Occupational Health in U.S. Transit Agencies: Trends in OSHA-Reportable Illnesses and Injuries (2016–2023)

Hunter David Solaro

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Marissa Baker

Marty Cohen

Lucia Luu

Program Authorized to Offer Degree:

Department of Environmental & Occupational Health Sciences

©Copyright 2025

Hunter David Solaro

University of Washington

Abstract

Occupational Health in U.S. Transit Agencies: Trends in OSHA-Reportable Illnesses and Injuries (2016–2023)

Hunter D. Solaro

Chair of Supervisory Committee:

Marissa Baker

Department of Environmental & Occupational Health Sciences

Abstract

Introduction

Transit workers in the United States face a host of occupational health hazards including chronic exposure to physical, chemical, and psychosocial agents. A systematic review of 187 studies identified that all workers suffering hazardous occupational noise exposure can cause permanent auditory threshold shifts, though the evidence for elevated risk among railway workers is mixed, with some studies showing minimal or no increased loss relative to reference populations. Respiratory illness is a recognized occupational concern for transit workers, driven in part by elevated exposures to airborne pollutants in enclosed and high-density transit environments. Measured concentrations of fine particulate matter ($PM_{2.5}$) inside buses and subways, coupled with inadequate ventilation or filtration, contribute to inflammation and increased risk of chronic respiratory conditions, and musculoskeletal disorders driven by static postures and vibration.[1,2,3,4,5] These exposures may contribute to a higher prevalence of hypertension, diabetes, and mental health conditions among transit employees compared to the general workforce.[6] Despite the severity and persistence of these conditions, occupational

health surveillance and regulatory prioritization in the transit sector remain fragmented and underdeveloped.[6,7,8] Efforts to quantify these burdens are further complicated by systemic underreporting, definitional inconsistencies, and jurisdictional barriers to case recognition.[8,9,10] This thesis explores these gaps by examining national Occupational Safety and Health Administration's (OSHA's) Injury Tracking Application (ITA) data from 2016–2023[11], with analyses by agency size, North American Industry Classification System (NAICS) classification[12], and geographic context. The ITA dataset offers establishment-level detail enabling trend analysis by agency characteristics. The findings aim to inform targeted policy responses and advance data-driven protections for transit workers across the country.

Our hypothesis is that OSHA-reportable rates of hearing loss, poisonings, respiratory disorders, skin disorders, days away from work, other illnesses, deaths, and total illness, differ significantly over time and by agency subgroup, particularly size and geographic classification. This will be evaluated through:

Aim 1: Assess temporal trends in illness rates across the study period (2016–2023).

Aim 1A: Identify statistically significant differences between individual years.

Aim 1B: Examine whether respiratory illness rates increased during the COVID-19 pandemic (2020–2021).

Aim 2: Compare illness rates across transit agency subgroups.

Aim 2A: Assess statistically significant differences in illness rates by establishment (agency) size classification.

Methods:

A retrospective descriptive study design was applied using OSHA's ITA data by NAICS codes 4851 for urban transit systems and 4852 for interurban and rural bus transportation and

establishment level information such as, total hours worked at the agency in the previous year, agency size, and geographic location. The ITA is composed of OSHA-reportable cases, which include any work-related fatality, loss of consciousness, days away from work, restricted duty, job transfer, or injuries and illnesses requiring medical treatment beyond first aid. Also included are diagnosed cases of cancer, chronic irreversible diseases, fractured bones or teeth, punctured eardrums, and cases meeting specific recording criteria such as needlestick injuries, medical removal, hearing loss, tuberculosis, and work-related Covid-19 reported as respiratory conditions.[13] Rates were calculated per full-time equivalent (FTE) employees to normalize across agency size based on hours worked. Variables such as geographic region, agency size, and transit classification based on NAICS codes for urban transit vs. interurban and rural bus transportation, were incorporated into the analysis. Analytical methods included Mann-Kendall trend tests to assess changes in illness and injury rates from 2016 to 2023. Year-to-year differences were evaluated using Kruskal-Wallis tests, and one-way ANOVA was used to compare rates across pandemic periods. Linear regression models were applied to examine associations between illness rates and agency-level characteristics such as size, region, and NAICS classification. A hierarchical clustering method was used to normalize and merge establishment records across years by grouping similar company names and addresses, improving consistency in longitudinal analysis.

Results:

OSHA-reportable illness and injury rates in U.S. transit agencies remained largely stable between 2016 and 2023, with no apparent trends observed. While respiratory conditions and poisonings showed increases in 2021 and 2022 respectively, these increases were not statistically significance. Only injury rates varied significantly by year, spiking in 2021, though these were not statistically significant after adjustment. Linear regression models, excluding state-level predictors, identified agency size was a modest but statistically significant predictor

of hearing loss (β = 0.0925, 95% CI [0.0667, 0.1183]), though the overall model fit was poor (adjusted R^2 = 0.026).

Conclusions:

This thesis presents a national, retrospective multi-year assessment of OSHA-reportable illness and injury trends among reporting U.S. transit employers from 2016 to 2023. While rates remained generally stable over the study period, the ability to predict illness and injury patterns based on agency-level characteristics captured in ITA data was limited. This limitation likely reflects the narrow scope of available variables, which exclude key factors such as worker demographics, job classifications, and exposure conditions. These limitations emphasize the need for improved occupational health surveillance in both urban and rural transit sectors. A more robust and integrated surveillance system would better equip health and safety practitioners to identify patterns early and implement preventive strategies. Future research should incorporate workers' compensation records, Federal Transit Administration (FTA) datasets, and proactive hazard monitoring, with a focus on high-risk roles and standardized reporting practices.

Acknowledgements

I would like to express my deepest gratitude to Dr. Marissa Baker, who not only served as the chair of my thesis committee but has also been a major influence on my educational journey and professional development. Her guidance, care, and knowledge have shaped my path in ways that extend far beyond the scope of this research. I am profoundly thankful for her continued support and mentorship, which directly contributed to my growth and the opportunity to transition into full-time work in the field.

I would also like to thank Dr. Marty Cohen for his insight, encouragement, and enthusiasm for occupational hygiene, especially during a time of uncertainty and transition within the United States. His ability to spark interest in the field and foster student engagement has left an impression on me.

Finally, I am sincerely grateful to Lucia Luu, who has been both a mentor and a close friend in the world of occupational hygiene beyond academia. Her guidance, integrity, and dedication to the field have set a powerful example of what it means to be a practicing Certified Industrial Hygienist (CIH). I could not have asked for a better role model as I began to define my own path in this profession.

Research reported in this thesis was supported by the National Institute for Occupational Safety and Health (NIOSH) under Federal Training Grant 2T42OH008433-16 and 5T42OH008433-17. The content is solely the responsibility of the author and does not necessarily represent the official views of NIOSH.

Research reported in this thesis was also supported by the Department of Environmental and Occupational Health Science (DEOHS) Construction Industrial Hygiene Research Fund scholarship provided by Don Garvey. The content is solely the responsibility of the author and does not necessarily represent the official views of the scholarship funders.

Introduction

Occupational health and safety within transit agencies is a present public health concern due to the diverse range of hazards faced by both urban and rural transit workers, which includes all direct employees of a transit agency. As of 2024 there were forty-four thousand three-hundred and ten urban transit systems workers[14] and fourteen thousand, seven-hundred and eighty interurban and rural bus transportation workers[15] in the USA working in jobs such as driving buses, operating light rail, operating subways, electromechanics, signals maintainers, custodial, administrators, etc... Transit workers experience noise-induced hearing loss[1], respiratory illness from airborne contaminants[2,4,5], musculoskeletal disorders (MSDs) from repetitive or static physical demands[3], and stress-induced cardiovascular conditions[6]. In addition to these recordable health outcomes, transit workers routinely encounter a range of psychosocial and environmental stressors that are often omitted from official surveillance systems like OSHA's ITA. These include, but are not limited to, workplace violence, verbal abuse, and confrontations with passengers.[16] A recent study also found detectable levels of fentanyl and methamphetamine from passenger use, which can impact operator and rider safety.[17] These risks, though well-documented, remain underprioritized in occupational health research and policy. The absence of comprehensive surveillance systems limited federal funding, and weak regulatory accountability have contributed to persistent gaps in intervention and oversight.[1,2,3,6.]

