, http://www.entrepreneur.com/article/234305. Also, Williams and Barrett find a negative relationship between a firm's reputation and the number of OSHA and Environmental Protection Agency (EPA) violations at the firm. See, e.g., Robert J. Williams, and J. Douglas Barrett, "Corporate Philanthropy, Criminal Activity, and Firm Reputation: Is There a Link?" *Journal of Business Ethics* 26 (2000): 341-350.

[44] As noted in a *New York Times* article summarizing the findings of a GAO report, "[m]any employers fear that reporting numerous injuries will prompt a full-scale OSHA inspection." See Greenhouse (2009). OSHA's site-specific targeting program, for example, targets firms for inspection based on their reported incidence rates. See U.S. Department of Labor, Occupational Safety and Health Administration, *Site-Specific Targeting 2014 (SST-14),* Directive Number: 14-01 (CPL 02), March 6, 2014, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02-14-01.pdf. High rates of injury or illness may also subject mines for increased enforcement by MSHA. See US DOL, OIG, *MSHA Has Taken Steps* (2014).

Yet failing to timely report an injury can expose a firm to financial penalties and reputational harms beyond those arising from the injury itself. Under OSHA's NEP program, for example, employers found to have violated their recordkeeping requirements were subjected to substantial monetary penalties, especially in cases where OSHA exercised its discretion to "stack" penalties for multiple violations.[45] In two well-publicized cases, for example, OSHA assessed firms $1,215,000[46] and $2,590,000[47] in penalties for their repeated failure to report occupational accidents.

If the frequency and stringency of regulatory inspections are largely exogenous and firms treat them as such, one would expect risk-neutral, profit-maximizing firms to consider both the costs of reporting and the costs of failing to report in choosing which, and how many, injuries to include in regulatory filings. Any regulatory change that lowers the risk that unreported injuries are detected and penalized, such as a cut in federal funding for recordkeeping audits, should increase the prevalence of underreporting. Conversely, any factor that *increases* the expected cost of underreporting – such as a rise in average penalties or an increase in the likelihood of detection – should lower the prevalence and magnitude of underreporting.

These simple predictions can be usefully extended, however, if one accounts for the fact that not all injuries are created equal. The likelihood that an injury can be successfully hidden from regulatory scrutiny varies by injury type. For example, occupational fatalities are very difficult for

---

[45] See Patrick Melfi, "New OSHA Initiative Targets Underreporting of Workplace Injuries," *New York Labor and Employment Law Report,* February 22, 2010, http://www.nylaborandemploymentlawreport.com/2010/02/articles/occupational-safety-and-health/new-osha-initiative-targets-underreporting-of-workplace-injuries.

[46] In 2010, OSHA "issued Goodman Manufacturing Co. LP 83 willful citations for failing to record and improperly recording work-related injuries and illnesses at the company's Houston air conditioning cooling facility. Proposed penalties total $1,215,000." See U.S. Department of Labor, Occupational Safety and Health Administration, *US Department of Labor's OSHA Cites Houston Manufacturing Company for Hiding Work-Related Injuries and Illnesses; Fines Exceed $1.2 Million,* Release Number: 10-1179-DAL, September 1, 2010, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=NEWS_RELEASES&p_id=18261.

[47] In 1987, OSHA assessed a $2.59 million fine against IBP Inc., a large meatpacker, for failing to report more than a thousand job-related injuries and illnesses over a two-year period. See Philip Shabecoff, "OSHA Seeks $2.59 Million Fine for Meatpacker's Injury Reports," *The New York Times,* July 22, 1987, http://www.nytimes.com/1987/07/22/us/osha-seeks-2.59-million-fine-for-meatpacker-s-injury-reports.html.

firms to shield from regulatory purview. This is so not only because of their extreme salience to coworkers and surviving family members, but also because of the unparalleled intensity of regulatory scrutiny.[48] Conversely, cumulative low-back injuries whose underlying cause and work-relatedness are often difficult to ascertain may be quite easy for firms to underreport. Not only may managers question whether the injury is truly work-related,[49] but they can easily direct the injured worker to providers outside the occupational safety and health system, such as group health plan clinicians.[50] In the latter scenarios, little if any paperwork is typically filed by the employer. The only way the injury may come to light is if the employee persists in filing a workers' compensation claim, or if an auditor learns of the injury during an interview with an affected worker.

To illustrate the choice confronting an employer, it is helpful to visualize the full range of injuries as falling along an "injury detectability continuum," represented by the horizontal (*x*) axis in Figure 1. The term "detectability," as used throughout this paper, encompasses not merely the extent to which an injury is visible to onlookers, but more broadly, the likelihood that an employer could realistically, and with impunity, avoid disclosing it to regulators. The vertical (*y*) axis represents the probability that the employer will decline to report the injury.

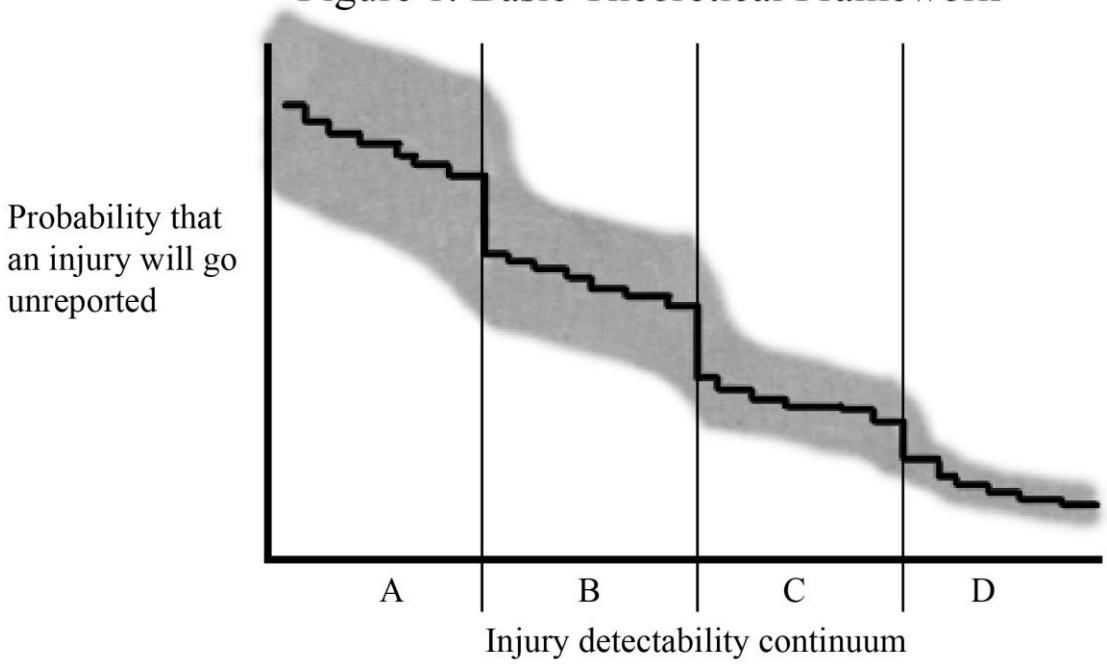
---

[48] The Census of Fatal Occupational Injuries (CFOI), conducted by the Bureau of Labor Statistics, uses extremely sophisticated surveillance techniques in an effort to obtain a comprehensive annual census of all workplace fatalities occurring nationwide. For a description of the program, see http://www.bls.gov/iif/oshfat1.html (last visited August, 13, 2014).

[49] States vary with regard to if injuries stemming from pre-existing conditions are compensable under workers' compensation. For example, in Oregon, a claim is only compensable if "the work injury/exposure was the major contributing (51% or more) cause of the disability or need for treatment,"(see Arthur W. Stevens, "New Comp Law Defines 'Preexisting Conditions'," *Black Chapman Webber & Stevens Attorneys* (blog), June 8, 2010, http://www.blackchapman.com/new-comp-law-defines-preexisting-conditions) while in California, any injury that aggravates a pre-existing condition is compensable (see California Department of Human Resources, *Workers' Compensation Preview* (January 2014), http://www.calhr.ca.gov/Documents/workers-compensation-preview.pdf).

[50] See *Increasing Access to Workers' Compensation Medical Benefits for Low-Income and Immigrant Workers*, NH Coalition for Occupational Safety and Health Research Report, March 31, 2008, http://www.nhcosh.org/pdfs/NHCOSH_WC_Report_3-31-08.pdf.

Figure 1: Basic Theoretical Framework



At the origin are less-detectable injuries that are relatively easy for an employer to underreport. At the far-right end of the *x*-axis are injuries that are very difficult for an employer to hide, such as fatalities. If all aspects of an injury were perfectly observable, the relationship between detectability and the likelihood of underreporting could appear linear, or at least smoothly continuous. In reality, however, many aspects of an injury that affect its detectability are unobservable, even to employers. Therefore, one would expect employers (and researchers) to group injuries into categories based on their observable characteristics. Figure 1 depicts a scenario in which injuries can be grouped into four discrete types, A, B, C and D, that vary in average detectability.

Importantly, within a given type, not all injuries are reported with equal probability. This is because employers have access to more granular information than researchers and are likely to use finer classification schemes. For example, even among nonspecific, cumulative low-back injuries, there will be variations in the likelihood that an employer who chooses not to report the injury will

be caught and penalized. (A highly vocal, unionized employee working full-time in a warehouse loading boxes onto trucks, for example, will have a more detectable injury claim than a part-time, non-unionized receptionist who plays amateur rugby.) The "mini-steps" *within* each injury type illustrate this point. For purposes of empirical analysis, however, the key assumptions are simply that average detectability varies across injury types, and that the types observable to researchers are equally observable to employers. Since the costs of underreporting increase with an injury's detectability, the likelihood of underreporting should peak at the origin and then decline monotonically, in stair-step fashion, as one proceeds along the *x*-axis.[51]

Another important feature of Figure 1 is the shaded area surrounding the staircase. This area represents variation across firms in the likelihood that an injury will be reported at each level of detectability. In other words, it represents heterogeneity in how different firms "descend the staircase." As is shown in the figure, the elevation and steepness of the staircase, and the relative steepness of each individual step, can vary widely across employers. Yet the amount of dispersion is also likely to vary by injury type. For less-detectable injuries, routine underreporting may be the profit-maximizing choice. Consequently, many firms will choose to underreport. Yet at the same time, non-economic factors such as social norms, or risk aversion on the part of individual managers, may induce some firms to report even those injuries they could successfully shield from regulatory scrutiny.[52] In short, for less detectable injuries, one would expect to see considerable dispersion across firms in the prevalence of underreporting. For highly detectable injuries, however, the opposite logic applies. Simple cost-benefit analysis will dictate the wisdom of reporting, so one is

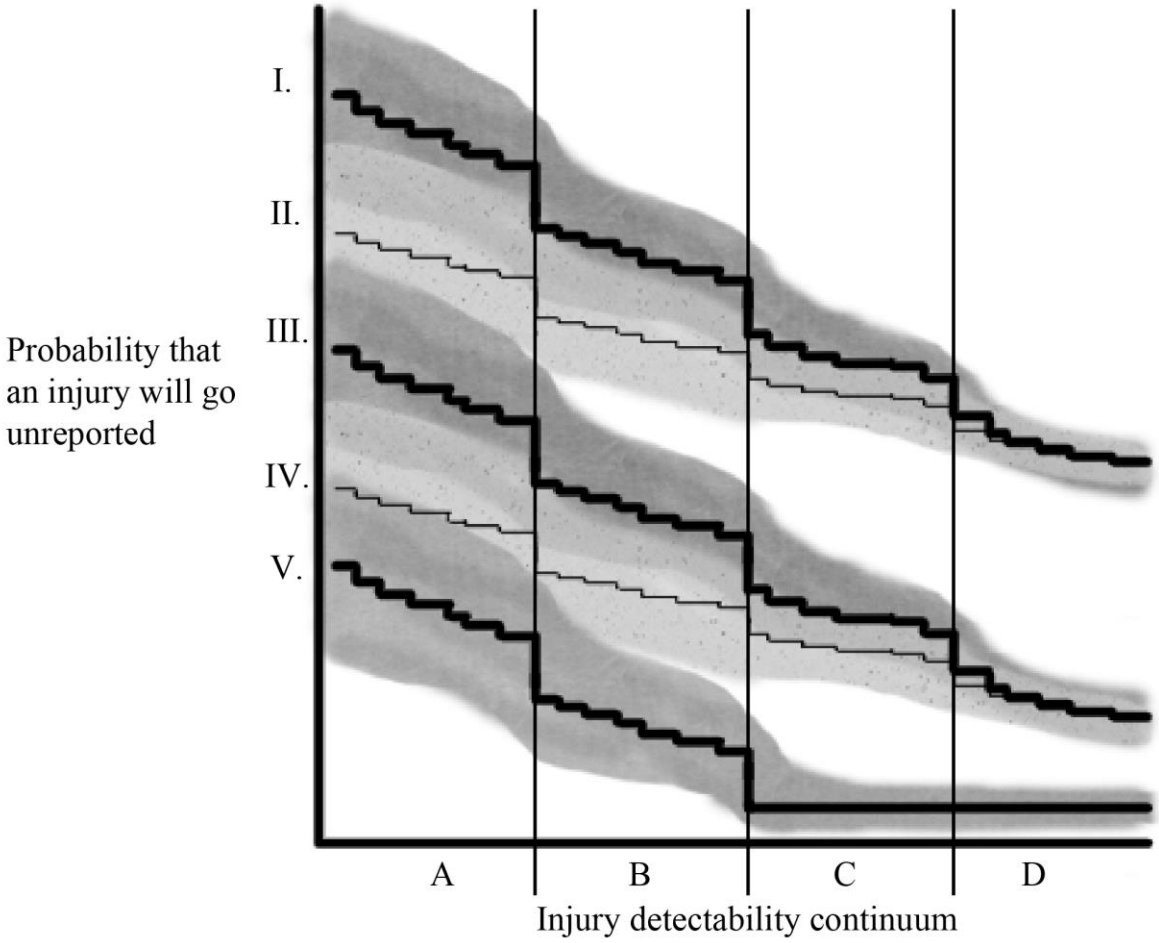
[51] This relationship need not, however, be linear, since the marginal difference in "detectability" may vary at different points along the continuum.

[52] For a discussion of this phenomenon in the OSHA context, see Weil (1996).

likely to see less variation across firms. Cross-firm dispersion in the likelihood of underreporting, then, should decline as detectability increases.

Figure 2 expands this simple framework by considering the effects of a change in regulatory intensity. An escalation in regulatory intensity can take two different forms: an increase in the *frequency* of audits, and/or an increase in the *stringency* (i.e., rigor and thoroughness) of each audit. Although in many contexts the two are largely interchangeable,[53] in this model they have different properties and different expected impacts on firm behavior.

Figure 2: Expanded Theoretical Framework



[53] If the probability of detecting unreported injuries is simply a function of the amount of time that an auditor spends onsite, then a 10% increase in total audit hours would have more or less the same effect, *ceteris paribus*, regardless of whether it results from an increase in the frequency of audits or in the number of hours devoted to each audit.

To see this, consider a regime that shifts from rudimentary audits (consisting exclusively of a comparison of injury filings with detailed injury logs, workers' compensation claims, and first-incident reports) to more thorough audits involving extensive employee interviews and scrutiny of hospital records. If rudimentary audits are reasonably successful at discovering more highly detectable injuries, such as amputations, but only rarely uncover less detectable injuries, like cumulative low-back injuries, the effects of the shift will vary widely by injury type. Whereas the likelihood of discovering highly detectable injuries may change very little, if at all, the odds of detecting less detectable injuries should increase substantially. This scenario is depicted by the shift from Staircase I to Staircase II in Figure 2.

Now consider the alternative scenario in which the stringency of each audit does not change, but audits occur with greater frequency. In this scenario, the staircase will shift in a more complex fashion that depends on the marginal increase in frequency and the (initial) height and position of each step in the staircase.[54] Figure 2 illustrates a scenario in which the probability of detection shifts by roughly equal amounts for all injury types, represented by the shift from Staircase I to Staircase III.[55] Moreover, if audit frequency or stringency increases to such a degree that the probability of underreporting certain types of injuries cannot fall any further, one will observe a "kink" in the

[54] Assuming that the likelihood of reporting a given injury is a linear function of the probability that the injury will be detected by auditors, the shift in the underreporting "staircase" will depend on the marginal increase in audit frequency; the probability of detecting the injury in each audit; and the dispersion across injury types ("steps") arrayed along the x-axis (injury detectability continuum). Intuitively, this is so because the probability of detection, like the probability of getting heads when flipping a coin, does not increase in a constant (linear) fashion as the number of inspections increases. Rather, the marginal increase in probability depends on the probability of detection during any given audit and on the number of audits already conducted. More precisely, the probability of an injury being detected (D) is described by the equation $D = 1-p^n$, where p is the probability that the injury is *not* detected in each audit, and n is the number of audits. Generally speaking, the smaller the increase in audit frequency, the closer p (and the probability of detection) are to 0.5, and the smaller the dispersion across injury types in the probability of detection, the greater the relative impact of increasing audit frequency, and the more closely the shift will resemble a constant (linear) shift like that depicted by the move from Staircase I to III in Figure 2.

[55] For example, a shift from 1 to 2 inspections, in which the initial detection probabilities were .45, .55, .65, and .75 (respectively) for injury types a, b, c and d, would resemble the shift from Staircase I to Staircase III in Figure 2.

staircase where the likelihood of underreporting reaches its lower bound.[56] This scenario is illustrated by the shift from I to V in Figure 2.[57]

In reality, of course, changes in audit frequency and in audit stringency are not mutually exclusive. A scenario in which they occur simultaneously is depicted by the shift from I to IV.

Within this simple theoretical framework, one can generate a number of empirical predictions about the relationship between injury type, regulatory intensity, and underreporting. Generally speaking, the laxer an inspection regime – in other words, the smaller the probability of detection times the size of the expected penalty – the higher the staircase. Moreover, the lower the expected likelihood that a given injury will be detected (and the lower the associated penalty), the higher the step associated with that injury.

Secondly, if employers underreport most (if not all) injuries below a given level of detectability, and report nearly all injuries above that threshold, then some steps in the staircase will be much steeper than others. Consider, for example, the simplest case in which an employer underreports *all* injuries of type A or B, but *no* injuries of type C or D. Under these conditions, the entire staircase will be descended in a single step at the juncture between B and C. This is the pattern one would expect to see if employers use rules of thumb to guide reporting decisions based on readily observable information that can be easily conveyed to mid-level managers.

---

[56] As long as a few employers – or the agents to whom they delegate decision-making authority – are heedless of the risk of detection, perhaps because of an idiosyncratically short time horizon or even a preference for risk, then the lower bound will be greater than zero.

[57] One aspect of reporting behavior that is not incorporated into the model is the possibility of "overreporting." Overreporting occurs when an employer reports an incident that does not meet the agency's recordability criteria – for example, reporting injuries that only require first aid or reporting injuries that do not result in medical treatment (although they may have received medical attention). Both RK and NEP OSHA audits make note of overreported cases as a part of the audit protocol, but these reporting inaccuracies do not result in citations. See U.S. Department of Labor, Occupational Safety and Health Administration, *1998 Audit and Verification Program of Occupational Injury and Illness Records*, Directive Number: 00-1 (CPL 2), December 2, 1999, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=2003, and US DOL, OSHA, Directive Number: 09-08 (CPL 02) (2009) (cited earlier). Approximately 10% of all injuries reported do not meet the recordability criteria, and approximately 5% of firms audited engaged in a significant amount of overreporting. (Note: all overreported cases have been dropped from the dataset used for the empirical analysis.)

The location of the steepest step(s) will likewise depend on the intensity of the inspection regime. In an extremely lax regime – in which employers fail to report all but the most severe injuries – the steepest step(s) will occur toward the far-right end of the *x*-axis. Conversely, in a very strict regime, in which all but the least detectable injuries are uncovered by auditors, the steepest step(s) will occur very close to the origin.

Third, the best predictors of underreporting will tend to align with the steepest step(s) in the staircase descended by the typical firm. For example, consider the simple case described above, in which the typical employer underreports *all* injuries of type A or B, and *no* injuries of type C or D. In that case, the best predictor of whether an injury is underreported will be *whether the injury is of type A or B versus type C or D*. Since the location of the steepest step(s) varies by the intensity of the inspection regime, the best empirical predictor(s) of underreporting will also vary with regulatory intensity. In a strict regime, the best predictor(s) will occur relatively close to the origin; in a lax regime, the best predictor(s) will occur much farther from the origin.

Finally, a worker protection agency's capacity to mimic the IRS – that is, to identify which employers are underreporting, merely by inspecting the injuries listed in mandatory filings – depends not only on the amount of heterogeneity across firms, but also on the intensity of the regulatory regime. To see this, consider a case in which the regulatory regime is so lax that the *only* injuries reported are those at the highest level of detectability. Under such conditions, there could still be considerable dispersion across firms in the magnitude of underreporting. For example, relatively compliant firms might report *all* injuries that are of Type D, whereas less compliant ones might only report those that fall into the top decile of the detectability continuum. Yet the composition of injuries in the two firms' injury filings would be identical. (Both would consist exclusively of Type D injuries.) In short, extreme laxity of enforcement compromises an agency's capacity to distinguish

typical firms from the worst violators. As long as a regulatory regime is sufficiently robust that many firms descend the staircase at the left and middle portions of the detectability continuum, this concern does not apply. Assuming a reasonable amount of heterogeneity across employers, a robust regulatory agency should be able to develop empirical red flags or algorithms to identify which firms underreport the most injuries merely by examining firm-level attributes and the composition of injuries in regulatory filings.

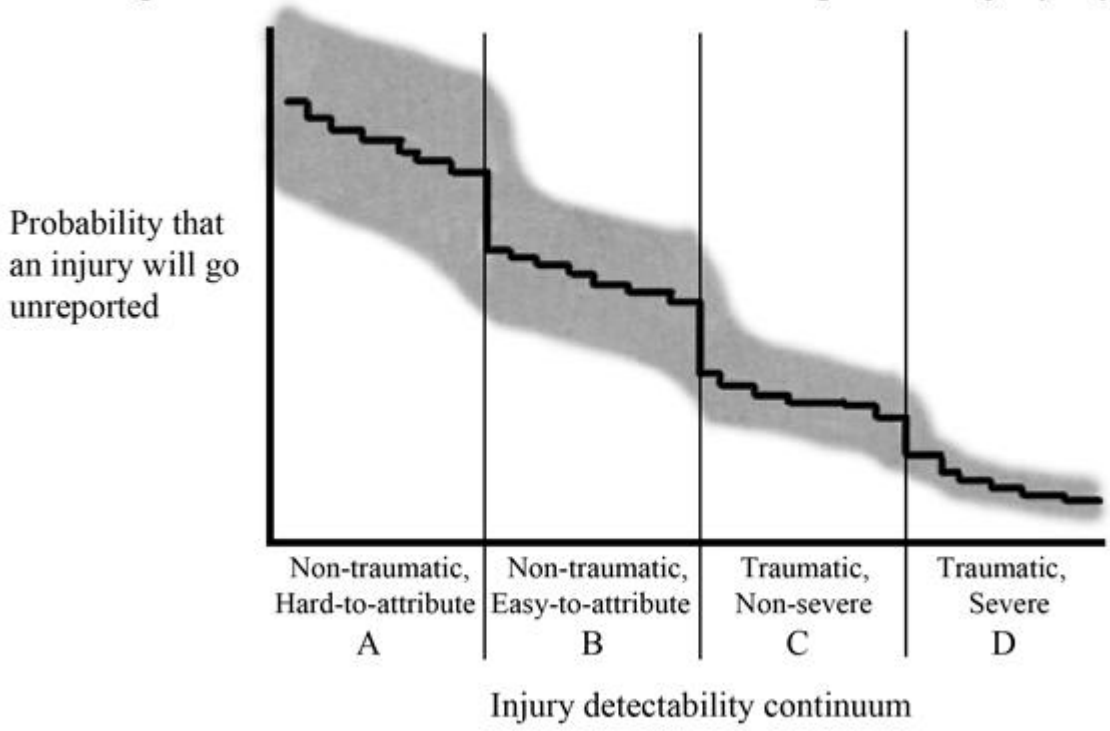
Although many of these predictions are amenable to empirical verification, further information is required to do so. Unless one can identify several discrete, observable types of injuries and place them correctly along the *x*-axis, one cannot test whether the hypotheses are consistent with available data. Drawing on expertise from the field of occupational medicine,[58] I introduce a four-part classification scheme for ordering injuries along the *x*-axis, depicted below in Figure 3. (Since the Census of Fatal Occupational Injuries and Illnesses detects virtually all occupational fatalities across all industries in the U.S., making detection a virtual certainty, I omit occupational fatalities from the empirical analysis and focus exclusively on non-fatal injuries.) There is nothing sacred about the scheme proposed. With perfect and costless information about every injury, one could calibrate the detectability of each individual injury with infinite precision, and line up all injuries along the detectability continuum in a smoothly continuous fashion. In reality, however, many aspects of an injury's detectability are not perfectly observable. Therefore, developing a categorical classification scheme reliant upon characteristics reported in injury filings is a critical simplification that helps me bridge the gap between theory and practice.

