The decline of U.S. workplace injuries, improvements in safety, and changes in occupational composition *

Jonathan B. Cohn
University of Texas at Austin

Benjamin W. Raymond
Bureau of Labor Statistics

Malcolm I. Wardlaw
University of Georgia

July 2026

Abstract

Using establishment-level administrative data and a novel empirical strategy, we show that changes in occupational composition overall did not contribute to the 57% decline in the U.S. workplace injury rates since 1999, contrary to intuition that the U.S. has shifted towards a safer job mix. Instead, injury rates declined substantially within most occupations, especially those involving work with machinery, vehicles, and other mechanized devices. While declining employment in production occupations and growing employment in high-skilled occupations lowered the aggregate injury rate, these effects are small and offset by other compositional changes. We also provide novel occupation-specific workplace injury rate estimates.

*Jonathan Cohn: jonathan.cohn@mccombs.utexas.edu, (512) 232-6827. Malcolm Wardlaw: malcolm.wardlaw@uga.edu, (706) 204-9295. We thank William Fuchs, Murray Gittleman, David Matsa, Nicole Nestoriak, Sheridan Titman, seminar participants at Rice University, the University of Texas at Austin, and Virginia Tech, and participants in the 2024 Fall Labor and Finance Group meeting and 2025 SOLE/JOLE World Labor Conference for valuable feedback. We also thank Mike Soloy from the Bureau of Labor Statistics for help understanding the Occupational Employment and Wage Statistics data. The views expressed here do not necessarily reflect the views of the Department of Labor or the Bureau of Labor Statistics.

The rate of workplace injuries and illnesses (hereafter, the "workplace injury rate") in the U.S. private sector fell by 75% between 1973 and 2022 (Figure 1), eliminating roughly 12 million injuries per year at 2022 employment levels. This is among the most consequential—and least understood—changes in the postwar American workplace, and it could reflect two very different forces. The first is a change in the *composition* of work: the U.S. labor market shifted markedly away from manual and goods-producing occupations and toward service- and knowledge-based ones, a reallocation that could lower the aggregate injury rate even if no individual job became any safer. The second is a decline in injury rates *within* occupations, holding the occupational mix fixed. The two have sharply different implications for how we interpret a half-century of apparent progress in workplace safety and what is achievable in settings where economies and regulatory institutions are less developed than in the contemporary U.S.

— Insert Figure 1 here —

This paper measures how much of the decline each force accounts for. We assemble establishment-year microdata that merge two Bureau of Labor Statistics (BLS) surveys: the Survey of Occupational Injuries and Illnesses (SOII), from which we obtain an establishment's injury rate, and the Occupational Employment and Wage Statistics (OEWS) survey, from which we obtain its employment breakdown across 867 *detailed* occupations. We focus on 1999–2022, the period over which the two surveys can be consistently merged and over which the injury rate fell by 57%. The merged data let us construct a counterfactual path for the aggregate injury rate that holds the occupational distribution of employment fixed, isolating the portion of the decline attributable to within-occupation changes in risk. The gap between the observed and counterfactual paths captures the portion attributable to changes in occupational composition. The large size and high occupational granularity of our sample allow us to estimate this counterfactual precisely.

Our central finding is that changes in occupational composition account for essentially *none* of the decline in the aggregate workplace injury rate. The counterfactual injury rate that holds the occupational mix fixed at its 1999 level tracks the observed series almost exactly, with the two paths nearly coinciding throughout 1999–2022. Contrary to the common presumption that the U.S. has shifted toward an intrinsically safer mix of jobs, the reallocation of employment across occupations did not, on net, make work safer. This conclusion is robust to defining occupations at coarse and fine levels of aggregation, restricting attention to severe injuries, and a range of other checks. In one test, we show that shifts in employment across wage bins *within* occupations are likewise unimportant, indicating that movement toward higher-wage, lower-risk roles *within* occupations also played essentially no role in the decline.

Having established that the decline in workplace injury rates is overwhelmingly within-occupation, we next analyze how it varies across occupations. We estimate occupation-year-specific injury rates for each of 22 broad occupation categories using year-to-year variation in the relationship between establishment-level injury rates and occupational employment shares. Declines are widespread: in all 14 occupation categories with an estimated 1999 injury rate of at least two injuries per 100 workers, injury rates fell by at least 34%. They are especially pronounced in historically hazardous categories such as *Production* (67%), *Construction and Extraction* (81%), and *Healthcare Support* (79%). Because of the scale of employment in *Production* occupations, the decline in that category alone accounts for more than 20% of the aggregate decline.

To learn what kinds of jobs experienced the largest declines, we merge our dataset with occupation-level task-intensity measures from O*NET. The steepest declines occur in occupations intensive in controlling machinery and operating vehicles and other mechanized devices, consistent with technological improvement as an important driver of the within-occupation decline. Declines are also larger in occupations with high routine-manual task

intensity and high offshorability, based on the measures of Acemoglu and Autor (2011)—a modest piece of good news for workers in the jobs most exposed to automation and offshoring—and in more heavily unionized occupations, though we cannot determine whether unionization *caused* the larger declines.

Finally, we return to the compositional result to ask why the reallocation of employment did not lower the aggregate injury rate. Declining employment in high-risk *Production* occupations and rising employment in low-risk *Business and Financial Operations* and *Computer and Mathematical* occupations did put modest downward pressure on the aggregate rate. But these effects are small, and they are offset by countervailing shifts: employment growth in high-risk *Healthcare Support* and *Installation, Maintenance, and Repair* occupations and declines in low-risk *Office and Administrative Support* and *Sales and Related* occupations. On net, the compositional shifts roughly cancel.

A measurement caveat is in order. Our decomposition separates changes in the aggregate injury rate due to occupational composition changes from changes in injury rates within occupations, but it does not reveal whether within-occupation declines reflect genuine reductions in injury risk or changes in reporting. Underreporting of workplace injuries is well documented (Boden and Ozonoff, 2008; Leigh, 2000; Welch et al., 2007), and increased underreporting would cause measured injury rates to fall even if true injury risk did not. Still, we are skeptical that reporting changes explain the entire decline: there are strong *a priori* reasons to believe that work has become safer; workplace fatalities—based on administrative rather than survey data—also declined during the sample period, and underreporting would need to have increased most in occupations where safety plausibly benefited most from technological progress to explain cross-sectional differences in the decline across occupations.[1] More importantly, even if within-occupation changes in injury rates are partly contaminated

[1]The BLS Census of Fatal Occupational Injuries (CFOI) is based on administrative data such as death certificates. The number of fatal work injuries reported in the CFOI fell from about 5,900 in 1999 to 5,283 in 2023, despite roughly 25% growth in employment over the period.

by reporting changes, the conclusion that changes in occupational mix did not meaningfully contribute to the decline in measured injury rates remains unchanged as long as changes in reporting are not systematically related to which occupations grew or shrank.

As a secondary contribution, and to better contextualize our setting, we produce estimates of occupation-specific average workplace injury rates over our full sample period. To our knowledge, these are the first comprehensive estimates of total occupation-specific injury rates.[2] Injury rates are generally highest in manually intensive occupation categories, moderate in service occupations, and lowest in knowledge-economy occupations. At a granular level, injury rates are especially high in occupations involving work with industrial machinery, animal handling, and food production, and in physically demanding healthcare occupations such as paramedic and occupational therapist.

This paper contributes to multiple literatures. First, it advances research on the determinants of workplace safety, which spans several disciplines. Prior studies link injury rates to technology (Evanoff et al., 2003; Chambers et al., 2015), regulation (Conway and Svenson, 1998; Friedman and Forst, 2007; Li and Singleton, 2019), and firm structure (Cohn and Wardlaw, 2016; Cohn et al., 2021). Several papers examine the effects of industry-level employment shifts, with mixed conclusions (Loomis et al., 2004; Morse et al., 2009; Ruser, 2014). However, industry is a coarse proxy for risk exposure, as hazardous and non-hazardous jobs coexist within industries (Viscusi, 1993). Indeed, we show formally that occupations are typically widely dispersed across many industries. By using detailed occupation-level data, we provide a more precise and meaningful decomposition of the sources of long-run declines in workplace injuries.

Second, the paper adds to the literature on occupational change and task reallocation in

[2]The BLS publishes annual estimates of rates of injuries specifically requiring days away from work by occupation during part of our sample period. However, no comparable long-run series exists for total injuries. Supporting the validity of our approach, we find a 0.92 cross-sectional correlation between average occupation-category injury rates in this BLS series and the corresponding days-away-from-work injury rate estimates over the same period in our dataset.

the U.S. labor market. Influential work by Autor et al. (2003), Acemoglu and Autor (2011), and Autor and Dorn (2013) documents the polarization of the U.S. occupational structure, with declines in routine-intensive and middle-skill jobs and growth in both low- and high-skill occupations. Related research by Hershbein and Kahn (2018) and Acemoglu and Restrepo (2020) highlights the evolving skill requirements and task content of jobs in response to technological change. Our paper adds a new dimension to that literature by showing that, despite well-documented shifts in occupational composition and skill demands, the aggregate decline in injury risk reflects changes within rather than between occupations. This finding challenges common intuition that there has been a structural shift toward safer occupations and highlights the role of occupation-specific technological and organizational improvements.

We contribute novel empirical tools and data to the study of occupational heterogeneity. While prior studies document differences in physical demands, job tasks, and exposure to risk using sources like O*NET (e.g., Autor and Handel, 2013; Deming, 2017), there has been limited direct measurement of injury risk at the occupation level over time. To our knowledge, our paper provides the first longitudinal estimates of total occupation-specific injury rates, spanning more than two decades, with potential applications to a wide range of research questions in labor, health, and public economics. These data complement recent efforts to improve occupational measurement (e.g., Hummels et al., 2001; Kuhn and Lozano, 2008) and may inform future work on working conditions, labor supply decisions, and the design of occupational safety policy.

Finally, we make a methodological contribution to the literature on decomposing aggregate outcome differences and trends. The classical approach, pioneered by Kitagawa (1955) and extended in regression form by Oaxaca (1973) and Blinder (1973), requires the group-specific outcome of interest to be directly observable, as do its many modern refinements; see Fortin et al. (2011) for a survey of regression-based methods, Oaxaca and Sierminska (2023) for formal equivalence results, and Riffe et al. (2026) for a recent extension to

life-expectancy comparisons. Recent work on difference-in-differences with compositional changes (Sant'Anna and Xu, 2025) and on outcomes that are themselves vectors of shares (Boussim, 2025) similarly assumes individual-level data or directly measured group outcomes. Our setting is different. The group-specific outcomes — occupation-specific injury rates — are latent: we observe only unit-level aggregate outcomes and group composition. We address this problem by combining the ecological-regression logic of Goodman (1953), which recovers group-specific rates as coefficients on group composition shares, with year fixed effects that trace within-group changes holding composition fixed. To our knowledge, this approach provides the first regression-based Kitagawa decomposition over time for settings in which group-specific outcomes are unobservable.

1 Data and Sample

This section describes the data and sample that we use to conduct our analysis.

1.1 Data

We obtain workplace injury data from the Bureau of Labor Statistics' (BLS') annual Survey of Occupational Injuries and Illnesses (SOII) microdata. The BLS surveys approximately 230,000 establishments each year and uses the data to produce aggregate workplace injury statistics at the national, state, and industry levels. For the purpose of the SOII, a "case" is any injury or illness requiring treatment beyond first aid, days away from work, or temporary job transfer or restriction. The "summary" data contains establishment-level case counts and employment information. The "case" data contains details about any case requiring days away from work ("DAFW" cases), including the occupation of the injured employee. For reasons we discuss shortly, we rely primarily on the summary data in our analysis and use the case data only for a robustness test. We use data only for privately-owned establishments,

though the survey also covers state and local government-owned establishments.

For each establishment-year in the SOII summary data, we obtain counts of each of three classes of injury: total cases; DAFW cases; and "DART" cases. DART cases are those involving days away from work, restricted work activity, and/or transfer to another job. DAFW cases are a subset of DART cases, which are in turn a subset of total cases. We also obtain the establishment's reported average employment for the year. We use this data to construct three injury-rate measures: total cases per 100 workers, DAFW cases per 100 workers, and DART cases per 100 workers. We focus on total cases and DART cases in most of our analysis. In addition, we capture the establishment's industry as reflected in its 6-digit NAICS code.

We obtain establishment-year data on employment by occupation from the BLS' *Occupational Employment and Wage Statistics* (OEWS) survey microdata. Twice per year, the OEWS surveys approximately 200,000 U.S. establishments and collects employment counts by occupation–hourly wage interval cell. The data cover 867 *detailed* occupations defined under the Standard Occupational Classification (SOC) system and 12 narrowly defined wage intervals (e.g., less than \$9.25, \$9.25–\$11.99, \$12.00–\$15.49). An example of a detailed SOC occupation is *Welders, Cutters, Solderers, and Brazers*. The BLS uses the data to construct annual estimates of aggregate employment and average wages by occupation. Appendix A provides more background on both the SOII and OEWS.[3]

Since we conduct our analysis at the annual level to match the frequency of the SOII, we assign each half-year OEWS cohort to its calendar year. For most of our analysis, we collapse the OEWS data across wage intervals to focus on total employment by occupation. We compute, for each establishment-year, the share of employees in each occupation. By

[3]The SOC code classifications were revised in 2010 and 2018. Because SOC codes both split and merged in these revisions, there is no way to homogenize the SOC codes throughout our sample period, and we therefore use the SOC codes as reported in the OEWS microdata. Effects of changes in SOC code classifications appear unlikely to materially affect our results. We discuss these revisions in more detail in Appendix A.

construction, these shares sum to one within each establishment-year. While most of our analysis uses the 867 detailed SOC occupations, we also aggregate to the 22 *major* occupation groups defined by the SOC hierarchy for broader categorization of employment composition in some tests.

