Injury Tracking Application (ITA) Data Users Guide

Last Updated: September 25, 2024

Table of Contents

Who Should Use This Guide

This guide is intended to assist anyone using OSHA's Injury Tracking Application (ITA) data to analyze workplace injuries and illnesses. The ITA is OSHA's data collection system for establishments that are required to electronically submit certain information to OSHA once a year about recordable injuries and illnesses entered on their OSHA Form 300A Summary of Work-Related Injuries and Illnesses (Form 300A), Form 300 Log of Work Related Injuries and Illnesses (Form 300 or OSHA Log), and Form 301 Injury and Illness Incident Report (Form 301), as required by OSHA's recordkeeping regulation 29 CFR 1904. This guide covers ITA data available on the OSHA ITA Data webpage, which includes ITA datasets at the establishment level (Summary Data) and incident level (Case Detail Data). This guide also provides supporting information for data users and examples of possible uses of ITA data.

What is Not in This Guide

This document is not intended to provide guidance for establishments on how to electronically submit data to the ITA, or information about how establishments comply with OSHA recordkeeping requirements. For information about submitting data to ITA, visit OSHA's Injury Tracking Application page. For information about recordkeeping, visit OSHA's Injury and Illness Recordkeeping and Reporting Requirements page.

How ITA Data Is Collected

Scope of Data Collected

OSHA covers most private sector workers in all 50 states, the District of Columbia, and other United States territories and jurisdictions either directly through OSHA or through an OSHA-approved State Plan. Workers at state and local government agencies are not covered by OSHA, but have occupational safety and health protections if they work in states with a State Plan. Most states with State Plans have ITA reporting requirements identical to OSHA. When state and local government workers are covered by a State Plan, that State Plan dictates which establishments must report recordkeeping data to OSHA through the ITA. ITA data includes establishments covered by OSHA and establishments covered by State Plans.

OSHA collects data recorded on the Form 300A from (1) all establishments with 250 or more employees in industries required to keep OSHA injury and illness records (i.e., not on the list of Exempt Industries in 29 CFR 1904 Subpart B, Appendix A) and from (2) establishments with 20-249 employees listed in the 300A Table (29 CFR 1904 Subpart E, Appendix A).

OSHA collects data recorded on the Form 300 and Form 301 from establishments with 100 or more employees in industries listed in 29 CFR 1904 Subpart E, Appendix B. These industries meet the following criteria described in the Preamble to the Final Rule for Improve Tracking of Workplace Injuries and Illnesses:

- Average total case rate of at least 3.5 per 100 FTE (2017-2019 BLS SOII), or
- Average Days Away, Restriction, and Transfer (DART) rate of at least 2.25 per 100 FTE (2017-2019 BLS SOII), or
- Fatality rate of at least 5.7 per 100,000 FTE (2019 BLS CFOI), and

- In an industry in which establishments with 20-249 employees have been required to electronically submit their Form 300A Annual Summary, since 2017).

OSHA does not remove data submitted by establishments that were not required to submit data.

Recording or reporting a work-related injury, illness, or fatality does not mean that the employer or employee was at fault, that an OSHA rule has been violated, or that the employee is eligible for workers' compensation or other benefits.

Time Frame

Summary Data from establishments' submission of OSHA Form 300A information were first collected for incidents recorded during calendar year 2016, and Case Detail Data from establishments' submission of OSHA Forms 300 and 301 information were first collected for incidents recording during calendar year 2023. Employers covered by OSHA's electronic reporting requirements are required to submit data between January 2 and March 2 of each year for incidents recorded in the previous calendar year. Employers may also submit data after the deadline for timely submission. OSHA periodically updates data posted on the ITA website to include newly-received data.

How ITA Data is Shared and Organized

OSHA provides ITA data in a Comma Separated Values (CSV) file format. Before 2024, there was one CSV file for each year. This file contained Summary Data from the Form 300A. Starting in 2024, OSHA provides two CSV files: one with Summary Data from the Form 300A, and one with Case Detail Data from the Form 300 and Form 301.