As depicted in Figure 3, I hypothesize that non-fatal injuries can be meaningfully grouped

[58] This classification scheme would not have been possible without the help of Dr. Mark Cullen. See notes 1 and 59.

into four categories of increasing detectability: nontraumatic and hard-to-attribute (A); nontraumatic and easy-to-attribute (B); traumatic but not severe (C); and traumatic and severe (D).[59] Type D (traumatic and severe) injuries are those that usually require urgent care, such as fractures, crushing, concussions and amputations. Type C (traumatic but non-severe) injuries, such as lacerations, contusions, burns, and non- eye-related abrasions, also occur instantaneously but are generally less serious and require less acute care. Type B (non-traumatic and easy-to-attribute) injuries include cumulative impairments whose underlying cause is relatively easy to detect, such as hearing loss, eye-related abrasions, hernias and heatstroke.

Figure 3: Theoretical Framework with Specific Injury Types



[59] Conversation with Mark Cullen, M.D., Professor of Medicine, Stanford University. Non-traumatic, hard-to-attribute (A) injuries include sprains and strains (mostly of the back or shoulders) as well as joint, tendon, and muscle inflammation such as tendonitis and carpel tunnel (mostly of the wrist or elbow). Non-traumatic, easy-to-attribute (B) injuries encompass sprains and strains (mostly of the knee or ankle), hearing loss, eye-related abrasions, skin-related injuries, hernias, heatstroke, and poisoning. Traumatic but not severe (C) injuries include lacerations, contusions, foreign objects in eye, burns, non-eye-related abrasions, and electric shocks. Traumatic and severe (D) injuries include fractures, crushing, dislocations, amputations, enucleation, and concussions. See companion website for more detailed description of categorization scheme, available at LINK TO WEBSITE TO BE INSERTED HERE ONCE IT IS ONLINE.

Type A, the final category, poses especially complex policy challenges. Non-traumatic and hard-to-attribute injuries encompass musculoskeletal disorders such as carpal tunnel syndrome and nonspecific strains or sprains of the back and shoulders. Although their root cause is often difficult to detect using existing medical technologies, such injuries can be severely disabling and impose enormous burdens on the labor market and health care system. One study estimated the annual treatment costs and lost wages associated with musculoskeletal diseases to be $849 billion, or 7.7% of Gross Domestic Product,[60] and another observed that "[m]ore U.S. health care dollars are spent treating back and neck pain than almost any other medical condition." [61]

Given their nontraumatic nature and the difficulty of pinning down whether they were caused by activities performed at work, such injuries are especially prone to underreporting. Not only may insured workers prefer to seek treatment from their primary care providers, but clinicians themselves have strong incentives *not* to probe the work-relatedness of the injury.[62] Indeed, the only way for regulators to uncover hard-to-attribute injuries during an audit may be to ask workers directly about their health status or pain they experience while performing job-related duties.

To test the model's predictions about how regulatory intensity affects reporting behavior, one must also discern which regulatory regimes are stricter than others. Because MSHA inspects regulated establishments much more frequently than OSHA and also collects injury data with more frequency and granularity, I take MSHA to be the stricter regime. I further subdivide my data in two ways. First, since MSHA is statutorily required to inspect underground mines twice as often as

---

[60] See "The Burden of Musculoskeletal Diseases in the United States," executive summary, *Bone and Joint Decade* (2011), http://www.boneandjointburden.org/pdfs/bmus_executive_summary_low.pdf. See also Deborah P. Lubeck, "The Costs of Musculoskeletal Disease: Health Needs Assessment and Health Economics," *Best Practice & Research Clinical Rheumatology* 17 (2003): 529–539.

[61] See Salynn Boyles, "$86 Billion Spent on Back, Neck Pain," *WebMD Health News,* February 12, 2008, http://www.webmd.com/back-pain/news/20080212/86-billion-spent-on-back-neck-pain.

[62] See, e.g., Azaroff, Levenstein, and Wegman (2002).

surface mines, I assume that underground mines are subject to greater regulatory scrutiny than surface mines.[63] Secondly, as explained above, the scope and stringency of OSHA's auditing protocols increased with the transition from the RK auditing program (1996-2006) to the short-lived NEP program (2007-09). Thus within the OSHA regime, I assume that *ceteris paribus*, the intensity of regulatory scrutiny increased with the advent of NEP audits.

IV. Methodology

The goal of the empirical analysis is to test whether the model's theoretical predictions are borne out by available data. Two real-world obstacles, however, complicate several portions of the identification strategy.

First, my estimates of the probability that an injury is not reported (or of the prevalence of underreporting) are susceptible to what I call "audit stringency bias." In addition to capturing the likelihood that an injury is actually underreported, the coefficients reflect the likelihood that an unreported injury is *brought to light during an audit, and thus included in the dataset*. Suppose, for example, that a new statute triples the amount of funding dedicated to audits. The law will have two countervailing impacts on the (observable) frequency of underreporting. If firms become aware of the change and believe it will raise the odds that unreported injuries are discovered, they will respond by underreporting fewer injuries. Yet if that the agency uses the extra resources to conduct more stringent audits, the number of unreported injuries discovered during each audit will also surely rise. The relative magnitudes of these two effects will determine the net effect of the law.

Importantly, the magnitude of audit stringency bias should generally decline as injury detectability increases. This is so because even a low-intensity audit is likely to catch highly

[63] The Mine Act requires MSHA to inspect underground mines four times per year, whereas surface mines need only be inspected twice per year.

detectable injuries like amputations, whereas only a rigorous one is likely to uncover an unreported cumulative back strain. In general, one might expect audit stringency bias to predominate in the short run, before firms can respond fully to a regulatory change. Over time, as they become more familiar with the altered incentive scheme, firms' behavioral responses to the change should mitigate audit stringency bias.

The second problem is the model's (implicit) simplifying assumption that employers are fully capable of adjusting their reporting behavior. In reality, the choice of whether to report an injury does not lie exclusively with employers; it also lies to some extent with injured workers themselves. If a worker decides that the cost of reporting an injury outweighs its benefits, she may decline to report it to her employer and simply seek treatment from her primary care physician. Instead of filing a workers' compensation claim, she may use up sick days (or vacation days) while she recuperates, or even take unpaid sick leave.[64] From an employer's perspective, encouraging employees not to report injuries in the first place may be far more appealing than omitting injuries from regulatory filings. After all, if an employee declines to report an injury, her employer can (credibly) maintain that it had no idea it even occurred. It also lessens the likelihood that the injury will ever come to light, since it is not recorded in any permanent record and may never be discovered unless the affected employee (or a coworker) brings it up during an interview.

The concern that some workers hide workplace injuries from their employers is more than theoretical. The rise of a popular risk-management philosophy known as "behavior-based safety" (BBS) has heightened concerns that such behavior is commonplace. Widely embraced by employers as a means to induce workers to play a more active role in creating a strong safety culture, BBS

[64] Under the Family and Medical Leave Act, "eligible employees of covered employers [are entitled] to take unpaid, job-protected leave for specified family and medical reasons with continuation of group health insurance coverage under the same terms and conditions as if the employee had not taken leave. " See http://www.dol.gov/whd/fmla (last visited August 1, 2014).

involves the identification of target behaviors that impact safety; the definition of such behaviors with sufficient precision to permit data collection; the tracking of targeted behaviors over time so as to chart progress toward clearly defined goals; the ongoing provision of feedback to workers; and the reinforcement of progress toward goals. Reinforcement systems commonly include the provision of tangible rewards (such as monetary bonuses or valuable prizes) to individuals or groups that achieve safety goals (such as an "injury free month"), and sometimes even include the application of penalties against individuals or groups that fail to meet specific targets. [65] Although such programs may succeed in encouraging workers to take greater care on the job, they may also discourage workers from reporting on-the-job injuries.

DOL officials have repeatedly voiced concerns about the deleterious effects of BBS-inspired incentive systems ("incentive programs") on injury reporting. On June 29, 2011, OSHA issued guidance materials for its Voluntary Protection Program discussing the danger of offering financial incentives for the non-reporting of injuries, and suggesting alternative ways to encourage safe practices without incentivizing workers to hide injuries.[66] The concern was reiterated in subsequent reports and memoranda.[67] Meanwhile, a report issued by the Department of Labor's Office of the Inspector General in early 2014 expressed nearly identical concerns about the prevalence of such programs in the mining industry.[68]

---

[65] Beth Sulzer-Azaroff and John Austin, "Does BBS Work? Behavior-Based Safety & Injury Reduction: A Survey of the Evidence," *Professional Safety* (July 2000): 19-24.

[66] U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from David Michaels to Regional Administrators, *Revised VPP Policy Memorandum #5: Further Improvements to the Voluntary Protection Programs (VPP)*, June 29, 2011, https://www.osha.gov/dcsp/vpp/policy_memo5.html.

[67] See Nancy Smith, et al., *OSHA's Voluntary Protection Programs (VPP) Review: Findings and Recommendations*, a Report Submitted to David Michaels, Assistant Secretary for OSHA, (November 2011): 21-22, https://www.osha.gov/dcsp/vpp/vpp_report_nov_2011_rev_7-11-12.pdf; and U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from Richard E. Fairfax, OSHA Deputy Assistant Secretary, to Regional Administrators, "Employer Safety Incentive and Disincentive Policies and Practices," March 12, 2012, https://www.osha.gov/as/opa/whistleblowermemo.html.

[68] See US DOL, OIG, *MSHA Has Taken Steps* (2014): 8-10.

For purposes of this study, the key point is not simply that incentive programs tend to deter the reporting of injures, but that their effects probably differ by injury type. Highly detectable injuries, especially severe ones, may be largely impervious to the influence of incentive programs. On the other hand, the reporting of less detectable injuries may be quite sensitive to their presence. In effect, even employers that *want* to fully report less-detectable injuries in the wake of an escalation in audit stringency may be unable to do so if they operate incentive programs that discourage employees from reporting injuries in the first place. Although in principle employers could suspend these programs, the benefits of doing so may be outweighed by the (perceived or actual) benefits of incentivizing workers to take greater care on the job. In short, the prevalence of incentive programs may cause some "stickiness" in the responsiveness of nontraumatic injury reporting to shifts in regulatory intensity.

To set the stage for the empirical analysis, I first manipulated the dataset to make it more amenable to statistical modeling. I dropped all fatal injuries, all incidents not resulting in an injury, all illnesses[69], all incomplete observations, and all "over-reported" injuries (i.e., those that were reported erroneously because they were not subject to OSHA reporting requirements) from the sample.[70] Using the injury-level fields that are contained in both datasets – source of injury, degree

[69] Although the theoretical model applies equally to injuries and illnesses, the difficulty of tracking illnesses in my dataset made them poorly suited for inclusion in the study.

[70] Besides fatalities, the following illnesses and injuries were categorically excluded from the OSHA data: work-related sicknesses; tuberculosis, bloodborne, and chronic obstructive pulmonary disorders; eczema; asthma; dust diseases of the lungs. The OSHA dataset also contains 98 injuries labeled "other occupational illnesses." All but 13 of these observations could be identified as more specific injuries and re-categorized using information in three text fields: injury description, injury narrative, and inspector comments. The following injuries were re-categorized: 34 instances of chipped or fractured teeth were marked as fractures, 20 instances of rashes, insect bites, scabies, and hives were marked as skin diseases, 11 respiratory conditions due to toxic agents, 7 bee stings, 6 eye injuries, 2 sprains and strains, 1 welding flash was categorized as disorders due to physical agents, 1 instance of hearing loss, 1 burn, 1 dislocation, and 1 puncture wound. The remaining 13 injuries could not be clearly identified and were dropped from the dataset. (The exclusion of these 13 observations did not substantively change any results.) The OSHA data contained 158 observations that were missing injury categories: 86 cases were categorized based on the three text fields (same as above) and 72 were dropped because they could not be categorized due to a lack of information on the injury, because they concerned pre-existing injuries, or because they were illnesses (infections, funguses, cysts, etc.) The 86 injuries

of injury, nature of injury, and affected body part – I then constructed four different indicator variables to capture injury characteristics: "severe," "traumatic," "easy-to-attribute," and "intermediate." The first three characteristics, as noted above, were developed through consultation with an expert in occupational medicine.[71] The fourth, which pertains exclusively to mining, was added because of its inclusion in a landmark historical study of mine safety.[72]

The injury characteristics upon which the empirical analysis rests – traumatic, severe, and hard-to-attribute – are defined in ways that are partially overlapping, so as to isolate different segments of the injury detectability continuum. The most fundamental distinction, traumatic/nontraumatic, divides the continuum into a pair of less-detectable categories (Types A and B) and a pair of more-detectable categories (Types C and D). The remaining categories further subdivide each respective half of the continuum. Among nontraumatic injuries, those that are hard-to-attribute (Type A) are assumed to be the least detectable of all. Meanwhile, among traumatic injuries, those that are also severe (Type D) are hypothesized to be the most detectable of all.

I also created a continuous variable ("employment") to reflect the number of workers employed (in hundreds), and four interaction terms between employment and injury type. To control for geographic variation and variation over time, I created dummy variables for the state and year in which the injury occurred. I also added a dummy indicating whether the establishment was unionized. (Although union status is available for all OSHA-regulated establishments, it is only

---

were categorized as follows: 72 sprains and strains, 6 instances of skin-related injuries, 3 eye injuries, 2 bee stings, 2 instances of hearing loss, and 1 instance of heat exhaustion. (The inclusion of these 86 observations did not significantly alter the results.) For MSHA, the following illnesses and injuries were excluded (in addition to fatalities) due to their questionable work-relatedness or their inability to be clearly categorized: pneumoconiosis/black lung; silicosis; occupational diseases, not elsewhere classified (NEC); unclassified, not determined; heart attack; cerebral hemorrhage; and other injury not elsewhere classified (NEC).

[71] See *supra* notes 1 and 59.

[72] See National Research Council, *Toward Safer Underground Coal Mines* (1982). Injuries classified as "intermediate" are those involving permanent total or permanent partial disability and those resulting from entrapment, falling/sliding/rolling materials, roof falls, haulage and machinery, electrical accidents, explosions, hoisting, impoundment, fire, or inundation.

available for coal mines.) Industry dummies were created using 2-digit SIC code for each non-mining establishment, and canvass code[73] for each mining establishment. Finally, in the MSHA-specific models, I created a dummy variable indicating whether or not the mine operates underground, and (in most models) interaction terms between underground status and injury type.

The empirical analysis unfolds in three stages. In Stage One, I test the model's predictions regarding which injury characteristics, if any, significantly predict the likelihood of underreporting and how these results vary across regimes. Specifically, I fit probit models in which the unit of observation is the injury and the (binary) dependent variable indicates whether the injury was reported. I estimate four different specifications: one pertaining exclusively to OSHA-reportable injuries and three pertaining exclusively to MSHA-reportable injuries. Of the three MSHA-specific models, one includes all mines, one restricts the sample to underground mines, and the third restricts the sample to coal mines. (The latter specification is included to test the effect of union status, which is unavailable for non-coal mines.) Most of the covariates used – severe, severeXemployment, traumatic, traumaticXemployment, easy-to-attribute, easy-to-attributeXemployment, employment, state dummies, year dummies and industry dummies – are common to all models. However, three independent variables (intermediate, intermediateXemployment, and underground) apply only to mining establishments. Standard errors are clustered on audit in the OSHA models and on mine identification number in the MSHA models.

In Stage Two, I explore several more nuanced theoretical predictions. First, I probe the height and topography of the underreporting staircase descended by a typical firm in both regulatory regimes. The goal is to pin down where the steepest step(s) of the staircase occur in each regime. Although the theoretical model does not indicate where they will occur, it does predict that *they will*

[73] The five canvass codes include Coal, Sand and Gravel, Stone, Nonmetal, and Metal.

tend to occur at lower levels of detectability in the stricter regime. In other words, relative to the OSHA regime, the steep step(s) in the MSHA regime should occur closer to the origin. In the most intensely inspected environment of all, underground mines, the steep step(s) should occur closer still to the origin. Of course, given the discrete nature of the categories analyzed and the possibility that some shifts will occur within (instead of between) categories, not all such shifts may be detectable. Yet by and large, as regulatory intensity increases, the staircase should tend to "migrate" toward the origin. Finally, since underground mines are inspected twice as frequently as surface mines but otherwise are highly comparable, one would also expect the likelihood of underreporting (i.e., the height of the staircase) to be lower overall in the underground environment.[74]

To estimate the height of each step, I estimate a probit model in which the dependent variable indicates whether the injury was reported This time, however, the main covariates of interest across all models are the *marginal* declines in reporting probability associated with a change in injury type from traumatic/severe to traumatic/nonsevere (D to C); from traumatic/nonsevere to nontraumatic/easy-to-attribute (C to B); and from nontraumatic/easy-to-attribute to nontraumatic/ hard-to-attribute (B to A). In the MSHA models, additional covariates of interest are the dummy on underground mine and the interaction terms between each of the injury type dummies and underground mine. All models control for employment, employmentXinjury type, state, year, and industry. Standard errors are clustered on audit in the OSHA models and on mine ID in the MSHA models.

Stage Two of the analysis also tests several predictions of the model that pertain exclusively to OSHA. Specifically, I examine the impact of OSHA's transition from the RK program, which

[74] In principle, one would expect the likelihood of underreporting to be lower overall in the MSHA regime as compared to the OSHA regime. However, this prediction cannot be meaningfully tested because of the different content and construction of the datasets obtained from the two agencies, including the inability to collect data at the audit level in the MSHA context and the omission from the MSHA sample of all audited mines that underreported zero injuries. Therefore, I do not attempt to test the prediction that underreporting overall is lower in the MSHA environment.

audited injury logs from the years 1996-2006, to the short-lived NEP program, which spanned logs from the years 2007-2009. As noted earlier, NEP audits were significantly more stringent than the RK audits that preceded them, but they also audited a somewhat different mix of firms. Specifically, firms with DART rates below the mean for high-hazard industries (4.2 injuries and illnesses per 100 full-time employees) were initially targeted under the NEP program on the suspicion that they were most likely to be underreporting. Given the different mix of firms audited and the escalation in audit stringency bias, one would expect the overall frequency of unreported OSHA injuries to have spiked in 2007. If the model's assumptions regarding the relationship between audit stringency and injury detectability are correct, hard-to-attribute injuries should have been affected more than severe ones.[75]

To test the latter (OSHA-specific) predictions, I visually compare trends in the mean number of total, severe and hard-to-attribute injuries omitted from each year's logs. I also examine these trends more formally by estimating three negative binomial models, one for each injury category, in which the dependent variable is the number of injuries of that type not reported to OSHA and employment is used as an exposure term. The variables of interest are the year dummies. Since the only year to which valid comparisons can be made is 2007, the first year affected by the NEP program, that is the year omitted from the models. The unit of observation is the injury set, and all models include robust standard errors and control for employment, union status, state, and industry.

In Stage Three, I commence the hunt for red flags. Statistically speaking, my goal is to determine whether particular (observable) characteristics of establishments or injury sets help predict the frequency and/or percentage of unreported injuries. Although there are grounds for optimism that

[75] I cannot rule out the possibility that the high-hazard firms audited in the first year of the NEP program (i.e., those reporting a DART rate less than 4.2) under-reported different types of injuries than those audited under the RK program, and therefore that my findings regarding the varying effects of audit stringency bias on different types of injuries are spurious. However, given the substantial overlap in the industries audited (both programs audited high-hazard industries included in the OSHA Data Initiative, although the NEP program additionally audited firms in NAICS codes 311614 and 115210), this scenario seems unlikely. Even if the *overall* spike in unreported injuries is partly attributable to a change in the sample of firms selected for audits, it is hard to imagine why selection bias would also affect the composition of injuries that went unreported or the composition of injuries that were uncovered during audits.

such red flags exist, two factors may weaken my identification strategy in the MSHA context. First, unlike OSHA, no *bona fide* injury sets are available for MSHA at the firm level. As noted earlier, I created "quasi injury sets" by bundling together all injuries from a given calendar year that a given firm *untimely* reported to MSHA to all injuries in the same calendar year that were *timely* reported. Firms for which no unreported injuries ever came to light are omitted altogether. This approach may have introduced some measurement error into my estimates. Secondly, very few quasi injury sets are available for analysis. In contrast to the thousands of injury sets available for OSHA-specific analysis, the MSHA dataset contains only 342 quasi injury sets, and neither subsample contains more than 175. (Even these numbers overstate the effective variation in the sample since many mines appear in the dataset more than once.) In short, the MSHA-specific results from Stage Three should be regarded as tentative and preliminary.

For all firms reporting at least one injury, I estimate a negative binomial model in which the dependent variable is the number of unreported injuries in the (quasi) injury set, with the number of employees used as an exposure term. As always, I include employment, state, year, industry, underground (for MSHA models), and unionization (in all models except those including non-coal mines) as covariates. OSHA models include robust standard errors, and standard errors are clustered on mine ID in MSHA models. Since predicting percentages requires one to calculate a numerator as well as a denominator, targeting injury rates is generally more difficult than predicting injury counts (frequencies).[76] Nevertheless, I also estimate an ordinary least squares model with robust standard errors in which the covariates are identical but the dependent variable is the *percentage* of total injuries that were untimely reported.

[76] See Alison Morantz, "Final Project Report: Designing a Pilot Program for Strategic Mine Safety and Health Improvements through the Use of Surveillance Data to Guide Targeted Inspection Activities," NIOSH Research Contract 200-2009-28820, September 28, 2012 (available from author upon request).

In all of the models just described, the primary covariates of interest are dummy variables indicating whether the respective *percentages of reported injuries* classified as severe, traumatic, or easy-to-attribute place the employer in the top quartile of the relevant sample.[77] Upon reflection, however, it is not obvious *a priori* which sample is the most pertinent. If the agency's goal is simply to identify as many underreported injuries as possible, regardless of where they occur,[78] there are two possible methods that one might use to calculate the quartile dummies. One approach would be to generate quartile rankings for *all* employers in each regulatory regime. (For example, MSHA quartiles would be calculated across *all* mines and OSHA quartiles would be calculated across all OSHA audits.) Alternatively, one could calculate quartile dummies separately for each industry, and distinguish even further between surface and underground mines.

