

Detecting PPE concerns in OSHA complaints using machine learning to support infectious disease outbreak response

Nora Y. Payne & Emily J. Haas

To cite this article: Nora Y. Payne & Emily J. Haas (19 Nov 2025): Detecting PPE concerns in OSHA complaints using machine learning to support infectious disease outbreak response, Journal of Occupational and Environmental Hygiene, DOI: [10.1080/15459624.2025.2573665](https://doi.org/10.1080/15459624.2025.2573665)

To link to this article: <https://doi.org/10.1080/15459624.2025.2573665>



View supplementary material [↗](#)



Published online: 19 Nov 2025.



Submit your article to this journal [↗](#)



Article views: 162



View related articles [↗](#)



View Crossmark data [↗](#)



Detecting PPE concerns in OSHA complaints using machine learning to support infectious disease outbreak response

Nora Y. Payne  and Emily J. Haas 

National Personal Protective Technology Laboratory, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Pittsburgh, Pennsylvania

ABSTRACT

Workers frequently struggle to acquire, maintain, and use personal protective equipment (PPE) during infectious disease outbreaks. Strategic PPE distribution, guidance, and interventions can help address these challenges, but the effectiveness of these measures depends on timely characterization of how these challenges manifest across the U.S. workforce—data which no U.S. public health surveillance system currently provides. This article describes a mechanism of generating such data by using a machine learning model to detect various PPE concerns in workplace safety complaints submitted to the U.S. Occupational Safety and Health Administration (OSHA). A publicly available dataset of 78,770 OSHA complaints received during the COVID-19 pandemic was used to assess the feasibility of this approach. Results demonstrate that these OSHA complaints contained a substantial variety and number of PPE concerns, and that a machine learning model trained on these data was capable of detecting three types of PPE concerns with at least 90% precision and 90% recall: unavailable or inaccessible PPE, lack of PPE use among workers, and inadequate enforcement of PPE use. Furthermore, analyses of ML-facilitated detections were shown to elucidate national and industry-specific trends in worker PPE concerns. Although further development is needed to accurately detect a broader set of PPE concerns, the results of this study suggest that machine learning can help efficiently repurpose OSHA complaints to generate insightful real-time data on worker PPE concerns during future outbreaks.

KEYWORDS

Personal protective equipment; occupational surveillance; artificial intelligence; emergency preparedness; respiratory health

Introduction

Infectious disease outbreaks create widespread challenges for proper personal protective equipment (PPE) implementation, maintenance, and use among workers. During past outbreaks in 2003 (SARS) and 2014 (Ebola), healthcare personnel reported significant difficulty obtaining sufficient PPE amid global shortages (Murphy 2006; Northam et al. 2014). Conservation efforts often necessitated reuse. Inadequate training, insufficient respirator fit testing, and physiological discomfort (e.g., heat stress and breathing difficulties caused by extended use of gowns and respirators, respectively) further complicated proper PPE use (Campbell 2006; Shaw 2006; World Health Organization 2016). These challenges reemerged during the 2020 COVID-19 pandemic at a much greater scale, as the need for PPE expanded beyond

healthcare personnel to include individuals working in industries where PPE use is less typical (Houghton et al. 2020; Cash et al. 2021; Gaitens et al. 2021; Justie et al. 2022; van Kampen et al. 2023). Increased PPE demand exacerbated shortages and caused a proliferation in the manufacture and distribution of counterfeit PPE (Plana et al. 2021). Where PPE was available, workers reported difficulty adhering to workplace PPE requirements due to evolving guidance and inconsistent enforcement by employers (Ho et al. 2020; Meyersohn 2021). Additional challenges, such as an inability to access PPE in the correct size (Hignett et al. 2021) and retaliation for voicing PPE concerns (Sfeir 2021) were also observed during this time. Future outbreaks will likely present similar challenges for workers as well as the governmental agencies and employers responsible for their protection.

CONTACT Nora Payne  NoraPayne@cdc.gov  National Personal Protective Technology Laboratory, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Pittsburgh, PA.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15459624.2025.2573665>. AIHA and ACGIH members may also access supplementary material at <http://oeh.tandfonline.com>.

This work was authored as part of the Contributor's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

No national data source currently exists to help federal agencies and other impacted parties directly monitor these occupational PPE challenges in real-time. Without these data, federal agencies are limited in their ability to rapidly identify emerging PPE challenges, allocate resources, formulate guidance, and coordinate response activities amid fast-changing conditions. State and local public health departments, businesses, and individuals also lack key evidence needed to inform PPE-related decision making at more local levels. This data gap is particularly consequential for non-healthcare establishments, which often lack pandemic preparedness plans and infection control experience, increasing their vulnerability during an outbreak (Rebmann et al. 2013). To most effectively address occupational PPE needs during an infectious disease outbreak, the ability to take a “quantitative, real-time snapshot” is needed (Gondi et al. 2020).