Summary Data

Summary Data is organized by establishment. Each establishment has one record each year. Data reported are totals for the establishment. For additional information about what variables are collected and their descriptions, see the Summary Data Dictionary.

Case Detail Data

Case Detail Data is organized by incident. Each recorded work-related injury or illness is one record. Data reported include free form text description of the injury or illness that occurred, the occupation of the injured worker, the result of the incident, and a checkbox for the type of incident (injury, skin disorder, respiratory condition, poisoning, hearing loss, all other illness), and the time of the incident. OSHA intends to make free-form text data publicly available after using automated information technology and some manual review to detect and remove information that could reasonably be expected to identify an individual directly. The Case Detail Data also includes some data from the summary data file, such as establishment name, address, and establishment size. This file also includes the total number of hours worked for all employees at the establishment and the annual average employees working at the establishment. Note that these two fields should not be summed across all cases reported from the same establishment since they reflect summary data for the entire establishment, rather than incident-specific data. For additional information about what variables are collected and their descriptions, see the Case Detail Data Dictionary.

Linking Summary Data to Case Detail Data

Establishment information from the Summary Data can be linked to the Case Detail Data using the field establishment_ID. When ITA users create a new establishment in the ITA, the system assigns a unique establishment_ID to the establishment. However, some records may not link between the Summary Data and the Case Detail Data for any of the reasons below:

1) The establishment is required to submit Summary Data but is not required to submit Case Detail Data.
2) The establishment is required to submit Case Detail Data but had no recordable injuries or illnesses.
3) The establishment is required to submit Case Detail Data but did not submit required data.
4) The establishment submitted Case Detail Data to the ITA but did not submit Summary Data.

In addition, if an ITA user creates a new establishment profile for the same establishment in a different year, OSHA will assign a new establishment_ID. For this reason, the establishment_ID may not be a reliable indicator of a single establishment over time.

Data Cleaning and Coding

Data Cleaning

Several checks are built into the ITA system to improve the quality of the collected data, including validating dates and requiring totals to equal the sum of injuries and illnesses reported in the Summary Data. OSHA also performs several data cleaning steps on submitted data. For example, OSHA removes test data and duplicate records where identified, and the most recent record is kept when an establishment submits a data update. As another example, the ITA contains checks to ensure that the number of reportable injuries and illnesses is consistent within the Form 300A, and that Form 300A totals are consistent with the number of incidents reported on the establishment's Forms 300 and 301, if they are required to report.

For most fields, OSHA reports data as it is submitted by establishments. While OSHA takes steps to help ensure the data collected are accurate, problems and errors invariably exist for some establishments. Efforts are made during the collection cycle to correct submission errors; however, there may be some errors or inaccuracies in the public facing database for some establishments. For example, addresses are not standardized and company names are not validated. The ITA does not prevent establishments from submitting data when they are not required to do so. If an establishment submits an inaccurate NAICS code, incomplete injury description, or inaccurate address, OSHA does not edit the data. OSHA also does not validate the counts of workers, hours, or injury and illness counts reported by establishments. Each year, OSHA performs some data quality outreach to establishments that report implausible or incorrect data. However, concluding that establishments are the "most dangerous" or the "least dangerous" solely based on whether they have the highest or lowest rates from these data would be inappropriate.

Protecting Personally Identifiable Information (PII)

OSHA is taking multiple steps to prevent disclosure of personally identifiable information (PII) submitted to the ITA. OSHA is not collecting employee names or addresses. The OSHA

recordkeeping forms include warnings to employers not to include PII (e.g., employee names, phone numbers, social security numbers) in the free form text descriptions of the incident (Questions 14-17 on Form 301). OSHA also includes reminders on the ITA website to not electronically submit information that could reasonably be expected to identify individuals directly.