Which set of dummies will be the most predictive is an open empirical question. If the true underlying distribution of injuries varies significantly across industries, then opting for industry-specific dummies should minimize measurement error.[79] On the other hand, if average reporting rates also differ significantly across industries, then calculating quartile dummies across the *entire* sample should enhance their explanatory power. The latter approach also reduces sampling error stemming from the fact that some industries comprise only a few observations. Since the net effect of these factors is uncertain, I calculate three sets of dummies – one for each regulatory regime in its

[77] As a robustness check, I estimated models with dummies for the top 10%, 15%, and 20% instead of the top quartile. These estimates exhibited patterns quite similar to those presented here.

[78] Alternatively, an agency could decide to target firms that underreport a number/proportion of injuries that is disproportionate by the standards of their respective industries. The downside of this approach is that assuming there is some variation across industries in the prevalence of underreporting, some firms that were *not* targeted in highly noncompliant industries would actually be underreporting more injuries than firms from highly compliant industries that *were* targeted. In the analysis that follows, I assume that the agency's goal is to identify firms that underreport the most injuries, regardless of the industries to which they belong.

[79] Intuitively, using dummies that do not account for industry-specific variation could increase the likelihood of false positives and false negatives in the calculation of quartiles. For example, a relatively compliant firm in an industry with an inherently high proportion of traumatic injuries could be erroneously included in the top quartile, whereas a relatively noncompliant firm in an industry with an inherently low proportion of traumatic injuries could be erroneously excluded.

entirety, one for each industry, and one that further subdivides each mining industry into surface and underground mines – and select the one with the most predictive power.

Another definitional ambiguity involves the precise definition of "quartile." Depending on whether there are many tied values across observations – and if so, where these tied values occur across the frequency distribution – one might define "quartile" as either strictly greater than, or greater than or equal to, the frequency observed at the 75th percentile. In some contexts, one of these definitions may not be feasible at all. (For example, if observed frequencies range from 0 to 2, and one third of all observations report a frequency of 2, defining quartile as strictly greater than the 75th percentile would flag no observations at all.) To resolve this problem, I attempted to use a "greater than or equal to" definition in all models, but if this approach was not feasible because it flagged all observations, I used the "strictly greater than" definition instead.

The empirical analysis draws to a close by considering the case of firms that reported no injuries at all ("nonreporters"), for which no reported injury-type dummies could be calculated. The key policy question is how highly an agency should prioritize the inspection of nonreporters as compared to reporters (i.e., firms that report at least one injury). I approach this question in two ways. First, to explore the importance of observable establishment-level or mine-level characteristics for this group, I confine the sample to nonreporters and estimate a Poisson model in which the dependent variable is the number of injuries that are not reported to regulatory officials. The usual set of covariates (employment, industry, etc.) are included. Secondly, although my small sample sizes preclude me from drawing any firm conclusions, I use several descriptive techniques to compare the number of unreported injuries among reporters and nonreporters in each industry examined.

V. Results

Table 1 presents average marginal effects from the first stage of the analysis, in which my goal is to identify the most important injury-level predictors of underreporting.

For OSHA, the laxer inspection regime, the best single predictor of underreporting is the severity (or lack thereof) of the injury. The statistically significant and *positive* average marginal effect of the "easy-to-attribute" dummy, however, seems to contradict the theory. Although initially counterintuitive, this coefficient makes sense when one takes into account the presence of audit stringency bias. As noted earlier, audit stringency bias is likely to affect nontraumatic injuries, especially hard-to-attribute ones, the most. The fact that this dummy is positive and statistically significant only for OSHA, the laxer regime in which rigorous audits were only conducted for a brief period, lends credence to this hypothesis.

Similar logic could explain the statistical insignificance of the union dummy, which appears to contradict prior literature suggesting that unionized establishments underreport fewer nontraumatic injuries.[80] Although unions may increase the likelihood that injuries are initially reported, they might also increase the chances that unreported injuries are discovered during an audit (for example, by empowering workers selected for interviews to speak more candidly with auditors). In other words, the presence of a union could exacerbate the confounding influence of audit stringency bias.

Most of the other findings echo prior literature. As expected, the likelihood of underreporting declines uniformly and significantly with firm size. I also find significant variations by state, region, and industry.[81]

---

[80] See Morantz (2013) and National Research Council, *Toward Safer Underground Coal Mines* (1982).

[81] Notably, injuries taking place in New York establishments are significantly more likely to go unreported across the different models (except Model 4). In Model 1, injuries taking place in Arizona and Montana establishments are significantly more likely to go unreported while those taking place in California, Kentucky, Missouri, Mississippi, New

Although injury severity is also a significant predictor of underreporting in the main MSHA model (Model 2), the other two characteristics examined – whether the injury is traumatic and whether it is easy to attribute – also have significant predictive value. In the subsample that is subject to the greatest regulatory scrutiny of all, underground mines, the pattern is similar but the characteristic that lies farthest from the origin (injury severity) loses statistical significance. The coal-only model displays identical patterns to the underground-mine model, which is not surprising since most of the sample (about 91% of the observations and 70% of the clusters) consists of underground mines. In all of these regards, the MSHA findings are consistent with the model's predictions. Also as expected, the likelihood of underreporting also declines with firm size and I detect significant variations by state, region, and industry.[82]

Jersey, Nevada, South Dakota, and Virginia establishments are significantly less likely to go unreported. In the three MSHA models, injuries at West Virginia and Kentucky establishments are significantly more likely to go unreported. Pennsylvania is excluded (used as the basis for comparison) for all models. OSHA-regulated industries associated with an increased probability of underreporting include *Building Construction General Contractors and Operative Builders (15), Chemicals and Allied Products (28), Electronic and Other Electrical Equipment and Components, Except Computer Equipment (36),* and *Local and Suburban Transit and Interurban Highway Passenger Transportation (41)*, while *Construction Special Trade Contractors (17)* is associated with a decreased probability of underreporting. SIC Major Group 80 (Health Services) is excluded from Model 1 because it is the factor level appearing most frequently in the OSHA injury set level data. For OSHA, one time trend emerges from the injury-level analysis: the year 2006 is associated with a significant decrease in the probability that an injury went unreported, while 2007 is associated with a significant increase in the likelihood of underreporting. Interestingly, these two years are the last audited log-year of the RK program and the first audited log-year of the NEP program, respectively. Across MSHA models, 2001 is associated with a significant decrease in the likelihood of underreporting. The year 1996 is excluded as the baseline in all models.

[82] Notably, injuries taking place in New York establishments are significantly more likely to go unreported (0.05 to 0.19 higher probability) across the different models (except Model 4). In Model 1, injuries taking place in Arizona and Montana establishments are significantly more likely to go unreported while those taking place in California, Kentucky, Missouri, Mississippi, Nevada, South Dakota, and Virginia establishments are significantly less likely to go unreported. In the three MSHA models, injuries at Kentucky and West Virginia establishments are significantly more likely to go unreported. Pennsylvania is excluded (used as the basis for comparison) for all models. OSHA-regulated industries associated with an increased probability of underreporting include *Building Construction General Contractors and Operative Builders (15), Chemicals and Allied Products (28), Electronic and Other Electrical Equipment and Components, Except Computer Equipment (36),* and *Local and Suburban Transit and Interurban Highway Passenger Transportation (41)*. SIC Major Group 80 (Health Services) is excluded from Model 1 because it is the factor level appearing most frequently in the OSHA injury set level data. According to Models 2 and 3, injuries at stone mines are associated with a 0.05 to 0.08 increased probability of underreporting. Coal is the excluded industry for MSHA. For OSHA, one time trend emerges from the injury-level analysis: the year 2006 is associated with a significant decrease in the probability that an injury went unreported, while 2007 and 2008 are associated with a significant increase in the likelihood of underreporting. Interestingly, these years are the last audited log-year of the RK program and the first two audited log-year of the NEP program, respectively. Across MSHA models, there is not notable time trend. The year 1996 is excluded as the baseline in all models.

Interestingly, the marginal effect of the "intermediate" dummy is statistically insignificant in the first two MSHA models but significant and *positive* in the coal-only model. Given its failure to predict underreporting in any other model, combined with the absence of any prior literature confirming its hypothesized negative correlation with underreporting, the "intermediate" dummy is excluded from all subsequent models. [83]

The MSHA results do, however, exhibit two perplexing patterns. First, since underground mine are inspected twice as often as surface mines, the statistical insignificance of the "underground" dummy does not support the expectation that the *overall* likelihood of underreporting (i.e., the height of the staircase) declines in the stricter regime. This puzzling finding could indicate the combined effects of audit stringency bias and incentive programs. As noted earlier, the most stringent audits that MSHA conducts, its "PPOV" audits, have taken place almost exclusively at underground mines. Thus one would expect audit stringency bias to be the strongest in underground mines. The fact that underground mine operators have not responded to this increase in regulatory scrutiny by reporting more nontraumatic injuries (i.e., lowering the height of the staircase) could be due to the existence of incentive programs that constrain employers' capacity to report less-detectable injuries. In apparent corroboration of this hypothesis, the negative marginal effect of underground mines attains statistical significance if one drops all observations from 2006 (the first year of injury logs audited under the PPOV program) and after.

Stage Two of the analysis tests more detailed predictions of the model. Although the probit models presented in Table 2 superficially resemble those presented in Table 1, the injury-type variables are constructed differently. This time, each injury of a specified type and each injury whose detectability equals or is less than that of the specified type are coded as 1, while all injuries

[83] See note 10 and accompanying text for an explanation for why the variable was included.

of higher detectability than X are coded as 0.[84] In effect, this coding scheme enables one to examine *marginal differences between* adjacent injury categories.

As expected, the steepest step in the OSHA model lies at the juncture between traumatic nonsevere and severe injuries (Types C and D). Once again, the negative and significant marginal effect of Type A (hard-to-attribute) injuries likely reflects the predominance of audit stringency bias, i.e., the fact that unreported hard-to-attribute injuries are only rarely being uncovered by OSHA auditors.

The MSHA models in Table 2 bring to light additional subtleties. Once again, I estimate three specifications: one for the entire sample that differentiates between surface and underground mines, one for underground mines, and one for coal mines that differentiates between surface and underground mines and also includes a union dummy. As expected, the step between nontraumatic, easy-to-attribute and traumatic, nonsevere injures (C→B) is large and significant for all mines across all models. However, the step between nontraumatic, easy-to-attribute injuries and nontraumatic, hard-to-attribute ones (B→A) is also positive and significant for underground mines, although not among surface mines. In other words, there is *an additional step* between Type A and Type B injuries, but only for underground mines. All of these findings accord with the model's core predictions. As before, the union dummy is uniformly insignificant, and the data exhibit significant variation by year, state and industry.[85]

---

[84] In Table 2, the "specified type" in question refers to the type to which the arrow is pointing. For example, in the first and sixth lines of Table 2, the reference to "Type D→C" implies that each injury of Type C as well as each injury of Types A and B is coded as 1, while each injury of Type D is coded as 0.

[85] Again, injuries taking place in New York establishments are significantly more likely to go unreported across the different models (except Model 4). In Model 1 (OSHA), injuries occurring in Arizona and Montana are significantly more likely to go unreported, whereas injuries occurring in California, Kentucky, Missouri, Mississippi, Nevada, South Dakota, and Virginia are significantly less likely to go unreported. In Models 2-4 (MSHA), injuries occurring in Kentucky and West Virginia are significantly more likely to go unreported. For OSHA, several industries – *Building Construction General Contractors And Operative Builders (15); Chemicals And Allied Products (28); Electronic And Other Electrical Equipment And Components, Except Computer Equipment (36); and Local, Suburban Transit And*

Turning to changes over time, Figure 4 explores OSHA's transition from the RK program to the NEP auditing program. The trend lines show annual means averaged across all injury sets. The top line represents total unreported injuries; the middle line represents unreported hard-to-attribute injuries; and the bottom line depicts unreported severe injuries. As noted earlier, the rigor of each audit increased markedly under the NEP program, although the average annual frequency of audits slightly declined. A shift of this type not only should have increased the overall frequency of unreported injuries because of audit stringency bias, but also should have had a larger impact on hard-to-attribute injuries than severe ones. The trends displayed in the figure align well with the model's predictions. Although the mean number of hard-to-attribute (and total) injuries that went unreported rose dramatically in 2007, the rise in unreported severe injuries was negligible. Table A (included as an appendix) formalizes these comparisons by estimating negative binomial models in which the unit of the analysis is the injury set and the dependent variable is the frequency of (severe, hard-to-attribute, and total) unreported injuries, respectively.[86] The models confirm that relative to prior years, the transition to the NEP program in 2007 had a statistically significant impact on the frequency of unreported hard-to-attribute (and total) injuries, but not on the frequency of unreported severe injuries.

In Stage Three, the key question becomes whether scrutinizing establishment-level characteristics and the composition of reported injuries could help regulators identify which firms underreport the most. This portion of the analysis relies exclusively on the injury set as the unit of analysis.

Interurban Highway Passenger Transport (41) – predict a significantly increased probability of underreporting. The MSHA Models 2 and 3 show that injuries occurring at stone mines are significantly more likely to be underreported. The time trends in Table 2 mirror those in Table 1. The year 1996 is excluded as the baseline in all models. Pennsylvania is excluded from the state dummies; coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

[86] All models include robust standard errors, and the number of employees is used as an exposure term. Other model covariates include the number of employees (in hundreds), an indicator of union status, and state and industry dummies.

Based on the results presented in Table 3, the answer is a tentative yes. Having a relatively high percentage (within the top quartile) of reported injuries that are highly detectable is a statistically significant predictor of underreporting in both regimes. As one would expect given the findings from Stages 1 and 2, the precise nature of the red flag varies across models. In the OSHA regime, a reported percentage of *severe* injuries that places the employer in the top quartile of the sample predicts a large (roughly 57%) and statistically significant increase in unreported injuries. Meanwhile, in the full MSHA sample, a reported percentage of *easy-to-attribute* injuries that places the employer in the top quartile of the sample predicts a sizable (roughly 54%) increase in unreported injuries.[87] The top-quartile easy-to-attribute dummy loses statistical significance in both MSHA subsamples, however, presumably due to the small number of observations. Employment retains its robust negative association with the frequency of unreported injuries across all models. As before, there is significant variation across states, industries, and years.[88]

The union dummy displays an erratic pattern: it is significant and *greater* than 1 in the OSHA model, but statistically insignificant in the MSHA coal model. Given its uniform statistical

[87] These findings are consistent regardless of whether one calculates the top-quartile dummies across the whole sample or separately for each industry, although the coefficient estimates fluctuate slightly. (The coefficients presented in Table 3 define top-quartile dummies across the entire sample.) The precise definition of quartile also varies slightly across samples. In most models, an observation is flagged as being in the top quartile if the fraction of reported injuries is *greater than or equal to* that of the 75th percentile. Defining the dummy in this manner (as opposed to defining it as strictly greater than the 75th percentile) was the only way to avoid flagging zero injuries. The sole exception to this rule is the top-quartile dummy for severe injuries in Model 1, which was defined to include only observations that are *strictly greater than* the 75th percentile so as to avoid flagging all observations. At least 20 percent and no more than 30 percent of all observations were flagged in every model.

[88] Establishments in Missouri (except in Model 4) and establishments in Mississippi (except in Model 3) underreport a significantly smaller number of injuries. The MSHA models show that mines in Indiana underreport a significantly smaller number of injuries. In Model 1, several industries – *Building Construction General Contractors And Operative Builders (15); Food and Kindred Products (20); Furniture And Fixtures (25); Chemicals And Allied Products (28); Primary Metal Industries (33); Fabricated Metal Products (34); Industrial And Commercial Machinery And Computer Equipment (35); Electronic, Electrical Equipment And Components, Except Computer Equipment (36); Transit And Interurban Highway Passenger Transportation (41)* – predict a significantly greater number of unreported injuries. Models 2 and 3 predict that stone mines underreport a significantly greater number of injuries. In Model 1, the 2006 coefficient is less than one and significant at the .1% level, and the 2007 and 2008 coefficients are greater than one and significant at the .1% level. No clear time trends are apparent in the MSHA models. As before, Pennsylvania is excluded from the state dummies; 1996 is excluded from the year dummies; and coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

insignificance in Stages 1 and 2 of the analysis, the surprise is that the dummy gains significance (and carries a positive sign) in the OSHA model. It is important to bear in mind that although the prior models compare all injuries to one another, Table 3 makes comparison across injury *sets*. Perhaps in the OSHA-regulated economy, unions play a vitally important role in enhancing the scope and rigor of audits, so that *ceteris paribus*, auditors find more violations at unionized establishments. If the increase in audit intensity bias is sufficiently large, it could outweigh unions' (presumed) salutary effect on the overall reporting of workplace injuries. Alternatively, perhaps incentive programs are more prevalent at unionized establishments.

Table 4 presents results from an OLS model in which the dependent variable is the *percentage* of total injuries that went unreported among establishments that reported at least one injury. This time, none of the covariates in the OSHA model has any predictive value. In the MSHA model, on the other hand, the coefficient on the easy-to-attribute top-quartile dummy remains sizable and statistically significant in the main model and in the coal subsample. For the first time, the coefficient on underground mine also attains statistical significance. Employment remains negative and significant in all three MSHA models, and the union dummy is uniformly insignificant. Although a few of the state, year, and industry dummies attain statistical significance within models, none is robust across all models.[89]

[89] No clear state trends are apparent across models although there is significant variation across states *within* each of Models 1-4. In Model 1, two industries – *Chemicals And Allied Products (28); Local, Suburban Transit And Interurban Highway Passenger Transport (41)* – predict a significantly higher percentage of unreported injuries, whereas several other industries – *Agricultural Production Crops (1); Agricultural Production Livestock And Animal Specialties (2); Agricultural Services (7); Water Transportation (44); Transportation Services (47); Communications (48)* – predict a significantly lower percentage of unreported injuries. Models 2 and 3 predict that stone mines underreport a significantly higher percentage of injuries. Similar to previous OSHA models, the 2006 coefficient is negative and significant at the 1% level in Model 1 while the 2007 coefficient is positive and significant at the 5% level. The MSHA models do not show clear time trends. As before, Pennsylvania is excluded from the state dummies; 1996 is excluded from the year dummies; and coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

The final models presented in Table 5 again estimate the number of unreported injuries per (quasi) injury set, but this time only among "nonreporters," i.e., establishments that timely reported zero injuries. As expected, the frequency of unreported injuries declines uniformly with establishment size. There is some robust variation by industry, although the findings do not vary consistently by state or year.[90] The more policy-relevant question, however, is how the prevalence of underreporting among nonreporters compares to the prevalence of underreporting among firms that reported at least one injury ("reporters"). I approach this question in two ways. First, I compare the respective percentages of firms that were highly noncompliant, defined as failing to report more than one injury. Secondly, I compare the mean number of unreported injuries across the two groups. In making these comparisons, I divide each group into small, medium and large firms, and calculate mean values for each subgroup. The results of these analyses, presented in Table 6, suggest that nonreporters are generally no more likely than reporters to engage in underreporting. Indeed, by many measures, nonreporters in the OSHA sample actually underreport *fewer* injuries than reporters. In the MSHA sample, there are no statistically significant distinctions at all between the two groups.

Notwithstanding these broad trends, however, there are a few industries in which nonreporters *do* appear less compliant than their peers. In the OSHA setting, for example, non-unionized, medium-sized nonreporters in the Transportation, Communications, Electric, Gas and Sanitary Services industry (SIC Major Groups 41-45, 47-49) and the Wholesale Trade Industry (SIC Major Groups 50-51) performed worse than other firms in the same industries. Meanwhile, in the MSHA sample, nonreporters in the nonmetal mining sector appear to be less compliant than other

[90] Again, no clear state trends are apparent across models although there is significant variation across states *within* each of Models 1-4. In Model 1, several industries – *Heavy Construction, Except Building Construction – Contractors (16); Food And Kindred Products (20); Lumber And Wood Products, Except Furniture (24); Furniture And Fixtures (25); Motor Freight Transportation (42); Transportation By Air (45)* – predict a significantly greater number of unreported injuries. The MSHA models do no display clear industry trends. No clear time trends emerged in the results. As before, Pennsylvania is excluded from the state dummies; 1996 is excluded from the year dummies; and coal and SIC Major Group 80 - Health Services are the excluded industry dummies.

mines. Yet these findings should be viewed with caution. Given the small sample sizes, they could merely be statistical artifacts. Moreover, since the MSHA sample includes only those mines that untimely reported at least one injury, all of the MSHA comparisons should be construed as "worst-case depictions" of the relative prevalence of underreporting among nonreporters.[91]

VI. Practical Challenges Posed by Policy Implementation

My empirical findings suggest that regulatory intensity, injury detectability, and underreporting relate to one another in systematic and predictable ways that worker protection agencies could profitably explore. Because of the confounding effects of audit stringency bias and the "stickiness" of firms' reporting behavior due to incentive programs, testing the predictions of the theoretical model is not entirely straightforward. Nevertheless, my findings generally bear out its core predictions. More-detectable injuries are generally less likely to be underreported than less-detectable injuries, and the single best predictor of underreporting relates in coherent ways to the overall rigor of the inspection regime. The reporting of less-detectable injuries is likewise far more responsive to fluctuations in regulatory stringency. Stage Three of the analysis further suggests that analyzing the *composition* of injuries in mandated filings could help regulators anticipate which employers are most likely to underreport.

Yet my theoretical model is relatively static in the sense that regulatory behavior is treated as largely exogenous, and firms are assumed to respond in straightforward, predictable ways to changes in the likelihood of detection. In other words, the model assumes that the interaction between agency and employer closely resembles a one-shot game. If one were to relax this assumption and assume that *both* sides in the enforcement game behave in an iterative and strategic fashion, how might this

[91] This is so because the sample excludes mines that actually had, and likewise reported, zero injuries during a given fiscal year.

affect the model's predictions? For example, if firms learned of the precise algorithm being used, could they respond in ways that would thwart the agency's capacity to implement this approach?

Although development of a full game-theoretic model is beyond the scope of this paper, several practical considerations suggest that employers could not easily strip the agency of its targeting capacity. Suppose, for example, that an employer who engages in extensive underreporting, and therefore would ordinarily fall into the top quartile of reported traumatic injuries, has just learned of the new targeting strategy and wishes to avoid being targeted. Realistically, there are only three strategies at its disposal. First, the employer could start to fully report all nontraumatic injuries. If this occurs, the program will achieve precisely its intended effect of deterring underreporting. The more important question, therefore, is whether an employer can evade detection in ways that thwart the policy's intended goals. Two possibilities come to mind. First, an employer could choose to report even fewer *traumatic* injuries. Second, an employer could report "extra" nontraumatic injuries, such as those that do occur but are not reportable.[92] Either approach would lower the percentage of nontraumatic injuries reported, and thereby lessen the likelihood that an employer will be audited.

If employers responded in this way, however, the agency could develop counter-measures to deter such behavior. For example, if employers responded by reporting even fewer traumatic injuries, the agency could also target employees whose total injury count fell below industry norms[93], or increase the penalties for underreporting traumatic injuries. If some employers responded by reporting injuries it knew were not reportable, the agency could respond by penalizing

[92] See note 57 for a discussion of "overreporting" in the OSHA dataset.