While the SOII summary and OEWS microdata are both available starting in 1997, the OEWS only adopted the SOC occupation classification scheme starting with its 1999 survey. We therefore construct our main sample by merging the SOII and OEWS data for the period 1999–2022.[4] This merge yields a sample of 535,436 unique establishment-years, with an average of 22,310 establishments per year.

To supplement the BLS data, we merge in occupational characteristics from O*NET and other publicly available sources. Using the shares of each occupation in an establishment's total employment, we compute weighted averages of these characteristics for each establishment-year. These characteristics include: (i) measures of work activity intensity from O*NET (e.g., machine operation), (ii) task-based characteristics from Acemoglu and Autor (2011), and (iii) annual unionization rates by occupation from public microdata. For O*NET variables, which are updated periodically by occupation, we use the most recent available value for each year of observation and rescale all variables to lie between 0 and 1.

1.1.1 Weighting

Finally, in computing all statistics involving workplace injury rates—whether summary statistics or regression coefficients—we weight each observation by the product of the establishment's SOII sampling weight and its reported employment. We do so to ensure that our estimates reflect aggregates, consistent with the aggregate nature of the stylized fact

[4]We link the two surveys using a common establishment identifier that both derive from the Quarterly Census of Employment and Wages sampling frame, matching each establishment's OEWS employment by occupation to its SOII injury and employment records within a given calendar year. Because the OEWS is fielded semiannually and distributed in overlapping panels, we assign each establishment's most recent non-overlapping semiannual cohort to its calendar year before merging. Appendix A describes the identifiers, the cohort assignment, and the resulting sample construction in detail.

we seek to explain: the long-run decline in the national workplace injury rate. Because the SOII employs disproportionate stratified sampling, the sampling weights are necessary to make the data representative of the full population of U.S. establishments. Weighting by employment further ensures that larger establishments contribute proportionately more to the estimated injury rate, reflecting their greater role in the aggregate statistic. This weighting approach allows us to replicate the official aggregate workplace injury rate reported by the BLS each year exactly using the full SOII sample and ensures that our decompositions are interpretable as explaining changes in the aggregate injury rate.[5]

The OEWS is also a stratified sample, and we would ideally account for the OEWS sampling weights in our weighting scheme. However, the OEWS only provides sampling weights in the microdata prior to 2020. As we explain in more detail in Appendix A, the BLS changed the confidential OEWS microdata from a weighted sample of responding establishments to a file representing the full establishment universe, where employment by occupation-wage bin is imputed for nonresponding establishments. In one robustness test, we reestimate the decomposition based on Regression 2 using composite SOII-OEWS sampling weights for the period 1999–2019, when OEWS sampling weights are available.

1.2 Sample

We next describe our sample. Table 1 presents a breakdown of the number of establishment-year observations in our sample by year. The number of observations is smaller in 2002 than in other years because the OEWS changed the timing of its survey that year and only surveyed one half-year cohort. The number of observations begins to shrink starting in 2019 because the OEWS started surveying fewer establishments that year.

[5]The BLS scales injuries by hours worked rather than number of employees in its official injury statistics, reporting injuries per 100 full-time equivalent workers, where a full-time equivalent employee is equal to 2,000 hours worked. In matching the aggregate injury rate, we use full-time equivalent workers to compute weights. In our analysis, we focus on injuries per 100 workers, which we find slightly easier to interpret, and thus use total workers to compute weights.

— Insert Table 1 here —

Table 2 presents a breakdown of the percentage of establishment-year observations in our sample in each 2-digit NAICS code industry, without any adjustment for sample weighting. For comparison, it also provides the percentage of establishment-years in each 2-digit NAICS code industry as reported in the Census Bureau's Statistics of U.S. Business annual datasets over the years in our sample period. Differences in the two percentages reflect the oversampling of certain industries in the SOII. Specifically, the SOII oversamples establishments in the *manufacturing* industry and, to a lesser extent, the *transportation and warehousing* and *health care and social assistance* industries. With the caveat that we are not using OEWS sampling weights in our main line of analysis, the use of the SOII sampling weights throughout our analysis allows us to recover representative statistics that correspond to aggregates.

— Insert Table 2 here —

Figure 2 plots injuries per 100 workers in our sample (solid blue line) by year, weighting observations as described in Section 1.1.1. It also plots the official aggregate injury rate each year as reported by the BLS (dashed orange line). Note that the numbers are not exactly comparable, as the BLS reports injuries per 100 *full-time equivalent* workers (equal to 200,000 reported hours worked), while we use injuries per 100 workers based on reported average employment. Nevertheless, differences should be and are small. The injury rate is slightly higher in our sample than the official rate from 1999 to 2001, but the difference between the two is never more than 0.2 injuries per 100 workers after 2001.

— Insert Figure 2 here —

Table 3 presents statistics on employment by major SOC occupation category. Specifically, the table reports the percent of all employees in each occupation category for all years in the sample, both unweighted and weighted. It also reports the official published employment shares by occupation category.

— Insert Table 3 here —

The largest occupation categories in our sample are *Office and Administrative Support* (18.7% of employment) and *Sales and Related* (14.6%). Most of the shares in our sample are close to the shares in the official data – for 13 of the 22 major SOC codes, deviations are one percentage point or less. The discrepancy in the Educational Instruction and Library category arises from the fact that most employment in the educational sector is in public schools and state-run universities, and we exclude state and local government-owned establishments from our analysis. Other discrepancies may be caused by the fact that, for reasons described above, we only use the SOII sampling weights and thus do not account for the stratified nature of the OEWS sample.

One strength of our analysis is that we are able to distinguish injury risk by occupation rather than just by industry. To better understand the connection between occupations and industries, we compute the Herfindahl-Hirschman Index (HHI) of the share of each detailed SOC code occupation's employment across different 4-digit NAICS code industries. Letting $emp_{j,k}$ denote the number of employees in occupation j at establishments in industry k, the HHI for a given occupation j across industries $k \in K$ is

$$HHI_j = \sum_{k \in K} \left(emp_{j,k} / \sum_{k \in K} emp_{j,k} \right)^2 . \tag{1}$$

Figure 3 plots a histogram of HHI values. The bin width is 0.01. Each bar depicts the number of occupations in a bin.

— Insert Figure 3 here —

The figure shows that most occupations span many industries. Fewer than 3% of SOC code occupations are represented in only a single 4-digit NAICS code industry. The median HHI is 0.33, and more than 30% of occupations have HHI values less than 0.2.

2 Methodology

This section describes how we use establishment-year workplace injury rate data and matched establishment-year-occupation employment data to decompose changes in the aggregate injury rate into two complementary components: one reflecting changes in occupational composition and the other within-occupation changes in injury rates.

Throughout, we use i to index establishments, $j = 1, \ldots, J$ to index occupations, and $t = 0, \ldots, T$ to index years. Let $IR_{i,t}$ denote the observed injury rate (injuries per 100 workers) for establishment i in year t. Let

$$\mathbf{s} = (s_{i,1,t}, \ldots, s_{i,J,t})'$$

be the $J \times 1$ vector of occupation employment shares for establishment i in year t, with $\sum_{j=1}^{J} s_{i,j,t} = 1$.

2.1 Average occupation-specific workplace injury rates

As a preliminary step, we estimate the average occupation-specific injury rates over the sample period by running the following linear regression:

Regression 1. Average occupation-specific injury rates

$$IR_{i,t} = \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{i,t},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)'$ is a $J \times 1$ vector of coefficients. Coefficient β_j can be interpreted as the average injury rate for occupation j. To see this, suppose that the establishment-year

injury rate can be written as the share-weighted average of latent occupation-level rates:

$$IR_{i,t} = \sum_{j=1}^{J} s_{i,j,t}\, IR_{i,j,t},$$

and assume

$$IR_{i,j,t} = \overline{IR}_j + \nu_{i,j,t}, \qquad E[\nu_{i,j,t} \mid \mathbf{s}] = 0.$$

Then

$$E[IR_{i,t} \mid \mathbf{s}] = \mathbf{s}'\,\overline{\mathbf{IR}},$$

where $\overline{\mathbf{IR}} = (\overline{IR}_1, \ldots, \overline{IR}_J)'$ is a $J \times 1$ vector of mean latent occupation-specific injury rates. Under these assumptions, the population regression coefficient vector satisfies $\boldsymbol{\beta} = \overline{\mathbf{IR}}$. Note that an unweighted version of Regression 1 estimates simple average injury rates for different occupations across establishment-years. However, because we weight by employment, our estimates recover average aggregate rates.

2.2 Decomposition of decline in workplace injury rates

To decompose the aggregate decline in workplace injury rates into components reflecting changes in occupational composition and within-occupation changes in injury rates, we estimate the counterfactual trend in the aggregate injury rate *holding occupational employment mix constant*. We implement this approach by augmenting Regression 1 with a $T \times 1$ vector of year indicators, $\mathbf{d} = (d_{t,1}, \ldots, d_{t,T})'$, that excludes year 0:

Regression 2. Trend in injury rate holding occupational mix constant

$$IR_{i,t} = \mathbf{s}'\boldsymbol{\beta} \;+\; \mathbf{d}'\boldsymbol{\lambda} \;+\; \varepsilon_{i,t},$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ is a $T \times 1$ vector of coefficients on the year indicators. With year 0 omitted, the coefficient λ_t captures the change in the injury rate from year 0 to year t, *holding occupational composition fixed*. Because the regression conditions on establishment-level occupation shares, λ_t isolates the portion of the change in injury rates between year 0 and year t that is attributable to *within-occupation changes* in injury risk. The coefficients $\boldsymbol{\beta}$ now represent time-invariant differences in injury rates across occupations.

After estimating Regression 2, we compute the fraction of the total change in injury rates from year 0 to year t reflecting within-occupation injury rate changes as

$$f_t = \frac{\lambda_t}{\Delta IR_t}, \tag{2}$$

where ΔIR_t denotes the observed change in the aggregate injury rate from year 0 to year t. The complement, $1 - f_t$, represents the fraction attributable to *changes in occupational composition*.

2.3 Within-occupation changes in injury rates by occupation

We next estimate within-occupation changes in injury rates by occupation, which we then use to estimate the contribution of the change in each occupation's injury rate to the aggregate decline in injury rates. To do so, we estimate the following specification, which allows each occupation to have its own injury rate in each year:

Regression 3. Occupation–year specific injury rates

$$IR_{i,t} = \mathbf{s}' \, \mathbf{B} \, \mathbf{d} \; + \; \varepsilon_{i,t},$$

where $\mathbf{d} = (d_0, \ldots, d_T)'$ is a $(T+1) \times 1$ vector of year indicators and $\mathbf{B}$ is a $J \times (T+1)$

matrix of occupation–year coefficients,

$$\mathbf{B} = \begin{pmatrix} \beta_{1,0} & \beta_{1,1} & \cdots & \beta_{1,T} \\ \beta_{2,0} & \beta_{2,1} & \cdots & \beta_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{J,0} & \beta_{J,1} & \cdots & \beta_{J,T} \end{pmatrix}.$$

Because an establishment's injury rate in a given year is the employment-weighted average of its occupation-specific injury rates, $\beta_{j,t}$ provides an estimate of the injury rate for occupation j in year t. The change in the injury rate for occupation j between years 0 and T is $\beta_{j,T} - \beta_{j,0}$. Let $S_{j,t}$ denote the aggregate share of occupation j in year t. Holding the aggregate occupational employment mix fixed at its year-0 composition, the contribution of the change in occupation j's injury rate to the change in the aggregate injury rate from year 0 to year T is:

$$D_j^{IR} = S_{j,0}(\beta_{j,T} - \beta_{j,0}). \tag{3}$$

2.4 Changes in injury rates and occupation characteristics

To further investigate the types of jobs where workplace injuries fell the most between 1999 and 2022, we next map occupations into bundles of observable characteristics. Consider a set of observable establishment-year characteristics $c = 1, \ldots, C$ associated with an occupation. Let $x_{i,c,t}$ denote the employment-weighted average value, for establishment i, of characteristic c in year t, and define the $C \times 1$ vector of characteristic values $\mathbf{x} = (x_{i,1,t}, \ldots, x_{i,C,t})'$. We estimate regressions of the following form:

Regression 4. Change in injury rates and occupation characteristics

$$IR_{i,t} = \mathbf{d}'\,\boldsymbol{\lambda} \;+\; \mathbf{x}'\,\boldsymbol{\phi} \;+\; \mathbf{x}'\,\mathbf{G}\,\mathbf{d},$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_C)'$ is a $C \times 1$ vector of occupation-level coefficients and $\mathbf{G}$ is a $C \times T$ matrix of occupation–year coefficients,

$$\mathbf{G} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,T} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{C,1} & \gamma_{C,2} & \cdots & \gamma_{C,T} \end{pmatrix}.$$

We estimate the change in injury rate from year 0 to year T associated with a one-unit increase in a given characteristic c as $\gamma_{c,T}$.