In addition, OSHA manually reviewed and removed such information from the job description field. OSHA intends to use automated information technology and some manual review to detect and remove information that could reasonably be expected to identify individuals directly from the other open text fields before making that data publicly available. These fields are:

- Where the event occurred (Form 300)
- Description of the injury or illness (Form 300)
- What was the employee doing just before the injury or illness (Form 301)
- What happened? Tell us how the injury occurred (Form 301)
- What was the injury or illness? (Form 301)
- What object or substance directly injured the employee? (Form 301)

Occupation Coding

Standard Occupation Classification (SOC) Codes

The 2018 Standard Occupational Classification (SOC) system is a federal statistical standard coding system used by federal agencies to classify workers into occupational categories for the purpose of collecting, analyzing, and disseminating data. The SOC classifies all workers into one of 867 detailed occupations according to their occupational definition. To facilitate classification, detailed occupations are grouped into 459 broad occupations, 98 minor groups, and 23 major groups. The SOC groups together similar job duties, and in some cases, similar skills, education, and/or training.

NIOSH Industry and Occupation Computerized Coding System

The NIOSH Industry and Occupation Computerized Coding System (NIOOCS) is a free, web-based application that uses machine learning, a form of Artificial Intelligence (AI), to assign industry and occupation codes to text descriptions. Although NIOCCS can assign both occupation codes and industry codes, the ITA already has NAICS codes reported by the establishment, so OSHA only uses NIOCCS to assign occupation codes. The occupation codes assigned by NIOCCS are a subset (n=808) of the 2018 SOC codes. The occupation code with the highest probability is selected, and NIOCCS returns the probability that the assigned code was correct.

Following occupation coding conventions, NIOCCS uses both the industry description and occupation description to assign a SOC code. The ITA collects industry as both a numeric NAICS code and a text industry description (often one of the NAICS index titles). NIOCCS coded each ITA record twice, once using the NAICS code and once using the text industry description. This duplicate coding acts as a check on the NIOCCS coding, and OSHA used the duplicative coding as a check for the accuracy of the assigned code.

Accuracy Cutoff

When NIOCCS could not assign a SOC code, due to incomplete or vague job descriptions, or when NIOCCS indicated an unlikely industry and occupation combination, OSHA assigned a SOC code of

"9999" to indicate the case did not have a SOC code assigned. Cases with a probability score below 85% were also assigned a score of "9999" when OSHA did not manually review the case and confirm it's code.

Human review

In 2024, OSHA reviewed 50% of NIOCCS-assigned SOC codes to determine the accuracy of NIOCCS coding and used the fields SOC_reviewed and SOC_probability to document the review of codes (See Table 1). The variable SOC_reviewed indicates when OSHA reviewed the code assigned by NIOCCS, where 0 are SOC codes that were not reviewed by OSHA, 1 are codes that were reviewed by OSHA, and 2 are cases that were not SOC coded. In most cases, the SOC_probability is assigned by NIOCCS. If during review, OSHA determined that another code was more applicable, it assigned a new SOC code. SOC codes that were reassigned by OSHA are assigned a SOC_probability of 5. The SOC_probability of 5 does not reflect a probability score; it is an indicator variable. If NIOCCS could not assign a SOC code, or if, during review, OSHA determined that a SOC code could not be accurately assigned (such as due to vague job description) these entries were assigned SOC code "9999" and a SOC_reviewed score of 2. The code "9999" is assigned to 18% of job descriptions in the full ITA dataset, including both instances where NIOCCS did not assign a code and OSHA did not assign a code.

Table 1: SOC Probability Values and SOC Reviewed Values

Circumstance	SOC_probability value	SOC_reviewed value
NIOCCS assigned an SOC code and OSHA reviewed it and agreed with the coding	<0 and <= 1	1
NIOCCS assigned an SOC code and OSHA did not review	<0 and <= 1	0
NIOCCS assigned an SOC code and OSHA reviewed it and assigned a more applicable SOC code	5	1
NIOCCS assigned an SOC code and OSHA determined a code could not be accurately assigned	0	2
NIOCCS did not assign an SOC code	0	1

Data Uses

OSHA collecting data electronically through the ITA to increase knowledge of workplace hazards, in order to improve worker safety and health and reduce the number of occupational injuries and illnesses. OSHA will use the collected data to identify workplace hazards and target specific establishments for its enforcement inspection and compliance assistance programs, as well as to better analyze illness and injury trends related to specific industries, processes, or hazards. Most of

the collected data will be made available to the public to allow employers, employees, and others to increase their knowledge about workplace hazards, make more informed decisions about workplace safety and health at a given establishment, and to enable researchers to better identify patterns of injuries, illnesses and hazardous workplace conditions.