[93] This was precisely the targeting criteria used during much of OSHA's NEP program, which targeted establishments in high-hazard industries that participated in the OSHA Data Initiative (ODI and reported a DART (days away, restricted, or transferred) injury rate between 0 and 4.2. See US DOL, OSHA, Directive Number: 09-08 (CPL 02) (2009), and US DOL, OSHA, Directive Number: 10-07 (CPL 02) (2010).

persistent over-reporting. In short, relaxing the assumption of exogeneity does not automatically doom a targeting strategy such as that outlined here. It simply means that worker protection agencies, like the IRS, would have to continuously adjust their targeting criteria (and perhaps also their penalty structure) in response to observed changes in industry behavior.

Another important issue that bears on policy implementation is the likelihood that the *social cost* of underreporting varies by injury type. For example, the choice of whether to devote the marginal auditing dollar to increasing the frequency of inspections, or to increasing the rigor of inspections, depends in part on the *social cost* associated with undercounting different types of injuries. Since musculoskeletal diseases are the costliest injuries to treat, an agency might reasonably devote its marginal inspection dollar to increasing inspection rigor and thereby improving the reporting of hard-to-attribute (Type A) injuries. Alternatively, the agency might choose to conduct more frequent inspections to improve the measurement of traumatic yet nonsevere (Type C) injuries, whose work-relatedness is usually easier to ascertain. The optimal balance between inspection frequency and rigor cannot be determined in a vacuum, but only with reference to such broader policy considerations.

VII. Conclusion

Like the Internal Revenue Service, several of the Department of Labor's worker protection agencies receive self-reported filings from U.S. establishments on the quantity and magnitude of regulated activities.[94] Each agency relies, at least in part, on this self-reported information to calibrate the frequency and intensity of its regulatory inspections and audits. Yet self-reported injury filings – like income tax filings – are prone to underreporting. Unlike the IRS, which has developed

[94] It should be noted that in contrast to the IRS and MSHA, which receive information at least once per year from all regulated establishments, OSHA only receives such information from a sizable subset of establishments in high-hazard industries. Nevertheless, the same essential principles apply.

proprietary algorithms to detect telltale patterns or red flags of underreporting in income tax filings, neither OSHA nor MSHA has developed robust statistical techniques to ferret out the underreporting of occupational injuries and illnesses. An important first step toward this goal is determining which *types* of injuries are most (or least) likely to evade detection.

I take up this challenge by completing four interrelated tasks. First, I develop an informal theoretical model of the relationship between regulatory intensity, injury type, and underreporting. The model yields concrete predictions about how injury reporting behavior will respond to changes in the frequency and/or stringency of auditing. Secondly, I propose a scheme whereby different types of injuries can be classified according to their relative detectability. Third, using a dataset that encompasses thirteen years of granular audit data obtained from OSHA and MSHA, I test the model's predictions regarding which types of injuries will be underreported the most, both across regimes and over time. Finally, I explore whether any readily observable, establishment-level covariates – such as the percentage of reported injuries that are highly detectable – could be used by labor regulators to identify which employers are underreporting the most injuries.

By and large, my findings provide considerable grounds for optimism that the model reasonably approximates reality. The best predictors of underreporting in less rigorous regimes differ from the best predictors of underreporting in more rigorous regimes in precisely the ways that theory predicts. The model's more nuanced predictions are sometimes difficult to test because of the confounding influence of audit stringency bias, the simultaneous adoption of multiple regulatory reforms, and the prevalence of incentive programs that induce workers to hide their injuries. Nevertheless, my findings still largely bear out the model's theoretical predictions. For example, OSHA's transition to a more rigorous auditing regime around 2007 had a dramatic effect on the frequency of unreported hard-to-attribute injuries, but no statistically significant effect on the

frequency of unreported severe injuries. Most importantly, at least one formulation of the red flag explored – a reported percentage of highly-detectable injuries that places a firm within the top quartile of its respective sample – does have significant predictive value. Specifically, having a high percentage of reported injuries that are *severe* is a significant predictor of the number of underreported injuries in the OSHA regime, whereas having a high percentage of reported injuries that are *easy to attribute* is a significant predictor of both the frequency and percentage of underreported injuries in the MSHA regime. Overall, my findings provide grounds for optimism that DOL's worker protection agencies could use data-mining techniques to combat underreporting in a manner similar to their counterparts at the IRS.

The paper also briefly discusses several challenges associated with making the leap from theory to practice. I suggest that even if one relaxes the assumption of exogeneity, so that both regulators and firms behave in an iterative and strategic manner, the use of targeting techniques to identify the most likely violators could significantly improve upon the status quo. I also argue that to find the optimal tradeoff between the frequency and stringency of audits, regulators should weigh the respective social costs associated with underreporting different types of injuries.

My findings point to several promising areas for future research. First, the theoretical model proposed here could be tested in other settings to explore whether it is generalizable to other time periods and enforcement regimes. Finding new datasets on which to test its core predictions is particularly vital, since the datasets used here were prone to both measurement error and selection bias. Conducting audits on a stratified random sample of *all* establishments within a particular industry would yield the most valuable insights.[95] Secondly, the injury classification scheme proposed, including the categorization of various injuries as severe, traumatic, and/or hard-to-

[95] In a similar vein, a report issued on 2014, the Department of Labor's Office of the Inspector General emphasized the importance of "deriving betters estimates of [underreporting's] overall occurrence" by performing random audits from one or more sectors of the mining industry. See US DOL, OIG, *MSHA Has Taken Steps* (2014).

attribute, could potentially be refined in ways that would enhance its predictive value. Finally and most importantly, it remains to be seen whether red flags similar to those developed here – such as the respective percentages of severe or easy-to-attribute injuries reported to regulatory officials – could become tools of practical value to DOL regulators. Using field experiments and other experimental techniques to explore their predictive accuracy would give inspectors new leverage in their efforts to combat underreporting, and stimulate the development of more sophisticated targeting methods.

Table 1: Average Marginal Effect of Injury⊕ and Establishment Characteristics on the Probability that an Injury is Not Reported: Probit Models Testing Basic Injury Characteristics

	1. OSHA	2. MSHA	3. MSHA (Underground Only)	4. MSHA (Coal Only)
Severe Inj. (Type D)	-0.066***	-0.020*	-0.018	-0.016
	(0.01)	(0.01)	(0.01)	(0.01)
Traumatic Inj. (Types C, D)	-0.016	-0.073***	-0.070***	-0.091***
	(0.01)	(0.01)	(0.02)	(0.02)
Easy-to-Attr. Inj. (Types B, C, D)	0.036**	-0.024*	-0.037**	-0.035**
	(0.01)	(0.01)	(0.01)	(0.01)
Intermediate Inj.		0.016	0.016	0.024*
		(0.01)	(0.01)	(0.01)
Employment	-0.000***	-0.017**	-0.022***	-0.024***
	(0.00)	(0.01)	(0.00)	(0.00)
Underground Mine		-0.025		-0.021
		(0.02)		(0.02)
Union	0.009			-0.017
	(0.01)			(0.01)
Injury Char.×Employment Interactions	Yes	Yes	Yes	Yes
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of Injuries	12093	6415	3995	3666
Number of Clusters	2523	291	104	117
Mean of Dependent Variable	0.127	0.096	0.070	0.080

⊕ Injury types are as follows: Type A is nontraumatic and hard-to-attribute; Type B is nontraumatic and easy-to-attribute; Type C is traumatic and nonsevere; and Type D is traumatic and severe.

Model: The model is a probit regression. Average marginal effects are presented. Standard errors are clustered on the audit identification number (OSHA) or on the mine identification number (MSHA). Significance levels: *** .1%, ** 1%, * 5%.

Dependent Variable: The dependent variable is set to 1 if an injury was not recorded on the audited OSHA 200 or 300 Log or if an injury was reported to MSHA after the year close for the calendar year of the injury, and 0 otherwise.

Covariates: All models include indicators of whether an injury is classified as severe, traumatic, or easy-to-attribute; the number of employees at the establishment (in hundreds); interactions between each injury type indicator and employment; and state and year dummies. Models 1-3 include controls for industry. For MSHA models, an indicator of whether an injury is intermediate is included, and industries are the canvas codes (coal, sand and gravel, stone, nonmetal, and metal). Models 1 and 4 include an indicator of union status at the injury site for the calendar year of the injury. Models 2 and 4 include an indicator of whether the injury occurred at an underground mine.

Unit of Observation: The unit of observation is the injury.

Sample: In Model 1, the sample includes reported and unreported injuries (excluding occupational illnesses) collected during audits conducted under the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009), covering injury logs from the years 1996-2006, and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012), covering injury logs from the years 2007-2009. In Model 2, the sample includes MSHA reported and late-reported injuries (excluding occupational illnesses, fatalities, and accidents not resulting in injuries) for any mine-year in which at least one injury was reported after the year close. Model 2 includes injuries from 1992, 1994, 1996-1998, and 2000-2012. Before 2000, the data only includes one mine for each year. Model 3 is limited to injuries in the MSHA data that occurred at underground mines and includes injuries from 2000-2012. Model 4 is limited to injuries in the MSHA data that occurred at coal mines and includes injuries from 2000-2001 and 2003-2012. For details on sample construction and omitted observations, see footnote 70. In Models 1-3, some observations were dropped due to their having attributes that perfectly predict success or failure. Following is the number of observations dropped in each model; Model 1: 135 obs, Model 2: 7 obs, Model 3: 3 obs.

Table 2: Average Marginal Effect of Injury Type⊕ and Establishment Characteristics on the Probability that an Injury is Not Reported: Probit Models Testing Marginal Differences Between Injury Types

	1. OSHA	2. MSHA	3. MSHA (Underground Only)	4. MSHA (Coal Only)
Type D→C [at Surface Mine in Models 2 & 4]†	0.069***	0.029		0.005
	(0.01)	(0.02)		(0.02)
Type C→B [at Surface Mine in Models 2 & 4]†	0.018	0.051*		0.146***
	(0.02)	(0.02)		(0.04)
Type B→A [at Surface Mine in Models 2 & 4]†	-0.038*	0.027		0.061
	(0.02)	(0.02)		(0.06)
Employment	-0.000***	-0.017**	-0.022***	-0.024***
	(0.00)	(0.01)	(0.00)	(0.00)
Underground Mine		[-0.046*]^		[0.010]^
		[(0.02)]		[(0.02)]
Type D→C at Underground Mine†		0.002	0.007	0.005
		(0.01)	(0.01)	(0.01)
Type C→B at Underground Mine†		0.076***	0.060***	0.066***
		(0.02)	(0.02)	(0.02)
Type B→A at Underground Mine†		0.042**	0.056***	0.061***
		(0.02)	(0.01)	(0.02)
Union	0.009			-0.017
	(0.01)			(0.01)
Injury Type×Employment Interactions	Yes	Yes	Yes	Yes
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of Injuries	12093	6415	3995	3666
Number of Clusters	2523	291	104	117
Mean of Dependent Variable	0.127	0.096	0.070	0.080

⊕ Injury types are as follows: Type A is nontraumatic and hard-to-attribute; Type B is nontraumatic and easy-to-attribute; Type C is traumatic and nonsevere; and Type D is traumatic and severe.

† These are not the covariates included in the model. See Covariates for a full list of the covariates used in each probit model.

∧ The marginal effect of underground at Type C is presented. The marginal effects of underground at other injury types is not shown and were not significant at the 5% level for either model.

Model: The model is a probit regression. Average marginal effects are presented. Standard errors are clustered on the audit identification number (OSHA) or on the mine identification number (MSHA). Significance levels: *** .1%, ** 1%, * 5%.

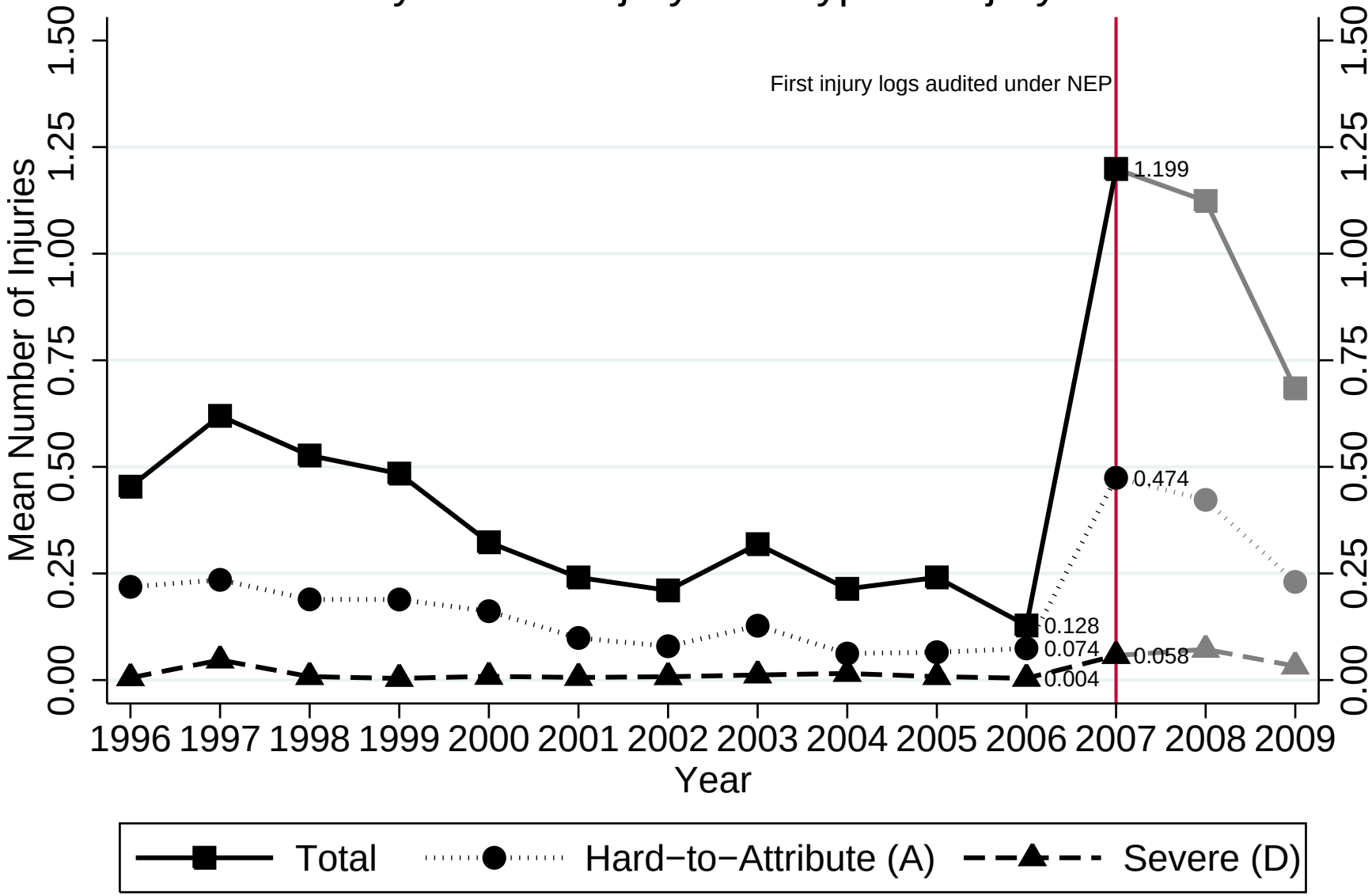
Dependent Variable: The dependent variable is 1 if an injury was not recorded on the audited OSHA 200 or 300 Log or if an injury was reported to MSHA after the year close for the calendar year of the injury, and 0 otherwise.

Covariates: While the table above presents marginal effects, the probit models include the following covariates. Each model includes injury type indicators for injury types C, B, and A, which are 1 for that injury type and also for injury types of lesser detectability (see Figure 3). Also included are the number of employees at the establishment (in hundreds), interactions between each injury type indicator and employment, and state and year dummies. Models 1-3 include controls for industry. For MSHA, industries are the canvass codes (coal, sand and gravel, stone, nonmetal, and metal). Models 1 and 4 include union status at the establishment for the calendar year of the injury. Models 2 and 4 include an indicator of whether the injury occurred at an underground mine, and interaction terms between each injury type indicator and the underground dummy.

Unit of Observation: The unit of observation is the injury.

Sample: In Model 1, the sample includes reported and unreported injuries (excluding occupational illnesses) collected during audits conducted under the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009), covering injury logs from the years 1996-2006, and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012), covering injury logs from the years 2007-2009. In Models 2-4, the sample includes MSHA reported and late-reported injuries (excluding occupational illnesses, fatalities, and accidents not resulting in injuries) for any mine-year in which at least one injury was reported after the year close. Model 2 includes injuries from 1992, 1994, 1996-1998, and 2000-2012. Model 3 is limited to underground mine injuries and includes injuries from 2000-2012. Model 4 is limited to coal mine injuries and includes injuries from 2000-2001 and 2003-2012. For details on sample construction and omitted observations, see footnote 70. In Models 1-3, some observations were dropped due to their having attributes that perfectly predict success or failure. Following is the number of observations dropped in each model; Model 1: 135 obs, Model 2: 7 obs, Model 3: 3 obs.

Figure 4: Mean Number of Unreported Injuries per Audit, OSHA, by Year of Injury and Type of Injury



Notes: Each point represents the mean across injury sets (i.e. audits) for each year. Each year represents the year of the injury logs that were audited (not the year in which the audit took place). NEP stands for National Emphasis Program. The years 2008 and 2009 are grayed out because the selection criteria for establishments and industries changed under NEP2 (2007 and 2008) and NEP3 (2008 and 2009); thus, these years are not direclty comparable to prior years. See Section II: Description of Data in the paper for more details.

Unit of Observation: The unit of observation is the injury set (i.e. audit).

Sample: The sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (1996-2009), collectively covering injury logs from the years 1996-2009. For more details on sample construction and omitted observations, see footnote 70.

Table 3: Effect of Reported Injury Distribution and Establishment Characteristics on Number of Unreported Injuries (Among Establishments and Mines Reporting at Least One Injury)

	1. OSHA	2. MSHA	3. MSHA (Underground Only)	4. MSHA (Coal Only)
Top Quartile, % Severe Inj.	1.569***	1.026	0.982	1.026
	(0.18)	(0.13)	(0.16)	(0.17)
Top Quartile, % Traum. Inj.	0.854	1.041	1.064	0.974
	(0.13)	(0.18)	(0.26)	(0.22)
Top Quartile, % Easy-to-Attr. Inj.	0.875	1.544**	1.154	1.432
	(0.14)	(0.20)	(0.22)	(0.27)
Employment	0.898***	0.786**	0.712***	0.709***
	(0.01)	(0.07)	(0.03)	(0.03)
Underground Mine		0.926		1.056
		(0.16)		(0.20)
Union	1.440**			0.879
	(0.17)			(0.15)
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of (Quasi) Injury Sets	2556	342	169	175
Number of Clusters		233	101	110
Mean of Dependent Variable (Raw)	0.523	1.541	1.627	1.640

Model: The model is a Negative Binomial, selected because the sample variance of the dependent variable far exceeds the sample mean. Coefficients are presented as Incident Rate Ratios (IRR). The number of employees is used as an exposure term. Robust standard errors are used for OSHA, and standard errors are clustered on the mine identification number for MSHA. Significance Levels: *** .1%, ** 1%, * 5%.

Important Caveat: The sample used for the OSHA model includes some establishments which do not underreport at all. This is not true of the MSHA samples used in Models 2, 3, and 4; all mines in these samples underreport at least one injury.

Dependent Variable: The dependent variable is the number of unreported injuries in the (quasi) injury set.

Covariates: Top-quartile dummies in each model indicate whether or not the reported fractions of severe, traumatic, and easy-to-attribute injuries in the firm's (quasi) injury set place it within the top quartile of the sample analyzed. Covariates invariably include the number of employees at the establishment (in hundreds), and state, year and industry dummies. In most models, an observation is flagged as being in the top quartile if the fraction of reported injuries is *greater than or equal to* that of the 75th percentile. Defining the dummy in this manner (as opposed to defining it as strictly greater than the 75th percentile) was the only way to avoid flagging zero injuries. The sole exception to this rule is the top-quartile dummy for severe injuries in Model 1, which was defined to include only observations that are *strictly greater than* the 75th percentile to avoid flagging all observations. In each model, at least 20 percent and no more than 30 percent of all observations are flagged. Models 1 and 4 include a union status indicator, and models 2 and 4 include an indicator for whether the injuries in the quasi injury set occured in an underground mine.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: In Model 1, the sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009 that included at least one reported injury. In Models 2, 3, and 4, the sample includes MSHA quasi injury sets for mine-years with at least one reported injury and one injury reported after the year close. Model 2 includes quasi injury sets from 1992, 1994, 1996-1998, and 2000-2012. Before 2000, the data only includes one quasi injury set for each year. Model 3 excludes all quasi injury sets that are not from an underground mine and includes quasi injury sets from 2000-2012. Model 4 excludes all quasi injury sets that are not from a coal mine and includes quasi injury sets from 2000-2001 and 2003-2012. For more details on sample construction and omitted observations, see footnote 70.

Table 4: Effect of Reported Injury Distribution and Establishment Characteristics on Percentage Unreported Injuries (Among Establishments and Mines Reporting at Least One Injury)

	1. OSHA	2. MSHA	3. MSHA (Underground Only)	4. MSHA (Coal Only)
Top Quartile, % Severe Inj.	0.008	0.018	0.018	0.009
	(0.01)	(0.02)	(0.02)	(0.03)
Top Quartile, % Traum. Inj.	-0.007	0.003	-0.011	0.007
	(0.01)	(0.02)	(0.03)	(0.03)
Top Quartile, % Easy-to-Attr. Inj.	0.003	0.127***	0.050	0.094**
	(0.01)	(0.02)	(0.03)	(0.03)
Employment	0.000	-0.015*	-0.028***	-0.026***
	(0.00)	(0.01)	(0.00)	(0.01)
Underground Mine		-0.061*		-0.100**
		(0.02)		(0.03)
Union	0.017			-0.047
	(0.01)			(0.02)
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	No
Number of (Quasi) Injury Sets	2556	342	169	175
Number of Clusters		233	101	110
Mean of Dependent Variable (Raw)	0.072	0.190	0.138	0.162

Model: The model is OLS. Robust standard errors are used for OSHA, and standard errors are clustered on the mine identification number for MSHA. Significance Levels: *** .1%, ** 1%, * 5%.

Important Caveat: The sample used for the OSHA model includes some establishments which do not underreport at all. This is not true of the MSHA samples used in Models 2, 3, and 4; all mines in these samples underreport at least one injury.

Dependent Variable: The dependent variable is the percentage of unreported injuries in the (quasi) injury set.

Covariates: Top-quartile dummies in each model indicate whether or not the reported fractions of severe, traumatic, and easy-to-attribute injuries in the firm's injury set place it within the top quartile of the sample analyzed. Covariates invariably include the number of employees at the establishment (in hundreds), and state, year and industry dummies. In most models, an observation is flagged as being in the top quartile if the fraction of reported injuries is *greater than or equal to* that of the 75th percentile. Defining the dummy in this manner (as opposed to defining it as strictly greater than the 75th percentile) was the only way to avoid flagging zero injuries. The sole exception to this rule is the top-quartile dummy for severe injuries in Model 1, which was defined to include only observations that are *strictly greater than* the 75th percentile to avoid flagging all observations. In each model, at least 20 percent and no more than 30 percent of all observations are flagged. Models 1 and 4 include a union status indicator, and models 2 and 4 include an indicator for whether the injuries in the quasi injury set occured in an underground mine.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: In Model 1, the sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009 that included at least one reported injury. In Models 2, 3, and 4, the sample includes MSHA quasi injury sets for mine-years with at least one reported injury and one injury reported after the year close. Model 2 includes quasi injury sets from 1992, 1994, 1996-1998, and 2000-2012. Before 2000, the data only includes one quasi injury set for each year. Model 3 excludes all quasi injury sets that are not from an underground mine and includes quasi injury sets from 2000-2012. Model 4 excludes all quasi injury sets that are not from a coal mine and includes quasi injury sets from 2000-2001 and 2003-2012. For more details on sample construction and omitted observations, see footnote 70.