Noise-induced hearing loss is a documented concern for many high-noise occupations. A systematic review of 187 studies found that while occupational noise exposure can cause permanent auditory threshold shifts, the evidence for elevated risk among railway workers is mixed though, with some studies showing minimal or no increased loss relative to reference populations.[1] Respiratory illness is also a concern among transit workers, particularly due to occupational exposures to airborne pollutants in enclosed or high-traffic environments. Studies have documented elevated concentrations of fine particulate matter (PM2.5) in public

transportation settings such as buses and subways.[2] Inhalation of these particles has been associated with inflammation and oxidative stress.[18] Recent modeling studies have further shown that inadequate HVAC filtration or poor ventilation on transit vehicles can substantially elevate the relative risk of airborne disease transmission, especially in high-occupancy scenarios.[19]

MSDs are another concern among transit operators. A cross-sectional study of (n=254) metropolitan bus drivers showed nearly half of respondents (49.2%) reported musculoskeletal discomfort, with significantly higher prevalence of low back, hip, and knee pain compared to office staff.[3] These effects were significantly associated with risk factors such as Body Mass Index (BMI) and tobacco use, reinforcing the need for ergonomic interventions and evaluation of health promotion programs targeting drivers' behaviors.[3] Overall transit workers exposures drive chronic illnesses like hypertension, diabetes, and mental health disorders at rates higher than the general workforce. [6]

Compounding these challenges are stressors such as the COVID-19 pandemic. Transit workers, positioned on the front lines of public mobility, faced heightened viral exposure, often in confined, poorly ventilated environments without timely access to personal protection equipment (PPE).[20] These conditions exposed systemic failures in preparedness and workforce protection, adding urgency to calls for more resilient occupational health systems.

According to the National Academies through TCRP F-26 with Econometrica, Inc. and the Transportation Learning Center using Bureau of Labor Statistics (BLS) data, there are six chronic health conditions are more prevalent among transit workers compared to the general workforce, including musculoskeletal disorders, respiratory illness, cardiovascular disease, diabetes, hypertension, and adverse mental health.[6] Various studies have also noted transit workers suffer higher rates of hearing loss, cancer, and obesity compared to the general population.[21-25] Addressing these overlapping risks requires a systemic, data-driven, and worker-

informed approach to occupational health. Despite these elevated risks, transit-focused occupational health remains under-researched. An absence of comprehensive injury and illness surveillance, inadequate federal funding, and limited regulatory requirements and accountability have all contributed to this visibility gap.[6] Moreover, public perceptions that minimize the physical and psychological toll of transit work continue to fuel cycles of underinvestment in workforce protections.[6] Compounding these issues is the fragmented regulatory environment in which many transit agencies operate. Transit systems often span multiple jurisdictions and involve collaborations across local, state, and federal entities, as well as coordination with private contractors, infrastructure owners, and other transportation operators. This operational complexity can lead to inconsistent application of occupational health standards, unclear lines of accountability, and challenges in implementing uniform worker protections.

Public health surveillance is foundational to effective intervention, The Centers for Disease Control and Prevention (CDC) defines public health surveillance as, "The ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know."[26] The Committee on Developing a Smarter National Surveillance System for Occupational Safety and Health in the 21st Century outlines the importance of systems that extend beyond outcome tracking to include exposures, hazards, and social determinants of health.[7] Their vision calls for a "system of systems" that incorporates disaggregated data, improves coordination across agencies, and enhances the representativeness of surveillance for diverse worker populations, including those in nontraditional and high-risk occupations.[7] Yet, current occupational illness surveillance often stops short of capturing the full scope of risk. U.S. employers are required to record and report work-related injuries and illnesses that meet OSHA's criteria, such as those involving medical treatment beyond first aid, lost work time, or restricted duty. These data are submitted via Form

300A and compiled in the ITA, which offers establishment-level insight into OSHA-recordable outcomes. While the ITA supports surveillance of OSHA-recordable outcomes, it has notable limitations: it excludes non-reporting establishments, lacks worker-level detail, and may be affected by underreporting or inconsistent interpretation of what constitutes a reportable case. A 2002 study in the American Journal of Public Health outlined the many “filters” preventing recognition and reporting of work-related illness, ranging from employer incentives to suppress injury reports to physicians’ lack of training in occupational diagnosis.[8] These filters are especially obstructive for marginalized workers, who face greater structural barriers to diagnosis, treatment, and compensation.[8]

The Federal Transit Administration’s 2024 Transit Advisory Committee for Safety (TRACS) report further highlights fragmented reporting systems within transit agencies. Inconsistent definitions of assault, injury, or exposure across departments have resulted in incomplete and non-comparable data submissions to the National Transit Database.[9] TRACS has recommended unified definitions, mandatory reporting, and stronger transparency mechanisms to improve national surveillance and oversight.[9] These issues extend well beyond incident reporting, reflecting a deeper need for consistent occupational health metrics across the transit industry.[9] This need is echoed in recent analytic work using OSHA’s Severe Injury Reports dataset. A 2023 time series analysis of seven years of national data revealed meaningful seasonal and industry-specific injury trends, yet also identified limitations in coverage, data completeness, and standardization.[10] The authors emphasize that while administrative data can reveal national patterns, it remains subject to significant underreporting and definitional ambiguity.[10]

To explore these gaps, this thesis draws on data from the ITA, a national reporting system that captures annual counts of workplace injuries and illnesses submitted by employers to OSHA. These categories offer insight into chronic health outcomes relevant to the transit workforce:

Hearing Loss

Hearing loss among transit workers may develop due to prolonged exposure to engine noise, rail braking systems, or loud public environments. Operators, maintenance staff, and inspectors may work in settings where average noise levels exceed recommended thresholds.

Respiratory Conditions

Respiratory conditions may emerge from long-term exposure to diesel exhaust, brake dust, cleaning chemicals, or ambient urban pollutants. Enclosed or poorly ventilated environments are particularly susceptible.

Skin Disorders

Skin disorders among transit personnel may result from contact with solvents, degreasers, fuels, and other irritants. Maintenance and custodial staff are especially at risk. These exposures may trigger contact dermatitis or, in more severe cases, chemical burns or systemic allergic responses.

Poisonings

Occupational poisonings may occur through inhalation or dermal absorption of hazardous substances, often during maintenance or emergency response activities. Though less frequent, such events may involve accidental chemical mixing, inadequate ventilation, or improper personal protective equipment (PPE) use.

These categories provide a window into chronic occupational health outcomes that may otherwise remain underreported or overlooked, particularly in high-risk sectors like transit. This looks to see if there are differences in report injury and illness rates, with attention to both organizational and geographic context. Results from this analysis may inform targeted intervention efforts.

In response, this thesis presents a comparative, multi-year evaluation of OSHA-reportable illness and injury rates among transit agencies, using federal datasets and validated statistical methods. The results are intended to support informed policy, promote targeted investment in worker protections, elevate occupational health as a critical component of transit system resilience, and to expose challenges in existing data collection and surveillance methods.

Methods

Study Overview

This retrospective descriptive study analyzes OSHA-reportable cases, which include any work-related fatality, loss of consciousness, days away from work, restricted duty, job transfer, or injuries and illnesses requiring medical treatment beyond first aid. Also included are diagnosed cases of cancer, chronic irreversible diseases, fractured bones or teeth, punctured eardrums, and cases meeting specific recording criteria such as needlestick injuries, medical removal, hearing loss, tuberculosis, and work-related Covid-19 reported as respiratory conditions.,[13] illness and injury counts from 2016 to 2023 among transit agencies in the United States classified under NAICS codes 4851 (Urban Transit Systems) and 4852 (Interurban and Rural Bus Transportation). The analysis focused on evaluating changes in illness and injury rates over time, assessing differences between subgroups by establishment size (as per OSHA thresholds), NAICS classification (urban vs rural), and geographic region (FTA region and state), and modeling outcome variation using regression techniques. Eight outcomes were examined: hearing loss, respiratory conditions, skin disorders, poisonings, days away from work (DAFW) cases, fatalities, other injuries, and other illnesses. All outcomes were normalized by calculating full-time equivalents (FTE), per 10,000 FTE for deaths and per 100 FTE for all other outcomes. The unit of analysis was the agency-year, representing a single agency's reported data for one calendar year.

Data Sources

Publicly available data were sourced from the OSHA ITA and FTA National Transit Database (NTD). ITA data included the annual OSHA Form 300A counts of recordable injuries and illnesses reported by transit employers. This data provided both raw counts and total hours worked per establishment, allowing rates to be calculated. It is important to note that the ITA dataset does not represent a statistically sampled cross-section of all establishments within each NAICS code. Instead, it includes data only from establishments that meet specific reporting thresholds based on size and industry classification. As a result, the findings from this study reflect trends among reporting agencies and should not be generalized to the entire transit industry without caution. NTD data was used exclusively for cross-referencing agency names and addresses to support entity resolution and normalization of agency identification of ITA records across years. The study period spanned 2016 through 2023, encompassing all available reporting years at the time of analysis.