2.5 Changes in occupational employment shares

Finally, we estimate the contribution of the change in the aggregate employment share of each occupation from year 0 to year T to the total change in the aggregate injury rate. We do so holding injury rates fixed at their year-0 levels based on estimates of $\beta_{j,0}$, $j = 1, \ldots, J$, from Regression 3. We assume the change in occupation j's aggregate employment share, $\Delta S_j = S_{j,T} - S_{j,0}$, is offset by changes in the opposite direction in every other occupation $-j$ proportionate to that occupation's year 0 aggregate employment share $S_{-j,0}$, such that the total sum of changes $\Delta S_j + \sum_{-j \in \{1,\ldots,J\} \setminus j} \Delta S_{-j} = 0$, ensuring that the shares continue to sum to 1. That is, we assume that $\Delta S_{-j} = -\Delta S_j \cdot S_{-j,0}/(1 - S_{j,0})$. With this assumption, the estimated contribution of the change in the employment share of occupation j between years 0 and T to the total change in aggregate injury rate is:

$$D_j^s = \frac{\Delta S_j}{1 - S_{j,0}}(\beta_{j,0} - IR_0), \tag{4}$$

where IR_0 is the aggregate year 0 injury rate. This aggregate injury rate is equivalent to

$$IR_0 = \sum_j S_{j,0}\beta_{j,0},$$

The contribution D_j^s is naturally positive when a high-injury rate occupation ($\beta_{j,0} > IR_0$) gains employment share or a low-injury rate occupation ($\beta_{j,0} < IR_0$) loses share, and negative in the reverse case.

2.6 Discussion

Note that in all the regression specifications described in this section, we are assuming a linear relationship between an establishment's injury rate and its occupation shares. Implicitly, we are assuming that each worker in an occupation faces an occupation-specific injury risk that is independent of both the number of coworkers in that occupation and the number of coworkers in other occupations. For example, our approach does not allow for the possibility that having more information technology workers or occupational health and safety specialists might reduce the risk incurred by an establishment's production workers. While this assumption could bias estimates of occupation-specific injury rates, such biases seem likely to be small. Moreover, it is unclear what bias it would cause in our decomposition tests. Nevertheless, in one robustness test as part of our decomposition analysis, we control for all possible interactions of major SOC code shares.

3 Average Occupation-Specific Injury Rate Estimates

Before decomposing the decline in workplace injury rates into portions due to within-occupation changes in injury rates and changes in employment composition, we first provide estimates of average occupation-specific workplace injury rates for our full sample period by

running Regression 1. We start by defining occupations at the *major* SOC category level. This level of aggregation allows us to report results for an exhaustive set of occupations in a single table. Table 4 presents the results.

— Insert Table 4 here —

The estimates in Table 4 paint an intuitive picture of how workplace injury rates vary across occupations. Injury rates are high in categories that typically involve substantial physical labor, including Production (7.6 total cases per 100 workers), Transportation and Material Moving (7.2), Construction and Extraction (5.9), and Installation, Maintenance, and Repair (5.8). Perhaps less appreciated is the workplace injury risk that healthcare workers face. The Healthcare Practitioner and Technical category has the third highest injury rate (6.7 total cases per 100 workers), and the Healthcare Support category also has an above-median rate (4.4). Predominantly office-based occupations (e.g., Architecture and Engineering; Management; Office and Administrative Support) tend to have the lowest workplace injury rates. Service jobs (e.g., Personal Care and Service; Community and Social Service; Food Preparation and Serving Related; Sales and Related) tend to have intermediate workplace injury rates.

Workers in the Transportation and Material Moving occupation category appear to experience the most severe injuries, with DART cases of 5.0 per 100 workers – more than 2/3 of the total cases in that occupation category. Almost 60% of these severe injuries result in days away from work. Severe injury rates are also high in the Production occupation category, with DART cases of 4.1 per 100 workers. However, only slightly more than 1/3 of these cases involve days away from work. This relatively small proportion may reflect the feasibility of temporary accommodations allowing a seriously injured worker in a production setting to continue working in some capacity.

Note that our approach does not constrain occupation-specific injury rate estimates to

be non-negative. Our estimated injury rates for three categories – Legal, Business and Financial Operations, and Computer and Mathematical – are slightly negative (total cases per 100 workers of -0.2, -0.5 and -0.5, respectively). We omit injury rate estimates for these occupation categories from the table. While these negative estimates may suggest that our assumption that injury rates are a separable linear function of employment shares is violated, these categories likely have true injury rates close to 0. Thus, these negative estimates do not imply that bias due to violations of the assumed relationship form is likely to be large.

We next estimate occupation-specific workplace injury rates for *detailed* SOC occupations using Regression 1. While there are too many detailed occupations for us to present all of these estimates in a simple table, Table 5 presents injury rate estimates for detailed occupations with annual rates of 10 or more total injury cases per 100 workers. We limit the list to detailed occupations representing at least 0.01% of total employment in the sample period to focus on materially important occupations.

— Insert Table 5 here —

13 of the 33 occupations listed in Table 5 are in Production occupations (major SOC category 51). These occupations tend to involve physically-intensive work, often using industrial equipment. Three healthcare occupations involving work with medical equipment (Cardiovascular Technologists and Technicians, Medical Equipment Preparers, Nuclear Medicine Technologists) also make the list. Occupations involving work with animals, including Slaughterers and Meat Packers, Animal Trainers, Zoologists and Wildlife Biologists, and Veterinary Technologists and Technicians, are well represented, as are service occupations that involve heavy lifting (Baggage Porters and Bellhops; Refuse and Recyclable Material Collectors). Also noteworthy is the presence of Special Education Teachers, Kindergarten and Elementary School. Special Education Teachers, Secondary School (not listed) also have a high injury rate, at 8.8 cases per 100 workers. The highest injury rate, at 15.2 cases

per 100 workers, is among paramedics, who often interact with people experiencing mental health crises, lift heavy objects, ride in vehicles at high speeds, and use potentially dangerous medical equipment.

While these estimates are, to our knowledge, the first estimates of occupation-specific total workplace injury rates, the BLS did publish such estimates for DAFW injuries beginning in 2006. We use these rates to help validate our general empirical strategy. Specifically, for each major occupation category, we compute the DAFW injury rate for the same period in our sample and compare it to the annual average for that category in the public BLS figures. We produce this comparison in Appendix Table B1. We find a comfortingly high correlation between the estimates in our data and the public figures of 0.92.[6] Moreover, the three categories where we estimate slightly negative total injury rates correspond to three categories with near-zero DAFW injury rates in the public data, suggesting that our negative estimates indeed represent minor distortions in our estimation procedure and not significant problems with our overall methodology.

4 Main Results

This section presents the paper's main results. We first decompose the change in aggregate injury rate from 1999 to 2022 into complementary portions representing the effects of within-occupation changes in injury rates and changes in occupational employment shares. We then analyze changes in injury rates by occupation, how these changes vary with occupational characteristics, and how changes in occupational shares for each occupational category affect the aggregate injury rate.

[6]Our employment share-weighted average DAFW injury rate is slightly lower than the average based on the public figures – 0.95 injuries per 100 workers vs. 1.00 injuries per 100 workers.

4.1 Decomposition of decline in workplace injury rates

We begin by estimating the counterfactual trend in workplace injury rate holding occupational mix constant using Regression 2. We measure occupation employment shares at the *detailed* SOC occupation level, which allows us to account for changes in occupational employment mix at a granular level. Figure 4 plots the actual trend in the aggregate injury rate, based on our data, against the counterfactual trend, starting from the aggregate 1999 level. The figure shows separate panels for total cases and DART cases.

— Insert Figure 4 here —

In each plot, the vertical distance between the flat dashed green line and dotted orange line in a given year represents the change in aggregate workplace injury rate due to within-occupation changes in injury rate from 1999 to that year. The distance between the dotted orange line and the solid blue line is the estimated change due to changes in occupational composition. These distances as a fraction of the total distance between the dashed green line and the solid blue line are f_t from equation (2) and $1 - f_t$, respectively.

Figure 4 shows that effectively all of the decline in workplace injury rates between 1999 and 2022, for both total and DART cases, is driven by within-occupation changes in injury rates and effectively none by changes in occupational composition. Moreover, the solid blue line and dotted orange line largely coincide throughout the entire 1999-2022 period, suggesting that almost all of the movement in the aggregate injury rate throughout this period is driven by within-occupation injury rate changes.[7] Table 6 presents the decomposition formally, using equation (2) to compute the fraction of the aggregate decline attributable to within-occupation changes in injury rate.

— Insert Table 6 here —

[7]The uptick in DART cases starting in 2020 reflects the impact of COVID-19 (see Cohn et al., 2022).

In addition to presenting results based on our baseline specification of Regression 2, Table 6 also presents results for a variety of alternatives. The first two rows of the table shows results for total and DART cases where we measure occupational employment at the detailed SOC code level. The numbers in the final three columns correspond to the vertical distance from the dashed green line to the solid blue line in 2022 in the two panels of Figure 4, the vertical distance from the dashed green line to the dotted orange line, and the fraction of the former distance that the latter accounts for. This last number is 101.0% for total injuries and 97.5% for DART injuries. These estimates imply that changes in occupational composition account for 2.5% of the decline in DART injuries from 1999 to 2022 and actually cause a slight increase in total injuries.

The third and fourth rows repeat the total-injury analysis from the first row but define occupational employment at different levels. The third row defines it at the major SOC code level. The fourth row defines it at the detailed SOC code-by-wage bin level. This latter definition allows us to account for the possibility that changes in injury rates within occupation reflect movements toward higher-wage, potentially lower-risk roles *within* occupations. The fifth row controls for the *interactions* of employment shares in each pair of major SOC codes ($23 \times 23 = 529$ pairs). This specification allows us to account for the possibility of spillovers from employment in one occupation to the injury rate in a different occupation. The within-occupation decline in injury rates accounts for 99.3%, XX.X%, and 97.8%, respectively, of the total change in injury rates from 1999 through 2022.

The sixth row shows results where we use composite SOII-OEWS sampling weights rather than just the SOII sampling weights. As we note in Section 1.1.1, the OEWS weights are not available after 2019, so the numbers in the last three columns reflect changes in injury rates from 1999 to 2019. The seventh row shows results from unweighted regressions. Note that the total change in injury rates differs from the total in the first row in these regressions because of the differences in weighting. These rows show...

The eighth row shows results from a specification where we include establishment fixed effects. Because the data is from random surveys and was not intended for use in longitudinal analysis, many establishments appear in the data only once and therefore drop out of the sample when we include establishment fixed effects. This reduces our usable sample size by almost half, from 535,436 to 277,818. Larger establishments are more likely to appear in the data more than once and are therefore overrepresented in the remaining sample. In this specification, within-occupation changes in injury rates explain 113.8% of the total change in injury rate, implying that changes in occupational composition substantial *increased* injury rates...

The ninth row shows results where we exclude observations for which the total employment figures reported in the SOII and OEWS for a given establishment-year differ by 30% or more to test whether mismatches in employment totals between the SOII and OEWS in an establishment-year contaminate our estimates. As we note in Section **??**, these employment totals may reflect differences in how employment is reported in the SOII and the OEWS. Within-occupation changes in injury rates explain 97.8% of the total decline in injury rates in this specification. We have experimented with tighter and looser screens, and the fraction of the total change in injury rates explained by within-occupation changes is consistently between 97% and 102%, giving us confidence that differences in employment in the SOII and OEWS are not driving our conclusions.

Finally, the tenth row shows results based on the alternative approach using case-level data on Days Away from Work Injuries that we describe in Section **??**. Note that the unit of observation in this approach is an establishment-year-major SOC occupation category for any occupation in which an establishment reports at least one employee in the OEWS in a given year. As a result, the sample is much larger than in our main approach (more than 3.7 million observations). Within-occupation declines in injury rates explain 97.4% of the total decline in injury rates when we use this alternative approach.

Overall, the results in Table 6 confirm our conclusion from Figure 4 – that the decline in the aggregate workplace injury rate between 1999 and 2022 is due almost solely to within-occupation declines in injury rates. In fact, XXX of the ten rows in the table show within-occupation changes in injury rates explaining more than 100% of the total decline in injury rates during the period, suggesting that employment may have, on net, shifted slightly towards rather than away from more hazardous occupations. The lack of explanatory power of changes in employment mix measured at the occupation-wage interval level in row 4 suggests that even compositional changes *within* occupations, such as movement into higher-wage, lower-risk roles, appear to have played no meaningful role in the aggregate injury rate decline.

4.2 Changes in workplace injury rate by occupation

We next estimate Regression 3 to analyze changes in workplace injury rates in different occupations during our sample period. In addition, we compute the implied effect that the change in each occupation has on the aggregate injury rate, holding employment fixed at its 1999 level. We focus on major occupation categories here rather than detailed occupations to make the results easier to summarize. Table 7 presents this analysis.

— Insert Table 7 here —

For each occupation category, the table reports the estimated change in injury rate from 1999 to 2022, the percentage change in injury rate for each occupation category with an average of at least 2.0 cases per 100 workers in our sample, aggregate 1999 employment share, and the effect of the change in injury rate on the aggregate injury rate computed using equation (3). Reductions in injury rates during this period are widespread across occupations: All 14 occupational categories with at least 2.0 injuries per 100 workers in our sample period experience declines of at least 34%. Declines are especially large in tradi-

tionally hazardous occupation categories, including *Production* (67% decline), *Construction and Extraction* (81% decline), and *Healthcare Support* (79% decline). The decline in injury rates in *Production* occupations has the largest effect on the aggregate injury rate, causing a decline of 0.90 injuries per 100 workers, which is 22% of the aggregate decline.

4.3 Changes in workplace injury rates and occupation characteristics

We further unpack the heterogeneity in changes in injury rate across different occupations by examining how injury rate changes between 1999 and 2022 vary with occupational characteristics. We do so by estimating Regression 4. The occupational characteristics we analyze include *Work activities*, *Automation & offshoring exposure* measures, and union membership. We estimate this regression separately for each group of occupation characteristics. Table 8 reports the results. We also show the mean and standard deviation of each characteristic.

— Insert Table 8 here —

A few notes on units of measurement are in order. The *Work activities* and *Automation & offshoring exposure* variables are measured as intensities on a scale from 0 to 1, with the exception of *Working with animals*, which is the percentage of jobs in occupations that involve working with animals. Other than *Working with animals*, these variables have cardinal meaning but not defined units. *Union membership* is the fraction of workers in an occupation who report belonging to a union.