Table 5: Effect of Establishment Characteristics on Number of Unreported Injuries (Among Establishments and Mines Reporting Zero Injuries)

	1. OSHA	**2. MSHA**	**3. MSHA** (Underground Only)	**4. MSHA** (Coal Only)
Employment	0.624***	0.242***	0.004***	0.004***
	(0.08)	(0.05)	(0.00)	(0.00)
Underground Mine		0.537***		
		(0.10)		
Union	1.255			
	(0.47)			
State and Year Dummies	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	No	No
Number of (Quasi) Injury Sets	755	72	7	7
Number of Clusters		66	7	7
Mean of Dependent Variable (Raw)	0.253	1.292	1.286	1.000

Model: The model is a Poisson regression. Coefficients are presented as Incident Rate Ratios (IRR). The number of employees is used as an exposure term. Robust standard errors are used for OSHA, and standard errors are clustered on the mine identification number for MSHA. Significance Levels: *** .1%, ** 1%, * 5%.

Important Caveat: The sample used for the OSHA model includes some establishments which do not underreport at all. This is not true of the MSHA samples used in Models 2, 3, and 4; all mines in these samples underreport at least one injury.

Dependent Variable: The dependent variable is the number of unreported injuries in the (quasi) injury set.

Covariates: All models include the number of employees at the establishment (in hundreds), and state and year dummies. Models 1-3 include controls for industry. For MSHA, the industry controls are the canvas codes (coal, sand and gravel, stone, nonmetal, and metal). Models 1 and 4 include an indicator of union status at the audited establishment or mine. Models 2 and 4 include an indicator of whether the injuries in the quasi injury set occurred in an underground mine.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: In Model 1, the sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009 that reported zero injuries for the audited year. In Models 2, 3, and 4, the sample includes MSHA quasi injury sets for mine-years with zero reported injuries and at least one injury reported after the year close. Model 2 includes quasi injury sets from 2001-2012. Model 3 excludes quasi injury sets that are not from underground mines and includes quasi injury sets from 2003, 2005-2006, 2008, and 2011. Model 4 excludes quasi injury sets that are not from coal mines and includes quasi injury sets from 2003-2004 and 2006-2007. For more details on sample construction and omitted observations, see footnote 70.

Table 6: Relationship Between Nonreporter Status on the Number of Unreported Injuries by Regime, Industry, and Size

OSHA (Sample Includes All Audited Firms)

Industry, Size	Number of Injury Sets in Sample	Number of Nonreporters	Percentage of Injury Sets with Multiple Unreported Injuries		Mean Number of Unreported Injuries	
			Nonreporter	Reporter	Nonreporter	Reporter
All Industries, Small	1112	386	4.15%	4.55%	0.22	0.19
All Industries, Medium	1101	229	4.80%*	9.06%	0.23*	0.40
All Industries, Large	1100	142	4.23%***	15.14%	0.39^	0.88
Transportation, Communications, etc., Medium (Non-Union)	79	20	10.00%*	0%	0.45*	0.10
Wholesale Trade, Medium (Non-Union)	51	10	10.00%*	0%	0.50**	0.07

MSHA (Sample Only Includes Firms That Reported At Least One Late Injury)

Industry, Size	Number of Quasi Injury Sets in Sample	Number of Nonreporters	Percentage of Quasi Injury Sets with Multiple Unreported Injuries		Mean Number of Unreported Injuries	
			Nonreporter	Reporter	Nonreporter	Reporter
All Industries, Small	136	68	14.71%	16.18%	1.28	1.22
All Industries, Medium	140	3	33.33%	23.36%	1.67	1.59
All Industries, Large	138	1	0%	26.28%	1.00	1.65
Nonmetal, All	23	6	66.67%**	5.88%	2.50**	1.06

$\wedge$ While the p-value for a two-tailed t-test is 0.062, the p-value for a one-tailed t-test is .031. Thus, we can reject the hypothesis that large OSHA reporters do not have more unreported injuries on average than non-reporters.

Significance: For percentage of injury sets with multiple unreported injuries, I compare the respective proportions among nonreporters and reporters using a two-tailed two-sample z-test of proportions. For mean number of unreported injuries, I compare the respective means among nonreporters and reporters using a two-tailed two-sample t-test. Significance Levels: *** .1%, ** 1%, * 5%.

Nonreporter: A nonreporter is defined as an establishment-year without any reported injuries. An establishment may be audited in multiple years, and its status as a reporter or nonreporter may change between years.

Size: Size is defined by the number of employees: "small" represents the bottom third of the distribution, "medium" represents the middle third of the distribution, and "large" represents the top third of the distribution. For OSHA, size is defined separately based on industry groupings (the data include the following seven of the ten OSHA "divisions": Agriculture, Forestry, and Fishing (SIC Major Groups 01, 02, 07); Construction (SIC Major Groups 15-17); Manufacturing (SIC Major Groups 20, 22-39); Transportation, Communications, Electric, Gas, and Sanitary Services (SIC Major Groups 41-45, 47-49); Wholesale Trade (SIC Major Groups 50-51); Retail Trade (SIC Major Groups 52-54, 59); and Services (SIC Major Groups 76, 80, 87)). For OSHA, statistics are first calculated separately for each industry grouping and size, and then averaged across all industry groupings. For MSHA, size and statistics are calculated across all mines.

Unit of Observation: The unit of observation is the (quasi) injury set.

Sample: The OSHA sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009. The MSHA sample includes quasi injury sets for mine-years with at least one injury reported after the year close from the years 1992, 1994, 1996-1998, 2000-2012. Before 2000, the data only includes one quasi injury set for each year. For more details on sample construction and omitted observations, see footnote 70.

Appendix

Table A: Effect of Year of Injury on the Number of Unreported Injuries, by Type of Injury (Among All OSHA Establishments)

	1. Total Num. Unreported Inj.	2. Num. Unreported Severe Inj.	3. Num. Unreported Hard-to-Attribute Inj.
1996 Dummy	0.457**	0.180	0.602
	(0.11)	(0.18)	(0.17)
1997 Dummy	0.625	1.107	0.742
	(0.16)	(0.61)	(0.22)
1998 Dummy	0.494**	0.160*	0.523*
	(0.12)	(0.14)	(0.16)
1999 Dummy	0.516**	0.111*	0.565
	(0.13)	(0.12)	(0.17)
2000 Dummy	0.265***	0.181	0.385**
	(0.07)	(0.16)	(0.12)
2001 Dummy	0.270***	0.152	0.309***
	(0.07)	(0.18)	(0.09)
2002 Dummy	0.263***	0.200**	0.251***
	(0.07)	(0.12)	(0.08)
2003 Dummy	0.298***	0.292*	0.327***
	(0.07)	(0.17)	(0.10)
2004 Dummy	0.248***	0.513	0.202***
	(0.07)	(0.35)	(0.07)
2005 Dummy	0.257***	0.189	0.218***
	(0.07)	(0.16)	(0.08)
2006 Dummy	0.134***	0.131	0.210***
	(0.04)	(0.16)	(0.07)
2008 Dummy	0.831	1.229	0.942
	(0.20)	(0.62)	(0.25)
2009 Dummy	0.468**	0.480	0.451**
	(0.12)	(0.31)	(0.14)
Employment	0.896***	0.907**	0.904***
	(0.01)	(0.03)	(0.01)
Union	1.451**	0.566	1.545**
	(0.17)	(0.19)	(0.22)
State and Industry Dummies	Yes	Yes	Yes
Number of Injury Sets	3311	3311	3311
Mean of Dependent Variable (Raw)	0.461	0.020	0.179

Model: The model is a Negative Binomial, selected because the sample variance of the dependent variable far exceeds the sample mean. Coefficients are presented as Incident Rate Ratios (IRR). The number of employees is used as an exposure term. The year dummy for 2007 was excluded. Each year represents the year of the injury logs that were audited (not the year in which the audit took place). The first year that injury logs were audited under the National Emphasis Program (NEP) is 2007. All models use robust standard errors. Significance Levels: *** .1%, ** 1%, * 5%.

Dependent Variable: For Model 1, the dependent variable is the total number of unreported injuries in the injury set. For Model 2, the dependent variable is the number of unreported severe injuries in the injury set. For Model 3, the dependent variable is the number of unreported hard-to-attribute injuries in the injury set.

Covariates: Each model includes the number of employees at the establishment in hundreds; an indicator of union status at the audited establishment; and state, year, and industry dummies.

Unit of Observation: The unit of observation is the injury set.

Sample: The sample includes all injury sets documented in the OSHA Audit and Verification Program of Occupational Injury and Illness Records (1997-2009) and the Injury and Illness Recordkeeping National Emphasis Program (2009-2012) audit data covering injury logs from the years 1996-2009. For more details on sample construction and omitted observations, see footnote 70.

References

Azaroff, Lenore S., Charles Levenstein, and David H. Wegman. "Occupational Injury and Illness Surveillance: Conceptual Filters Explain Underreporting." *American Journal of Public Health* 92 (2002): 1421-1429.

Barling, Julian, E. Kevin Kelloway, and Roderick D. Iverson. "Accidental Outcomes: Attitudinal Consequences of Workplace Injuries." *Journal of Occupational Health Psychology* 8 (2003): 74-85.

Boden, Leslie, and Alexander Ozonoff. "Capture-Recapture Estimates of Nonfatal Workplace Injuries and Illnesses.*" Annals of Epidemiology* 18 (2008): 500-506.

Boyles, Salynn. "$86 Billion Spent on Back, Neck Pain." *WebMD Health News*, February 12, 2008, http://www.webmd.com/back-pain/news/20080212/86-billion-spent-on-back-neck-pain.

California Department of Human Resources. *Workers' Compensation Preview.* January 2014, http://www.calhr.ca.gov/Documents/workers-compensation-preview.pdf.

Daniels, Christine, and Peter Marlow. "Literature Review on the Reporting of Workplace Injury Trends." *Health and Safety Laboratory* 36 (2005).

Dionne, Georges, and Pierre St-Michel. "Workers' Compensation and Moral Hazard." *The Review of Economics and Statistics* 73 (May 1991): 236-244.

Dong, Xiuwen S., Alissa Fujimoto, Knut Ringen, Erich Stafford, James W. Platner, Janie L. Gittleman, and Xuanwen Wang. "Injury Underreporting Among Small Establishments in the Construction Industry." *American Journal of Industrial Medicine* 54 (2011): 339-349.

Eastern Research Group. *Final Report (Revised): Evaluation of the Accuracy and Completeness*

of Nonfatal Injury and Illness Reporting in the Mining Industry. Lexington, MA: Eastern Research Group, June 11, 2013, http://www.dol.gov/asp/evaluation/reports/MSHA-Part50-Underreporting.pdf.

Fan, Z. Joyce, David K. Bonauto, Michael P. Foley, and Barbara A. Silverstein. "Underreporting of Work-Related Injury or Illness to Workers' Compensation: Individual and Industry Factors." *Journal of Occupational and Environmental Medicine* 48 (2006): 914-922.

Galizzi, Monica, Petra Miesmaa, Laura Punnett, Craig Slatin, and The Phase in Healthcare Research Team. "Injured Workers' Underreporting in the Health Care Industry: An Analysis Using Quantitative, Qualitative, and Observational Data." *Industrial Relations* 49 (2010): 22-43.

Gray, Wayne B., and John T. Scholz. "Does Regulatory Enforcement Work? A Panel Analysis of OSHA Enforcement." *Law and Society Review* 27 (1993): 177-214.

Greenhouse, Stephen. "Work-Related Injuries Underreported." *The New York Times.* November 16, 2009, http://www.nytimes.com/2009/11/17/us/17osha.html.

Harcourt, M., H. Lam, and S. Harcourt. "The Impact of Workers' Compensation Experience-Rating on Discriminatory Hiring Practices." *Journal of Economic Issues* 41 (2007): 681-699.

Hunting, Katherine L., and James L. Weeks. "Transport Injuries in Small Coal Mines: An Exploratory Analysis." *American Journal of Industrial Medicine* 23 (1993): 391-406.

Increasing Access to Workers' Compensation Medical Benefits for Low-Income and Immigrant Workers. NH Coalition for Occupational Safety and Health Research Report. March 31, 2008, http://www.nhcosh.org/pdfs/NHCOSH_WC_Report_3-31-08.pdf.

"Improving Construction Safety Performance." Report A-3, *The Business Roundtable* (January

1982).

Kniesner, Thomas J., and John D. Leeth. "Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States." *Journal of Risk and Uncertainty* 4 (1991): 75-90.

Kralj, Boris. "Employer Responses to Workers' Compensation Insurance Experience Rating." *Industrial Relations* 49 (Winter 1994): 41-61.

La Duke, Phil. "What Every Entrepreneur Should Know about Worker Safety." *Entrepreneur*. May 30, 2014, http://www.entrepreneur.com/article/234305.

Laffont, Jean-Jacques and David Martimort. *The Theory of Incentives: The Principal-Agent Model.* Princeton, NJ: Princeton University Press, 2002.

Lewis-Beck, S., and John R. Alford. "Can Government Regulate Safety? The Coal Mine Example." *The American Political Science Review* 74 (1980): 745-756.

Lubeck, Deborah P. "The Costs of Musculoskeletal Disease: Health Needs Assessment and Health Economics." *Best Practice & Research Clinical Rheumatology* 17 (2003): 529–539.

Machlup, Fritz. "Theories of the Firm: Marginalist, Behavioral, Managerial." *The American Economic Review* 57 (1967): 1-33.

McCurdy, S. A., M. B. Schenker, and S. J. Samuels. "Reporting of Occupational Injury and Illness in the Semiconductor Manufacturing Industry." *American Journal of Public Health* 81 (1991): 85-89.

Melfi, Patrick. "New OSHA Initiative Targets Underreporting of Workplace Injuries." *New York*

Labor and Employment Law Report. February 22, 2010, http://www.nylaborandemploymentlawreport.com/2010/02/articles/occupational-safety-and-health/new-osha-initiative-targets-underreporting-of-workplace-injuries.

Morantz, Alison. "Final Project Report: Designing a Pilot Program for Strategic Mine Safety and Health Improvements through the Use of Surveillance Data to Guide Targeted Inspection Activities." NIOSH Research Contract 200-2009-28820, September 28, 2012 (available from author upon request).

Morantz, Alison. "Coal Mine Safety: Do Unions Make a Difference?" *Industrial and Labor Relations Review* 66 (2013): 88-116.

Morantz, Alison and Alexandre Mas. "Does Post-Accident Drug Testing Reduce Injuries? Evidence from a Large Retail Chain." *American Law and Economics Review* 10 (2008): 246-302.

National Research Council. *Toward Safer Underground Coal Mines.* Washington, D.C.: National Academy Press, 1982.

Oleinick, A., J. V. Gluck, and K. E. Guire. "Establishment Size and Risk of Occupational Injury." *American Journal of Industrial Medicine* 28 (1995): 1-21.

Parker, David L., William R. Carl, L. Ronald French, and Frank B. Martin. "Characteristics of Adolescent Work Injuries Reported to the Minnesota Department of Labor and Industry." *American Journal of Public Health* 84 (1994): 606-611.

Paul, Leigh J., James P. Marcin, and Ted R. Miller. "An Estimate of the US Government's Undercount of Nonfatal Occupational Injuries." *Journal of Occupational and Environmental Medicine* 46 (2004): 10-18.

Polich, Steven. "Do Not Underestimate the Importance of OSHA Incidence Rates." *Lockton*

Companies. December 2012, http://www.lockton.com/Resource_/PageResource/MKT/Polich_OSHA%20incidence%20rates_Dec%2012%20update.pdf.

Probst, Tahira M., Ty L. Brubaker, and Anthony Barsotti. "Organizational Injury Rate Underreporting: The Moderating Effect of Organizational Safety Climate." *Journal of Applied Psychology* 93 (2008): 1147-54.

Robertson, Leon S., and J. Philip Keeve. "Worker Injuries: The Effects of Workers' Compensation and OSHA Inspections." *Journal of Health Politics, Policy and Law* 8 (1983): 581-597.

Rosenman, K. D., J. C. Gardiner, J. Wang, J. Biddle, A. Hogan, M. J. Reilly, K. Roberts, and E. Welch. "Why Most Workers with Occupational Repetitive Trauma Do Not File for Workers' Compensation." *Journal of Occupational and Environmental Medicine* 42 (2000): 25–34.

Rosenman, Kenneth D., Alice Kalush, Mary Jo Reilly, Joseph C. Gardiner, Mathew Reeves, and Zhewui Luo. "How Much Work-Related Injury and Illness Is Missed By the Current National Surveillance System?" *Journal of Environmental Medicine* 48 (2006): 357-365.

Ruser, John W. "Examining Evidence on Whether BLS Undercounts Workplace Injuries and Illnesses." *Monthly Labor Review* (August 2008): 20-32.

Shabecoff, Philip. "OSHA Seeks $2.59 Million Fine for Meatpacker's Injury Reports." *The New York Times.* July 22, 1987, http://www.nytimes.com/1987/07/22/us/osha-seeks-2.59-million-fine-for-meatpacker-s-injury-reports.html.

Shapiro, Sidney A. "Occupational Safety and Health Regulation." *Encyclopedia of Law and Economics, Cheltenham, Edward Elgar* 5540 (2000): 596-625.

Siddharthan, Kris, Michael Hodgson, Deborah Rosenberg, Donna Haiduven, and Audrey Nelson. "Under-Reporting of Work-Related Musculoskeletal Disorders in the Veterans Administration." *International Journal of Health Care Quality Assurance* 19 (2006): 463-476.

Smith, Nancy, Christi Griffin, John Newquist, Kris Hoffman, Mark Briggs, and Diane Price. *OSHA's Voluntary Protection Programs (VPP) Review: Findings and Recommendations*, a Report Submitted to David Michaels, Assistant Secretary for OSHA, (November 2011): 21-22, https://www.osha.gov/dcsp/vpp/vpp_report_nov_2011_rev_7-11-12.pdf.

Stevens, Arthur W. "New Comp Law Defines 'Preexisting Conditions'." *Black Chapman Webber & Stevens Attorneys* (blog). June 8, 2010, http://www.blackchapman.com/new-comp-law-defines-preexisting-conditions.

Sulzer-Azaroff, Beth and John Austin. "Does BBS Work? Behavior-Based Safety & Injury Reduction: A Survey of the Evidence." *Professional Safety* (July 2000): 19-24.

"The Burden of Musculoskeletal Diseases in the United States." Executive Summary. *Bone and Joint Decade* (2011), http://www.boneandjointburden.org/pdfs/bmus_executive_summary_low.pdf.

U.S. Department of Labor, Occupational Health and Safety Administration. *Injury and Illness Recordkeeping National Emphasis Program (RK NEP)*. Directive Number: 10-07 (CPL 02), September 28, 2010, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_10-07.pdf.

U.S. Department of Labor, Occupational Safety and Health Administration. *Audit and Verification Program of Occupational Injury and Illness Records.* Directive Number: CPL 02-00-138, January 12, 2006,

https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=3329.

U.S. Department of Labor, Occupational Safety and Health Administration. *Injury and Illness Recordkeeping National Emphasis Program (RK NEP).* Directive Number: 09-08 (CPL 02), September 30, 2009, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02_09-08.pdf.

U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from David Michaels to Regional Administrators. *Revised VPP Policy Memorandum #5: Further Improvements to the Voluntary Protection Programs (VPP).* June 29, 2011, https://www.osha.gov/dcsp/vpp/policy_memo5.html.

U.S. Department of Labor, Occupational Safety and Health Administration, Memorandum from Richard E. Fairfax, OSHA Deputy Assistant Secretary, to Regional Administrators. *Employer Safety Incentive and Disincentive Policies and Practices*. March 12, 2012, https://www.osha.gov/as/opa/whistleblowermemo.html.

U.S. Department of Labor, Occupational Safety and Health Administration. *Site-Specific Targeting 2014 (SST-14).* Directive Number: 14-01 (CPL 02), March 6, 2014, https://www.osha.gov/OshDoc/Directive_pdf/CPL_02-14-01.pdf.

U.S. Department of Labor, Occupational Safety and Health Administration. *US Department of Labor's OSHA Cites Houston Manufacturing Company for Hiding Work-Related Injuries and Illnesses; Fines Exceed $1.2 Million.* Release Number: 10-1179-DAL, September 1, 2010, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=NEWS_RELEASES&p_id=18261.

U.S. Department of Labor, Occupational Safety and Health Administration. *1998 Audit and Verification Program of Occupational Injury and Illness Records.* Directive Number: 00-1 (CPL 2), December 2, 1999, https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=DIRECTIVES&p_id=2003.

U.S. Department of Labor, Office of the Inspector General (OIG). *MSHA Has Taken Steps to Detect and Deter Underreporting of Accidents and Occupational Injuries and Illnesses, But More Action Is Still Needed.* March 31, 2014, http://www.oig.dol.gov/public/reports/oa/2014/05-14-001-06-001.pdf.

Weil, David. "If OSHA Is So Bad, Why Is Compliance So Good?" *The Rand Journal of Economics* 27 (1996): 618-640.

Welch, Laura, and Katherine Hunting. "Injury Surveillance in Construction: What Is an 'Injury', Anyway?" *American Journal of Industrial Medicine* 44 (2003): 191-196.

Wiatrowski, William J. "Examining the Completeness of Occupational Injury and Illness Data: An Update on Current Research." *Monthly Labor Review* (June 2014).

Wiatrowski, William J. "Using Workplace Safety and Health Data for Injury Prevention." *Monthly Labor Review* (October 2013).

Williams, Robert J., and J. Douglas Barrett. "Corporate Philanthropy, Criminal Activity, and Firm Reputation: Is There a Link?" *Journal of Business Ethics* 26 (2000): 341-350.

Effects of Unionization on Workplace Safety:

Regression Discontinuity Evidence from OSHA's Enforcement Data

By Aaron Sojourner[1]

Carlson School of Management

University of Minnesota

"The empirical literature on the relationship between unionization and workplace safety presents a curious puzzle. On the one hand, scholars have documented numerous ways unions help to promote safe work practices…. Yet most empirical studies of the relationship between unionization and important safety outcomes, such as injuries and fatalities, have failed to find statistically significant evidence of a `union safety effect.'" (Morantz, 2012)

This paper was prepared with funding from the U.S. Department of Labor. The views expressed are those of the authors and should not be attributed to the Federal Government or the Department of Labor.

The empirical literature on union effects on OSH within firms struggles with two primary obstacles to credibly estimating the effect of unionization on workplace safety. First, unionized employees may be more likely to *report* occupational risks to OSHA, inducing greater rates of inspection and citation of unionized firms for violations than occurs in otherwise similar nonunion firms. This is a kind of measurement error in commonly-used workplace safety outcomes that is positively correlated with unionization. It would lead to a spurious negative association between unionization and safety. Second, less safe industries and less safe firms within industry may be more likely to unionize than more safe ones. This selection bias also would generate a negative association between unionization and safety.