To produce data on hazards or injuries for which robust surveillance is limited or non-existent, and for which the cost of data collection is high, occupational health and safety researchers have often sought to repurpose existing data sources. Mining free text narratives using machine learning (ML) has emerged as a popular method of doing so. With adequate training on appropriate datasets, ML models can efficiently and accurately process large volumes of text, transforming unstructured data into a form that is more relevant and more easily quantified, summarized, and analyzed. Researchers have successfully employed ML to produce novel information from workers’ compensation records, safety reports, and electronic health records (Lincoln et al. 2004; Vallmuur 2015; Vallmuur et al. 2016; Scott et al. 2021). The potential of ML to enhance occupational surveillance, combined with increasing information needs, has led to numerous calls for federal agencies such as the National Institute for Occupational Safety and Health (NIOSH), Bureau of Labor Statistics (BLS), and Occupational Safety and Health Administration (OSHA) to continue developing modern approaches to processing free text data (National Institute for Occupational Safety and Health 2019; Tamers et al. 2021).

This article describes a study undertaken to assess the feasibility of mining OSHA complaints using ML to generate real-time data on PPE challenges faced by the U.S. workforce during an infectious disease outbreak. In the U.S., any worker or workers’ representative witnessing unsafe working conditions has the right to file a complaint with OSHA per 29 CFR §1903.11, Complaints by Employees. The complaint

process produces a free text narrative describing the hazard in addition to auxiliary information about the establishment where unsafe working conditions are reported to have occurred (e.g., industry sector, address). Although these complaints are not collected for the express purpose of monitoring specific occupational PPE challenges, they may contain relevant information for doing so. Using a publicly available dataset of OSHA complaints received during the COVID-19 pandemic, this study:

1. created a labeled dataset suitable for training ML models by developing, validating, and manually applying a labeling scheme that defines specific PPE concerns to be monitored;
2. trained an ML model to automatically detect PPE concerns using the labeled dataset and evaluated its detection accuracy for each PPE concern; and
3. Identified temporal trends in resulting detections across industry sectors to assess the value of the generated data for outbreak response.

Results suggest that an ML-based approach can reliably detect certain PPE concerns in OSHA complaints and that the generated data may facilitate improved outbreak response.

Data and methods

Creation of a labeled dataset

This feasibility study necessitated the development of a labeled dataset in which each OSHA complaint was accompanied by a vector label indicating the types of PPE concerns expressed within it. The following subsections describe how PPE-related complaints were identified and sampled from a source dataset, how a labeling scheme was developed and validated to facilitate reliable detection of various PPE concerns in complaints, and how a sample of complaints was manually labeled and reconciled by two human annotators to form a labeled dataset suitable for training an ML model.

Source dataset

A publicly available dataset of 78,770 COVID-19-related OSHA complaints from January 25, 2020 through July 15, 2022 served as the source dataset from which the labeled dataset was developed. OSHA advises filing a complaint as soon as possible after noticing a hazard and provides several avenues for reporting, including a web-based form (Occupational Safety and Health Administration). Therefore, the

"Employee is concerned the mask provided by the employer is ineffective, unapproved, and there is no direction on replacement or sanitization. Employee is concerned of the result of breathing restriction cause by the mask. There is no significant training regarding the effectiveness or use of the mask."

"1. Employer does not provide N95 respirators to patient transporters who are exposed to patients with COVID-19. 2. Employer does not enforce social distancing or the use of face coverings per Executive Order # 20-81. 3. Employer does not provide safe patient handling training to patient transporters."

"1. Office has not been cleaned since the Pandemic started. 2. Too many employees in office. 3. Coworkers have gotten sick. 4. Temperatures are not being taken. 5. We are fully operational meaning we are in contact with tenants on a daily basis and our schedules are full from 9-5."

Figure 1. Examples of three hazard narratives from the source dataset of 78,770 closed COVID-19 related OSHA complaints.

complaint receipt date was assumed to closely approximate the time when unsafe working conditions were encountered.

Each complaint generally includes a free text description of the alleged hazard (a *hazard narrative*), in addition to auxiliary information about the location of the alleged hazard, such as the establishment name, address, and industry classification as given by North American Industry Classification System (NAICS) codes. Examples of three hazard narratives are given in [Figure 1](#). Sixty complaints lacked a hazard narrative and were excluded, resulting in 78,710 valid complaints. Hazard narratives were converted to lower-case. Approximately 1,328 Spanish narratives were identified and translated into English using LibreTranslate, an open source translation engine (LibreTranslate Authors 2025).

Identifying PPE-related complaints

To focus manual labeling efforts, PPE-related hazard narratives were identified. First, a comprehensive list of PPE-related keyword strings was compiled to include various PPE (e.g., "ppe," "respirator," "gloves") and terms related to PPE use (e.g., "fit test," "fit check"). This list is available via the project's code repository; see the Data Availability Statement for details. Terms for face-worn products and other items (e.g., "masks," "face cover," "bandana," "coverall") were also included, as PPE terminology varied greatly during the COVID-19 pandemic. The term "PPE" was broadly defined to encompass all items workers may have received, procured, or worn to reduce COVID-19 risk.