Among *Work activities* characteristics, we find the largest injury rate declines in occupations with high intensities of tasks involving *controlling machines and processes* and *Operating vehicles and mechanized devices*. A one-standard deviation increase in *Controlling machines and processes* intensity is associated with a decline of 1.30 injuries per 100 workers,

while a one-standard deviation increase in *Operating vehicles and mechanized devices* intensity is associated with a decline of 0.70. A natural interpretation of these findings is that technological improvements have substantially increased the safety of operating machinery and equipment.

Among *Automation & offshoring exposure* characteristics, declines are largest in occupations with high *Routine manual* task intensity and high *Offshorability*. A one-standard deviation increase in *Routine manual* task intensity is associated with a decline of 1.15 injuries per 100 workers, while a one-standard deviation increase in *Offshorability* is associated with a decline of 0.29. These findings suggest a sliver of positive news for workers in the types of jobs most subject to automation and offshoring risk.

Finally, injury rates fell more in more-unionized occupations. A one-standard deviation increase in unionization is associated with a decline of 0.77 injuries per 100 workers. However, this association is difficult to determine: it is unclear how much of it represents the *causal* effects of unions and how much reflects declines in injury rates for other reasons in the types of jobs that tend to be more unionized.

Taken together, these patterns suggest a coherent, if necessarily tentative, interpretation of the within-occupation decline. The characteristics associated with the steepest declines—controlling machines and processes, operating vehicles and mechanized devices, and routine-manual task intensity—all describe physically hazardous work centered on interaction with equipment, and they are concentrated in the historically dangerous categories (*Production*, *Construction and Extraction*, *Transportation and Material Moving*) that account for most of the aggregate decline. The most natural reading is that technological and organizational improvements in how mechanized, manually intensive work is performed—safer equipment, automation of the most dangerous tasks, and better hazard control—are a leading driver of the fall in injury rates. Two caveats temper this interpretation. First, because we estimate each group of characteristics separately and these characteristics are strongly correlated

with one another across occupations, we cannot isolate the contribution of any single one; the unionization result, in particular, may reflect a causal effect of unions, the concentration of union representation in mechanized occupations that grew safer for other reasons, or both. Second, our characteristics are measured at the occupation level and are largely time-invariant, so they speak to *which kinds of jobs* became safer rather than identifying the specific practices that changed within them. What the analysis establishes is thus less the precise channel than its locus: the decline was realized inside jobs, and disproportionately inside the mechanized, manually intensive jobs where the scope for technology to improve safety was greatest.

4.4 Occupational composition changes

The fact that changes in occupational composition do not contribute to the decline in workplace injury rates between 1999 and 2022 is surprising. To understand why it does not, we next estimate the effects of the change in occupation share for each major occupation category on the aggregate injury rate, holding occupation category-specific injury rates fixed at their 1999 levels, using equation (4). We again focus on major occupation categories here to make the results easier to summarize. Table 9 presents the results of this analysis.

— Insert Table 9 here —

For each occupation category, the table reports the percentage point change in employment from 1999 to 2022, the estimated 1999 injury rate based on Regression 3, and the implied effect of the change in employment on the aggregate injury rate. As expected, employment during this period shrinks in *Production* occupations, where injury rates tend to be high, and grows in *Business and Financial Operations* and *Computer and Mathematical* occupations, where injury rates tend to be low. However, while these changes are large in percentage terms, they are modest in absolute terms. As a result, their effects on the aggre-

gate workplace injury rate are small. Moreover, employment share increases in *Healthcare Support* and *Installation, Maintenance, and Repair* occupations, which have relatively high injury rates, and employment share decreases in *Sales and Related* and *Office and Administrative Support* occupations, which have relatively low injury rates, have *negative* effects on the aggregate injury rate. The effects of the different changes in employment composition mostly balance out, explaining why the overall set of changes in occupational composition has effectively no effect on the aggregate injury rate.

4.5 A measurement limitation

Our analysis is subject to an important limitation that bears on how the within-occupation decline should be interpreted. The SOII measures *recorded* injuries, and we cannot fully separate genuine improvements in safety from changes in reporting. Reporting error runs in both directions. Underreporting arises when workers decline to report injuries or when employers discourage reporting to limit workers' compensation costs or regulatory scrutiny, and employers may systematically omit injuries to contract workers (Nestoriak and Ruser, 2007). Overreporting arises when workers seek paid leave or compensation for injuries that are minor or not work-related, and employer confusion about what constitutes a recordable case can push in either direction. The weight of the evidence indicates that underreporting is the more serious problem (Boden and Ozonoff, 2008; Leigh, 2000; Leigh et al., 2004; Welch et al., 2007), though some studies find evidence of overreporting as well (Biddle and Roberts, 2003; Ruser, 2008).[8] It is unlikely, however, that reporting changes account for the *entire* within-occupation decline: work has plausibly become safer on *a priori* grounds, administratively measured fatalities also fell, and a purely reporting-based account would require

[8]Features of workers' compensation programs can also affect reported rates. Boden and Ruser (2003) document declines in reported lost-time injuries following state law changes in the 1990s that made compensation claims harder to file, though whether this reflects reduced overreporting or increased underreporting is unclear.

underreporting to have grown most precisely in the occupations where technology most plausibly improved safety. It is in any case unclear why reporting behavior would have shifted so dramatically since the late 1990s across so many different occupations. Most importantly, this limitation does not touch our central result: whatever fraction of the within-occupation decline is real, changes in occupational composition contributed little to the decline in measured injury rates, so long as reporting changes were not systematically related to which occupations grew or shrank.

5 Conclusion

The 57% decline in U.S. workplace injury rates since 1999 is a large and consequential change that has received surprisingly little attention. A natural conjecture is that it reflects the most familiar account of how the U.S. economy has changed: automation, offshoring, and the shift from goods production toward service- and knowledge-based work reduced employment in physically hazardous jobs. That account captures something real. Declining employment in high-risk occupations and rising employment in low-risk knowledge-economy occupations did push the aggregate injury rate downward, but only modestly, as large percentage changes in employment in these occupations are less impressive in terms of absolute changes. Moreover, the broader transformation of the U.S. economy also changed the labor market in offsetting ways. Population aging and the rise of Internet-mediated commerce are reflected in rising employment in relatively hazardous healthcare support occupations and declining employment in relatively safe retail and administrative occupations. The net effect is that employment has not, in fact, shifted toward safer occupations. The decline in workplace injury rates instead occurred within occupations.

This distinction matters for how we understand economic progress and workplace safety policy. If falling injury rates mainly reflected occupational upgrading, then safer work would

be a mechanical consequence of development, deindustrialization, or movement into higher-skill jobs. Instead, our findings suggest that large safety improvements are possible even within historically hazardous occupations. The relevant margins are therefore likely to be the design of tasks, machinery, training, monitoring, regulation, workplace organization, and incentives. This interpretation is especially important for developing economies, where dangerous occupations remain central to employment: progress in worker safety appears to be possible even without waiting for a shift to a developed-economy occupational structure.

The results also narrow the agenda for future work. Because occupational composition explains little of the aggregate decline in injury rates, the key open question is what changed inside occupations. Our evidence points toward technological improvement and changes in work practices as plausible forces, but the relative importance of technology, regulation, unionization, employer incentives, and reporting behavior remains unresolved. Distinguishing among these mechanisms is essential for understanding which safety improvements are reproducible, which depend on broader institutional conditions, and whether and how further reductions in workplace injury risk can be achieved.

References

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, Volume 4, pp. 1043–1171. Elsevier.

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy 128*(6), 2188–2244.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review 103*(5), 1553–1597.

Autor, D. H. and M. J. Handel (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of labor Economics 31*(S1), S59–S96.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics 118*(4), 1279–1333.

Biddle, J. and K. Roberts (2003). Claiming behavior in workers' compensation. *Journal of Risk and Insurance 70*(4), 759–780.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources 8*(4), 436–455.

Boden, L. I. and A. Ozonoff (2008). Capture–recapture estimates of nonfatal workplace injuries and illnesses. *Annals of Epidemiology 18*(6), 500–506.

Boden, L. I. and J. W. Ruser (2003). Workers' compensation "reforms," choice of medical care provider, and reported workplace injuries. *Review of Economics and Statistics 85*(4), 923–929.

Boussim, J. (2025). Compositional difference-in-differences. Working paper, arXiv:2510.11659.

Chambers, A., C. A. Mustard, and J. Etches (2015). Trends in needlestick injury incidence following regulatory change in ontario, canada (2004–2012): an observational study. *BMC Health Services Research 15*, 1–6.

Cohn, J., N. Nestoriak, and M. Wardlaw (2021). Private equity buyouts and workplace safety. *Review of Financial Studies 34*(10), 4832–4875.

Cohn, J. B., L. Guo, and Z. Wang (2022). Corporations and covid-19 in the workplace. *Available at SSRN 4177494*.

Cohn, J. B. and M. I. Wardlaw (2016). Financing constraints and workplace safety. *Journal of Finance 71*(5), 2017–2058.

Conway, H. and J. Svenson (1998). Occupational injury and illness rates, 1992-96: why they fell. *Monthly Lab. Rev. 121*, 36.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The quarterly journal of economics 132*(4), 1593–1640.

Evanoff, B., L. Wolf, E. Aton, J. Canos, and J. Collins (2003). Reduction in injury rates in nursing personnel through introduction of mechanical lifts in the workplace. *American Journal of Industrial Medicine 44*(5), 451–457.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4A, pp. 1–102. Elsevier.

Friedman, L. S. and L. Forst (2007). The impact of osha recordkeeping regulation changes on occupational injury and illness trends in the us: a time-series analysis. *Occupational and Environmental Medicine 64*(7), 454–460.

Goodman, L. A. (1953). Ecological regressions and behavior of individuals. *American Sociological Review 18*(6), 663–664.

Hershbein, B. and L. B. Kahn (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review 108*(7), 1737–1772.

Hummels, D., J. Ishii, and K.-M. Yi (2001). The nature and growth of vertical specialization in world trade. *Journal of international Economics 54*(1), 75–96.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association 50*(272), 1168–1194.

Kuhn, P. and F. Lozano (2008). The expanding workweek? understanding trends in long work hours among us men, 1979–2006. *Journal of Labor Economics 26*(2), 311–343.

Leigh, J. P. (2000). *Costs of occupational injuries and illnesses*. University of Michigan Press.

Leigh, J. P., G. Waehrer, T. R. Miller, and C. Keenan (2004). Costs of occupational injury and illness across industries. *Scandinavian Journal of Work, Environment & Health*, 199–205.

Li, L. and P. Singleton (2019). The effect of workplace inspections on worker safety. *ILR Review 72*(3), 718–748.

Loomis, D., D. B. Richardson, J. F. Bena, and A. J. Bailer (2004). Deindustrialisation and the long term decline in fatal occupational injuries. *Occupational and Environmental Medicine 61*(7), 616–621.

Morse, T. F., A. Deloreto, T. St. Louis, and J. D. Meyer (2009). Are employment shifts into non-manufacturing industries partially responsible for the decline in occupational injury rates? *American Journal of Industrial Medicine 52*(10), 735–741.

Nestoriak, N. and J. Ruser (2007). Emerging labor market trends and workplace safety and health. In *CRIW Conference on Labor in the New Economy*.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*(3), 693–709.

Oaxaca, R. L. and E. Sierminska (2023). Oaxaca-blinder meets kitagawa: What is the link? IZA Discussion Paper 16188, Institute of Labor Economics (IZA).

Riffe, T., R. Tursun-zade, and B. Llimós (2026). Arriaga meets kitagawa. *Geographical Analysis*. Forthcoming.

Ruser, J. W. (2008). Examining evidence on whether bls undercounts workplace injuries and illnesses. *Monthly Lab. Rev. 131*, 20.

Ruser, J. W. (2014). Industry contributions to aggregate workplace injury and illness rate trends: 1992–2008. *American Journal of Industrial Medicine 57*(10), 1149–1164.

Sant'Anna, P. H. C. and Q. Xu (2025). Difference-in-differences with compositional changes. *Journal of Econometrics*. Forthcoming. arXiv:2304.13925.

Viscusi, W. K. (1993). The value of risks to life and health. *Journal of Economic Literature 31*(4), 1912–1946.

Welch, L. S., X. Dong, F. Carre, and K. Ringen (2007). Is the apparent decrease in injury and illness rates in construction the result of changes in reporting? *International Journal of Occupational and Environmental Health 13*(1), 39–45.

Figure 1: Official injuries per 100 full-time equivalent workers by year

This figure presents total workplace injuries per 100 full-time equivalent workers by year as reported by the Bureau of Labor Statistics in its official annual releases. A full-time equivalent worker is defined as 2,000 hours worked.