This study will address both of these obstacles and deliver evidence about the effect of unionization on workplace safety. To overcome the first obstacle, the primary outcome will be a measure of workplace fatalities, which are subject to less measurement error and differential reporting than are less severe occupational injuries and illnesses. To overcome the second obstacle, I will use a regression discontinuity design (RDD) comparing establishments where unions just won National Labor Relations Board (NLRB) union-certification elections to establishments where unions just lost such elections (DiNardo &

[1] asojourn@umn.edu. I would like to thank Jooyoung Yang and Suyoun Han for excellent research assistance on this project and to the U.S. Department of Labor for financial support.

Lee, 2004; Frandsen, 2012; Sojourner, Town, Grabowski, Chen, & Frandsen, 2013). Rather than comparing union to nonunion establishments generally and relying on statistical controls and untestable identifying assumptions vulnerable to selection bias, I will restrict attention only to establishments where employees indicated an interest in unionizing such that the NLRB held a union-certification election. At the time of the election, establishments where the union just won are very similar to establishments where the union just lost. After the election, unions are certified as collective bargaining agents in the former set of establishments but not in the latter set. Around the 50% vote-share threshold, this generates quasi-random assignment of unionization to establishments and overcomes the selection problem.

Frandsen (2013) raises concerns about the validity of RDD designs using NLRB elections due to potential post-election, legal manipulation in very close elections. He suggests two basic ways to address this and concludes that these methods provide the best available design to estimate certification effects. The current study uses his suggestions. First, rather than using post-election levels as the outcome, we use changes in levels (post-election level minus pre-election level) as the outcome. This adjusts for any pre-election differences across the threshold and combines both difference-in-difference and RDD logic. Second, he suggests assessing results' sensitivity to deletion of cases where a very small number of votes could change the election outcome because these type of elections are most vulnerable to post-election manipulation.

I will study whether the two sets of establishments experienced different changes in their measure of occupational fatality rates post-election. Any difference can be interpreted as the effect of union certification. I will also study whether the establishments experienced different numbers of occupational fatalities and other observable characteristics *prior* to the election. This falsification exercise allows testing of a key identifying assumption of RDD.

Design and Data

Sample

The population is all U.S. private-sector establishments on the margin of unionization between 1962 and 2009 as measured by experiencing at least one NLRB certification election during this period. For each of the 253,449 elections with valid election month and year, the establishment name, city, state, 2-digit standard industrial classification (SIC) code, date of

election, and the numbers of employees voting in favor of unionization and against unionization are observed. In some cases, establishment street address is available.[2]

Because any establishment may have multiple NLRB elections, I construct longitudinal unique establishment identifiers using a fuzzy-matching algorithm. Across the set of establishments in the same state, city, and industry, the algorithm links establishments with similar names and addresses. This identifies 212,101unique establishments represented in the 253,677 elections with valid election month and year.[3]

To measure workplace safety, I use occupational fatalities because these are relatively well-measured. I draw on the OSHA enforcement database, which records occupational fatalities in U.S. private sector establishments back to 1970 (U.S. Department of Labor, 2014). This period included 103,768 accidents, of which 48,275 involved 55,058 fatalities. For each accident, the establishment's name, address, city, state, and SIC code are observed, as well as the accident date, number of workers killed, and other details. The database also includes records from all OSHA inspections, not just those triggered by accidents. It includes the same kind of establishment information. We will focus only on OSHA records from establishments that experienced NLRB elections.

NLRB and OSHA records are linked at the establishment level using fuzzy-matching based on establishment name, address, city, state, and industry. For each OSHA record, we look for a match among all the NLRB election records. Using the NLRB-based establishment identifiers, this yields a longitudinal database of all NLRB elections and all of their OSHA enforcement data, including reported occupational fatalities. This produces links to 135,366 OSHA records at 38,047 unique establishments that underwent NLRB certification elections. This implies that 18.0% of such establishments are linked to any OSHA record. Of these, 1,450 records in 1,154 unique establishments involve fatal accidents.

Design

[2] I integrate two databases that compile and standardize NLRB election records: one from Holmes (2006) that includes elections during 1977 to 1999 and which includes many establishments' street address and a second provided by Hank Farber covering 1962 to 2009 but lacking any street addresses.

[3] The algorithm requires the researcher to choose weights to value matching characters and penalize mismatches across records. Parameters were chosen by choosing different values and inspecting samples of the results by hand to assess which delivered the best quality matches. To assess sensitivity of final estimates to these nuisance parameters, I use two alternative algorithm parameterizations to generate two alternative sets of establishment identifiers. Results will be re-estimated using these alternative identifier sets.

For a given outcome (Y_i), the main treatment of interest is whether the union won the election or not, indicated by D_i. The forcing variable is the election's pro-union vote-share, which is centered at zero and adjusted to provide a support invariant to the number of voters (DiNardo & Lee, 2004). The basic model is:

$$Y_i \;\; = \; D_i\tau + f(X_i) + W_i\delta + U_i$$

Unionization depends deterministically on vote share, $D_i = 1[X_i > 0]$ and f is assumed to be continuous at 0. The causal effect of unionization near the certification threshold is identified by τ under the following continuity assumption on unobservable influences (U_i):

$$lim_{x\uparrow 0}E[U_i|W_i, X_i = x] = lim_{x\downarrow 0}E[U_i|W_i, X_i = x]$$

This condition means that no unobserved factor influences the outcome in a discontinuous manner across the election victory threshold. The only factor that shifts discontinuously at the threshold is unionization (D) so any observed differences in outcomes across the threshold after the election can be attributed to the causal effect of unionization.

The main regression discontinuity analysis will focus on NLRB elections that meet the following criteria:

1) *At least 20 individuals voted*: a vote-total-floor minimizes the risk that the exact outcome could be manipulated by the company, the union, or workers, which would somewhat undermine the quasi-randomization across the vote-share threshold (Frandsen, 2012).

2) *Election occurs after 1970:* ensures post-election fatalities may be observed in OSHA data.

3) *First election observed in an establishment*: Considering multiple elections for the same establishment raises a number of conceptual questions about whether an establishment should be considered as treated (union wins) or control (union loses). Focusing on only the first election in each establishment sidesteps these thorny issues. This election is termed the establishment's *focal election* (Sojourner, Town, Grabowski, Chen, & Frandsen, 2013).

Another issue raised by multiple elections is the possibility that unions or management learn enough through recently-past elections to manipulate the outcome of the election in such a way as to introduce systematic differences across the threshold in unobservables and, thereby, to invalidate the identifying assumption. This concern diminishes as the time between elections extends. Therefore, using the NLRB data between 1966 and 1970, we will also exclude any establishments that experienced an NLRB election regardless of outcome in the 5 years immediately prior to the focal election.

4) *No evidence of prior unionization:* Using the NLRB data back to 1962, we exclude all establishments where a union was certified prior to the focal election. This clarifies the interpretation of the treatment as a contrast between establishments with no unions certified as bargaining agents and any union so certified (Sojourner, Town, Grabowski, Chen, & Frandsen, 2013).

Filtering on criteria 1 and 2 reduces the number of unique establishments and focal elections to 103,918. These account for 77,979 OSHA records in 20,872 unique establishments including 938 fatal accidents in 739 unique establishments. After implementing criteria 3 and 4, the number of unique firms shrinks to 94,430. Among these, 93,309 have a recorded number of eligible voters (bargaining-unit size) no greater than the recorded number of total votes. This is our analytic sample.

Measures

The effect of establishment unionization on fatality rates will be analyzed in the cross-section. Assume that each establishment-*i* had exactly one NLRB election. To measure post-election fatality rate at each establishment, we construct the following variable:

$$Y_i^{post} \equiv \frac{(\#\ post\text{-}election\ fatalities)_i * 100{,}000}{[2013 - (election\ year)_i] * [2.43 * (\#\ in\ bargaining\ unit)_i]}$$

The key information is the number of post-election occupational fatalities observed at each establishment across all post-election years, which is measured directly from the matched NLRB-OSHA panel constructed above. Because there should be no systematic differences in the size of the establishments or the timing of elections across the 50% vote-share threshold, we could analyze this outcome directly. However, occupational fatality rates are

conventionally measured per 100,000 employees per year and measuring union effects in similar units will facilitate quality checking and interpretation. Dividing by the number of years of OSHA data available post-NLRB-election, (2013 – year of establishment's NLRB election), yields each establishment's occupational fatalities per year. Dividing through by the number of employees in the NLRB bargaining unit, that is the number eligible to vote in the election, adjusts for differences in establishment size.[4] To convert from the establishment's number of employees in the bargaining unit to the establishment's number of employees, we use a scaling factor of 2.43 based on auxiliary data.[5] Then, we multiply times 100,000. This is our measure of each establishment's post-election fatality rate per 100,000 employees per year.[6] Analogously, we construct an establishment-level pre-election fatality rate (Y_i^{pre}) using years between the election date and the first year of the OSHA data (1970) to count the number of years of pre-election observation and the number of pre-election, rather than post-election, fatal injuries observed in the numerator. Finally, we compute an establishment level change in fatality rate for use as our primary outcome (ΔY_i = Y_i^{post}- Y_i^{pre}).[7]

Table 1 provides summary statistics on change in fatality rate. Across the 93,309 establishments with focal NLRB elections, changes in fatal injury rate between post- and pre-election is average 0.12 per 100,000 employees per year with standard deviation 3.01. Pre-

[4] We cannot use OSHA variables to measure the number of workers per establishment because many establishments with NLRB elections never show up in the OSHA data. If unionization causes firms to shrink headcount or to go out of business, this study's estimates will be biased towards a negative effect of unionization on occupational fatality risk. Unionized firms will have less occupational fatalities partly because they have less employees and we lack a good measure of the number of post-election employees to adjust for this directly. However, Frandsen (2013) uses a very similar sample and design as the current study to study the question of union impacts on head count directly. He estimates a positive, though imprecise, effect of unionization on head count. Available evidence also suggests that unionization does not drive establishments out of business (Freeman & Kleiner, 1999; Sojourner, Town, Grabowski, Chen, & Frandsen, 2013).

[5] Frandsen (2013) linked NLRB election data to the Census Longitudinal Business Database over a similar period, which gave him a measure of each establishment's number of employees. He reports that the average number of voters is 93 and the average number of employees is 254. He does not report turn-out rates. In our, very-similar sample, among focal elections with more than 20 voters, the average turnout rate ≡ #voters/#eligible = 0.89. So, putting these facts together, we compute the scaling factor as #employees/#eligible = #employees*(turnout rate/#voters) = 254 *(.89/93) = 2.43.

[6] This is certainly not a perfect measure of occupational fatalities at each establishment. For years that they both exist, the OSHA enforcement data contain many fewer reported fatal injuries than the Census of Fatal Occupational Injuries (CFOI), which was developed to do a better job than OSHA enforcement data was doing. However, it is impossible to link CFOI data to establishment or firm identities so impossible to use in a RDD. The general implications of using a noisy outcome measure has been extensively studied. This issue will be discussed further in the context of the results.

[7] In robustness analysis, we use the same approach to measure pre-, post-, and change in establishment non-fatal injury rates requiring hospitalization and those not requiring hospitalization. Total injury rates are the computed using the sum of all three injury types: 1) fatal, 2) non-fatal requiring hospitalization, and 3) non-fatal and not requiring hospitalization.

election fatal injury rate will be used as a conditioning variable and for falsification testing. Its mean is 0.04 per 100,000 employees per year with standard deviation 1.42.

In all these models, covariates fixed at the time of the election (W_i) which may also influence the likelihood of post-election fatalities can be included in order to increase precision. Under RDD assumptions, they are not necessary to eliminate bias although they can help do so if these assumptions do not hold. I assembled a rich set of conditioning variables that help explain variation in establishment occupational safety. The most basic is a set of indicators of the establishment's broad industry, 1-digit SIC (Table 10). We also construct a measure of each establishment's narrow industry (2-digit) occupational fatality risk.[8] Across establishments in our sample, the average narrow-industry risk level is 0.06 fatalities per thousand FTE with standard deviation of 0.08 (Table 1), equivalent to 6.0 fatalities per 100,000 FTE per year. The average annual occupational fatality rate in the U.S. economy broadly fell from 5.0 per 100,000 FTE in 1992 to 4.2 in 2002. So, the establishments that have union elections tend to be from slightly riskier-than-average industries. Including both of these types of industrial variables compares outcomes in establishments in the same broad industry while controlling for differences in narrow-industry risk.

The fatal injury risk measure that I construct from the OSHA data has a far lower average than expected based on the sources above: 0.04 per 100,000 employees per year in the pre-election period and 0.15 in the post-election period. Further, this upward trend goes against what we know we know occurred in the overall economy, where fatality rates have fallen across these decades. This evidence is consistent with under-reporting of fatalities to OSHA, which improved somewhat over time. My measure of post-election rates is only about 2.5% of what is expected given average narrow-industry rates. This suggests real caution is warranted when interpreting effects on fatal rates, especially the magnitude of effects.

Within narrow industry, each establishment's own pre-election history of OSHA enforcement records (or absence thereof) should contain information about its idiosyncratic occupational injury risk. Pre-election OSHA enforcement actions are likely to be predictive

[8] The occupational fatality rate for each SIC2 industry each year is measured by the ratio of a) fatal occupational injuries from the Bureau of Labor Statistics (BLS) Census of Fatal Occupational Injuries per b) thousands of employees in the industry from the BLS Current Employment Statistics. To reduce measurement error, eleven annual rates, from 1992 to 2002, are averaged within SIC2 industry. This is based on all establishments in the U.S. economy, not just those in our sample and derived from sources completely outside the OSHA data.

of post-election injury risk so including measures of these experiences may also reduce bias and improve precision. For each establishment and pre-election inspection, OSHA assigns a number of current violations, initial penalty, current penalty, and failure-to-abate (FTA) penalty amount for each of 5 types of violations: Serious, Willful, Repeated, Other, and Unclassified. Appendix Figure 3 contains an example of the kind of data used. Because Other and Unclassified violations are very rare, I focus only on Serious, Willful, and Repeated violations. By dividing through by the appropriate number of pre-election years, I derive establishment-level measures of violation rates per pre-election year. Similarly, I construct measures of average annual penalties of 9 types, {serious, willful, repeated}x{initial, current, FTA}. A similar process gives to post-election and overall measures for each establishment. Appendix Table 11 provides summary statistics for the level of annualized violations and penalties across establishments in the sample.[9] This is rich measurement but these variables are highly collinear. Entering them all together as predictors in a regression might reduce precision.

To aggregate the information in the complete OSHA-inspection histories of each establishment and reduce multi-collinearity, I use factor analysis on pre-election measures to extract the single latent factor that explains the most variance in each set of measures. I use this approach to construct an index of each establishment's pre-election history of OSHA violations and an index of its history of OSHA penalties.[10] The pre-election index variables that are used as control variables are summarized in the bottom panel of Table 1. Establishment-level changes in these indices will be used as alternative outcomes and are summarized in the middle panel.

Finally, each OSHA inspection record contains a measure of whether a union representative was present during the inspection. From this, I construct a measure of the share of pre-election and post-election inspections that were attended by a union representative as well as a measure of the change in this share. In the pre-election period, an average of 1 percent of inspections were attended by union representatives. Because the sample is constructed to try to focus on establishments without union representation in the pre-election

[9] Column 1 presents overall establishment averages (SD) without respect to election date. Column 2 presents pre-election statistics and Column 3 presents post-election statistics.

[10] Column 4 of Table 11 reports the scoring coefficients used to aggregate the measures into each index. For instance, each establishment's pre-election violation index is measured as 0.52*(number of serious violations per pre-election year)+0.227*(number of willful violations per pre-election year)+0.258*(number of repeated violations per pre-election year). These same coefficients are used to score the post-election index. The penalty index is based on the first latent factor from analysis of the 9 pre-election penalty measures.

period, it is no surprise the share is so low.[11] This variable is useful for two purposes. First, the pre-election level can be used as a control variable that might pick up differences in unionization across establishments that are missed via the NLRB data. Second, by analyzing the effect of certification on the change in share of inspections attended by union representatives, we can get empirical verification of whether certification drives unionization, whether the RD design appears valid, and a measure of the effect of unionization on employee exercise of a federal right to representation during OSHA inspections.

Analysis

Assessing validity

We present evidence from two falsification tests of the validity of the RD identifying assumption, no pre-election discontinuity across the 50% vote-share threshold. First, Figure 1 presents a histogram of binned vote-shares across the sample of establishment NLRB elections. Most elections are close, giving a large share of the sample close to the threshold. However, we reject the null of no discontinuity in the density of vote shares across the threshold (t=4.137) (McCrary, 2008), consistent with concern about possible post-election manipulation (Frandsen, 2013). Second, we test for discontinuity across the threshold in the distribution of pre-election observables. I implement the test with a seemingly-unrelated regression model (Lee & Lemieux (2010): Section 4.4.2). Outcomes are pre-election fatality rate, violations index, penalty index, union-representative share, total number of focal NLRB-election votes, and narrow industry fatality rate. A piecewise linear function of vote-share that allows for different intercepts and slopes on either side of the certification threshold predicts. The set of establishments with vote-shares within a given bandwidth of the threshold (h) is the sample. Table 2 presents the results. The first column of results uses only establishments with vote shares within 0.10 of the threshold, those with 40-60% vote share, and presents the estimated discontinuity coefficients for each outcome with standard errors. Vote-share coefficients are not displayed. After estimation, a test is performed of the joint null hypothesis that, for all outcomes, there is no discontinuity across the threshold. Results are presented in the bottom row at p-values. For bandwidths 0.10, 0.20, and 0.30, the joint null is not rejected at the conventional 5% significance level. At data from establishments with less-close elections are included, the null is rejected due to a discontinuity in the total

[11] A union representative might be present if the employees organized outside the NLRB process.

number of votes. For the analysis, I will focus attention on elections within the 0.2 bandwidth where the assumptions appear valid and analyze results' sensitivity to alternative bandwidths.

Results

To start the analysis of effects, consider Figure 2(a). It plots the average change in establishment fatality rates among establishments in each of 20 vote-share bins of width 5%. There is not evidence of a discontinuity across the threshold in the average change in fatality rates. This change depends on pre-election levels (Figure 4(a)) and any post-election levels (Figure 5(a)). In inspecting these figures, no discontinuity is apparent in levels or changes.

To obtain an estimate and enable a statistical test, I estimate the effect of union certification on change in fatality rate among establishments with NLRB election vote shares within 0.2 of the certification threshold (pro-union vote share between 30% and 70%) controlling for a piece-wise linear function of vote share. The coefficient on the "Union certified" indicator measures the effect of union certification on the outcome. As displayed in specification (1) of Table 3, the estimated effect is 0.012 per 100,000 employees per year with a standard error of (0.031). This implies that a 95% confidence interval on the effect of union certification on post-election fatalities of [-0.060, 0.072]. The average post-election fatality rate across establishments in our sample is 0.15. The interval rules out effects with magnitude larger than half the base rate.

In specification (2) of Table 3, the establishment's pre-election violation index, penalty index, and union-representative share are added as controls. Each of these predictors has a strong relationship with changes in fatality rates but, most importantly, the estimated union certification effect is stable. Finally, the broad industry indicators and narrow industry fatal risk measure are added in specification (3). Again, the estimate remains very stable despite the fact that average narrow-industry fatality rate is a strong predictor. This stability suggests that the union certification "treatment" is uncorrelated with all observable measures of risk, as predicted by a valid RD design and similar to a randomly-assigned treatment. It is also consistent with an outcome that is just noise.[12]

Heterogeneity of effects?

[12] The OSHA-based fatality rate outcome measure's strong relationship with predictors, particularly with narrow-industry fatality rate which is derived from independent sources (COFI and CES) suggests that the outcome measure constructed here is not just noise, that it contains information about establishment fatality rates. However, the fact that narrow-industry levels predicts changes in OSHA-based rates bears further inquiry.

Effects of unionization may differ between industries with different levels on inherent risk. In riskier industries, occupational safety may be more of an issue in labor-management relations and this may change the effect of union certification. To explore this possibility, we divide the sample into quartiles based on the occupational fatality risk experience of all establishments nationally in the same 2-digit SIC code. The risk level topping the first (second) (third) quartile is 0.025 (0.035) (0.078) fatal occupational injuries per 100,000 employees. Table 4 presents estimates analogous to those in Table 3, column (3) but within each narrow-industry risk quartile. For each risk quartile, the effect of union certification is not significant and the point estimates are small and not significant.

Robustness

I assess the robustness of the main result. In all the results presented here, I use the richest specification, (3). However, for compactness, only the estimated coefficient (SE) on the union certification effect is presented in each case.

First, I vary the bandwidth from the baseline case of 0.2. Column (1) of Table 5 presents the estimate based on a bandwidth of 0.1, looking only at establishments with elections with pro-union vote shares between 40% and 60%. The estimate is -0.0261 (0.043). Column (2) reproduces the baseline result from Table 3 for comparison. Column 3 (4) (5) uses bandwidth 0.3 (0.4) (0.5). In every case, the effect of union certification is not significant.

Second, I look at effects on occupational injury outcomes other than fatal injury rate.[13] Column (2), (3), and (4) of Table 6 present the estimated effect of union certification on the risk of injuries requiring hospitalization, injuries not requiring hospitalization, and total injuries, respectively. In every case, the effect of union certification is not significant.

Third, as mentioned earlier, a recent working paper reports evidence that, in very close elections, post-election legal maneuvering may undermine the key identifying assumption of the RDD (Frandsen, 2013). In the elections with the narrowest margins of victory (the smallest difference between the number of pro-union votes cast and the number of pro-union votes necessary for the union to win certification), incentives for manipulation are strongest and there is compelling evidence that management and unions are able to manipulate final vote counts in the elections with the narrowest margins of victory. In fact,

[13] In all these and subsequent analyses, the baseline bandwidth of 0.2 is used.

evidence from the McCrary test is consistent with this kind of violation in our data. Analyzing changes is one way of dealing with this issue. Here, I present evidence using a second way. I use a donut regression discontinuity design, which involves excluding cases very close to the threshold (Barreca, Guldi, Lindo, & Waddell, 2011), which are most likely to reflect manipulation. In this case, the exclusion is made on the basis of margin of victory (MOV) rather than vote share because the ease of manipulation depends on the former rather than the latter. Column (1) of Table 7 excludes establishments where the election outcome would be changed if 1 vote switched sides. Moving right-ward across the table's columns, progressively more establishments are excluded with progressively larger margins of victory. The estimated effects of union certification are never significant.[14]

Our analysis is dependent on the algorithm that we used to measure records that belong to the same establishment within and across the NLRB and OSHA datasets. Within state and SIC2 code, the matching algorithm penalizes mismatched string values in establishment name, address, and city and matches with quality above a given threshold are retained. We constructed two alternative measures of records belonging to the same establishment by varying the threshold up and down. Each defines a somewhat different set of establishments with associated NLRB and OSHA records. Consequently, all variables defined at the establishment level vary somewhat. Table 8 presents estimates based on these two alternative penalization weights, along with our baseline estimate (NLRB_id2). Results are quite stable.

In conclusion, our evidence on the effect of union certification on occupational fatality rates is consistent with two interpretations. First, there is little effect. Second, our measure of the outcome is very noisy and the estimated effect suffers from attenuation bias. Future work will develop evidence to try to disentangle these stories.