Next, the keywords list was used to identify each of the 78,710 valid complaints as either PPE-related, if it included at least one PPE keyword string, or not PPE-related. Simple random samples of narratives identified as PPE-related and not PPE-related were taken and reviewed for false positives and false negatives. If any were observed, the keywords list was updated by

modifying existing keyword strings or adding new ones, including observed misspellings. This process was repeated until no false positives or false negatives were observed in the samples drawn for inspection.

Using the final list of PPE-related keyword strings, 31,018 (39.4%) valid complaints were identified as PPE-related. Approximately 10% of valid hazard narratives were randomly sampled for manual review ($n=3,200$). It was observed that distinct complaints, despite describing independent events, occasionally contained identical free text narratives, which could cause these narratives to be overweighted when training and evaluating an ML model. Therefore, only distinct hazard narratives were retained for subsequent manual labeling ($n_{distinct}=3,121$).

Developing the labeling scheme

A hazard narrative may contain several types of PPE concerns, each of which may be expressed in unique ways while reflecting the same underlying concern. For example, a concern about PPE availability may manifest as "no PPE available" in one complaint and "employer refused to provide masks" in another. Manual detection of these PPE concerns is thus subject to judgment and potentially error prone. To help human annotators more accurately and consistently detect the specific types of PPE concerns expressed in each narrative, a standardized set of PPE concerns with accompanying operational definitions (a *labeling scheme*) was developed. After reviewing a subset of the 3,121 distinct sampled narratives, the authors developed an initial labeling scheme that included a broad range of concerns related to PPE availability, use, acceptability, and accessibility (see [Table A1 of the Supplemental Material](#)). Certain concerns, such as lack of PPE use, were split into more granular sub-concerns to assess labeling consistency and concern frequency at a more detailed level (e.g., lack of PPE use among employees, non-employees, or by an

unspecified individual that could be either an employee or non-employee).

Validating the labeling scheme

The initial labeling scheme was then validated to identify labels that could not be consistently applied and therefore needed further attention. Inconsistent application suggested a PPE concern was ill-defined, too similar to another PPE concern, or inadequately captured a coherent construct. Validation occurred in two stages: (1) an informal pilot to identify major inadequacies in the labeling scheme, which also served to train the annotators; and (2) a formal intercoder agreement analysis to quantify the degree to which trained annotators could detect each PPE concern within hazard narratives.

For the pilot, two annotators (the authors) used the labeling scheme to detect the PPE concerns contained within 200 narratives randomly sampled from the 3,121 retained for manual labeling. For each PPE concern, each annotator reviewed the hazard narratives and recorded whether the concern was present or absent in each. The number of times a PPE concern appeared within a given narrative was not assessed, only whether it was expressed. Intercoder agreement was informally assessed by calculating and evaluating Krippendorff's alpha (K_α) without a strict cutoff. The annotators discussed PPE concerns with low agreement and then refined the labeling scheme by revising operational definitions and combining, adding, renaming, or removing PPE concerns from the labeling scheme. Following the pilot, the annotators manually detected PPE concerns in each of the remaining 2,921 distinct hazard narratives using the refined labeling scheme. Intercoder agreement was formally assessed using Krippendorff's alpha. Concerns were a priori determined to exhibit sufficiently high agreement if $K_\alpha > 0.80$.

Results of the formal intercoder agreement analysis can be found in [Table A2 of the Supplemental Material](#). Most concerns exhibited sufficiently high agreement, indicating that they could be consistently detected by the two annotators. Concerns that did not exhibit sufficient agreement were either dropped from the labeling scheme, collapsed to a less granular level exhibiting sufficiently high agreement (e.g., combining "enforcement concern about employee PPE use" and "enforcement concern about non-employee PPE use" into "enforcement concern about PPE use"), or retained with a caveat. An additional concern pertaining to cross-contamination was added since it emerged as a significant issue that was not captured

by the original labeling scheme. All complaints were reviewed again to ensure this new label was assigned to all applicable complaints.

Finalizing the labeled dataset

The authors revisited all previously coded hazard narratives and jointly reconciled any labeling disagreements. This process resulted in a labeled dataset consisting of 3,121 distinct hazard narratives, each accompanied by a vector of 1's and 0's indicating the presence or absence of each PPE concern. A description of the PPE concerns present in the labeled dataset is provided in [Table 1](#).

Ninety-three percent (93%) of complaints in the labeled dataset expressed at least one PPE concern from the labeling scheme; the median and maximum number of PPE concerns reported per complaint was one and seven, respectively. Approximately 70% of all PPE-related complaints originated from five NAICS sectors: Health Care and Social Assistance (23%), Retail Trade (14%), Manufacturing (14%), Accommodation and Food Services (11%), and Transportation and Warehousing (8%).

Concerns appearing in less than 1% of labeled narratives were removed to focus subsequent ML training on concerns with a sufficient number of training examples.