Statistical Software

All analyses were conducted in R (version 4.3.2) using the University of Washington School of Public Health's cloud-hosted RStudio environment. Key packages included *dplyr*[27], *tidyr*[28], *readr*[29], and *purrr*[30] for data wrangling; *ggplot2*[31], *gridExtra*[32], and *patchwork*[33] for data visualization; and *kableExtra*[34] and *knitr*[35] for tabular reporting. The *trend*[36] package was used for Mann-Kendall trend testing, while *stats*[37], *modelr*[38], and *broom*[39] supported regression modeling. The *stringdist*[40] package enabled hierarchical clustering of company names for record normalization, and *EnvStats*[41], *Hmisc*[42], and *MKmisc*[43] assisted with descriptive statistics and non-parametric tests. All scripts were written, executed, and version-controlled within this environment to support reproducibility.

Data Cleaning and Preparation

Prior to analysis, all ITA records were filtered to retain only those corresponding to NAICS codes 4851 (Urban Transit Systems) and 4852 (Interurban and Rural Bus Transportation). Establishments outside of these classifications or with missing NAICS identifiers were excluded. To maintain consistency across years, agency names were standardized using a normalization process based on employer address, state, and partial string matching business names. This approach addressed inconsistencies in reporting, such as minor name variations or address formatting discrepancies, which could fragment longitudinal records.

Where multiple entries existed for the same agency-year combination, records were aggregated by summing numerator fields (e.g., number of poisonings, number of DAFW cases) and taking the maximum value for denominators (e.g., FTEs, total hours worked) when discrepancies occurred. This approach was chosen to mitigate underestimation due to partial or incomplete submissions. However, this assumes that case records reflect distinct events while hours worked may be redundantly or sequentially reported. No sensitivity testing was done, such as using alternative aggregation strategies (e.g., average, or minimum denominators) to assess the impact on rate calculations. This aggregation approach also resolved inter-year discrepancies where the same agency may have been reported with slightly different NAICS codes or establishment types, preserving longitudinal integrity of each agency. Establishments with missing or zero FTE values were removed, as they could not be reliably used for rate-based calculations. Records with total hours worked equal to zero were excluded because they produce undefined or infinite rate calculations. Although these cases sometimes included nonzero employee counts or reported injuries, the absence of reported hours worked made it impossible to compute rates. The inclusion of such records would distort the analysis and overinflate illness or injury rates. All illness and injury counts were converted into rates per 100 FTE or 10,000 FTE employees using OSHA standard methodology.[44]

The NTD was used as a reference to confirm agency identities and support resolution of ambiguous or incomplete records. Linking between these two datasets was developed to align NTD agency names and locations with ITA entries where reporting inconsistencies prevented direct matching. While the NTD was not used analytically, its use was for validating and cleaning the ITA dataset prior to statistical testing.

Hierarchical clustering based on Jaro-Winkler string distance metrics[45] was used to group similar company names across reporting years. This reduced fragmentation due to reporting inconsistencies (e.g., minor spelling variations or address formatting), consolidating agency-level records into unified longitudinal profiles. No records were removed during this process; rather, establishments with similar names and identifiers (e.g., EIN, state) were grouped and aggregated to reduce duplication while preserving all original data points.

Table 1: Changes in Agency Record Counts Before and After Company Name Clustering, 2016–2023

Year	Original Rows	Merged Rows	Row Difference	Percent Combined
2016	351	173	178	49.29
2017	404	225	179	55.69
2018	675	333	342	49.33
2019	569	321	248	56.41
2020	567	277	290	48.85
2021	606	288	318	47.52
2022	473	261	212	55.18
2023	308	235	73	76.30

Analytic Approach

This study employed a multi-step analytical strategy to examine trends in OSHA-reportable illness and injury rates across required reporting U.S. transit agencies from 2016 through 2023. The agency-year served as the unit of analysis, enabling both longitudinal and cross-sectional comparisons while maintaining data independence. All illness and injury outcomes were

adjusted per 100 FTE employees to by the number of hours exposed except for deaths which was adjusted per 10,000 FTE.

Descriptive statistics were calculated for each outcome, including measures of central tendency (mean, median), dispersion (standard deviation, interquartile range), and frequencies of zero or missing values. To identify years with unusually high or low illness or injury rates, we computed standardized z-scores for the annual means of each outcome. Years were flagged as potential outliers if their annual mean rate for a given outcome deviated by more than 2 standard deviations from the multi-year average (i.e., z-score > |2|), consistent with conventional statistical thresholds. These outliers were retained for analysis and were used descriptively to identify anomalous reporting periods, not excluded. Normality was assessed using the Shapiro-Wilk test, and most outcome variables exhibited non-normal distributions. As a result, subsequent hypothesis testing relied on non-parametric methods (tests which don't assume a normal distribution) more appropriate for skewed or zero-inflated data. Temporal patterns were assessed using the Mann-Kendall trend test, a non-parametric rank-based method that detects monotonic trends (consistently moving up or down) without assuming linearity. The test was applied independently to each illness and injury outcome to identify significant directional shifts over the eight-year period. Year-over-year (YoY) percent changes were also calculated to examine short-term variability between consecutive years and identify spikes or declines that warranted further examination. Comparisons were calculated using unlagged denominators (i.e., using the current year's FTE as the base), with negative or undefined values retained where appropriate. Missing or zero values were preserved, and no imputation was performed.

Subgroup comparisons were conducted using the Kruskal-Wallis test to assess differences in median illness and injury rates by COVID-19 pandemic period (pre-pandemic: 2016–2019; pandemic: 2020–2021; post-pandemic: 2022–2023) and by agency size classification (small, medium, large) using ITA-defined thresholds.

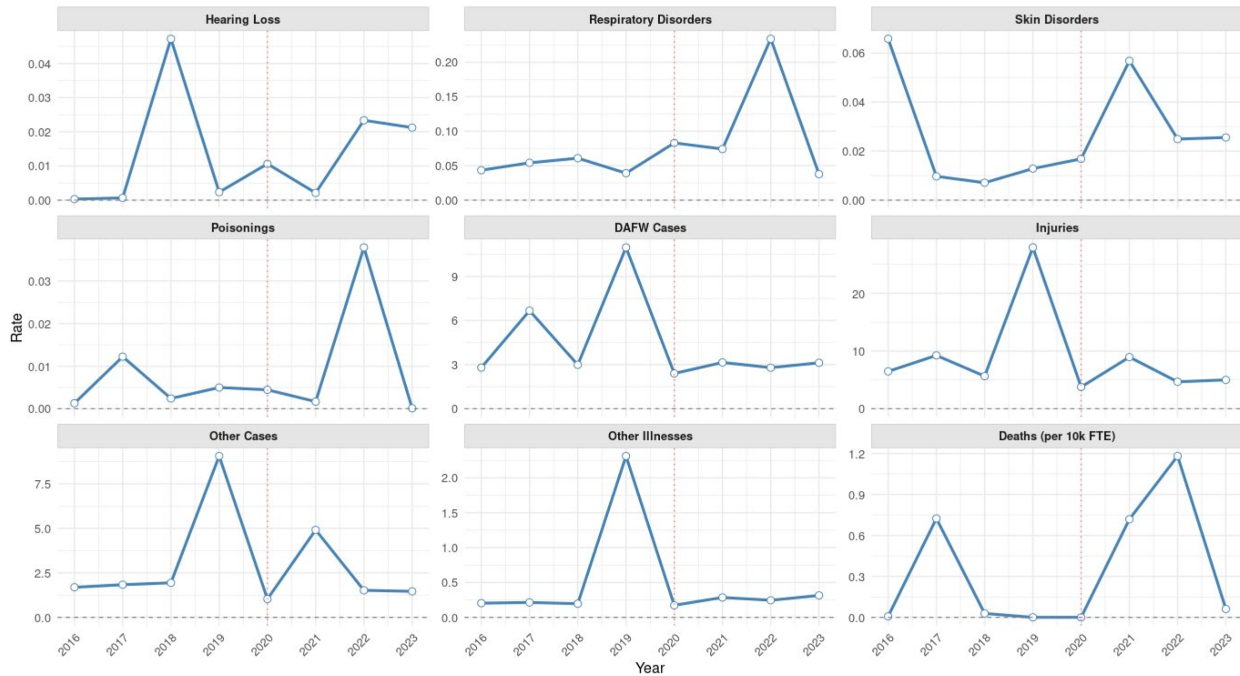
To evaluate the influence of agency-level predictors on outcome rates, multiple linear regression models were developed. Predictor variables were entered as main effects only; interaction terms were explored in sensitivity analysis but were excluded from final models due to overfitting and lack of significance. Models included predictors for agency size, NAICS classification, FTA region, establishment type, and state. Adjusted R^2 values were reported to assess model fit, and all predictors were retained regardless of statistical significance.