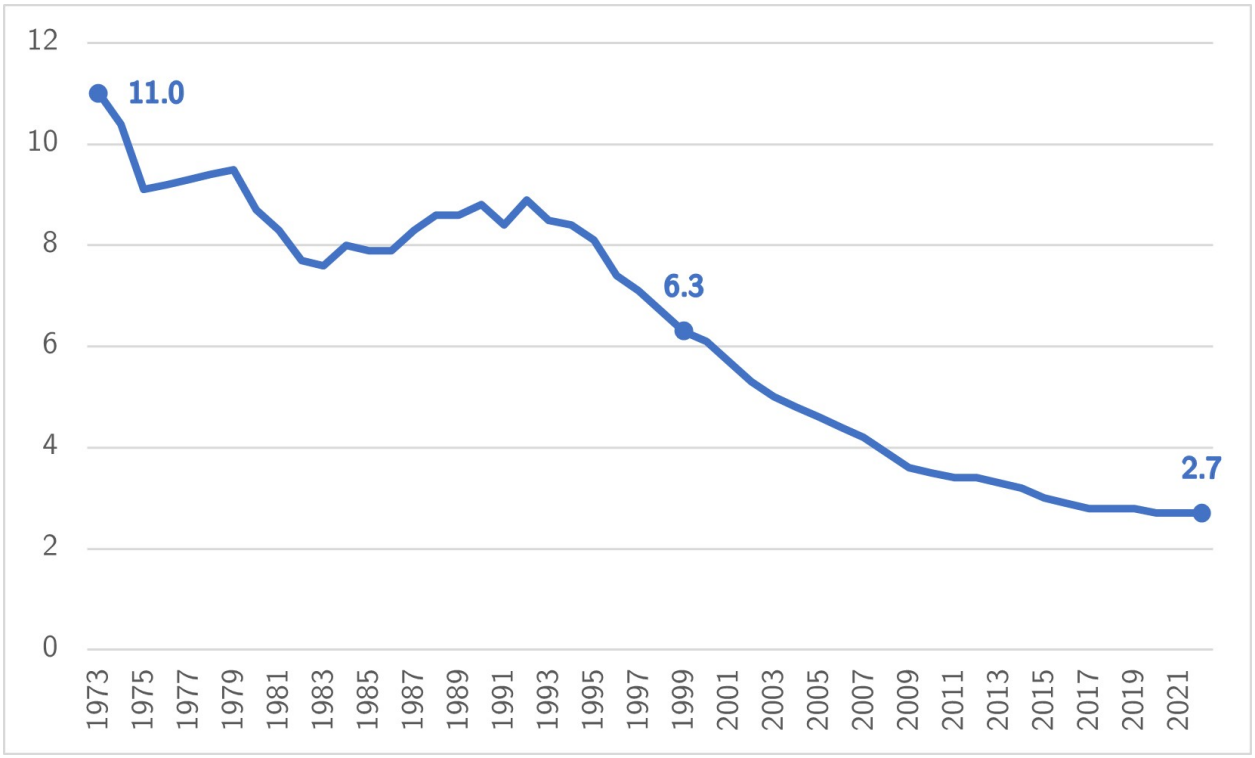


Figure 2: Sample and official workplace injury rates by year

This figure plots total workplace injuries per 100 workers for our sample and official aggregate workplace injuries per 100 full-time equivalent workers as reported by the Bureau of Labor Statistics by year for the period 1999–2022. A full-time equivalent worker is defined as 2,000 hours worked.

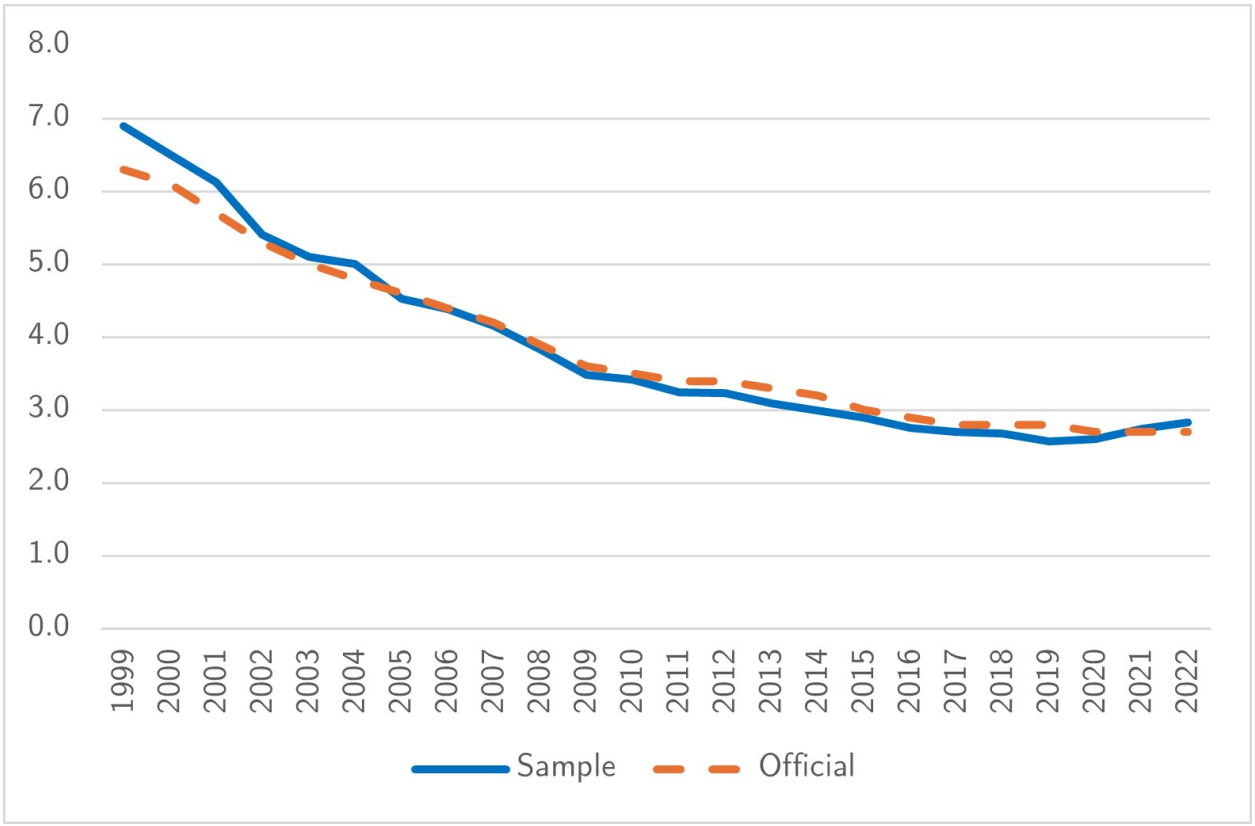


Figure 3: Concentration of occupational employment by industry

This figure presents a histogram of the distribution of the Herfindahl-Hirschman Index (HHI) of the share of each detailed SOC code occupation's employment across different 4-digit NAICS code industries. The HHI for a given occupation j across industries $k \in K$ is

$$HHI_j = \sum_{k \in K} \left(emp_{j,k} / \sum_{k \in K} emp_{j,k} \right)^2,$$

where $emp_{j,k}$ is the number of employees in occupation j in industry k in the sample. The bin width is 0.01. Each bar depicts the number of occupations in a bin.

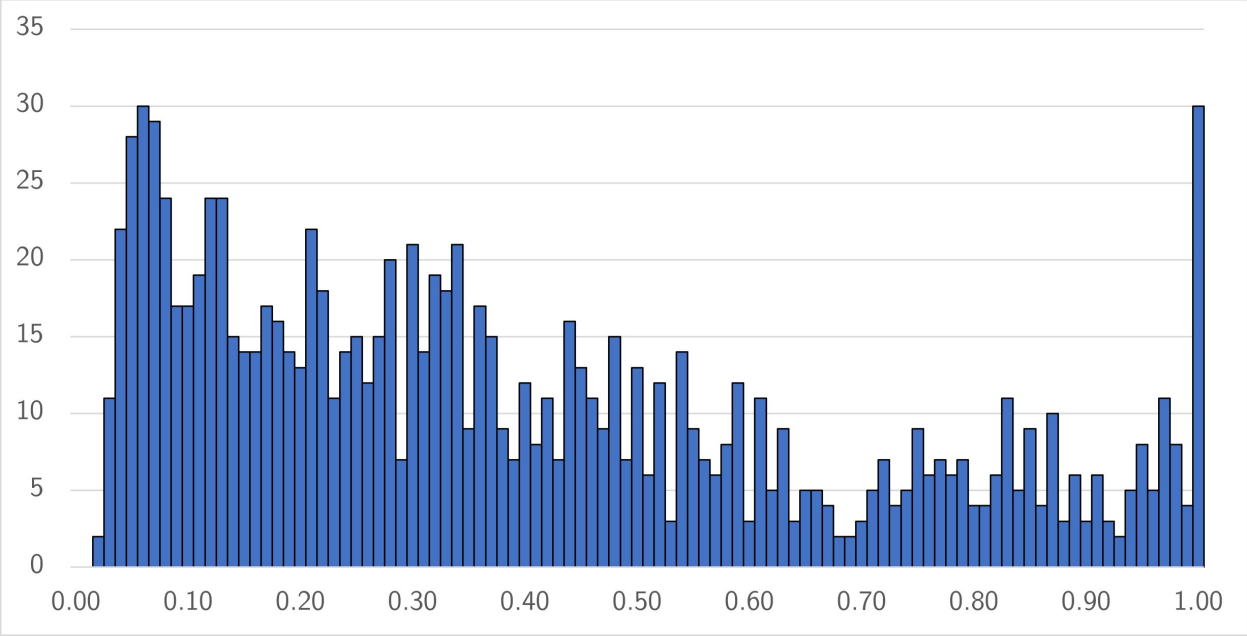


Figure 4: Average within-occupation trends in workplace injury rates

This figure presents a decomposition of the change in workplace injury rate between 1999 and 2022 into the change due to within-occupation changes in injury rates and the change due to changes in occupational composition. Subfigures (a) and (b) present this decomposition for total cases and Days Away, Restricted or Transfer (DART) cases, each per 100 workers. The dashed green line in each subfigure depicts the 1999 injury rate. The solid blue line depicts the actual rate in our sample in each year. The dotted orange line depicts the counterfactual trend accounting only for within-occupation changes in injury rates, which changes relative to 1999 based on the estimated λ_t coefficient from Regression 2,

$$IR_{i,t} = \mathbf{s}'\boldsymbol{\beta} \; + \; \mathbf{d}'\boldsymbol{\lambda} \; + \; \varepsilon_{i,t},$$

where $IR_{i,t}$ is the establishment-year injury rate, $s_{i,j,t}$ is the share of employees in establishment i who are in detailed SOC occupation j in year t, $\mathbf{d}$ is a vector of year indicators, and the omitted base year is 1999. The vertical distance between the green and orange lines represents the change due to within-occupation changes in injury rates. The vertical distance between the orange and blue lines represents the change due to changes in occupational composition.

(a) Total cases per 100 workers

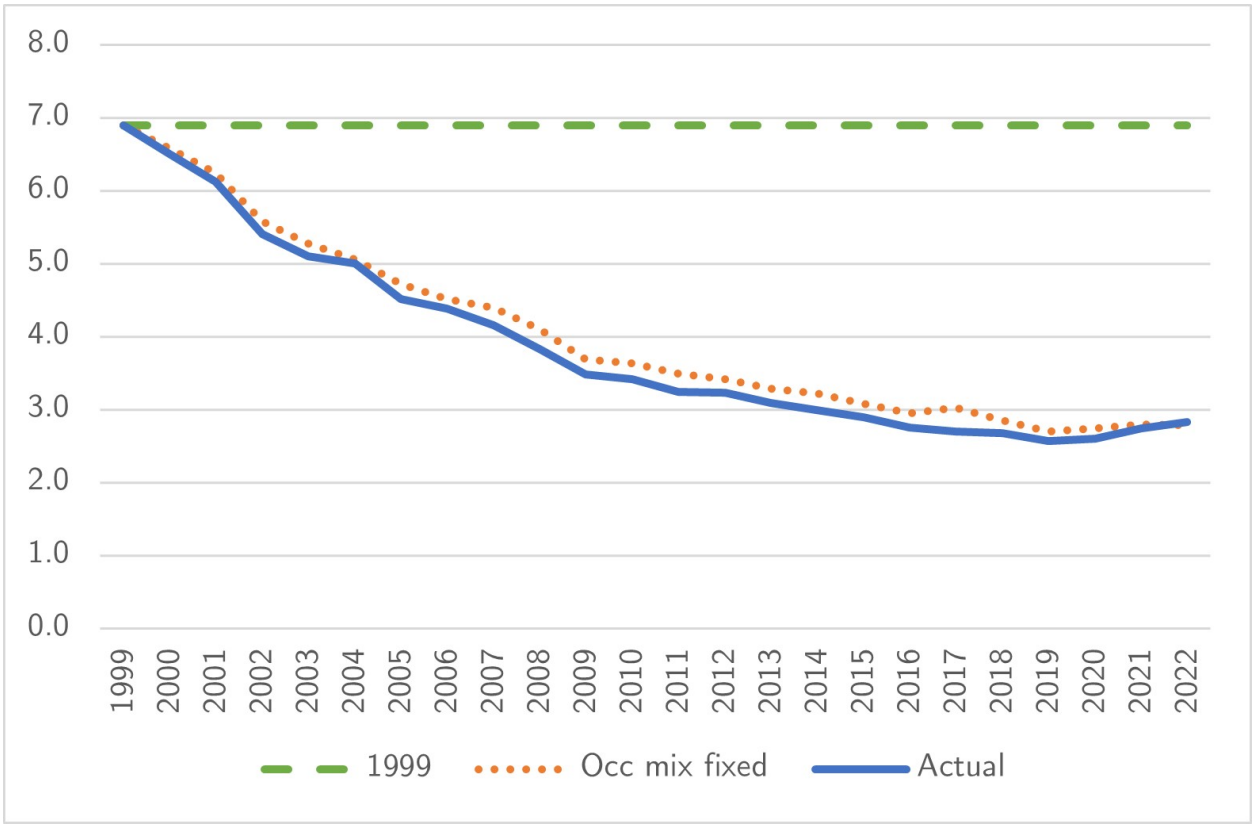


Figure 4: Average within-occupation trends in workplace injury rates (continued)