Alternative outcomes

We also analyze the effect of union certification on establishment's interactions with OSHA enforcement. These outcomes are more reliable because the data are very complete and accurate with respect to OSHA enforcement actions. Consider the results reported in Table 9. Column (1) reports that union certification is not significantly associated with increased violations rates, although the point estimate is positive. However, it does cause a

[14] Sample size drops rapidly because we are losing most of the small elections; if they have a large MOV, they must not be close. If they are close, they must have a small MOV.

positive change in penalties (Column (2)), the magnitude of which – 0.238 – is about a sixth of a standard deviation of the sample's distribution of changes in the penalty index. Finally, union certification causes a significant increase in the establishment's share of OSHA inspections performed in the presence of a union representative. This discontinuity shows up very clearly in Figure 2(d). The magnitude of the change, a 5.5 percent increase, may seem small. However, recall that only 18 percent of establishments with NLRB elections had any linked OSHA inspections. So 82 percent of the sample has a 0 share pre-election, post-election, and for its change. Scaling the 5.5 percent effect estimate up by a factor of 5.6 (=1/(1-.82)) yields an estimated effect of 30.5 percent effect among those with any inspections.

Limitations

First, because occupational fatalities are extreme events and thankfully rare, the outcome has less variance across establishments than nonfatal occupational accidents does. An outcome with little variance makes it difficult to generate power to detect effects and helps explain why an exclusive focus on occupational fatalities has been rare in the research literature. For this reason, it is essential to have a large sample. The ability to analyze all establishments with NLRB elections nationally over many decades is particularly useful here. However, the OSHA enforcement data may provide only a weak proxy true fatality rates and the linking process may weaken this further. Further work is needed to assess the value of this proxy.

Second, this design is most informative about the effects of the weakest unions, those that barely win elections. Further, NLRB certification of a union is not the same as persistent unionization, as only about of NLRB-certified bargaining units ever negotiate and sign first contracts. The largest studies that relied on NLRB regression discontinuity designs have found null effects on the outcomes they measured, including firm average wage, employment, survival, revenue, and profits (DiNardo & Lee, 2004; Lee & Mas, 2012).

Conclusion

Union certification increases the participation of worker representatives during OSHA inspections and to higher penalties being assessed for OSHA violations. We do not detect a significant effect on the number of OSHA violations or on our noisy measures of occupational injury risk.

Bibliography

Barreca, A., Guldi, M., Lindo, J., & Waddell, G. (2011). Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification. *Quarterly Journal of Economics, 126*(4), 2117-2123.

DiNardo, J., & Lee, D. (2004). Economic impacts of new unionization on private sector employers: 1984-2001. *Quarterly Journal of Economics, 119*(4), 1382-1441.

Frandsen, B. (2012). *Why Unions Still Matter: The Effects of Unionization on the Distribution of Employee Earnings.* Cambridge, Mass.: Massachusetts Institute of Technology manuscript.

Frandsen, B. (2013). *The Surprising Impacts of Unionization on Establishments: Accounting for Selection in Close Union Representation Elections.* Provo, Utah: Brigham Young University manuscript.

Holmes, T. (2006). *Geographic Spillover of Unionism.* Cambridge, Mass.: National Bureau of Economic Research Working Paper #12025.

Lee, D., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature, 48*, 281-355.

Lee, D., & Mas, A. (2012). Long-run impacts of unions on firms: New evidence from financial markets, 1961-1999. *Quarterly Journal of Economics, 127*(1), 333-378.

Morantz, A. D. (2012). Coal Mine Safety: Do Unions Make A Difference? *Industrial and Labor Relations Review, 66*(1), 88-116.

Sojourner, A., Town, R., Grabowski, D., Chen, M., & Frandsen, B. (2013). *Impacts of Unionization on Employment, Product Quality, and Productivity: Regression Discontinuity Evidence from Nursing Homes.* Minneapolis, Minn.: University of Minnesota manuscript.

U.S. Department of Labor. (2014, January). *Dataset Summary*. Retrieved January 2014, from OSHA Enforcement Data: http://ogesdw.dol.gov/views/data_summary.php

Tables

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
Vote share	0.49	0.22	0	1
Eligible voters	105.08	216.32	20	19000
Changes in				
Fatal injury rate	0.12	3.01	-143.9	182.6
Violation index	0.06	0.97	-159.0	42.8
Penalty index	0.12	1.72	-195.1	151.4
Union-rep share	0.05	0.23	-1	1
Pre-election				
Fatal injury rate	0.04	1.42	0	143.9
Violation index	0.00	0.80	0	161.1
Penalty index	0.00	0.99	0	199.5
Union-rep share	0.01	0.10	0	1
Average fatality rate in 2-digit SIC	0.06	0.08	0	0.4

Notes: Pre- and post-election injury rates – including fatality, hospitalized, non-hospitalized and total – were calculated by dividing each count by (100,000 /(2.43*eligible voters*number of years)). For violations and penalties, annual rates were calculated by dividing each count by number of years. Indexes of violation and penalty are predicted value of factor analysis of variables including serious willful and repeated violations. The number of observation of each variable is 93,309.

Table 2: Coefficients on union certified in Seemingly Unrelated Regression at varied bandwidth

Bandwidths: Outcome variables:	(1) 0.1	(2) 0.2	(3) 0.3	(4) 0.4	(5) 0.5
Pre-election fatality rate	0.00240	-0.0237	-0.00724	-0.00361	-0.0147
	(0.0263)	(0.0209)	(0.0183)	(0.0175)	(0.0165)
Pre-election violations index	0.00899	-0.00246	0.00196	0.0116	0.00687
	(0.0151)	(0.00991)	(0.00792)	(0.0106)	(0.00930)
Pre-election penalty index	0.00820	-0.00767	-0.00211	0.00667	-0.00112
	(0.0304)	(0.0172)	(0.0127)	(0.0127)	(0.0115)
Pre-election union-rep share	-0.000565	0.00123	0.00235	0.00189	-0.00170
	(0.00244)	(0.00162)	(0.00132)	(0.00118)	(0.00114)
Total number of votes	11.34*	4.380	-5.634*	-14.77***	-25.14***
	(5.365)	(3.526)	(2.762)	(2.432)	(2.160)
Average fatality rate in 2-digit SIC	0.00111	0.000662	0.000766	0.00194	0.00118
	(0.00194)	(0.00136)	(0.00113)	(0.00102)	(0.000940)
Observations	29,100	55,984	74,923	85,822	93,309
chi^2	5.49	4.02	8.23	43.6	139.99
p-value	0.4831	0.6738	0.2216	0.0000	0.0000

Note: Estimate (SE). ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level. All estimates shows coefficients for the union certified in seemingly unrelated regressions. chi^2 and p-value are from the joint hypothesis test of null discontinuity effects across pre-election variables.

Table 3: Effect of union certification on change in fatality rate at bandwidth 0.2 with varying sets of conditioning variables

VARIABLES	(1) Change in fatality rate	(2) Change in fatality rate	(3) Change in fatality rate
Union certified	0.012	0.008	0.010
	(0.031)	(0.031)	(0.031)
Vote share	-0.168	-0.163	-0.155
	(0.176)	(0.176)	(0.176)
Vote share*certified	0.263	0.279	0.294
	(0.290)	(0.290)	(0.290)
Pre-election violation index		-0.120***	-0.123***
		(0.0133)	(0.0134)
Pre-election penalty index		0.0174***	0.0181***
		(0.00508)	(0.00508)
Pre-election union-representative share		0.216***	0.222***
		(0.0696)	(0.0698)
1-digit SIC code 2 indicator			0.0986
			(0.0651)
1-digit SIC code 3 indicator			0.0806
			(0.0653)
1-digit SIC code 4 indicator			-0.0114
			(0.0634)
1-digit SIC code 5 indicator			0.0211
			(0.0690)
1-digit SIC code 6 indicator			0.00761
			(0.0772)
1-digit SIC code 7 indicator			0.0655
			(0.0731)
1-digit SIC code 8 indicator			0.0257
			(0.0706)
Average fatality rate in 2-digit SIC code			0.555***
			(0.148)
Constant	0.0527***	0.0533***	-0.0372
	(0.0204)	(0.0204)	(0.0696)
Observations	55,984	55,984	55,984
R^2	0.000	0.002	0.003
Adjusted R^2	-3.37e-05	0.00204	0.00250

Notes: Estimate (SE). ***Significant at 1% level, **Significant at 5% level, *Significant at 10% level.

Table 4: Effect of union certification on changes in fatality rate by establishments' occupational risk quartile at bandwidth 0.2

	(1)	(2)	(3)	(4)
Industrial Risk Quartile	Q1	Q2	Q3	Q4
Union certified	-0.0121	-0.000465	0.0509	0.0620
	(0.0191)	(0.0495)	(0.0997)	(0.104)
Observations	17,662	15,568	10,256	12,498
R^2	0.001	0.001	0.004	0.018
Adjusted R^2	0.000544	0.000692	0.00286	0.0170

Notes: Estimate (SE). ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level. All estimates based on specification (3) in Table 3 that also includes piecewise linear function vote share, establishment's pre-election occupational fatality, hospitalization, and other injury records, OSHA enforcement action histories, broad industry (SIC1) indicators, and narrow industry (SIC2) occupational fatality rate. Data is divided into four by average fatality rate by 2-digit SIC code. Q1 includes the establishments with the lowest risk while Q4 with the highest risk.

Table 5: Effect of union certification on occupational fatality rate at varying bandwidths

Bandwidth:	(1) 0.1	(2) 0.2	(3) 0.3	(4) 0.4	(5) 0.5
Union certified	-0.0261	0.00957	0.00481	0.00527	0.0227
	(0.0433)	(0.0312)	(0.0260)	(0.0237)	(0.0223)
Observations	29,100	55,984	74,923	85,822	93,309
R^2	0.009	0.003	0.002	0.004	0.004
Adjusted R^2	0.00823	0.00250	0.00135	0.00373	0.00360

Notes: Estimate (SE). ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level. All estimates based on specification (3) in Table 3 that also includes piecewise linear function vote share, establishment's pre-election occupational fatality, hospitalization, and other injury records, OSHA enforcement action histories, broad industry (SIC1) indicators, and narrow industry (SIC2) occupational fatality rate.

Table 6: Effect of union certification on various changes in occupational injury rates at bandwidth 0.2

Outcome:	(1) Change in Fatality	(2) Change in Hospitalized	(3) Change in Non-hospitalized	(4) Change in Total Injuries
Union certified	0.010	0.118	0.015	0.012
	(0.031)	(0.076)	(0.043)	(0.092)
Observations	55,984	55,984	55,984	55,984
R^2	0.003	0.003	0.001	0.003
Adjusted R^2	0.00250	0.00230	0.000397	0.00227

Notes: Estimate (SE). ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level. All estimates based on specification (3) in Table 3 that also includes piecewise linear function vote share, establishment's pre-election occupational fatality, hospitalization, and other injury records, OSHA enforcement action histories, broad industry (SIC1) indicators, and narrow industry (SIC2) occupational fatality rate.

Table 7: Effect of union certification on change in occupational fatality rate at bandwidth 0.2, varying exclusion of elections with narrow margins of victory (MOV) to account for possible post-election manipulation

Exclude elections that would change outcome if MOV votes switched sides	(1) MOV = 1	(2) MOV = 2	(3) MOV = 4	(4) MOV = 6
Union certified	0.00595	0.0256	0.0212	0.0116
	(0.0335)	(0.0359)	(0.0415)	(0.0458)
Observations	53,280	47,291	34,780	24,552
R^2	0.003	0.003	0.005	0.007
Adjusted R^2	0.00254	0.00309	0.00471	0.00652

Notes: Estimate (SE). ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level. All estimates based on specification (3) in Table 3 that also includes piecewise linear function vote share, establishment's pre-election occupational fatality, hospitalization, and other injury records, OSHA enforcement action histories, broad industry (SIC1) indicators, and narrow industry (SIC2) occupational fatality rate.

Table 8: Effect of union certification by strictness of matching-algorithm

VARIABLES	(1) NLRB id1 Post-election fatality	(2) NLRB id2 Post-election fatality	(3) NLRB id3 Post-election fatality
Union certified	0.0111	0.00957	0.00343
	(0.0313)	(0.0312)	(0.0317)
Observations	56,584	55,984	55,523
R^2	0.002	0.003	0.002
Adjusted R^2	0.00173	0.00250	0.00163

Note: Establishments were matched using three different matching algorithms by strictness. (NLRB id1: strgroup threshold 0.2, NLRB id2: strgroup threshold 0.25, NLRB id3: strgroup threshold 0.3). All estimates based on specification (3) in Table 3.

Table 9: Effect of union certification on OSHA-enforcement outcomes at bandwidth 0.2

Outcome:	(1) Change in violation index	(2) Change in penalty index	(3) Change in union-representative share
Union certified	0.031	0.238***	0.055***
	(0.021)	(0.053)	(0.004)
Observations	55,984	55,984	55,984
R^2	0.000	0.000	0.017
Adjusted R^2	4.75e-06	0.000386	0.0169

Notes: Estimate (SE). ***Significant at the 1% level, **Significant at the 5% level, *Significant at the 10% level. All estimates based on specification (3) in Table 3 that also includes piecewise linear function vote share, establishment's pre-election occupational fatality, hospitalization, and other injury records, OSHA enforcement action histories, broad industry (SIC1) indicators, and narrow industry (SIC2) occupational fatality rate.

Figures

Figure 1: Histogram of vote-shares across NLRB elections with at least 20 votes cast

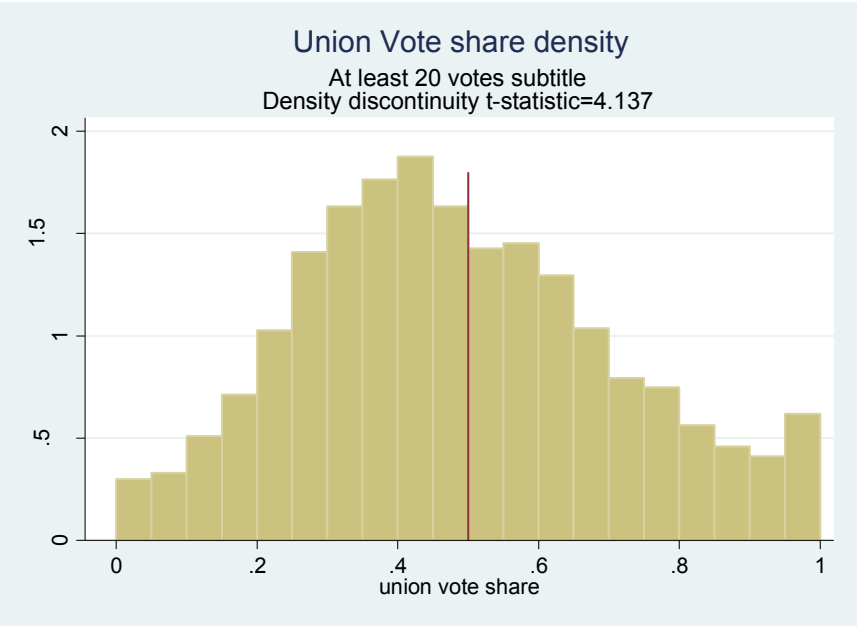
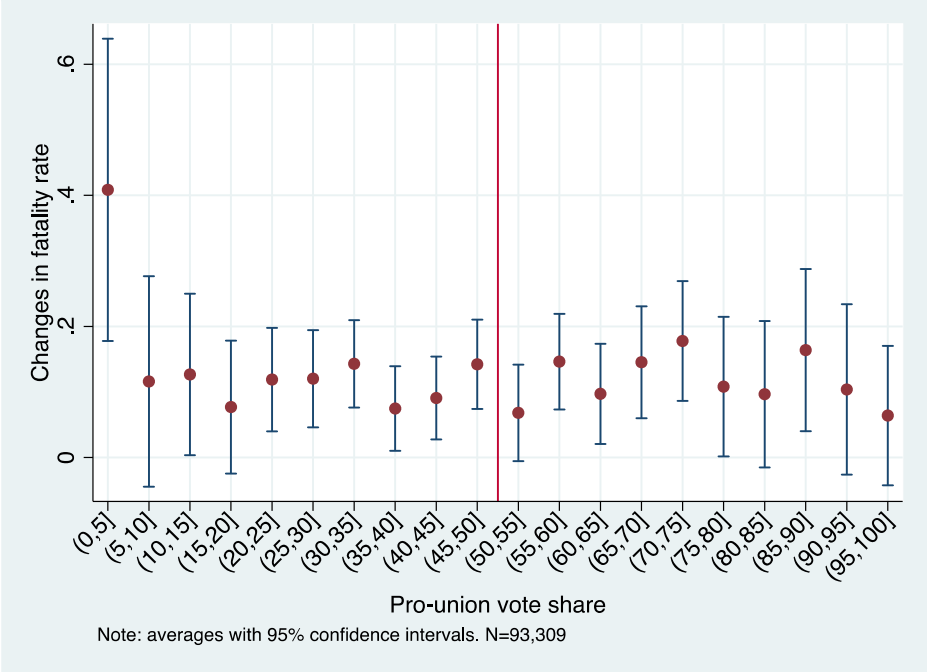
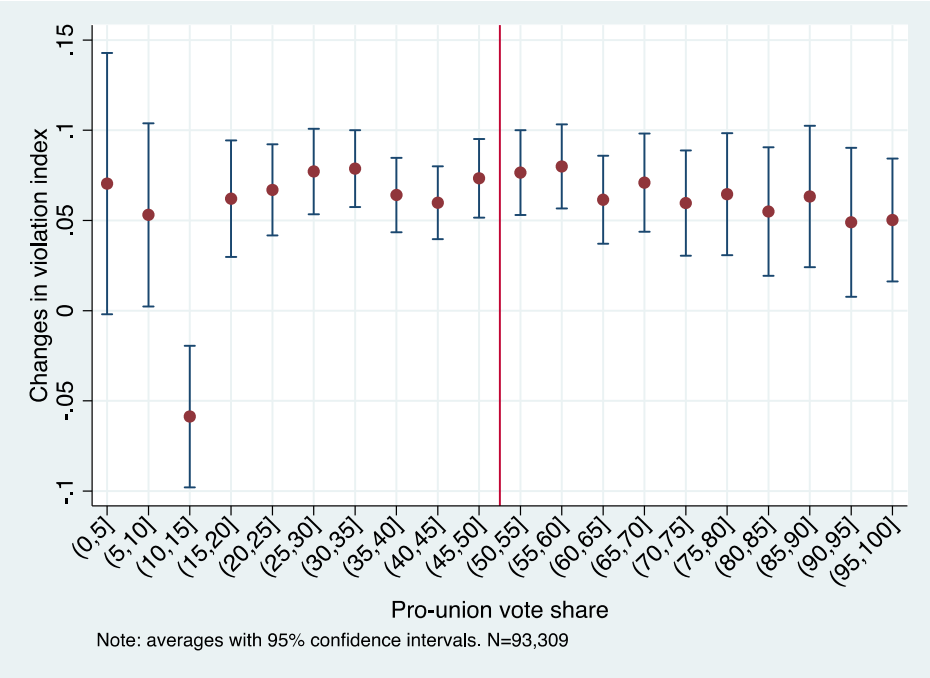


Figure 2: Average establishment (a) change in fatality rate (b) change in violation index, (c) change in penalty index, and (d) change in union-rep share by vote-share bin

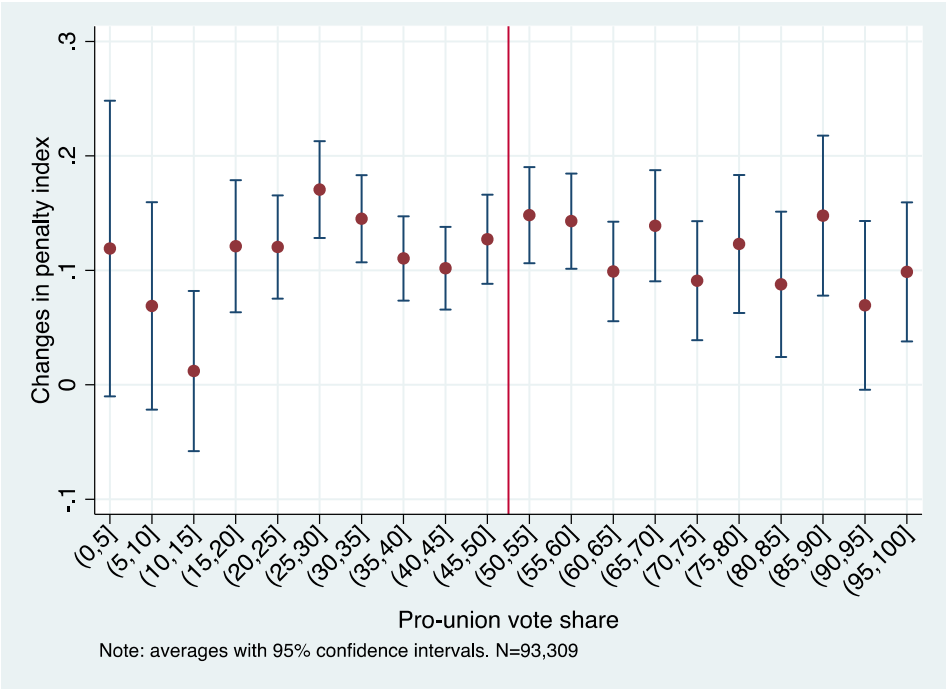
(a)



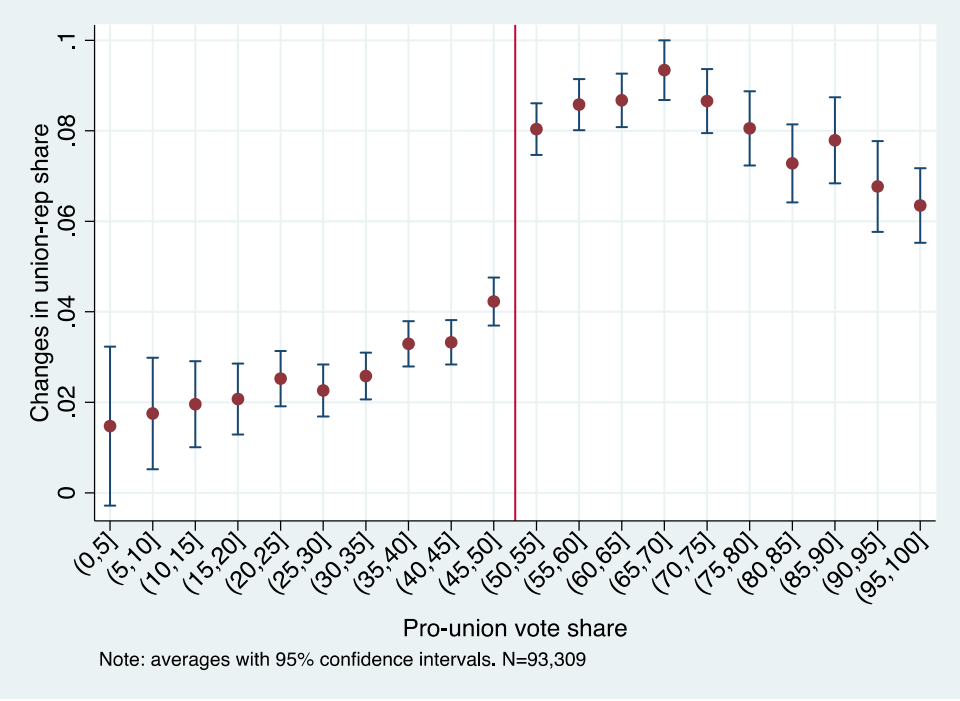
(b)



(c)


Changes in penalty index
-.1
0
.1
.2
.3
(0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
(40,45]
(45,50]
(50,55]
(55,60]
(60,65]
(65,70]
(70,75]
(75,80]
(80,85]
(85,90]
(90,95]
(95,100]
Pro-union vote share
Note: averages with 95% confidence intervals. N=93,309

(d)


Changes in union-rep share
0
.02
.04
.06
.08
.1
(0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
(40,45]
(45,50]
(50,55]
(55,60]
(60,65]
(65,70]
(70,75]
(75,80]
(80,85]
(85,90]
(90,95]
(95,100]
Pro-union vote share
Note: averages with 95% confidence intervals. N=93,309

Appendix Tables & Figures

Table 10: Distribution of elections across industries

1-digit SIC code	Freq.	Percent
1 (Mineral Industries/Construction Industries)	3,078	3.27
2 (Manufacturing)	18,351	19.50
3 (Manufacturing)	30,677	32.59
4 (Transportation, Communications, and Utilities)	10,309	10.95
5 (Whole sale Trade/Retail Trade)	13,604	14.45
6 (Finance, Insurance, and Real estate)	2,925	3.11
7 (Service Industries)	6,708	7.13
8 (Service Industries)	8,466	9.00
Total	**94,118**	**100**

Table 11: Summary stats and factor construction

Variable	(1) Overall Mean (SD)	(2) Pre-election mean (SD)	(3) Post-election mean (SD)	(4) Penalty index scoring coefficient
Violations; annual number of				
Serious violations	0.03 (0.14)	0.01 (0.14)	0.03 (0.17)	0.521
Willful violations	0.00 (0.02)	0.00 (0.04)	0.00 (0.02)	0.227
Repeated violations	0.00 (0.02)	0.00 (0.04)	0.00 (0.02)	0.258
Penalty measures; annual averages				
Initial Penalties for:				
Serious violations	25.5 (191.4)	5.8 (104.4)	32.8 (277.8)	0.145
Willful violations	5.5 (240.5)	2.1 (246.9)	5.8 (256.9)	0.107
Repeated violations	3.1 (78.2)	1.0 (69.9)	3.8 (102.4)	0.135
Current Penalties for:				
Serious violations	15.6 (113.4)	3.8 (62.5)	19.9 (162.1)	0.237
Willful violations	2.9 (138.3)	1.0 (117.7)	3.1 (143.0)	0.296
Repeated violations	1.8 (42.7)	0.7 (53.6)	2.1 (54.6)	0.283
Failure-to-abate Penalties for:				
Serious violations	0.4 (17.0)	0.2 (10.9)	0.4 (21.0)	0.019
Willful violations	0.1 (16.9)	0.1 (20.6)	0.1 (13.1)	0.004
Repeated violations	0.0 (3.3)	0.0 (2.4)	0.0 (5.1)	0.028

Notes: Annual rates of violations and penalties were calculated by dividing each count by number of years. The number of observation of each variable is 93,309.