Results

After cleaning and consolidation, the final dataset included 2,113 agency-year records from approximately 145 transit agencies spanning 2016 to 2023, with each record representing a unique agency-year report. Hierarchical clustering was used to reduce fragmentation due to inconsistent naming, resulting in an average row retention of approximately 55% across years (Table 1). Each agency-year entry included OSHA-recordable outcomes standardized per 100 full-time equivalent (FTE) workers (or per 10,000 FTE for fatalities). These reports captured thousands of illness and injury events across the transit workforce, encompassing hearing loss, respiratory disorders, skin disorders, poisonings, days away from work (DAFW) cases, other injuries, other illnesses, and fatalities.

Although these outcomes were consistently reported, no statistically significant monotonic trends were detected over the study period. Time series plots (Figure 1) illustrate generally stable or weakly fluctuating rates for all categories. Among them, respiratory conditions showed a visible but non-significant rise during 2021–2022, likely corresponding to COVID-19-related disruptions. .

Figure 1. Annual OSHA-Reportable Illness and Injury Rates Among Reporting U.S. Transit Agencies,2016–2023



Injury and DAFW case rates remained consistently higher than other categories throughout the study period. Hearing loss, respiratory conditions, and other illnesses exhibited relatively low and stable rates. Spikes were observed in respiratory disorders in 2021 and poisonings in 2022, though most outcomes showed minimal fluctuation year-to-year. Deaths remained rare across all years and agencies. Years with zero or missing data are indicated with vertical dashed red lines.

Table 2. Annual OSHA-Reportable Illness and Injury Rates per 100 FTEs Among Reporting U.S. Transit Agencies, 2016–2023

	Rates per 100 FTE (Deaths per 10,000 FTE)								
year	Hearing Loss	Respiratory Disorders	Skin Disorders	Poisonings	DAFW Cases	Injuries	Other Cases	Other Illnesses	Deaths (per 10k FTE)
2016	<0.001	0.043	0.066	0.001	2.796	6.479	1.692	0.203	0.007
2017	<0.001	0.054	0.010	0.012	6.669	9.243	1.835	0.214	0.724
2018	0.047	0.061	0.007	0.002	2.987	5.645	1.940	0.196	0.029
2019	0.002	0.039	0.013	0.005	10.970	27.960	9.060	2.313	<0.001
2020	0.011	0.083	0.017	0.004	2.398	3.769	1.029	0.173	0.000
2021	0.002	0.074	0.057	0.002	3.148	8.952	4.910	0.285	0.718
2022	0.023	0.234	0.025	0.038	2.793	4.668	1.520	0.245	1.181
2023	0.021	0.038	0.026	<0.001	3.124	4.999	1.460	0.314	0.061

No illness or injury rates showed statistically significant monotonic trends over time based on Mann-Kendall tests (all $p \geq 0.05$), although hearing loss ($\tau = 0.43$) and respiratory conditions ($\tau =$

0.21) demonstrated weak upward tendencies. Sporadic spikes were observed in 2018 (hearing loss), 2021 (respiratory disorders), and 2022 (poisonings), but these fluctuations did not represent consistent or sustained directional changes. Year-over-year (YoY) percent changes varied widely across outcomes, ranging from −100% (e.g., full drop-off to zero) to +248,511% (e.g., emerging from near-zero baseline rates). Such variability is characteristic of rare events with small denominators. Notably, 2019 showed elevated injuries and DAFW rates at several large agencies, contributing to localized rate inflation. These results emphasize the difference between short-term volatility and long-term trend, as summarized in Table 3.

Table 3. Mann-Kendall Trend Results and Year-over-Year Fluctuations for OSHA-Reportable Outcomes, 2016-2023

	Observations	Mann-Kendall Trend		Largest YoY Increase (%)
Outcome	n	Kendall's τ	p-value	Max YoY Change (%)
Hearing Loss	25	0.43	0.17	208.47
Respiratory Disorders	141	0.21	0.54	6490.49
Skin Disorders	58	0.29	0.39	110.87
Poisonings	22	-0.07	0.90	37.50
DAFW Cases	1176	-0.07	0.90	248510.84
Injuries	1353	-0.29	0.39	330973.90
Other Cases	917	-0.14	0.71	107275.32
Other Illnesses	275	0.29	0.39	7289.34
Deaths (per 10k FTE)	12	0.14	0.71	-100.00

* Note: Extreme YoY changes often reflect small baseline values in rare outcomes (e.g., 0.001 to 0.01 = +900%). Trends were tested using Mann-Kendall; none were statistically significant at α = 0.05.

Kruskal-Wallis tests identified statistically significant variation in injury rates across individual years (p = 0.0355), with a prominent peak in 2021. DAFW cases approached significance (p = 0.0547), indicating possible year-level fluctuation. However, when grouped into pre-pandemic (2016–2019), pandemic (2020–2021), and post-pandemic (2022–2023) periods, one-way ANOVA found no statistically significant differences in mean illness or injury rates across these intervals. This suggests that, despite pandemic-era disruptions, average reported rates remained relatively stable over time. Detailed annual differences are presented in Table 4, with summary comparisons across periods shown in Table 5.

Table 4. Kruskal–Wallis Test Results for Yearly Differences in OSHA-Reportable Outcomes, 2016–2023

	Observations	Kruskal-Wallis Test Results
Outcome	n	Kruskal-Wallis p-value
Hearing Loss	25	0.8099
Respiratory Disorders	141	0.5399
Skin Disorders	58	0.6344
Poisonings	22	0.6386
DAFW Cases	1176	0.0550
Injuries	1353	0.0353
Other Cases	917	0.1549
Other Illnesses	275	0.2074
Deaths (per 10k FTE)	12	NaN

Among the nine outcome categories tested, only injury rates showed statistically significant differences by year ($p = 0.0355$), driven primarily by a 2021 spike. DAFW cases approached significance ($p = 0.0547$), suggesting possible temporal variability. All other outcomes, including hearing loss, respiratory disorders, skin disorders, and poisonings, demonstrated no significant year-over-year differences in median rates ($p \geq 0.1$). These results show that while most health outcomes remained relatively stable across the study period, injuries, and potentially DAFW cases, exhibited greater temporal fluctuation.

Table 5: ANOVA Results for Differences in OSHA-Reportable Illness and Injury Rates by Pandemic Period

	Observations	ANOVA by Pandemic Period
Outcome	n	ANOVA p
Hearing Loss	1968	0.7913
Respiratory Disorders	1968	0.0578
Skin Disorders	1968	0.5214
Poisonings	1968	0.2099
DAFW Cases	1968	0.4925
Injuries	1968	0.5927
Other Cases	1968	0.7558
Other Illnesses	1968	0.6886
Deaths (per 10k FTE)	1968	0.3327