Training and evaluation of an ML model

A transformer-based language model, DistilBERT, was trained to detect PPE concerns in OSHA complaint narratives. Based on the popular BERT (Bidirectional Encoder Representations from Transformers) model, DistilBERT belongs to a class of ML models that use self-attention mechanisms to account for word sequence and context (Sanh et al. 2019). These models are pre-trained on large quantities of text data and then frequently further trained (i.e., *fine-tuned*) to perform a specific task using a user-supplied, domain-specific dataset. Due to their ability to capture information embedded in word sequences, these models excel at text classification tasks, frequently outperforming classifiers employing a "bag of words" approach, in which texts are regarded as an unordered collection of words. The DistilBERT model was chosen over BERT since it can be trained more quickly without an appreciable decrease in predictive performance.

To assess model performance, 150 random train-test splits of the labeled dataset were performed. In each split, 75% of the dataset was reserved for training and the remaining 25% for evaluation. A DistilBERT model was fine-tuned for 20 epochs on the training

Table 1. Description of PPE concerns present in the labeled dataset.

PPE concern	Definition	<i>n</i>	Frequency (%)
Availability	PPE not provided or made available in sufficient quantity, for employees or non-employees.	1,144	36.7
Enforcing Use	Employer fails to enforce PPE usage by employees or non-employees in the establishment.	1,052	33.7
Not Worn by Employees	PPE not worn by employees.	705	22.6
Worn Incorrectly by Employees	PPE not worn correctly by employees. For example, masks not covering the nose, masks worn on the chin, and masks pulled down.	136	4.4
Not Worn by Non-employees	PPE not worn by non-employees.	95	3.0
Not Worn by Unspecified*	PPE not worn by unspecified individual (i.e., unable to ascertain whether individual is an employee or non-employee).	79	2.5
Enforcing Correct Use	Employer fails to enforce the correct wearing of PPE by employees or non-employees in the establishment.	69	2.2
Cross-contamination	PPE used across multiple locations, tasks, or employees.	66	2.1
Discouraged or Prohibited	Employer discourages or prohibits PPE use.	63	2.0
Training	Lack of training on PPE usage, selection, or storage.	60	1.9
Fit Test	Fit testing not performed, performed improperly, or lack of alternatives provided for those failing.	50	1.6
Physiological	PPE causes physiological symptoms or discomfort, such as difficulty breathing, excessive moisture buildup, headaches, and overheating.	47	1.5
Disinfection and Maintenance	PPE not disinfected, improperly disinfected, or disinfected in a way that compromises its protective capabilities.	34	1.1
Size or Fit	PPE does not fit, or the worker does not have access to properly fitting PPE. To include PPE too large (e.g., falling off, slipping, or too loose), PPE too small (e.g., too tight, leaving marks on face), PPE needing to be modified to achieve proper fit, PPE not available in the worker's size.	27	0.9
Worn Incorrectly by Non-employees	PPE worn incorrectly by non-employees.	9	0.3
Worn Incorrectly by Unspecified*	PPE worn incorrectly by unspecified individuals (i.e., unable to ascertain whether individuals are employees or non-employees).	8	0.3
Respiratory Protection Program	Respiratory Protection Program deficiencies.	7	0.2
Counterfeit	PPE provided may be counterfeit.	5	0.2
Expired	PPE provided is expired.	2	0.1

For each PPE concern, the table indicates both the number and percentage of complaints in the labeled dataset expressing the concern. Definitions in this table are abbreviated for simplicity; detailed definitions for labels present in the original labeling scheme can be found in the [Supplemental Material](#). Asterisks (*) indicate concerns exhibiting insufficiently high agreement ($K_{\kappa} < 0.80$) which were retained with a caveat.

set using a learning rate of 2×10^{-5} and a weight decay of 0.01. A batch size of 16 was used for both training and evaluation. The precision, recall, and F1 score achieved by the model on the test set was recorded for each PPE concern. These metrics measure the proportion of narratives the model identified as expressing the concern that truly did so (precision), the proportion of narratives expressing the concern that were correctly identified by the model as doing so (recall), and a combination of the two (F1 score). The F1 score was computed by averaging two individual F1 scores, one assessing the model's ability to detect the concern's presence and the other assessing detection of the concern's absence, with equal weight given to each (i.e., macro-averaging) to more fairly assess detection ability when the concern appears infrequently. All three performance metrics were averaged over the train-test splits and corresponding standard errors were calculated.

Analysis of temporal trends

Trend analysis of specific PPE concerns required labeling all 31,018 PPE-related complaints. Labels for

3,200 complaints were either already manually assigned or inferred from a manually inspected complaint with an identical hazard narrative. An ensemble classifier approach was used to label the remaining 27,818 PPE-related complaints. Three of the 150 models previously trained to assess DistilBERT's performance were used to detect PPE concerns in each unlabeled narrative. For each PPE concern, the models' detections were combined via majority vote to form a final consensus prediction as to whether the concern was present or absent from the narrative.

Analysis focused on the four NAICS sectors with the highest number of PPE-related complaints: Health Care and Social Assistance, Retail Trade, Manufacturing, and Accommodation and Food Services. Seven-day lagged moving averages were calculated for (1) daily complaint counts, (2) daily PPE-related complaint counts, and (3) daily counts of complaints expressing specific PPE concerns. Seven-day lagged moving averages were also calculated within NAICS sectors to reveal potential differences in temporal trends across industries. COVID-19 hospitalization data from CDC COVID-NET was used to contextualize complaint counts (Centers for Disease Control and Prevention).