(b) DART cases per 100 workers

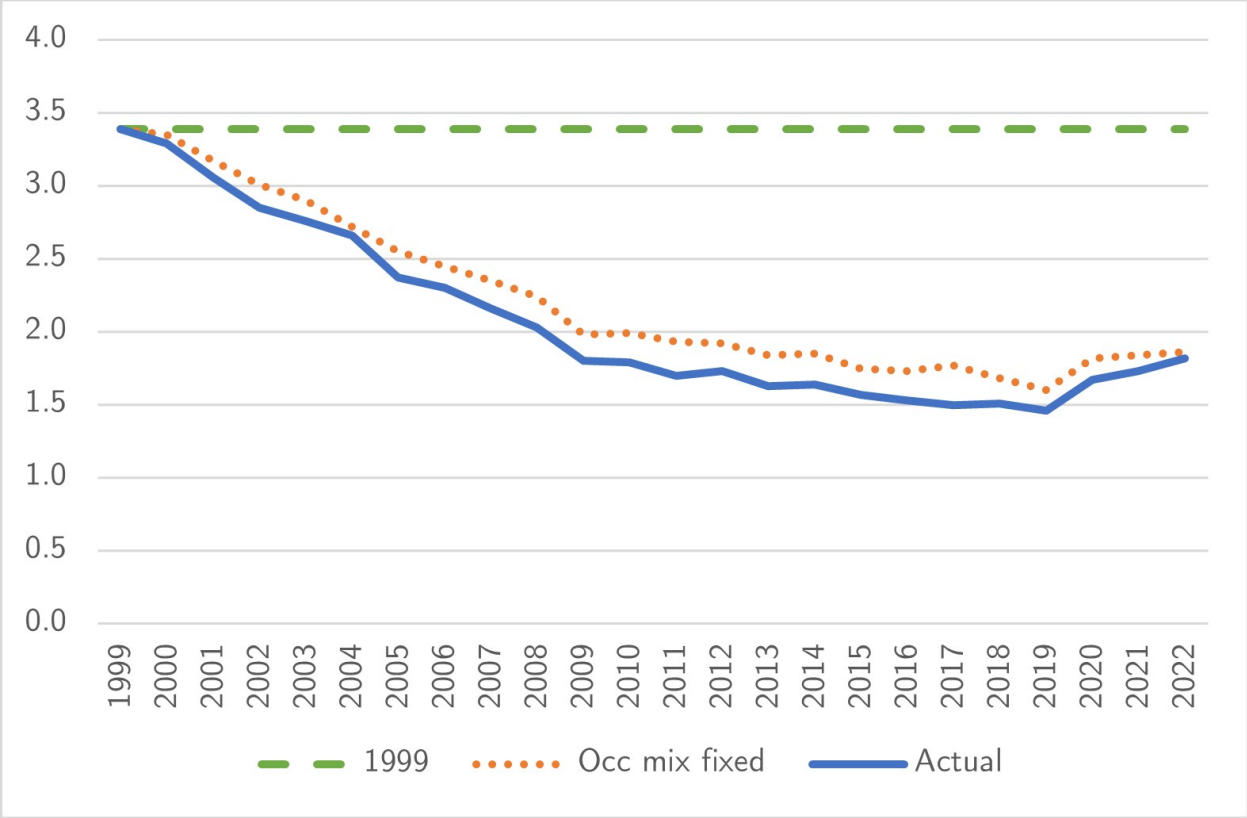


Table 1: Observations in sample by year

This table reports the number and percentage of establishment-year observations in the sample by year. Note that the number of observations is smaller in 2002 because the OEWS changed the timing of its survey that year and only surveyed half its normal number of establishments and in 2020–2022 because the OEWS surveyed fewer establishments in those years.

Year	Count	Percent
1999	24,564	4.6
2000	22,879	4.3
2001	20,707	3.9
2002	8,156	1.5
2003	22,284	4.2
2004	22,108	4.1
2005	18,581	3.5
2006	29,831	5.6
2007	26,845	5.0
2008	24,393	4.6
2009	25,330	4.7
2010	25,640	4.8
2011	23,620	4.4
2012	26,453	4.9
2013	24,141	4.5
2014	24,148	4.5
2015	25,462	4.8
2016	24,194	4.5
2017	23,328	4.4
2018	23,839	4.5
2019	21,618	4.0
2020	17,462	3.3
2021	14,888	2.8
2022	14,965	2.8
Total	535,436	100.0

Table 2: Observations by 2-digit NAICS code industries

This table presents the percentage of establishment-years in the sample analyzed in the paper in each 2-digit NAICS code industry (Sample Pct). For comparison, it also provides the average percentage of establishment-years in each 2-digit NAICS code industry as reported in the Census Bureau's Statistics of U.S. Business annual datasets (Census Pct) over the same years as in the sample period.

Industry	NAICS2	Sample Pct	Census Pct
Agriculture, forestry, fishing, and hunting	11	0.5	0.3
Mining, quarrying, and oil and gas extraction	21	0.5	0.3
Utilities	22	1.0	0.2
Construction	23	7.1	10.4
Manufacturing	31-33	21.8	4.0
Wholesale trade	42	5.7	5.4
Retail trade	44-45	13.7	13.2
Transportation and warehousing	48-49	5.1	2.9
Information	51	2.5	1.6
Finance and insurance	52	3.1	5.6
Real estate and rental and leasing	53	1.7	5.0
Professional, scientific, and technical services	54	4.0	12.1
Management of companies and enterprises	55	1.4	0.5
Admin, support, waste mgmt, remediation svcs	56	5.7	5.3
Educational services	61	2.1	1.3
Health care and social assistance	62	13.5	10.9
Arts, entertainment, and recreation	71	2.5	1.8
Accommodation and food services	72	4.0	8.5
Other Services (except Public Administration)	81	4.0	10.4

Table 3: Employment by occupation category in sample

This table reports the fractions of private-sector workers employed in each of 22 major SOC occupation categories. The first column reports the fractions in the sample, without applying sampling weights. The second column shows the fractions in the sample, applying sampling weights. The third column shows the average fractions over the sample period based on the official annual OEWS releases.

Major occupation	SOC code	Sample (unweighted)	Sample (weighted)	Official
Management	11	5.5	5.7	5.2
Business and Financial Operations	13	4.2	5.2	4.8
Computer and Mathematical	15	2.1	1.9	2.7
Architecture and Engineering	17	2.0	1.7	1.8
Life, Physical, and Social Science	19	0.7	0.7	0.9
Community and Social Service	21	1.2	1.4	1.4
Legal	23	0.4	0.6	0.8
Educational Instruction and Library	25	1.5	1.5	6.2
Arts, Design, Entertainment, Sports, and Media	27	1.5	2.3	1.3
Healthcare Practitioners and Technical	29	4.8	3.4	5.5
Healthcare Support	31	3.1	2.3	3.1
Protective Service	33	0.9	0.6	2.4
Food Preparation and Serving Related	35	4.3	5.1	8.5
Building and Grounds Cleaning and Maintenance	37	2.7	2.7	3.2
Personal Care and Service	39	2.9	2.9	2.6
Sales and Related	41	11.2	14.6	10.3
Office and Administrative Support	43	15.4	18.7	16.2
Farming, Fishing, and Forestry	45	0.4	0.5	0.3
Construction and Extraction	47	5.7	6.5	4.4
Installation, Maintenance, and Repair	49	5.1	5.9	4.0
Production	51	14.0	8.2	7.2
Transportation and Material Moving	53	10.5	7.8	7.4

Table 4: Workplace injury rates by occupation category, 1999–2022

This table presents estimates of average annual workplace injuries per 100 workers by major SOC occupation category for the period 1999–2022. The estimates are obtained from Regression 1,

$$IR_{i,t} = \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{i,t},$$

where $IR_{i,t}$ denotes the injury rate for establishment i in year t, and $\mathbf{s} = (s_{i,1,t}, \ldots, s_{i,J,t})'$ is the vector of occupation employment shares for establishment i in year t. The coefficient vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)'$ therefore provides estimates of occupation-specific injury rates. The table presents estimates for total cases and Days Away, Restricted or Transfer (DART) cases. Robust standard errors are reported in parentheses.

	Cases / 100 workers			
Major SOC occupation category	Total	(SE)	DART	(SE)
Production	7.63	(0.07)	4.12	(0.04)
Transportation and Material Moving	7.21	(0.06)	5.01	(0.05)
Healthcare Practitioners and Technical	6.72	(0.07)	2.58	(0.04)
Construction and Extraction	5.88	(0.08)	3.11	(0.05)
Installation, Maintenance, and Repair	5.80	(0.12)	3.09	(0.07)
Building and Grounds Cleaning & Maintenance	4.92	(0.11)	2.84	(0.07)
Farming, Fishing, and Forestry	4.44	(0.18)	2.66	(0.12)
Community and Social Service	4.41	(0.17)	2.02	(0.09)
Healthcare Support	4.41	(0.13)	2.86	(0.07)
Food Preparation and Serving Related	3.80	(0.07)	1.50	(0.04)
Personal Care and Service	3.47	(0.09)	1.90	(0.06)
Sales and Related	3.32	(0.04)	1.73	(0.03)
Protective Service	2.38	(0.12)	1.25	(0.08)
Office and Administrative Support	2.25	(0.04)	1.24	(0.03)
Management	1.89	(0.13)	0.68	(0.08)
Arts, Design, Entertainment, Sports, & Media	1.60	(0.13)	0.69	(0.09)
Educational Instruction and Library	1.26	(0.10)	0.43	(0.06)
Architecture and Engineering	0.58	(0.11)	0.09	(0.05)
Life, Physical, and Social Science	0.09	(0.13)	0.00	(0.07)
Legal	-0.21	(0.12)	-0.18	(0.08)
Computer and Mathematical	-0.48	(0.05)	-0.25	(0.03)
Business and Financial Operations	-0.49	(0.06)	-0.36	(0.04)

Table 5: Workplace injury rates for high-injury rate occupations, 1999–2022

This table presents estimates of average annual workplace injuries per 100 workers by detailed occupation for the period 1999–2022 for occupations with injury rates of at least 10 total cases per 100 workers and that represent at least 0.01% of total employment in our sample. The estimates are obtained from Regression 1,

$$IR_{i,t} = \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{i,t},$$

where $IR_{i,t}$ denotes the injury rate for establishment i in year t, and $\mathbf{s} = (s_{i,1,t}, \ldots, s_{i,J,t})'$ is the vector of occupation employment shares for establishment i in year t. The coefficient vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)'$ therefore provides estimates of occupation-specific injury rates. Robust standard errors are reported in parentheses.

Detailed SOC occupation (SOC code)	Cases / 100 workers	(SE)
Paramedics (29-2043)	15.12	(1.92)
Slaughterers and Meat Packers (51-3023)	14.43	(0.97)
Cardiovascular Technologists and Technicians (29-2031)	14.08	(2.87)
Material Moving Workers, All Other (53-7199)	13.25	(1.65)
Metal Workers and Plastic Workers, All Other (51-4199)	13.18	(1.54)
Animal Trainers (39-2011)	12.52	(4.77)
Structural Metal Fabricators and Fitters (51-2041)	12.51	(0.96)
Zoologists and Wildlife Biologists (19-1023)	12.41	(2.95)
Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders (51-4122)	12.37	(1.44)
Tool and Die Makers (51-4111)	12.24	(1.50)
Automotive Glass Installers and Repairers (49-3022)	12.17	(3.04)
Medical Equipment Preparers (31-9093)	11.93	(3.47)
Conveyor Operators and Tenders (53-7011)	11.38	(1.27)
Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic (51-4034)	11.30	(1.28)
Baggage Porters and Bellhops (39-6011)	11.24	(1.57)
Veterinary Technologists and Technicians (29-2056)	11.21	(2.03)
Grinding, Lapping, Polishing, and Buffing Machine Tool Setters, Operators, and Tenders (51-4033)	11.20	(0.91)
Food Processing Workers, All Other (51-3099)	11.18	(1.39)
Welders, Cutters, Solderers, and Brazers (51-4121)	11.05	(0.53)
Cooks, Institution and Cafeteria (35-2012)	10.93	(1.30)
Sawing Machine Setters, Operators, and Tenders, Wood (51-7041)	10.93	(1.11)
Sheet Metal Workers (47-2211)	10.85	(0.73)
Fence Erectors (47-4031)	10.73	(2.51)
Healthcare Practitioner and Technical Workers, All Other (29-9099)	10.58	(2.58)
Bus and Truck Mechanics and Diesel Engine Specialists (49-3031)	10.56	(0.76)
Special Education Teachers, Kindergarten and Elementary School (25-2052)	10.49	(2.68)
Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic (51-4031)	10.39	(0.47)
Refuse and Recyclable Material Collectors (53-7081)	10.23	(0.81)
Recreational Vehicle Service Technicians (49-3092)	10.12	(1.82)
Occupational Therapists (29-1122)	10.11	(2.18)
Woodworking Machine Setters, Operators, and Tenders, Except Sawing (51-7042)	10.09	(0.61)
Butchers and Meat Cutters (51-3021)	10.07	(1.21)
Nuclear Medicine Technologists (29-2033)	10.04	(4.81)

Table 6: Fraction of aggregate injury rate change due to within-occupation injury rate changes

This table reports the portion of the change in aggregate workplace injury rate from 1999 to 2022 due to within-occupation changes in injury rate. The estimates are obtained from Regression 2,

$$IR_{i,t} = \mathbf{s}'\boldsymbol{\beta} \; + \; \mathbf{d}'\boldsymbol{\lambda} \; + \; \varepsilon_{i,t},$$

where $IR_{i,t}$ denotes the injury rate for establishment i in year t, $\mathbf{s} = (s_{i,1,t}, \ldots, s_{i,J,t})'$ is the vector of occupation employment shares for establishment i in year t, and $\mathbf{d} = (d_{t,1}, \ldots, d_{t,T})'$ is a vector of year indicators. The coefficient vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ therefore provides estimates of the change in injury rate from base year 0, *holding occupational composition fixed*. The table reports the level at which occupation is defined, the type of injury cases considered, the variant of the regression (if applicable), the total change in injury rate from 1999 to 2022 (ΔIR_T), the portion of the change due to within-occupation changes in injury rates (λ_T), and the share of the change due to within-occupation changes in injury rates ($f_T = \lambda_T / \Delta IR_T$). Among the variants, *Estab FE* indicates the inclusion of establishment fixed effects in the regression, *Emp screen* indicates that any establishment-year where total employment in the SOII and OEWS data differs by more than 30% is excluded from the sample, *Unweighted* indicates that sampling weights are not used, and *Case data* indicates that results are from the test based on estimation of Regression C1 using case-level data.