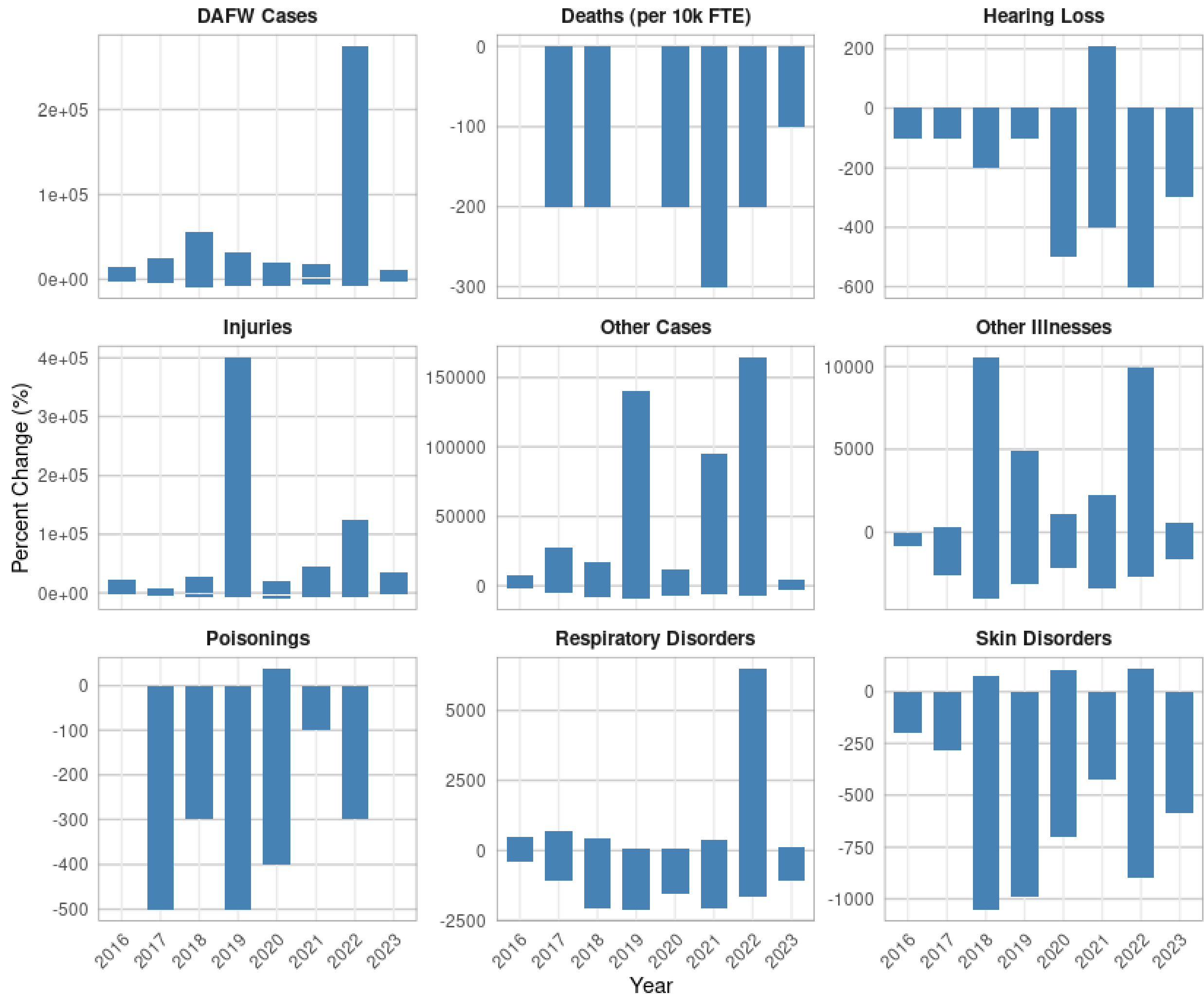
No outcome category demonstrated a statistically significant difference in mean rates across these periods (all $p \geq 0.05$). Respiratory conditions had the lowest p-value ($p = 0.0578$), approaching marginal significance but not meeting the threshold. All other conditions, including hearing loss, injuries, DAFW cases, and deaths, showed no meaningful variation in mean rates between time periods. These results suggest that average OSHA-reportable illness and injury rates among U.S. transit agencies did not differ substantially before, during, or after the COVID-19 pandemic.

Table 6: Summary of Linear Regression Results Predicting OSHA-Reportable Outcomes by Agency Characteristics

		Model Performance	Key Predictors
Outcome	n	Adjusted R^2	Significant Predictors ($p < 0.05$)
Hearing Loss	1968	0.019	size
Respiratory Disorders	1968	-0.015	
Skin Disorders	1968	-0.014	state
Poisonings	1968	-0.008	state
DAFW Cases	1968	0.006	state
Deaths (per 10k FTE)	1968	-0.016	
Injuries	1968	0.004	state
Other Cases	1968	0.003	state
Other Illnesses	1968	0.003	state

Across all outcomes, model fit was poor, with adjusted R^2 values ranging from approximately, 0.001 to 0.03, indicating that the included predictors explained little variance. Establishment size emerged as a statistically significant predictor of hearing loss rates ($p < 0.001$), though its practical effect was modest. A few models identified the NAICS code or FTA region as significant, but these associations were inconsistent and likely driven by administrative rather than exposure-related differences. State-level predictors were generally weak, although Indiana showed a recurring positive effect in several models. Overall, these regression models suggest that administrative classifications alone are insufficient to meaningfully predict illness or injury burden among transit agency workers.

Figure 2: Year-Over-Year Change in OSHA-Reportable Illness and Injury Rates, 2016–2023



Most outcomes showed modest or erratic YOY changes with no consistent direction. Notable spikes were observed in hearing loss, respiratory disorders, poisonings, injuries, and other cases, often driven by low prior-year values. These isolated surges are visually evident but do not reflect sustained trends. Due to prior data cleaning, no missing baseline values remained in the YOY dataset. Y-axis scales vary by outcome to preserve visibility across conditions with different incidence magnitudes.

Table 7: Outlier Years Detected for OSHA-Reportable Illness and Injury Rates

Outcome	Outlier Years
Hearing Loss	2018
Respiratory Disorders	2022
Skin Disorders	—
Poisonings	2022
DAFW Cases	2019
Injuries	2019
Other Cases	2019
Other Illnesses	2019
Deaths (per 10k FTE)	—

Across all illness and injury categories, only a few years emerged as statistical outliers. These spikes coincide with previously observed YOY volatility but do not reflect persistent trends or consistent directional change. The limited number of statistical outliers reinforces the broader display of stability seen across 2016–2023.

Discussion

This study provides a longitudinal assessment of OSHA-reportable occupational illnesses and injuries within the U.S. transit sector from 2016 to 2023. By using nationally reported ITA data across urban and rural transit agencies, the analysis explores patterns in workforce health and reporting practices. Rates remained largely stable over time, with some episodic fluctuations and only weak associations between outcomes and agency-level predictors such as size, region, and industry classification. These results suggest limited explanatory power in administrative variables alone and highlight the constraints of using self-reported data for public health surveillance. Prior work by Washington State's Safety and Health Assessment and Research for Prevention (SHARP) program has shown that workers' compensation claims and BLS surveillance data frequently diverge due to differences in reporting incentives, definitional criteria, and data completeness.[46] While ITA data offer broad national coverage, they share many limitations seen in other administrative datasets, including underreporting, inconsistent case classification, and lack of exposure context. In the absence of a more complete or

independently validated dataset, these issues likely obscure true burden estimates and impede timely intervention for at-risk worker populations.

Interpretation of Findings

This analysis did not identify statistically significant monotonic trends in these outcomes from 2016 to 2023. Additionally, reported illness and injury rates did not differ meaningfully by agency size, state, or year within the ITA dataset. These findings may reflect limitations in the dataset's ability to capture exposure variation rather than an absence of occupational risk. Respiratory conditions exhibited the most pronounced episodic increase, spiking in 2021, with smaller concurrent surges in poisonings and DAFW cases. These shifts temporally align with the COVID-19 pandemic and are consistent with the expected impacts of elevated exposure risk in enclosed, high-contact transit settings.[5] The rise in respiratory illnesses may reflect a combination of increased viral transmission, heightened symptom awareness, and expanded internal reporting during the pandemic. However, OSHA's reporting requirements did not presume all COVID-19 cases to be work-related, and employers were only required to record infections when there was objective evidence of workplace transmission. As a result, underreporting remains a plausible concern. By 2023, respiratory and other outcome rates had returned to pre-pandemic levels, and no sustained trends were detected across the study period. The absence of statistically significant trends should not be interpreted as a lack of occupational risk, particularly considering known reporting constraints and the episodic nature of rate spikes. This suggests that observed spikes were likely temporary disruptions to workforce health and operations rather than indications of long-term changes in occupational illness burden. While hearing loss rates also rose modestly in 2018 and 2022, no consistent pattern was found. These patterns were identified through outlier flagging using z-scores and were not sustained longitudinally. Additionally, hierarchical clustering revealed significant naming fragmentation that required normalization to enable valid trend comparison. The lack of

statistically significant trends may reflect actual stability in outcomes, but it may also point to the limitations of administrative datasets like OSHA ITA.

Agency size emerged as the most consistent predictor of hearing loss and injury rates in linear models, with larger systems associated with slightly elevated rates. However, interpretation of this relationship remains complex. While larger agencies operate more vehicles and facilities, which may increase opportunities for exposure, they may also have greater organizational capacity, more defined roles, and stricter task specialization, factors that could mitigate individual risk. The small effect sizes and low adjusted R^2 values (generally < 0.05) further suggest that agency size alone does not meaningfully explain variation in health outcomes, and that other unmeasured factors likely contribute to these patterns. Ultimately, the weak model fit across outcomes shows that structural agency variables like size, region, and NAICS classification explain little of the observed variation in reported rates.