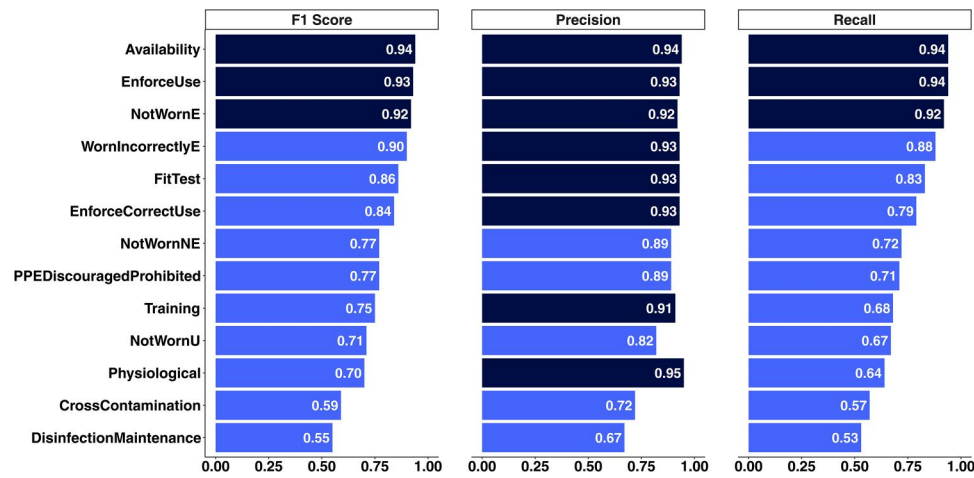


Figure 2. Mean precision, recall, and F1 score for each PPE concern over 150 train-test splits. Standard errors are given in [Table A3 of the Supplemental Material](#). Concerns are ordered by decreasing F1 score. Darker bars indicate metrics exceeding 0.90.

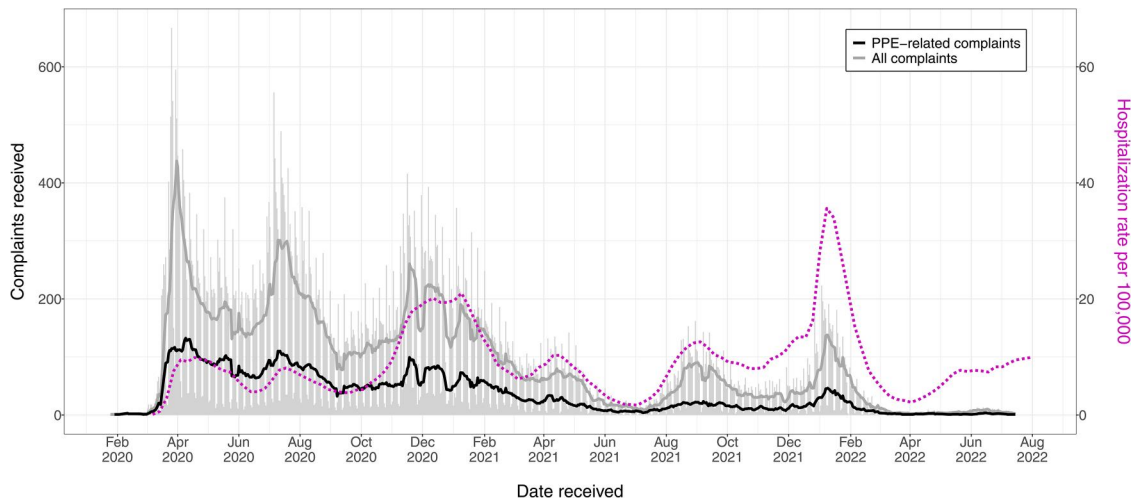


Figure 3. Number of closed OSHA COVID-19-related complaints by date received. Grey vertical bars indicate daily counts of complaints received. Solid lines depict 7-day lagged moving averages of all complaints received (grey) and PPE-related complaints received (black). Dotted line depicts weekly COVID-19 hospitalizations from CDC COVID-NET.

Only PPE concerns that the model detected with at least 90% average precision and 90% average recall during the evaluation phase were analyzed. As an additional check, trends reflected in the analysis of all PPE-related complaints were examined against those appearing in the subset of PPE-related complaints that were randomly sampled and manually labeled.

Results

Model evaluation results are given in [Figure 2](#) (corresponding standard errors appear in [Table A3 of the Supplemental Material](#)). Three PPE concerns were detected by the DistilBERT model with mean precision, recall, and F1 scores of at least 90%. These were concerns pertaining to: (1) PPE availability, (2) enforcement of PPE use, and (3) employee use of

PPE. Concerns about fit testing and incorrect usage of PPE by employees were detected by the DistilBERT model with mean precision, recall, and F1 scores between 80% and 90%. These concerns were likely well-predicted because they either appeared frequently in the dataset or were expressed similarly across narratives. Among concerns that were less well-predicted, recall was generally lower than precision, suggesting that the model was conservative in its predictions. That is, incorrect predictions tended to take the form of false negatives.