Row	Occ level	Cases	Variant	Obs.	ΔIR_T	W/in-occ ΔIR (λ_T)	% due to w/in-occ ΔIR (f_T)
(1)	Detailed	Total		535,436	-4.07	-4.11	101.0
(2)	Detailed	DART		535,436	-1.57	-1.53	97.5
(3)	Major	Total		535,436	-4.07	-4.04	99.3
(4)	Detailed × Wage Bin	Total		535,436			
(5)	Major × Major	Total		535,436	-4.07	-3.98	97.8
(6)	Detailed	Total	OEWS wts				
(7)	Detailed	Total	Unweighted	535,436	-3.24	-3.41	105.2
(8)	Detailed	Total	Estab FE	277,818	-4.70	-5.35	113.8
(9)	Detailed	Total	Emp screen	405,272	-4.55	-4.45	97.8
(10)	Major	DAFW	Case data	3,753,735	-0.38	-0.37	97.4

Table 7: Change in injury rate by occupation category and effect on aggregate injury rate

This table presents estimates of the change in the annual workplace injury rate per 100 workers from 1999 to 2022 for each major SOC occupation category. The estimates are obtained from Regression 3,

$$IR_{i,t} = \mathbf{s}' \mathbf{B} \mathbf{d} \; + \; \varepsilon_{i,t},$$

where $IR_{i,t}$ denotes the injury rate for establishment i in year t, $\mathbf{s} = (s_{i,1,t}, \ldots, s_{i,J,t})'$ is the vector of occupation employment shares for establishment i in year t, and $\mathbf{d} = (d_{t,0}, \ldots, d_{t,T})'$ is a vector of year indicators. Element $\beta_{j,t}$ of the $J \times (T+1)$ coefficient matrix $\mathbf{B}$ provides an occupation-year specific injury rate estimate. The table reports the change in occupation-specific injury rate from 1999 to 2022 ($\beta_{j,T} - \beta_{j,0}$), the percent change relative to 1999, 1999 employment share ($S_{j,0}$), and the effect of the change in occupation-specific injury rate on the total injury rate ($D_j^{IR} = S_{j,0}(\beta_{j,T} - \beta_{j,0})$). The percentage change is only reported for occupation categories with a total estimated injury rate of at least 2.0 cases per 100 workers (see Table 4).

Major occupation	ΔIR ($\beta_{j,T} - \beta_{j,0}$)	ΔIR%	1999 emp(%)	Effect on IR (D_j^{IR})
Production	-8.7	-67%	10.4	-0.90
Construction and Extraction	-9.7	-81%	5.9	-0.57
Transportation and Material Moving	-5.9	-47%	8.6	-0.51
Sales and Related	-2.3	-55%	18.0	-0.41
Office and Administrative Support	-2.3	-71%	18.0	-0.41
Installation, Maintenance, and Repair	-5.9	-59%	5.2	-0.31
Food Preparation and Serving Related	-4.2	-59%	3.8	-0.16
Healthcare Support	-9.0	-79%	1.7	-0.15
Building and Grounds Cleaning & Maintenance	-3.3	-49%	3.1	-0.10
Healthcare Practitioners and Technical	-3.0	-34%	2.7	-0.08
Personal Care and Service	-2.9	-51%	2.8	-0.08
Architecture and Engineering	-3.8	-	1.2	-0.05
Business and Financial Operations	-1.5	-	2.9	-0.04
Community and Social Service	-2.5	-41%	1.4	-0.04
Farming, Fishing, and Forestry	-3.5	-63%	0.5	-0.02
Protective Service	-1.5	-47%	0.6	-0.01
Management	0.0	-	8.3	0.00
Legal	0.9	-	0.4	0.00
Arts, Design, Entertainment, Sports, & Media	0.4	-	2.0	0.01
Computer and Mathematical	0.5	-	1.0	0.01
Life, Physical, and Social Science	1.1	-	0.5	0.01
Educational Instruction and Library	1.0	-	1.2	0.01

Table 8: Change in injury rates and occupation characteristics

This table presents estimates of changes in the relationship between annual workplace injuries per 100 workers and occupation characteristics over the period 2003–2022. The estimates are obtained from Regression 4,

$$IR_{i,t} = \mathbf{d}'\,\boldsymbol{\lambda} \;+\; \mathbf{x}'\,\boldsymbol{\phi} \;+\; \mathbf{x}'\,\mathbf{G}\,\mathbf{d},$$

where $IR_{i,t}$ denotes the injury rate for establishment i in year t, $\mathbf{x} = (x_{i,1,t}, \ldots, x_{i,C,t})'$ is a vector of employment-weighted average values of characteristics associated with an occupation, and $\mathbf{d} = (d_{t,1}, \ldots, d_{t,T})'$ is a vector of year indicators. Element $\gamma_{c,T}$ of the $C \times T$ coefficient matrix $\mathbf{G}$ provides an estimate of the change in injury rate associated with characteristic c. The first two columns show the mean and standard deviation of each characteristic in the sample. The last four columns show the estimated change in injury rates with each characteristic, for total injury cases and Days Away, Restricted or Transfer (DART) cases, with robust standard errors in parentheses. *Work activities* are based on the values of work activities as reported by O*NET, rescaled to lie between 0 and 1, with the exception of *Working with animals*, which is 1 for occupations in categories involving direct contact with animals and 0 otherwise. *Automation & offshoring exposure* variables are based on the categorization of Acemoglu and Autor (2011), rescaled to lie between 0 and 1. *Union membership* is the percent of workers in an occupation who report being union members, based on census data.

			Total		DART	
Characteristic	Mean	S.D.	Change	(SE)	Change	(SE)
Work activities						
General physical activities	0.496	0.154	4.49	(2.14)	4.36	(1.95)
Handling and moving objects	0.508	0.155	1.34	(1.83)	1.42	(1.59)
Operating vehicles, mechanized devices	0.326	0.184	-3.82	(0.84)	-3.02	(0.64)
Controlling machines and processes	0.420	0.157	-8.30	(1.05)	-6.30	(0.80)
Assisting and caring for others	0.465	0.161	-1.44	(0.80)	-1.95	(0.67)
Working with animals	0.003	0.042	1.30	(3.60)	0.44	(2.45)
Baseline change (λ_T)			-0.35	(0.43)	0.42	(0.32)
Automation & offshoring exposure						
Routine manual	0.406	0.123	-9.33	(1.39)	-4.82	(1.04)
Routine cognitive	0.536	0.061	5.24	(1.81)	2.82	(1.26)
Non-routine physical	0.371	0.125	-2.59	(1.72)	-1.11	(1.23)
Non-routine interpersonal	0.490	0.092	-2.27	(1.56)	-1.72	(1.16)
Non-routine analytical	0.480	0.107	0.89	(1.08)	0.04	(0.84)
Offshorability	0.482	0.068	-4.22	(1.86)	-4.10	(1.35)
Baseline change (λ_T)			2.00	(1.97)	3.04	(1.42)
Unionization						
Union membership	0.102	0.061	-12.61	(1.87)	-8.27	(1.21)
Baseline change (λ_T)			-0.51	(0.19)	0.23	(0.12)

Table 9: Change in employment by occupation category and effect on aggregate injury rate

This table presents estimates of the change in employment from 1999 to 2022 for each major SOC occupation category as well as its effect on the aggregate injury rate. The first three columns of the table report 1999 and 2022 employment shares and changes in shares from 1999 to 2022. The fourth column reports the 1999 injury rate (total cases per 100 employees), which is the estimated value of $\beta_{j,0}$ from Regression 3,

$$IR_{i,t} = \mathbf{s}' \mathbf{B} \mathbf{d} \; + \; \varepsilon_{i,t},$$

where $IR_{i,t}$ denotes the injury rate for establishment i in year t, $\mathbf{s} = (s_{i,1,t}, \ldots, s_{i,J,t})'$ is the vector of occupation employment shares for establishment i in year t, and $\mathbf{d} = (d_{t,0}, \ldots, d_{t,T})'$ is a vector of year indicators. Element $\beta_{j,t}$ of the $J \times (T+1)$ coefficient matrix $\mathbf{B}$ provides an occupation-year specific injury rate estimate. The final column reports the effect of the change in employment in the occupation on the aggregate injury rate, holding the injury rate in the occupation constant at its 1999 level (D_j^s). This effect is computed by assuming that any increase (decrease) in employment in an occupation is exactly offset by a decrease (increase) in employment in all other occupations proportionate to their 1999 employment.

Major occupation	1999 emp	2022 emp	Δemp	1999 IR ($\beta_{j,0}$)	Effect on IR (D_j^s)
Production	10.4	6.9	-3.5	13.1	-0.26
Business and Financial Operations	2.9	6.4	3.5	0.9	-0.20
Computer and Mathematical	1.0	2.3	1.3	-0.7	-0.09
Legal	0.4	1.1	0.7	-1.3	-0.05
Transportation and Material Moving	8.6	8.3	-0.3	12.4	-0.02
Life, Physical, and Social Science	0.5	0.7	0.2	-1.6	-0.02
Educational Instruction and Library	1.2	1.3	0.1	0.7	-0.01
Arts, Design, Entert., Sports, and Media	2.0	2.3	0.3	3.1	-0.01
Architecture and Engineering	1.2	1.6	0.4	3.9	-0.01
Building and Grounds Cleaning and Maint.	3.1	2.4	-0.7	6.7	0.00
Personal Care and Service	2.8	2.1	-0.7	5.8	0.00
Farming, Fishing, and Forestry	0.5	0.4	-0.1	5.6	0.00
Community and Social Service	1.4	1.7	0.3	6.0	0.00
Food Preparation and Serving Related	3.8	4.4	0.6	7.0	0.00
Protective Service	0.6	0.4	-0.2	3.2	0.01
Management	8.3	8.2	-0.1	1.0	0.01
Construction and Extraction	5.9	6.4	0.5	12.0	0.03
Healthcare Practitioners and Technical	2.7	3.7	1.0	8.9	0.03
Installation, Maintenance, and Repair	5.2	6.9	1.7	10.1	0.07
Office and Administrative Support	18.0	15.6	-2.4	3.3	0.09
Healthcare Support	1.7	3.5	1.8	11.4	0.09
Sales and Related	18.0	13.5	-4.5	4.3	0.12

Appendices

A Detailed Discussion of SOII and OEWS Microdata

A.1 Survey of Occupational Injuries and Illnesses (SOII)

The BLS conducts the Survey of Occupational Injuries and Illnesses (SOII) annually to produce aggregate workplace injury statistics at the national, state, and industry levels. The survey includes approximately 230,000 establishments each year, using disproportionate stratified sampling to select a random sample, oversampling some strata of establishments and undersampling others. The sampling frame for the survey is derived from the Quarterly Census of Employment and Wages (QCEW), which includes establishments subject to state unemployment insurance laws. It includes private industry establishments as well as state and local government workplaces across the United States.[9] While the BLS does not disclose its exact sampling methodology to protect the confidentiality of respondents, it divides establishments into strata based on industry classification, size, and geographic location.

The SOII collects data through a questionnaire sent to establishments either via mail or electronically. The BLS notifies establishments in advance of the year for which they will be required to provide data. Establishments can provide data by submitting a physical form via mail or fax, by phone to a BLS representative, or by secure web-based system. In 2022, approximately 77 percent of respondents provided data via the web-based system. Each establishment in the survey is asked to report summary information about

[9]Federal agencies, self-employed workers, farms with fewer than 11 employees, and private households are excluded from the survey.

work-related injury counts and employment and detailed information about any cases requiring days away from work. Most establishments are already required by law to record detailed information about work-related injuries in OSHA form 300 (Log of Work-Related Injuries and Illnesses) and to summarize their cases each calendar year in OSHA Form 300A (Summary of Work-Related Injuries and Illnesses). Cases include any work-related injuries or illnesses (we refer to injuries and illnesses collectively as "injuries" for brevity) requiring treatment beyond first aid, days away from work, or temporary job transfer or restriction. The response rate for the SOII in recent years ranges from 84% to 89%.[10] The summary microdata is available starting in 1997 and the case-level microdata starting in 1995. See https://www.bls.gov/opub/hom/soii/data.htm for a more detailed description of the SOII data collection process.

A.2 Occupational Employment and Wage Statistics (OEWS)

The BLS and state workforce agencies in each state are jointly responsible for conducting the Occupational Employment and Wage Statistics (OEWS) survey to produce aggregate employment and wage statistics for different occupations at the national, state, and industry levels each May.[11] Each half-year, in May and November, the BLS surveys a random, stratified sample of approximately 200,000 U.S. establishments, selected from a sampling frame derived from same Quarterly Census of Employment and Wages (QCEW) it uses for the SOII. It then uses data from the most recent six half-year cohorts, up to and including the May cohort from the current year, to compute aggregate statistics. As such, an establishment can generally only be surveyed at most once every three years. As with the SOII, establishments are stratified by industry, size, and location, and establishments in different strata are sampled with different probabilities.