The absence of strong state-level effects, despite known variation in OSHA enforcement capacity and the autonomy of state-plan states, was unexpected. While these effects were not seen broadly across all states, they suggest that localized factors may influence outcomes in ways not fully captured by national administrative data. Similarly, no statistically significant differences were detected in illness or injury rates across pre-, during-, and post-pandemic periods. However, modest increases were observed in respiratory conditions, poisoning, and DAFW cases during the pandemic years. Although these shifts lacked statistical robustness, they aligned with real-world disruptions faced by the transit workforce during COVID-19. In occupational health practice, decisions are often made based on imperfect or noisy data, and even non-significant changes may warrant precautionary planning or intervention, especially when they correspond to known risk environments or system stressors. Moreover, no clear regional clustering patterns or structural disruptions were observed over time, and pandemic-era

increases did not reach statistical significance in comparison across pre-, during-, and post-COVID periods.

Relevance to Occupational Health in Transit

Numerous conditions, such as latent disorders, chronic hazardous chemical exposures, and psychosocial impacts, are systematically underrepresented due to reporting thresholds, clinical ambiguity, or institutional disincentives to report. For example, respiratory illness from diesel exposure, heat-related illness, and musculoskeletal injuries from repetitive motion may not be recorded unless they result in days away from work or formal diagnosis. This reinforces long-standing concerns that OSHA-reportable data fail to reflect the full occupational burden and separation of work and nonwork factors. Current surveillance systems do not differentiate between frontline operational staff, support, and administrators' limiting ability to isolate high-risk worker subsets or evaluate intervention efficacy.

Implications for Policy and Practice

The findings of this study point to several practical implications for transit agency leadership, occupational health professionals, and policymakers. Although OSHA-reportable illness and injury rates among U.S. transit agencies remained largely stable between 2016 and 2023, this apparent steadiness may hide deeper issues in how workplace data are collected, reported, and interpreted. Many transit-specific health outcomes, especially those with gradual onset or ambiguous symptoms, may go unreported under current systems. To address this, agencies should prioritize more integrated and comprehensive surveillance practices by linking OSHA injury reports with internal incident logs, human resources records, and workers' compensation claims data. Such cross-linking can offer a fuller picture of workforce health and improve the identification of injury, illness, and risk factor patterns that single administrative datasets alone may fail to capture. At the national level, inconsistencies in reporting criteria and NAICS code

application hinder cross-agency and longitudinal comparisons. Policymakers and federal agencies should lead efforts to standardize classification frameworks and ensure more consistent reporting of occupational illnesses and injuries across transit environments.

In addition to strengthening surveillance systems, transit agencies must invest in targeted, exposure-specific monitoring strategies and adopt a more proactive and tailored approach to intervention. Routine assessment of environmental conditions, such as hazardous noise exposure, air quality indicators (e.g., carbon monoxide, $PM_{2.5}$, VOCs), and musculoskeletal risks, can help contextualize reported health outcomes and guide prevention efforts. These evaluations should focus on high-risk roles like operators, mechanics, security, and custodial staff, for whom routine exposure to physical, chemical, or environmental stressors is common. Simultaneously, agencies should foster a stronger culture of reporting injuries and illnesses by reinforcing non-punitive practices, expanding occupational health education, and encouraging early documentation of injuries, illnesses, hazardous conditions, and near-hit incidents through accessible and trusted reporting channels. Improving worker trust and engagement with safety systems may help address underreporting, though some cases, particularly illnesses that develop gradually or lack a clear exposure event, may go unrecognized entirely, highlighting the need for better exposure assessment and occupational health training alongside reporting reforms. Resource allocation should also reflect differentiated risk levels across job categories. For instance, hearing conservation initiatives may be most relevant for rail operators and heavy equipment users, while station personnel may require improved ventilation and respiratory protection. Taken together, these findings suggest that the lack of visible trends in reported illness and injury rates may reflect limitations in current data infrastructure and organizational practices, rather than a true decline in occupational risk. Advancing the health of transit workers will require proactive, role-specific interventions grounded in data that reflect real working conditions. An effective surveillance system should go beyond case counts to capture

contextual variables that help explain why incidents occur, including quality and enforcement of accident prevention plans, access to and training on respiratory protection, and availability of services like audiograms. If a system were developed at the agency level, it should collect data not only on outcomes but also on exposure sources, task type, duration of work, contributing environmental conditions, and whether symptoms were recognized and reported promptly. Surveillance systems are most impactful when they allow for grouping of similar cases, identification of root causes, and targeted prevention strategies, not just compliance tracking.

Contribution to the Field

This study explores a key gap in the literature by providing national, multi-year trend analysis on illness and injury rates among transit agency workers using metrics and a comprehensive dataset of OSHA-reportable outcomes. While previous studies have described occupational hazards in transit qualitatively, few have quantitatively assessed multi-year trends in reported health outcomes using national datasets. This study explores that gap while critically evaluating the limitations of administrative reporting systems.

This work also contributes to the discourse on occupational justice by examining a frequently overlooked public-sector workforce. Transit workers are critical infrastructure employees whose health is intertwined with environmental justice, urban planning, and public mobility systems. By documenting the limitations and gaps in current surveillance, this study provides a foundation for future mixed-method, exposure-based, or intervention studies within this sector.

Future Research Efforts

To advance characterization and prevention of occupational illness and injury in the transit sector, future research must extend beyond OSHA's ITA data. While ITA provides valuable insight into reportable events, its limitations, including underreporting, data coherency, narrow case definitions, and a lack of exposure detail, restrict its utility in identifying causal patterns or

emerging risks. The absence of robust associations in this study's regression models further emphasizes the need for more comprehensive data integration to contextualize incident trends.

Incorporating workers' compensation data into occupational health surveillance could allow for enhanced analytic depth. Compensation records often capture cases missed by OSHA reports, including chronic or cumulative conditions that evolve outside narrow regulatory definitions. Linking these data with OSHA logs could help validate illness and injury reports, improve case detection sensitivity, and facilitate root-cause analysis by connecting diagnoses to job roles and incident circumstances. For example, researchers linked workers' compensation claims with SOII records in Washington State to identify injuries that were eligible but unreported in federal logs. By following up with employers, they revealed that non-reporting often stemmed from misunderstanding recordkeeping rules, misclassification of workers, or communication breakdowns, demonstrating how integrated data sources can expose gaps in reporting practices and inform system-level prevention strategies.[47]

Additionally, integration with datasets maintained by the FTA, particularly NTD, offers unique potential. NTD provides detailed operational characteristics, ridership volumes, service area demographics, and safety incident data across agencies. By merging ITA and NTD data, researchers could examine whether factors such as vehicle miles traveled, mode type, service density, and fare enforcement policies influence occupational health outcomes. The 2024 TRACS report emphasizes the need for consistent, accurate data reporting across transit systems, particularly regarding assault incidents, which are often inconsistently coded or omitted altogether.[9] These recommendations could serve as handrail for improving occupational health surveillance in the broader transit context.

A future analytic framework should also include leading indicators such as hazard, and exposure surveillance. As emphasized by the National Academies in 2018, hazard surveillance systems, like noise monitoring, air sampling, and ergonomic assessments, offer proactive

insights into potential risks before injuries or illnesses occur.[7] Similarly, injury surveillance can be made proactive by capturing near-hit events, early symptom reports, and first aid-only incidents. Regular collection of exposure metrics such as diesel exhaust concentrations, whole-body vibration measurements, and shift duration logs would enable researchers to move beyond lagging indicators and construct more predictive models of occupational health. Greater attention should also be paid to psychosocial risk factors and workforce demographics. Surveillance systems must be capable of capturing the effects of shift work, public interactions, job control, and mental health exposures. Collecting disaggregated data by job type, age, gender, race, and tenure would enable a more equitable analysis of risk and ensure that prevention strategies are appropriately targeted.