Nationally, the average number of reported complaints rose sharply to a peak in March-April 2020 ([Figure 3](#)). Although this average showed a general downward trend after April 2020, local peaks were observed in subsequent months. These peaks appeared to coincide with increasing U.S. COVID-19 hospitalizations. The

average number of PPE-related complaints reported nationally exhibited similar trends, as did the averages within each of the four industry sectors examined, to varying extents (Figure 4A and B).

In July 2020, the average number of PPE-related complaints reported nationally reached the second highest peak since the beginning of the pandemic (Figure 4A). Two sectors saw an approximate doubling in the average number of PPE-related complaints reported: Accommodations and Food Services and Retail Trade (Figure 4B). For Retail Trade, this doubling represented a return to peak levels last observed in April 2020. For Accommodations and Food Services, average PPE-related complaint counts in July 2020 surpassed all previous levels. In the Health Care and Social Assistance and Manufacturing sectors, average PPE-related complaint counts were roughly one-third of peak levels last observed in March-April 2020.

Examination of specific PPE concerns within reported complaints revealed that the initial spike in PPE-related complaints in March-April 2020 was driven largely by PPE availability concerns (Figure 4A). This trend was observed in each of the four industry sectors examined, although it was less pronounced in Accommodations and Food Services (Figure 4B). PPE availability remained the top-reported concern nationally until July 2020, at which point PPE use enforcement became the predominant concern. This shift was observed in the three non-healthcare sectors examined. However, within Health Care and Social Assistance, PPE availability concerns remained the top-reported PPE concern. As with PPE use enforcement, the average number of reported concerns about employees failing to wear PPE increased nationally from March 2020 to July 2020 and exhibited similar or greater levels of reporting compared to PPE availability within the three non-healthcare sectors after July 2020.

Discussion

This study represents the first documented effort to generate real-time data on PPE challenges faced by U.S. workers during an infectious disease outbreak, filling critical information gaps unaddressed by existing occupational surveillance. It highlights the benefits of parsing worker complaints to uncover detailed information that may not otherwise be available and the advantages of using ML to do so. Furthermore, it establishes a focused foundation for the federal government to consider in the development of a national

surveillance system for occupational PPE concerns, akin to established disease surveillance systems. More broadly, this study's results underscore that the provision of PPE alone is insufficient to ensure worker safety during outbreaks; support to ensure proper implementation and enforcement is also needed.

OSHA complaints were found to contain substantial information about workers' PPE concerns, with manual coding suggesting that nearly 40% of the COVID-19-related OSHA complaints received between January 2020 and July 2022 expressed a PPE-related concern. Observed concerns reflect previous reports of PPE challenges experienced by healthcare workers during infectious disease outbreaks, while revealing similarly heightened concern in additional industries for which such information has been historically limited, such as food services and retail. Furthermore, average OSHA complaint counts appeared to change with the number of U.S. COVID-19 hospitalizations, suggesting that OSHA complaints were reported in a timely manner. Collectively, these results suggest that OSHA complaints contain sufficiently current and comprehensive information on real-time occupational PPE concerns.

Data generated by the ML-based approach further contextualize qualitative accounts of occupational PPE challenges reported during the COVID-19 pandemic. Generated data appeared to capture PPE procurement challenges faced by healthcare workers in March-April 2020. Furthermore, the persistence of PPE availability concerns in OSHA complaints through 2020 appear to reflect sustained concern about the ability to access necessary PPE, which had been previously reported by news media (Jacobs 2020). Generated data also revealed a national increase in the number of PPE-related complaints among food service establishments in July 2020, which may stem from the reopening of restaurants and bars for indoor dining in many U.S. states. ML-facilitated detections of specific PPE concerns in OSHA complaints suggest that this increase was driven largely by emerging concerns about employees failing to use PPE and inadequate enforcement of PPE use, the latter of which appears to confirm previous media reporting (Zarrollo 2020). This pattern was also observed in the Manufacturing and Retail Trade sectors but not in Health Care and Social Assistance, suggesting that non-healthcare establishments may face more challenges in enforcing PPE use. This pattern may also reflect greater familiarity with PPE among healthcare workers and more established rules around PPE use and enforcement in patient care environments.