Incidence Rate Calculation

Visit the Bureau of Labor Statistics (BLS) website for more information on injury and illness incidence rates. The most common formula to compute an incidence rate of non-fatal occupational injuries and illnesses is:

$$\frac{(Number\ of\ injuries\ and\ illnesses\ x\ 200{,}000)}{Employee\ hours\ worked} = \text{Incidence Rate}$$

An establishment's Total Case Rate (TCR) can be calculated using the Summary Data. The TCR equation includes all cases recorded on the OSHA Form 300 (Column H + Column I + Column J). The Days Away, Restriction, and Transfer (DART) Rate can be calculated using the Summary Dataset. DART rates are calculated using totals recorded in Column H + Column I.

Frequently Asked Questions

How complete are the ITA data?

Establishments that meet ITA electronic submission criteria are required to electronically report to OSHA certain information about the injuries and illnesses recorded on their OSHA recordkeeping forms. The ITA datasets reflect the scope of the ITA reporting requirements. The completeness and accuracy of the ITA data is limited by how well employers record and report injuries and illnesses under the regulation mediated by OSHA's compliance assistance, outreach, and enforcement efforts.

Is data from establishments covered by a State Plan included?

Yes.

How frequently are ITA data updated?

OSHA plans to update posted data quarterly, and to update occupation codes twice each year.

Can anyone use the ITA data?

Yes.

When will narrative fields be added to the ITA database?

OSHA will make narrative fields publicly available once it has completed the process of removing information that could reasonably be expected to identify an individual directly from ITA submissions.

Did OSHA review every instance where NIOCCS could not assign a SOC code?

OSHA did not review every instance when NIOCCS could not assign a code. Very few of these were reviewed. Of those OSHA did review, OSHA either assigned the instance a code or assigned 9999.

When did OSHA use "9999" as the SOC code?

OSHA used 9999 when NIOCCS did assign a code, but it had a very low probability score, or NIOCCS assigned an incorrect code and OSHA could not determine a correct code.

Did OSHA manually assign all the "9999" codes, or did the NIOCCS assign those itself?

NIOCCS assigned an insufficient information code, and for those cases where OSHA could not assign an SOC code, the case was assigned "9999."

If NIOCCS could not assign a code and OSHA did not review the case, what is indicated in the record?

In this situation, the variable SOC_reviewed is set to 2, and the SOC code is set to "9999." This is true for cases where NIOCCS could not assign a code and either 1) OSHA reviewed and agreed that there wasn't an appropriate code and 2) OSHA did not review and accepted the NIOCCS assessment that the record could not be coded.

Can I use ITA data to find information about young workers?

OSHA does not make information about age of workers publicly available.

When should I use OSHA ITA data rather than the data provided by the Bureau of Labor Statistics (BLS) Survey of Occupational Injuries and Illnesses (SOII)?

A comparison of OSHA ITA and BLS SOII data estimates is available here. Both OSHA ITA and the BLS SOII provide information based on worker injuries and illnesses that were recorded on OSHA recordkeeping forms. ITA data are collected and published for establishments that meet certain criteria regarding their industry classification and the number of workers they employ. ITA data may be used to learn more about the details of recordable injuries and illnesses at a specific establishment. ITA data may also be used to calculate occupational injury and illness incidence rates for individual establishments. However, ITA data should not be used to calculate incident rates for industry classifications because data are not statistically representative of all establishments in an industry.

The BLS SOII is the largest and most accurate survey available in the U.S. of non-fatal occupational injuries and illnesses. BLS SOII data may be used for all generalized statements about occupational injuries and illnesses in the US and to describe the rates of injuries at the industry level. BLS SOII data cannot be used to calculate incident rates for a particular establishment.