As national and regional efforts to modernize public transportation continue, investments in workforce safety must be prioritized alongside infrastructure upgrades. Researchers, transit agencies, and regulatory bodies must collaborate to design integrated, exposure-based surveillance systems that reflect the complexity of transit work. Doing so would not only improve analytic precision but also support evidence-based interventions that address the root causes of occupational illness and injury in this critical public service sector. Michaels and Wagner describe how establishment-specific OSHA data sets, such as those available through the Injury Tracking Application, can be cross-linked with external datasets like workers' compensation claims to explore causal relationships and improve prevention targeting. This approach enables researchers to examine establishments doing similar work but with differing injury rates, potentially revealing underlying organizational or environmental factors contributing to risk.[48]

Limitations

Despite the methodological framework and datasets utilized, this study faces several limitations that must be acknowledged to contextualize its findings appropriately. A primary limitation is the reliance on the ITA as a central data source. While the ITA provides invaluable nationwide data

on reportable workplace injuries and illnesses, its accuracy is contingent on the integrity of employer self-reporting. As noted by Michaels and Wagner, multiple studies commissioned by the Bureau of Labor Statistics have shown that a significant proportion of work injuries, ranging from 30% to more than 60%, were not included in employer reports, with underreporting varying by employer and type of injury.[48] These discrepancies highlight the potential for substantial surveillance blind spots, particularly for illnesses that are chronic, cumulative, or more difficult to attribute directly to workplace exposures.[48] Many transit agencies may underreport incidents, not necessarily out of intent, but due to factors such as misclassification of workers, limited administrative capacity, or confusion over reporting requirements. These challenges can lead to incomplete OSHA logs and underestimation of true injury and illness rates, particularly for conditions that are cumulative or not tied to a single identifiable event.[47] Additionally, the ITA dataset may not fully capture all relevant health outcomes. While the ITA includes any illness or injury deemed work-related and meeting OSHA-recordable criteria, including chronic conditions such as cancer or respiratory disease, many occupational illnesses may be underreported due to the difficulty of establishing a clear link to the workplace. Conditions with long latency periods or subclinical onset, as well as mental health outcomes, may go unrecognized or unrecorded, leading to underrepresentation of the true occupational health burden in the transit sector.

Another critical limitation lies in the variability of OSHA plan implementation across states. Jurisdictions operating under state OSHA plans may demonstrate higher injury and illness reporting due to more rigorous enforcement, targeted audits, or stronger reporting cultures. In contrast, states under federal OSHA oversight may exhibit artificially low rates stemming from inconsistent compliance or limited administrative oversight. This systemic variability complicates inter-state comparisons and may obscure real differences in workplace risk. For instance, Washington, an OSHA state-plan state, consistently ranks among those with the highest reported workplace injury rates in the nation, a trend likely influenced by stronger enforcement

and better reporting infrastructure rather than worse actual safety conditions.[49] As a result, occupational fatality data, less subject to underreporting, are often considered more reliable indicators of state-level safety performance than injury or illness rates.

Another key limitation is the absence of structured incident descriptors such as those provided by the Occupational Injury and Illness Classification System (OIICS) used in many workers' compensation datasets. The ITA does not include standardized codes to describe the nature, body part, source, or event leading to each injury or illness, limiting the ability to conduct detailed mechanistic analyses or compare datasets that do include such granularity. As a result, this analysis focuses on broad outcome categories (e.g., hearing loss, poisonings) without insight into causal pathways, contributing hazards, or physical mechanisms, reducing opportunities for targeted intervention planning.

The analysis assumes the accuracy and consistency of NAICS codes across years and establishments. Misclassification or changes in agency operations that are not reflected in NAICS designations can affect the identification of transit-related establishments. Similarly, while OSHA provides formal definitions for what constitutes a reportable illness or injury, variations in how transit agencies interpret and apply these definitions in practice can introduce inconsistencies in reporting.

Another methodological limitation includes potential errors or gaps introduced during the data merging and cleaning process. While best practices were employed, including normalization, clustering of establishment names, and rate calculations, these steps introduce inherent assumptions that may affect comparability and trend accuracy. Any misalignment between linked datasets or anomalies in hours worked reporting could skew illness rate estimates.

Lastly, this study's retrospective and observational design precludes any claims of causality. The findings demonstrate associations and trends but cannot account for confounding factors such

as the presence or absence of health and safety programs, socioeconomic differences among workers, local air quality. Future work incorporating qualitative data, direct workplace assessments, or longitudinal tracking would be necessary to deepen and validate these insights.

While these limitations do not diminish this study's findings, they emphasize the importance of cautious interpretation and the need for improved national data collection, consistent regulatory practices, and expanded research on transit worker occupational health.

Conclusion

This thesis provides a national, multi-year assessment of OSHA-reportable illness and injury trends among U.S. transit agency workers, addressing a gap in occupational health surveillance for this essential workforce. While the analysis revealed generally stable rates from 2016 to 2023, spikes, particularly during the COVID-19 pandemic, suggest localized disruptions, short-term disruptions, or changes in reporting, rather than systemic trends. Inconsistencies in illness classification, state-level reporting enforcement, and the absence of disaggregated worker data limit the interpretability and actionability of these findings.

To advance occupational health protections in public transportation, agencies and policymakers must expand their focus beyond compliance-based reporting and invest in integrated, exposure-informed surveillance systems. Future research should incorporate workers' compensation records, FTA datasets, and proactive hazard monitoring to develop a more complete picture of risk. Particular attention should be given to high-risk job categories and psychosocial stressors, which are often overlooked in traditional surveillance. Standardizing definitions and improving data transparency, consistent with recent recommendations by TRACS, will enhance comparability across agencies and time. Ultimately, protecting transit worker health requires

shifting from reactive reporting to proactive prevention, grounded in comprehensive, reliable data and informed by the realities of transit work environments.

Data & Code Availability

All datasets and R code used for data cleaning, analysis, and figure generation in this study will be made publicly available via GitHub at the time of the thesis submission. The repository will include annotated R scripts and supporting documentation to facilitate reproducibility. Raw OSHA ITA data and NTD reference data are publicly accessible through their respective official websites. The GitHub repository link is here.

AI Assistance Disclosure

This thesis was completed with the aid of OpenAI's ChatGPT (GPT-4), a language model trained on publicly available and licensed datasets. The tool was used exclusively to support technical and editorial tasks, including assistance with R code development, debugging, and simplification of technical jargon.

At no point did the AI generate or interpret original data independently. All analytical decisions, data processing steps, and final interpretations were made solely by the author. All AI-generated suggestions were critically reviewed, modified as needed, and validated by the author to ensure methodological accuracy and academic integrity. The use of AI in this project served to streamline code optimization and facilitate more efficient communication of technical results but did not replace any substantive scientific reasoning or human judgment.

The author assumes full responsibility for the accuracy, originality, and scholarly rigor of all content presented in this thesis.

Citations

1. Lie, A., Skogstad, M., Johannessen, H. A., Tynes, T., Mehlum, I. S., Nordby, K. C., ... & Moen, B. E. (2016). Occupational noise exposure and hearing: a systematic review. *International Archives of Occupational and Environmental Health*, 89(3), 351–372. https://doi.org/10.1007/s00420-015-1083-5
2. Nieuwenhuijsen, M. J., Gómez-Perales, J. E., & Colvile, R. N. (2007). Levels of particulate air pollution, its elemental composition, determinants and health effects in metro systems. *Atmospheric Environment*, 41(37), 7995–8006. https://doi.org/10.1016/j.atmosenv.2007.08.002
3. Sheth, A., Rao, A., Amin, A., & Kataria, A. (2023). Prevalence of work-related musculoskeletal disorders (WRMSDs) and its association with modifiable risk factors in metropolitan bus transit drivers: A cross-sectional comparison. *Journal of Family Medicine and Primary Care*, 12(8), 1673–1678. https://doi.org/10.4103/jfmpc.jfmpc_532_23
4. Lawin, H., Ayi Fanou, L., Hinson, A. V., Stolbrink, M., Houngbegnon, P., Kedote, N. M., ... & Mortimer, K. (2018). Health risks associated with occupational exposure to ambient air pollution in commercial drivers: A systematic review. *International Journal of Environmental Research and Public Health*, 15(9), 2039. https://doi.org/10.3390/ijerph15092039
5. Cummings, B. E., Haas, C. N., Lo, L. J., Sales, C. M., Fox, J., & Waring, M. S. (2024). Airborne disease transmission risks on public transit buses: Impacts of ridership, duration, and mechanical filtration using a relative risk metric. *Building and Environment*, 255, 111303. https://doi.org/10.1016/j.buildenv.2024.111303
6. National Academies of Sciences, Engineering, and Medicine. (2020). *Improving the health and safety of transit workers with corresponding impacts on the bottom line*. The National Academies Press. https://doi.org/10.17226/26022
7. National Academies of Sciences, Engineering, and Medicine. (2018). *A smarter national surveillance system for occupational safety and health in the 21st century*. The National Academies Press. https://doi.org/10.17226/24835
8. Azaroff, L. S., Levenstein, C., & Wegman, D. H. (2002). Occupational injury and illness surveillance: Conceptual filters explain underreporting. *American Journal of Public Health*, 92(9), 1421–1429. https://doi.org/10.2105/AJPH.92.9.1421
9. Federal Transit Administration, Transit Advisory Committee for Safety (TRACS). (2024). *Recommendations to improve data collection, analysis, and reporting to improve transit safety*. https://www.transit.dot.gov/sites/fta.dot.gov/files/2024-07/ARWS-TRACS-Final-Report.pdf
10. Gomes, H., Parasram, V., Collins, J., & Socias-Morales, C. (2023). Time series, seasonality and trend evaluation of 7 years (2015–2021) of OSHA severe injury data. *Journal of Safety Research*, 86, 235–244. https://doi.org/10.1016/j.jsr.2023.06.005
11. Occupational Safety and Health Administration (OSHA). (2025). *Injury Tracking Application (ITA)*. https://www.osha.gov/injuryreporting/ita
12. U.S. Census Bureau. (2022). *North American Industry Classification System (NAICS)*. https://www.census.gov/naics/