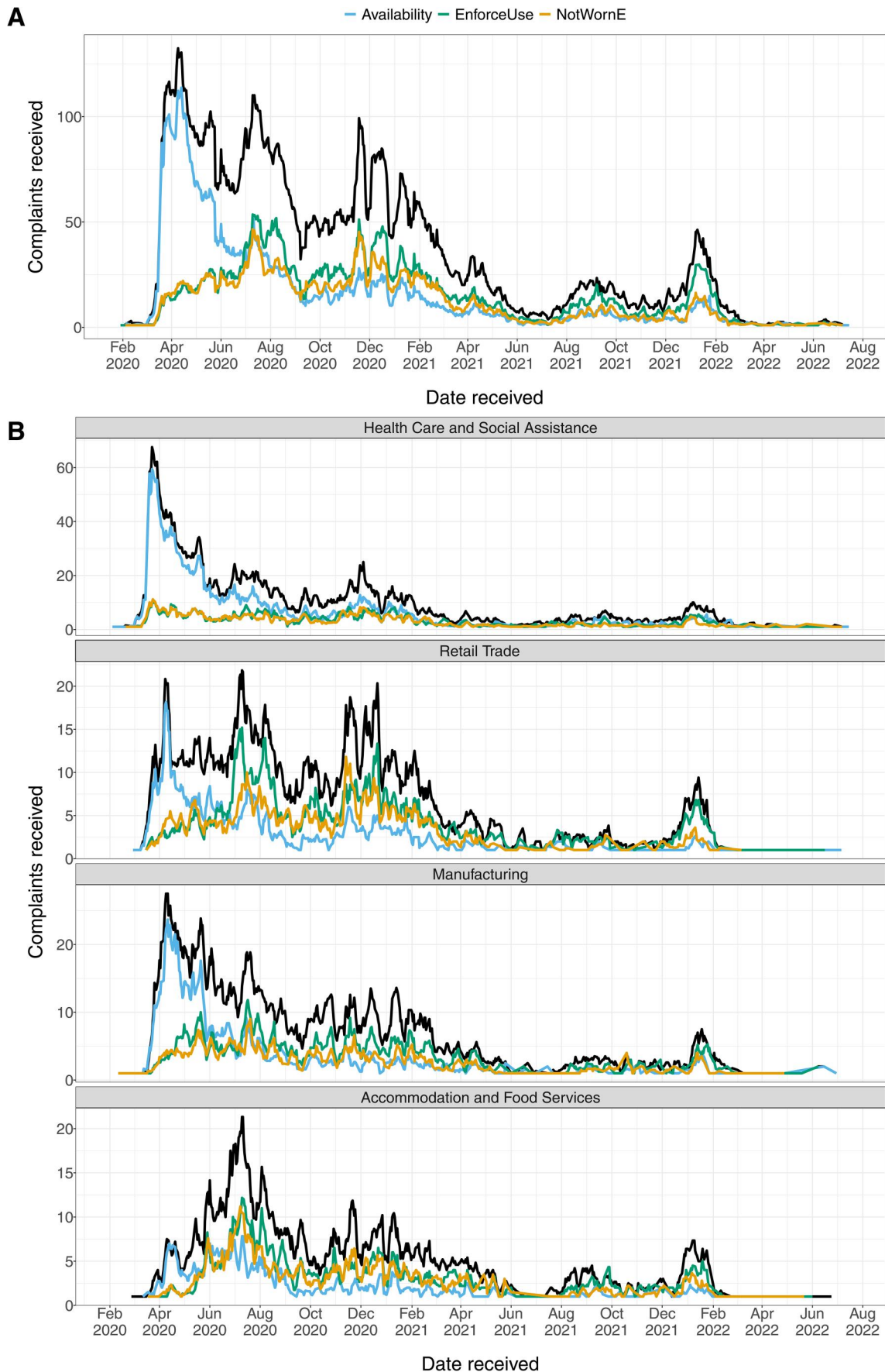


Figure 4. Seven-day lagged moving averages of the number of PPE-related complaints received and of the number of complaints expressing concerns about PPE availability, PPE use enforcement, and employees not wearing PPE, across all industries (Panel A) and within four industry sectors (Panel B).

Although the ML model detected only a subset of concerns with high accuracy, this result concurs with previously reported results from comparable multilabel classification studies (Jing et al. 2023). The highest F1 scores were generally achieved by the most frequently reported concerns and the lowest F1 scores by infrequently reported concerns. Consequently, augmenting the labeled dataset with additional examples for less frequently reported concerns may be beneficial in follow-on work. Additional opportunities to improve detection accuracy include exploring alternative models such as PHS-BERT, which is pre-trained on a health-specific corpus and may thus detect PPE concerns more accurately than the DistilBERT model employed here (Naseem et al. 2022). Leveraging additional features of text narratives, such as references to specific federal or state regulations (e.g., “29 CFR 1910.134(d)(1),” Respiratory Protection: Selection of Respirators – General Requirements) or specific types of PPE (e.g., respirators) where available, may also enhance detection accuracy.

Several modifications may improve the quality of data and granularity of insights generated by the proposed approach. In this study, several concerns in the original labeling scheme were collapsed; for example, a single label encompassed inadequate PPE use enforcement among both employees and non-employees. Refining the labeling scheme, so that the collapsed concerns can be consistently coded as standalone concerns, would likely improve the ML model’s detection accuracies while expanding the set of concerns it is able to detect. Ongoing manual annotation efforts and periodic model retraining could help maintain data integrity, counteract potential model drift, and improve detection accuracy. Additionally, exploration of trends within industry subsectors, as indicated by NAICS 3-digit codes, could yield more specific insights potentially obscured by analyzing data solely at the sector level. It is possible, for example, that workers in hospitals and residential care facilities encounter distinct PPE challenges, despite both being classified under the NAICS Health Care and Social Assistance sector.