[10]Source: https://www.bls.gov/osmr/response-rates/establishment-survey-response-rates.htm
[11]The OEWS was known as Occupational Employment Statistics (OES) prior to 2023.

Beginning with the May 2020 OEWS data, BLS changed the confidential microdata from a weighted sample of responding establishments to a file representing the full establishment universe from the QCEW frame. For establishments that did not provide usable OEWS responses, occupational employment and wage distributions are imputed using the program's donor-based nearest-neighbor modeling procedure. Because the file now represents the entire population—combining observed establishments and establishments with modeled staffing patterns—sampling weights are no longer required. Aggregates can therefore be obtained simply by summing employment or wage measures across establishments in the file. However, as a result, sampling weights are no longer available in the microdata after 2019. We therefore only use the SOII sampling weights and refrain from using the OEWS sampling weights in our main line of analysis.

As with the SOII, the OEWS collects data through a questionnaire sent to establishments either via mail or electronically, and establishments can also provide data through a secure web-based system. Each surveyed establishment is asked to provide counts of employees by Standard Occupation Classification (SOC) occupation-by-hourly-wage-interval-cell. The wage intervals, which the BLS uses to compute estimates of average wage by occupation, are granular (e.g., under $9.25 per hour, $9.25 to $11.99 per hour, $12.00 to $15.49 per hour, $15.50 to $19.74 per hour). The response rate for the OEWS survey in recent years ranges from 66% to 74%.[12] Since we conduct our analysis at the annual level to match the frequency of the SOII, we assign each half-year cohort to its calendar year. OEWS microdata is available starting in 1997. However, data prior to 1999 are based on a different occupation classification scheme and are therefore not directly comparable to the data for 1999 and later. We therefore start our sample period in 1999.

The SOC scheme, a federal statistical standard used by agencies to classify workers into occupational categories, was created by a consortium of federal agencies to ensure consistency

[12]Source: https://www.bls.gov/osmr/response-rates/establishment-survey-response-rates.htm

and comparability across various data sources and programs. The SOC includes 867 detailed occupations organized hierarchically into 459 broad occupations, 98 minor groups, and 22 major groups.[13] The SOC was revised in 2010 and again in 2018, with the BLS adopting the new codes in the OEWS starting in 2011 and 2019, respectively. Because the revisions involve both mergers and splits of existing SOC codes, there is no way to maintain a fully homogeneous set of SOC occupations in the data across revisions. We opt to use employment in the SOC codes reported in the OEWS in each year without attempting to homogenize across years.

To assess the potential importance of the SOC revisions, we use the official BLS crosswalks mapping 2000 SOC codes to 2010 SOC codes and 2010 SOC codes to 2018 SOC codes. In the 2010 revision, 62 of 821 detailed (six-digit) occupation codes were mapped to a different detailed code, and these occupations accounted for 7.94% of total employment in 2011. The 2018 revision was larger: 106 of 840 detailed codes changed, covering 13.10% of 2019 employment. Most of these changes involved reclassifications within the same major SOC group. At the major group level, only 11 codes changed in the 2010 revision, representing 1.19% of 2011 employment, and 14 codes changed in the 2018 revision, representing 2.13% of 2019 employment. In the 2010 revision, 30 codes split into multiple new codes and 13 new codes were formed by merging multiple old codes, while in the 2018 revision, 42 codes split and 31 merged.

A.3 Merging the SOII and OEWS Microdata

Both the SOII and the OEWS draw their samples from the same Quarterly Census of Employment and Wages (QCEW) sampling frame. As a result, establishments in the two surveys carry a common longitudinal establishment identifier, which is what makes the merge

[13]The SOC actually has 23 major occupational categories, but the 23rd is for military occupations, which have a separate set of detailed occupations not included in the 867 detailed occupations that we use.

possible. This appendix section describes the identifiers, the semiannual-to-annual cohort assignment, and the sequence of merges we use to construct the establishment-year analysis file.

OEWS file structure and identifiers

The confidential OEWS microdata are distributed as two sets of files. The *mastdata* files contain the core survey content: for each establishment and each detailed SOC occupation, employment is reported across twelve hourly wage intervals, which we label A through L.[14] Each mastdata file is released in support of a given May reference period and contains *six overlapping semiannual panels*—three years of data—identified by the panel field `yr` (`yr`= 0 denotes the most recent panel and `yr`= 5 the oldest) and the quarter field `qtr` (2 for a May cohort, 4 for a November cohort). Within a mastdata file, establishments are keyed by `schnum`, an eleven-digit OES identification number (a two-digit state FIPS code concatenated with a nine-digit schedule number). Critically, `schnum` uniquely identifies an establishment only within a mastdata file (and, at most, across a few adjacent files); it is not a stable longitudinal identifier and cannot by itself be linked to the SOII.

The second set of files, the *E0* files, contain establishment-level attributes not carried in the mastdata files. Each E0 file corresponds to a single semiannual cohort and is linked to the mastdata by `schnum`. For our purposes the essential E0 field is `udbnum`, the establishment's identifier on the QCEW Longitudinal Data Base. Unlike `schnum`, `udbnum` is stable across surveys and over time, and it is the identifier on which the SOII and OEWS can be joined: the SOII establishment identifier (`estab_id`) is the same `udbnum`.

[14]The OEWS collects employment by occupation-by-wage-interval cell rather than a single occupational headcount. Summing the twelve interval counts within an occupation recovers the establishment's total employment in that occupation. We collapse across the wage intervals for all analyses except those defined at the occupation-by-wage-bin level.

Selecting non-overlapping semiannual cohorts

Because each mastdata file stacks six overlapping panels, naively pooling the files would count most establishment-cohorts several times. To recover a single, non-overlapping observation for each establishment in each semiannual cohort, we retain only the two most recent panels from each May mastdata file. Specifically, from the May mastdata file for year i (for $i = 2004, \ldots, 2020$) we keep records with `yr`$= 0$, which constitute the May cohort of year i, and records with `yr`$= 1$, which constitute the November cohort of year $i-1$. We assign each retained cohort a label `cohort` equal to $10i + 2$ for the May cohort and $10(i-1) + 4$ for the November cohort. Using only `yr`$= 0$ and `yr`$= 1$ from consecutive May files reconstructs the full sequence of semiannual cohorts with no establishment-cohort appearing twice.

Because the SOII is annual, we then map each semiannual OEWS cohort to a calendar year, defined as

$$\texttt{survey_year} = \begin{cases} \texttt{cohort} & \text{if } \texttt{cohort} \leq 2001, \\ \lfloor \texttt{cohort}/10 \rfloor & \text{otherwise,} \end{cases}$$

so that the May cohort $10i + 2$ and the November cohort $10i + 4$ of year i are both assigned to calendar year i. Each calendar year in the merged data therefore draws on two OEWS cohorts: its own May cohort (the `yr`$= 0$ records from that year's May mastdata file) and its November cohort (the `yr`$= 1$ records from the following year's May mastdata file).

The earliest and transitional years require separate handling. For 1999–2001 we take the single available annual cohort from the corresponding mastdata file (`yr`$= 0$). The OEWS conducted only a November panel in 2002, so 2002 is represented by that single cohort; this is why the number of 2002 observations is roughly half that of adjacent years (Table 1). The 2003 data are taken from the May 2003 mastdata file.

Assembling the merged establishment-year file

We construct the merged data in three joins. First, from the retained mastdata records we assemble occupation-level employment counts keyed by `schnum` and `cohort`. Second, we build a crosswalk from the E0 files linking (`schnum`, `cohort`) to `udbnum`, and merge it onto the occupation counts, thereby attaching the stable QCEW identifier to every occupation record.[15] Third, we merge the resulting file to the SOII summary microdata on (`udbnum`, `survey_year`), retaining establishment-years present in both surveys. This final join is the SOII–OEWS link: the OEWS supplies the establishment's occupational (and wage-interval) employment composition, while the SOII supplies its injury and illness case counts, employment, hours, sampling weight, and industry.

We then reduce the merged file to one record per establishment-year. We drop establishment-years in which a `udbnum` maps to more than one `schnum` within a cohort (approximately 6,500 records), and we drop the small number of establishment-years left with overlapping cohorts after reshaping. We restrict the sample to privately owned establishments using the SOII ownership code. For each establishment-year we compute occupation employment shares by dividing the establishment's employment in each occupation by its total employment summed across all occupations; by construction these shares sum to one. We produce three versions of the file that differ only in the granularity of the occupation dimension—detailed SOC occupations, the 22 major SOC groups, and major SOC group $\times$ wage interval—corresponding to the specifications reported in the paper.[16] The merge yields

[15] A small number (on the order of twenty) of duplicate `schnum`–`cohort` pairs appear in the E0 files and are dropped before merging. For the 2019 and later E0 files, a data-entry change corrupted the `schnum` field; we reconstruct it from the state FIPS code and the preserved original schedule number so that it aligns with the mastdata `schnum`.

[16] Converting the raw BLS files to a usable form also requires routine cleaning that we do not detail here, including converting the SAS-format source files to Stata, harmonizing variable types across years (for example, industry is recorded in SIC through 2002 and in NAICS thereafter, and is stored as a string in some later years), and removing exact duplicate records. Stata code implementing the full procedure is available from the authors.

535,436 unique establishment-years over 1999–2022, an average of 22,310 establishments per year.

B DAFW Injury Rates: Our Estimates vs. Published

Table B1: Workplace injury rates by occupation category, 1999–2022

This table compares estimates of DAFW injuries per 100 workers by major SOC code occupation category using our methodology and sample to the average annual official DAFW injuries per 100 full-time equivalent employees published by the Bureau of Labor Statistics (BLS), both for the years 2006-2022. Our estimates are obtained from Regression 1,

$$IR_{i,t} = \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{i,t},$$

where $IR_{i,t}$ denotes the (DAFW) injury rate for establishment i in year t, and $\mathbf{s} = (s_{i,1,t}, \ldots, s_{i,J,t})'$ is the vector of occupation employment shares for establishment i in year t. Coefficient vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)'$ provides estimates of occupation-specific injury rates.

	DAFW Cases/100 workers	
	Estimated	Published
Management	0.47	0.40
Business and Financial Operations	-0.07	0.13
Computer and Mathematical	-0.10	0.07
Architecture and Engineering	0.05	0.18
Life, Physical, and Social Science	0.00	0.28
Community and Social Service	1.36	0.82
Legal	-0.05	0.11
Educational Instruction and Library	0.30	0.61
Arts, Design, Entertainment, Sports, and Media	0.49	0.46
Healthcare Practitioners and Technical	1.50	1.15
Healthcare Support	1.42	2.20
Protective Service	0.88	1.06
Food Preparation and Serving Related	0.79	0.97
Building and Grounds Cleaning and Maintenance	1.50	2.21
Personal Care and Service	0.99	1.03
Sales and Related	0.69	0.60
Office and Administrative Support	0.51	0.45
Farming, Fishing, and Forestry	1.38	1.37
Construction and Extraction	1.38	1.78
Installation, Maintenance, and Repair	1.84	1.90
Production	1.19	1.44
Transportation and Material Moving	2.51	2.45

C Case-Level Data Alternative Approach

This appendix describes how we use case-level data for DAFW injuries from SOII to implement an alternative approach to decomposing the decline in the aggregate injury rate into separate portions representing the effects of within-occupation injury rate declines and change in occupational composition. In addition to summary data containing information about injury rates at the establishment-year level, the SOII microdata also includes records of each case requiring days away from work for surveyed establishments. These records include an SOC code occupation encoded by the BLS based on job title and case narrative for the injured employee provided by the employer. We use this data to compute establishment-year-occupation workplace injury rates by scaling the count of injuries in an occupation for an establishment-year by the reported number of employees in that occupation in the OEWS data for that establishment-year.

We uncover some inconsistencies between the SOII case and OEWS data. For example, for 9.7% of establishment-year-occupations with at least one DAFW case, the OEWS reports zero employees in the major occupation category to which the occupation belongs. These discrepancies are not surprising since the classification mechanism differs between the OEWS survey and the BLS case data. For the OEWS, the employer maps internal job titles to SOC codes and reports employment counts by SOC code. BLS staff and state workforce agencies then review the reported information and ensure that workers are assigned the correct SOC code. For the SOII case data, the employer reports a job title and case narrative, and the BLS assigns an SOC code based on this information. For this reason, we use the case-level data only as a robustness test. To account for misclassifications, we only compute the establishment-year-occupation injury rate at the major SOC code level, and we exclude observations where OEWS reports zero employees in the corresponding major SOC code occupation.

Let $IR_{i,j,t}$ denote the (now observed) injury rate at establishment i for occupation j in year t. We estimate the counterfactual change in injury rate holding occupation fixed by running the following regression:

Regression C1. DAFW case data-based decomposition

$$IR_{i,j,t} = \alpha + \mathbf{h}'\boldsymbol{\psi} \; + \; \mathbf{d}'\boldsymbol{\lambda} \; + \; \varepsilon_{i,j,t},$$

where α is an intercept, $\mathbf{d} = (d_1, \ldots, d_T)'$ is a vector of year indicators and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ is the corresponding vector of coefficients, and $\mathbf{h} = (h_1, \ldots, h_J)'$ is a vector of occupation indicators and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_J)'$ is the corresponding vector of coefficients. With year 0 omitted, the coefficient λ_t captures the change in the injury rate from year 0 to year t, *holding occupation fixed*.

As in Regression 2, the coefficients λ_t in Regression C1 estimate within-occupation changes in injury rates relative to the year 0. We use these coefficients to compute the fractions of the aggregate change attributable to within-occupation changes in injury rates, f_T, and changes in occupational composition, $1 - f_T$. The results from this exercise are shown in the bottom row of Table 6.