13. Occupational Safety and Health Administration (OSHA). (2024). *29 CFR §1904.7 – General recording criteria*. https://www.ecfr.gov/current/title-29/subtitle-B/chapter-XVII/part-1904/subpart-C/section-1904.7
14. U.S. Bureau of Labor Statistics. (2024a). *Occupational employment and wage statistics: Urban transit systems (NAICS 485100)*. https://data.bls.gov/oes/#/industry/485100
15. U.S. Bureau of Labor Statistics. (2024b). *Occupational employment and wage statistics: Interurban and rural bus transportation (NAICS 485200)*. https://data.bls.gov/oes/#/industry/485200
16. International Labour Organization. (2003). *Working paper: Sectoral activities programme (WP.205)*. Geneva: ILO
17. Beaudreau, M., Srikanth, P., Zuidema, C., Cohen, M. A., Seto, E., Simpson, C. D., & Baker, M. G. (2025). Assessing fentanyl and methamphetamine in air and on surfaces of transit vehicles. *Journal of Occupational and Environmental Hygiene*, 22(4), 300–310. https://doi.org/10.1080/15459624.2024.2444430
18. Lawin, H., Ayi Fanou, L., Hinson, A. V., Stolbrink, M., Houngbegnon, P., Kedote, N. M., ... & Mortimer, K. (2018). Health risks associated with occupational exposure to ambient air pollution in commercial drivers: A systematic review. *International Journal of Environmental Research and Public Health*, 15(9), 2039. https://doi.org/10.3390/ijerph15092039
19. Cummings, B. E., Haas, C. N., Lo, L. J., Sales, C. M., Fox, J., & Waring, M. S. (2024). Airborne disease transmission risks on public transit buses: Impacts of ridership, duration, and mechanical filtration using a relative risk metric. *Building and Environment*, 255, 111303. https://doi.org/10.1016/j.buildenv.2024.111303
20. Federal Transit Administration. (2024). *Effects of the COVID-19 pandemic on transit ridership and accessibility* (FTA Report No. 0268). U.S. Department of Transportation. https://www.transit.dot.gov/sites/fta.dot.gov/files/2024-08/FTA-Report-0268-Effects-of-the-COVID-19-Pandemic-on-Transit-Ridership-and-Accessibility.pdf
21. Hansen, J., Raaschou-Nielsen, O., & Olsen, J. H. (1998). Increased risk of lung cancer among different types of professional drivers in Denmark. *Occupational and Environmental Medicine*, 55(2), 115–118. https://doi.org/10.1136/oem.55.2.115
22. Morikawa, Y., Nakagawa, H., Ishizaki, M., Tabata, M., Nishijo, M., Miura, K., ... & Kido, T. (1997). Ten-year follow-up study on the relation between the development of non-insulin-dependent diabetes mellitus and occupation. *American Journal of Industrial Medicine*, 31(1), 80–84. https://doi.org/10.1002/(sici)1097-0274(199701)31:1<80::aid-ajim12>3.0.co;2-3
23. Centre for Exploitation of Science and Technology. (1993). *The UK environmental foresight project: The future road transport noise agenda in the UK* (Vol. 3). H.M. Stationery Office.
24. Gubéran, E., Usel, M., Raymond, L., Bolay, J., Fioretta, G., & Puissant, J. (1992). Increased risk for lung cancer and for cancer of the gastrointestinal tract among Geneva professional drivers. *Occupational and Environmental Medicine*, 49(5), 337–344. https://doi.org/10.1136/oem.49.5.337
25. Siemiatycki, J., Gerin, M., Stewart, P., Nadon, L., Dewar, R., & Richardson, L. (1988). Associations between several sites of cancer and ten types of exhaust and combustion

products: Results from a case-referent study in Montreal. *Scandinavian Journal of Work, Environment & Health*, 14(2), 79–90. https://doi.org/10.5271/sjweh.1949
26. Centers for Disease Control and Prevention (CDC). (2022). *Introduction to public health surveillance*. https://www.cdc.gov/publichealth101/surveillance.html
27. Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* (Version 1.1.4) [R package]. https://dplyr.tidyverse.org
28. Wickham, H., Vaughan, D., & Girlich, M. (2024). *tidyr: Tidy messy data* (Version 1.3.1) [R package]. https://tidyr.tidyverse.org
29. Wickham, H., Hester, J., & Bryan, J. (2024). *readr: Read rectangular text data* (Version 2.1.5) [R package]. https://readr.tidyverse.org
30. Wickham, H., & Henry, L. (2025). *purrr: Functional programming tools* (Version 1.0.4) [R package]. https://purrr.tidyverse.org
31. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org
32. Auguie, B., & Antonov, A. (2017). *gridExtra: Miscellaneous functions for “grid” graphics* (Version 2.3) [R package]. https://CRAN.R-project.org/package=gridExtra
33. Pedersen, T. (2025). *patchwork: The composer of plots* (Version 1.3.0.9000) [R package]. https://patchwork.data-imaginist.com
34. Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax* (Version 1.3.4) [R package]. https://CRAN.R-project.org/package=kableExtra
35. Xie, Y. (2021). *knitr: A general-purpose package for dynamic report generation in R*. https://yihui.org/knitr/
36. Pohlert, T. (2023). *trend: Non-parametric trend tests and change-point detection* (Version 1.1.6) [R package]. https://CRAN.R-project.org/package=trend
37. R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
38. Wickham, H. (2023). *modelr: Modelling functions that work with the pipe* (Version 0.1.11) [R package]. https://modelr.tidyverse.org
39. Robinson, D., Hayes, A., & Couch, S. (2025). *broom: Convert statistical objects into tidy tibbles* (Version 1.0.8) [R package]. https://broom.tidymodels.org
40. Van der Loo, M. (2014). The stringdist package for approximate string matching. *The R Journal*, 6(1), 111–122. https://CRAN.R-project.org/package=stringdist
41. Millard, S. P. (2013). *EnvStats: An R package for environmental statistics*. Springer. https://www.springer.com
42. Harrell, F. E. Jr., & Dupont, C. (2024). *Hmisc: Harrell miscellaneous* (Version 5.1–1) [R package]. https://CRAN.R-project.org/package=Hmisc
43. Korthauer, K. (2021). *MKmisc: Miscellaneous functions from M. Kohl* (Version 1.8) [R package]. https://CRAN.R-project.org/package=MKmisc
44. Occupational Safety and Health Administration (OSHA). (2016). Interpretation of OSHA’s recordkeeping rule: Calculating rates per 100 full-time equivalent workers. https://www.osha.gov/laws-regs/standardinterpretations/2016-08-23
45. Winkler, W. E. (1990). String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

46. Wuellner, S. E., Bonauto, D. K., & Anderson, N. J. (2016). Unreported workers' compensation claims to the BLS survey of occupational injuries and illnesses: Establishment factors. *American Journal of Industrial Medicine*, 59(4), 274–289. https://doi.org/10.1002/ajim.22563
47. Rappin, C. L., Wuellner, S. E., & Bonauto, D. K. (2016). Employer reasons for failing to report eligible workers' compensation claims in the BLS survey of occupational injuries and illnesses. *American Journal of Industrial Medicine*, 59(4), 343–356. https://doi.org/10.1002/ajim.22582
48. Michaels, D., & Wagner, G. R. (2025). OSHA injury data: An opportunity for improving work injury prevention. *American Journal of Public Health*, 115(4), 588–595. https://doi.org/10.2105/AJPH.2024.307934
49. The Spokesman-Review. (2024, April 10). WA has one of the highest workplace injury rates in the U.S. Here's why. https://www.spokesman.com/stories/2024/apr/10/wa-has-one-of-the-highest-workplace-injury-rates-i/