Figure 3: Example of OSHA enforcement data for a particular establishment

UNITED STATES DEPARTMENT OF LABOR

SEARCH

A to Z Index | En español | Contact Us | FAQs | About OSHA

OSHA

OSHA QuickTakes Newsletter | RSS Feeds | Print This Page | Text Size

Occupational Safety & Health Administration We Can Help

What's New | Offices OSHA

Home | Workers | Regulations | Enforcement | Data & Statistics | Training | Publications | Newsroom | Small Business | Anti-Retaliation

Inspection: 316563352 - Ace Acme Septic Service Inc

Inspection Information - Office: Washington Region 1

Nr: 316563352 Report ID: 1055310 Open Date: 11/20/2012

Ace Acme Septic Service Inc
17924 67th Ave Ne
Arlington, WA 98223 Union Status: NonUnion
SIC: 7699/Repair Shops and Related Services, Not Elsewhere Classified
NAICS: 562991/Septic Tank and Related Services

Inspection Type: Referral
Scope: Partial Advanced Notice: N
Ownership: Private
Safety/Health: Health Close Conference: 12/11/2012
Close Case: 05/21/2013

Related Activity: Type	ID	Safety	Health
Referral	203171350		Yes

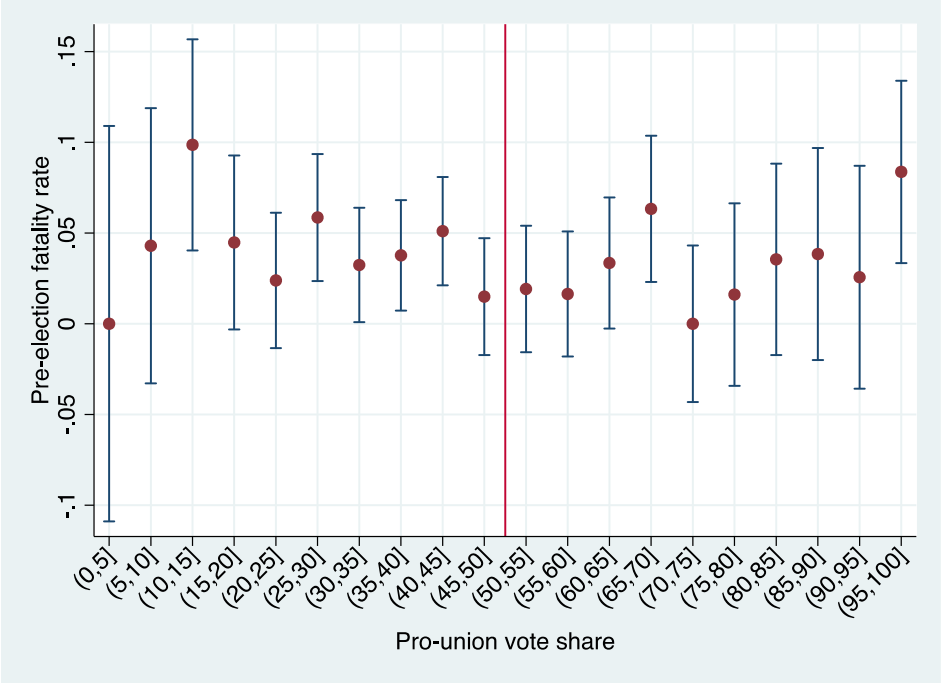
	Violation Summary					
	Serious	Willful	Repeat	Other	Unclass	Total
Initial Violations	5			6		11
Current Violations	5			6		11
Initial Penalty	1500					1500
Current Penalty	1500					1500
FTA Amount						

Violation Items

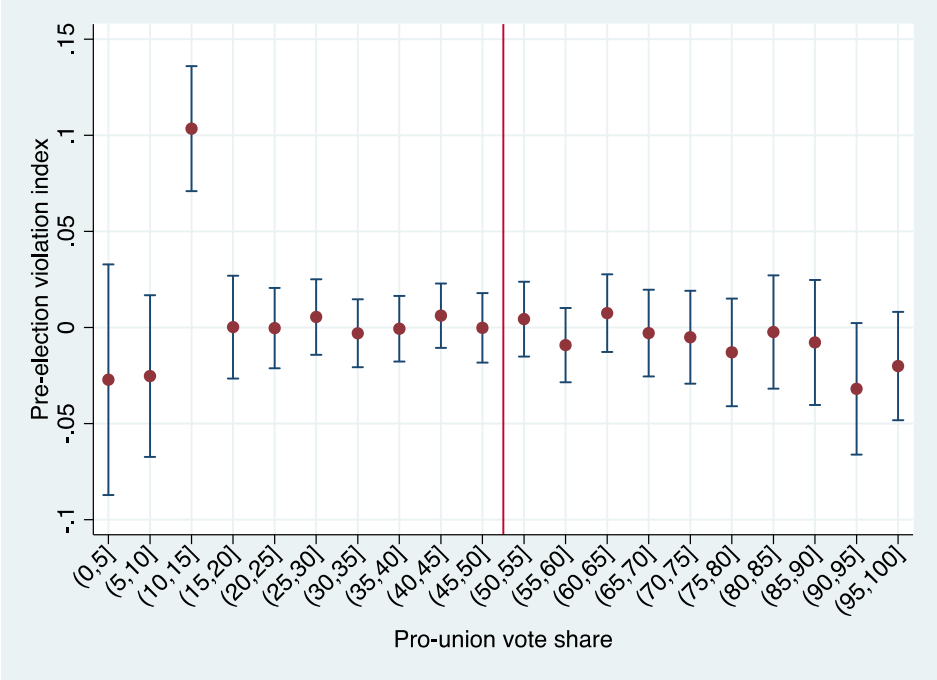
#	ID	Type	Standard	Issuance	Abate	Curr$	Init$	Fta$	Contest	LastEvent
1.	01001	Serious	80015030	12/27/2012	03/15/2013	$300	$300	$0		P - Petition to Mod Abatement
2.	01002	Serious	80017005	12/27/2012	04/30/2013	$300	$300	$0		P - Petition to Mod Abatement
3.	01003	Serious	80920002	12/27/2012	04/30/2013	$300	$300	$0		P - Petition to Mod Abatement
4.	01004	Serious	8092000401	12/27/2012	04/30/2013	$300	$300	$0		P - Petition to Mod Abatement
5.	01005	Serious	82311005	12/27/2012	04/30/2013	$300	$300	$0		P - Petition to Mod Abatement
6.	02001	Other	0244750910	12/27/2012	11/20/2012	$0	$0	$0		-
7.	02002	Other	8001302001	12/27/2012	03/15/2013	$0	$0	$0		P - Petition to Mod Abatement
8.	02003	Other	80014005	12/27/2012	03/15/2013	$0	$0	$0		P - Petition to Mod Abatement
9.	02004	Other	80016010	12/27/2012	04/30/2013	$0	$0	$0		P - Petition to Mod Abatement
10.	02005	Other	82311010	12/27/2012	04/30/2013	$0	$0	$0		P - Petition to Mod Abatement
11.	02006	Other	82312015	12/27/2012	03/15/2013	$0	$0	$0		P - Petition to Mod Abatement

Figure 4: Average establishment (a) pre-election fatality rate, (b) pre-election violation index, (c) pre-election penalty index, and (d) pre-election union-rep share by vote-share bin

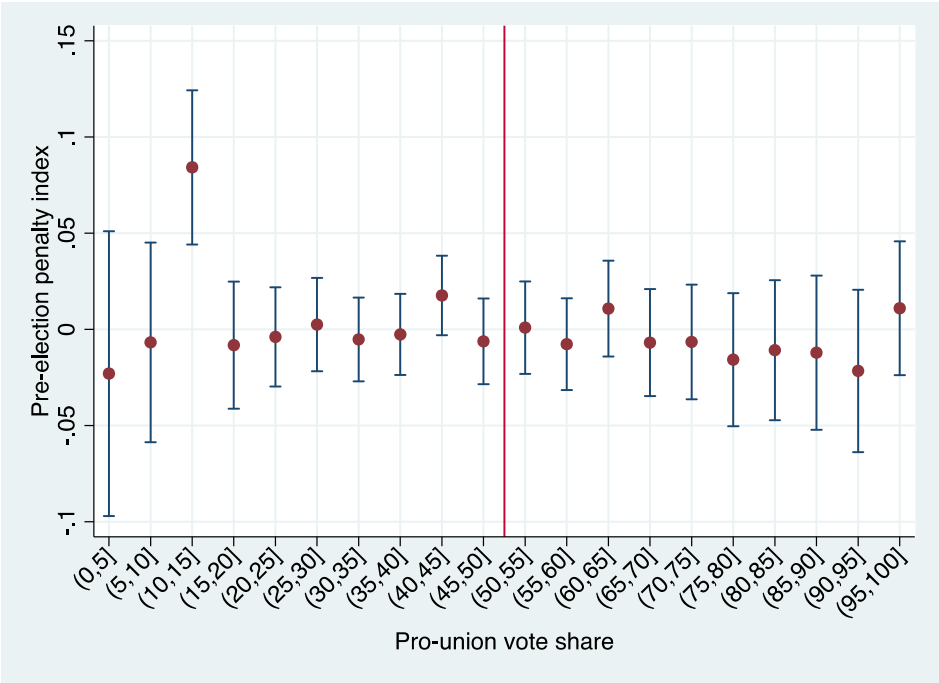
(a)



(b)



(c)


Pre-election penalty index
-.1
-.05
0
.05
.1
.15
(0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
(40,45]
(45,50]
(50,55]
(55,60]
(60,65]
(65,70]
(70,75]
(75,80]
(80,85]
(85,90]
(90,95]
(95,100]
Pro-union vote share

(d)

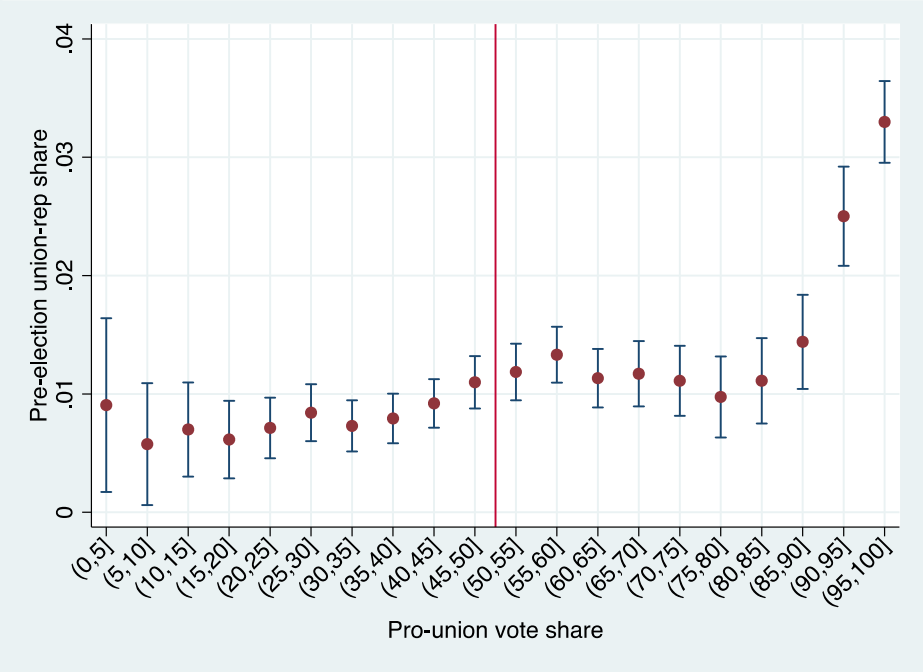
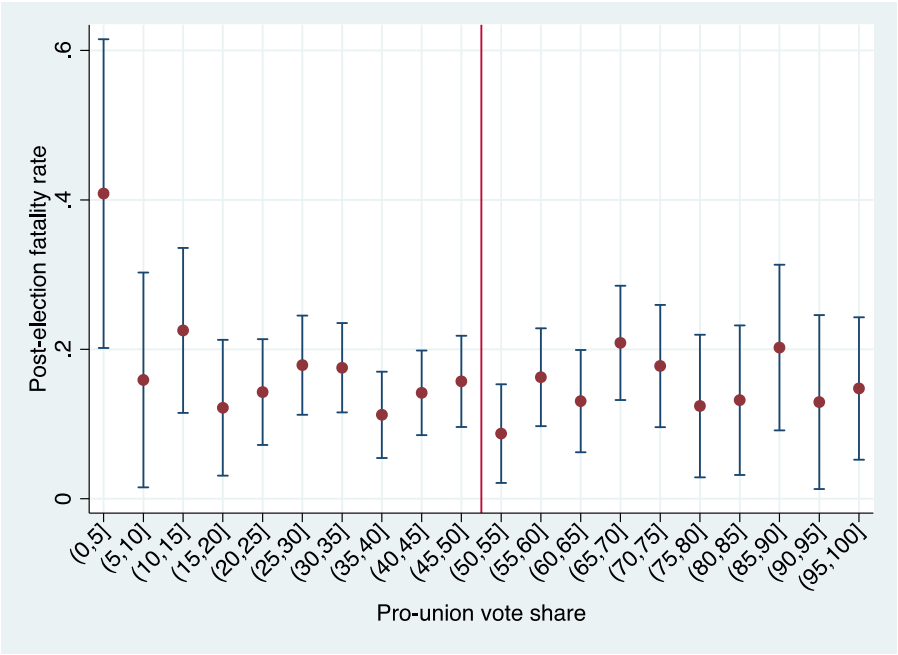
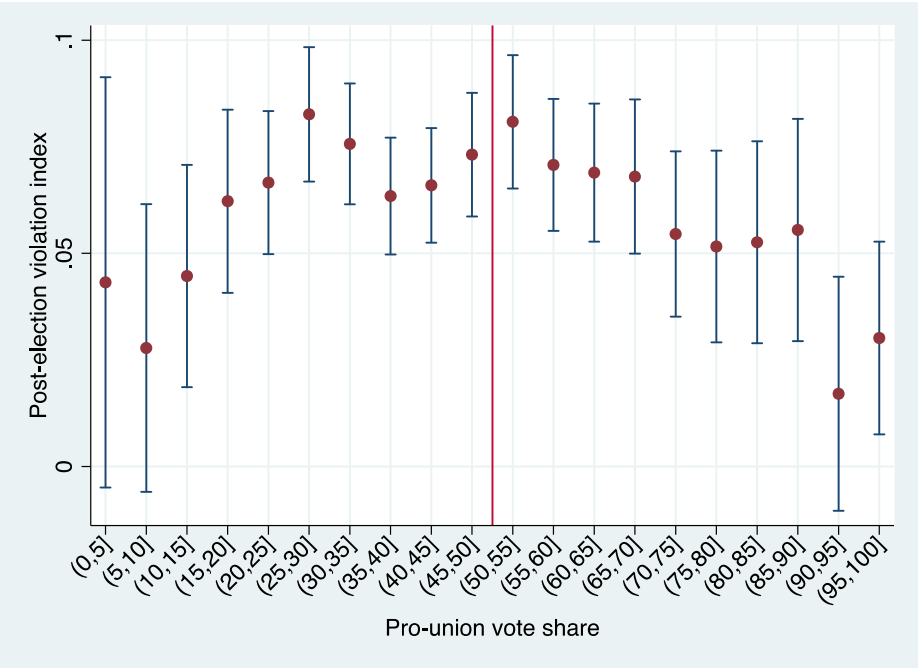

Pre-election union-rep share
0
.01
.02
.03
.04
(0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
(40,45]
(45,50]
(50,55]
(55,60]
(60,65]
(65,70]
(70,75]
(75,80]
(80,85]
(85,90]
(90,95]
(95,100]
Pro-union vote share

Figure 5: Average establishment (a) post-election fatality rate, (b) post-election violation index, (c) post-election penalty index, and (d) post-election union-rep share by vote-share bin

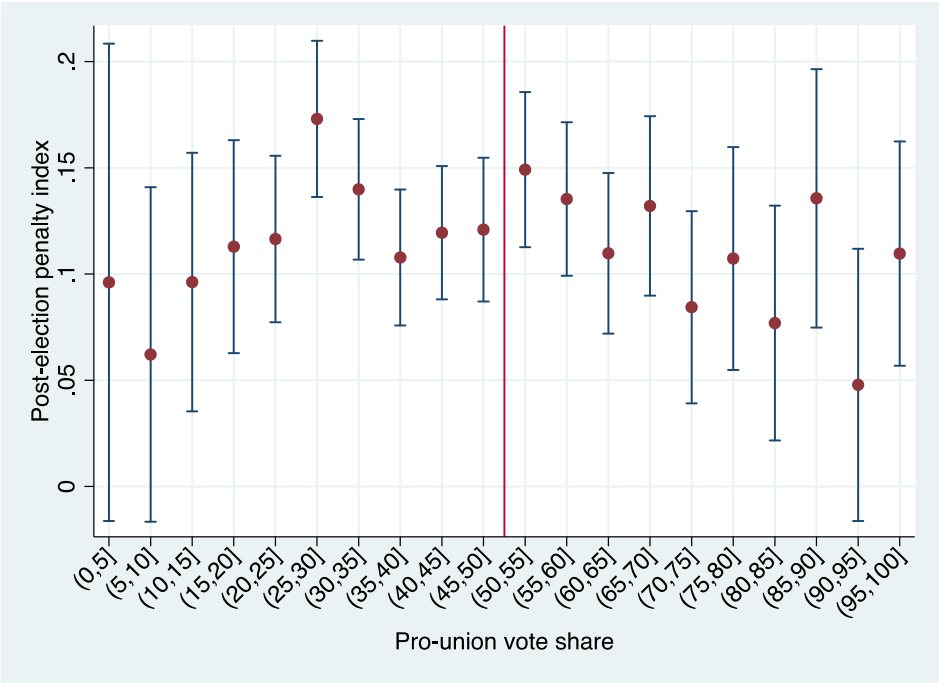
(a)



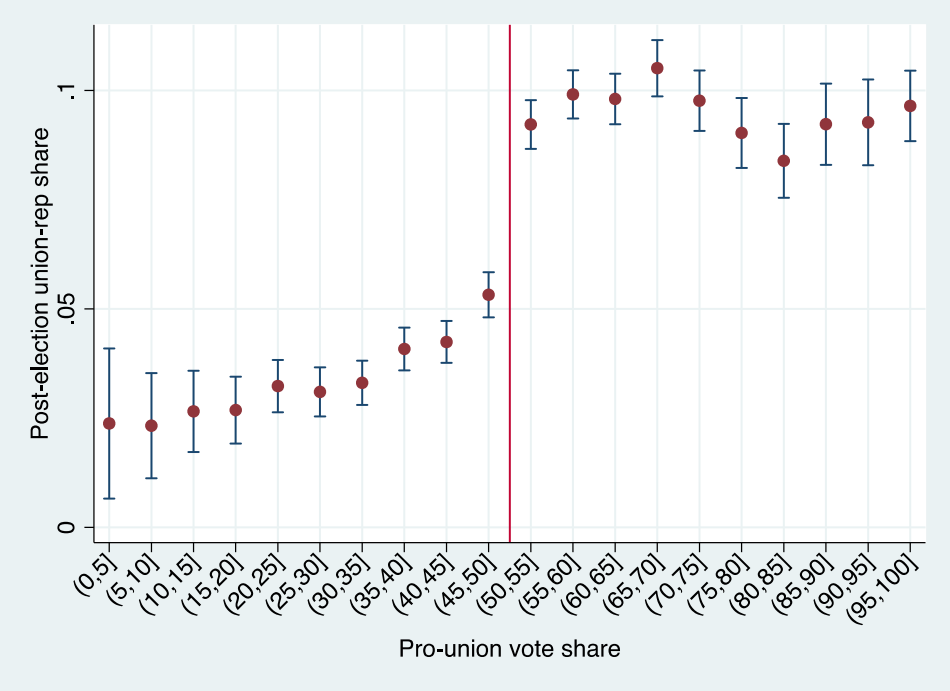
(b)



(c)


Post-election penalty index
0
.05
.1
.15
.2
(0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
(40,45]
(45,50]
(50,55]
(55,60]
(60,65]
(65,70]
(70,75]
(75,80]
(80,85]
(85,90]
(90,95]
(95,100]
Pro-union vote share

(d)


Post-election union-rep share
0
.05
.1
(0,5]
(5,10]
(10,15]
(15,20]
(20,25]
(25,30]
(30,35]
(35,40]
(40,45]
(45,50]
(50,55]
(55,60]
(60,65]
(65,70]
(70,75]
(75,80]
(80,85]
(85,90]
(90,95]
(95,100]
Pro-union vote share