Additional advantages of the ML-based approach presented here are worth noting. First, this approach avoids new data collection by repurposing OSHA complaints, rendering it highly cost-effective and placing no additional data reporting burden on workers. Cost-effectiveness is further increased when open-source models and tools are used to process these data, as was the case in this study. Second, because OSHA complaints are generally initiated by workers, this approach may capture PPE concerns more quickly than traditional data

collection mechanisms, such as agency-initiated surveys. Finally, if the need to monitor additional PPE concerns arises, the proposed approach can do so prospectively and retroactively. Unlike more traditional modes of data collection, such as surveys, which can generally only begin monitoring a new concern from the present onward, an ML model can be trained on re-annotated training data to detect a new concern in both past and future complaints.

Limitations

This study has several limitations, stemming primarily from the nature of the OSHA complaints dataset. First, the source dataset contained only COVID-19-related complaints, and it is unclear how this subset of complaints was obtained from the set of all OSHA complaints received during the same period. Second, although the mean length of a PPE-related complaint was 60 words (median = 44 words), several complaints were extremely brief (e.g., “no masks”), which created ambiguity during manual labeling that could have resulted in misinterpretation of the complainant’s intent. Third, workers may differ in their propensity to report workplace safety violations for various reasons, including fear of retaliation, limited familiarity with workers’ rights regarding PPE, and limited awareness of the OSHA complaint mechanism. Thus, findings only summarize reported complaints and cannot be used to establish population prevalences.

Conclusion

Mining OSHA complaints using ML can facilitate the monitoring of worker PPE concerns during infectious disease outbreaks. The results of this study suggest that PPE concerns appear frequently in OSHA complaints received during an outbreak and, for certain concerns, can be accurately detected by an ML model, enabling one to efficiently quantify the frequency with which they are reported. When analyzed with auxiliary information provided in the complaint, the generated data can be used to track how PPE concerns emerge and evolve over the course of an outbreak, nationally and within industry sectors. Further work may be conducted to refine and formally pilot a public health surveillance system based on the ML-based approach presented here. If established, such a system has the potential to enhance PPE-related decision-making for a wide range of end users. Access to real-time data on workers’ PPE concerns may empower industry organizations to gain a better understanding

of their sector's specific PPE needs, allowing them to implement more effective safety protocols to reduce infection risk, mitigate absenteeism, and maintain productivity; improve situational awareness among workers and the general public; and enable federal agencies to more rapidly identify and address emerging PPE challenges across the entire U.S. workforce during future outbreaks.

Acknowledgments

The authors gratefully acknowledge project funding from the National Occupational Research Agenda. Additionally, the authors would like to thank multiple reviewers for comments that improved the clarity and composition of the paper.

Disclaimer

The findings and conclusions are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention. Mention of any company or product does not constitute endorsement by the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Nora Y. Payne  <http://orcid.org/0000-0003-4935-6465>

Emily J. Haas  <http://orcid.org/0000-0001-5128-470X>

Data availability statement

Accompanying code and data are available at <https://github.com/CDCgov/npptl-ppe-concern-detection>.

References

- Campbell A. 2006. Spring of fear: the SARS commission final report. Government of Ontario.
- Cash RE, Rivard MK, Camargo CA, Jr, Powell JR, Panchal AR. 2021. Emergency medical services personnel awareness and training about personal protective equipment during the COVID-19 pandemic. *Prehosp Emerg Care*. 25(6):777–784. <https://doi.org/10.1080/10903127.2020.1853858>
- Centers for Disease Control and Prevention. Covid data tracker. U.S. Department of Health and Human Services, CDC; [accessed 2025 Apr 14]. <https://covid.cdc.gov/covid-data-tracker>
- Gaitens J, Condon M, Fernandes E, McDiarmid M. 2021. COVID-19 and essential workers: a narrative review of health outcomes and moral injury. *Int J Environ Res Public Health*. 18(4):1446. <https://doi.org/10.3390/ijerph18041446>
- Gondi S et al. 2020. Personal protective equipment needs in the USA during the COVID-19 pandemic. *Lancet*. 395(10237): e90–e91. [https://doi.org/10.1016/S0140-6736\(20\)31038-2](https://doi.org/10.1016/S0140-6736(20)31038-2)
- Hignett S, Welsh R, Banerjee J. 2021. Human factors issues of working in personal protective equipment during the COVID-19 pandemic. *Anaesthesia*. 76(1):134–135. <https://doi.org/10.1111/anae.15198>
- Ho H, Schneider D, Harknett K. 2020. COVID-19 safety measures update. Shift Project. <https://shift.hks.harvard.edu/covid-19-safety-measures-update/>
- Houghton C, Meskell P, Delaney H, Smalle M, Glenton C, Booth A, Chan XHS, Devane D, Biesty LM. 2020. Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: a rapid qualitative evidence synthesis. *Cochrane Database Syst Rev*. 2020(4):1–68. <https://doi.org/10.1002/14651858.cd013582>
- Jacobs A. 2020. Health care workers still face daunting shortages of masks and other P.P.E. The New York Times; [updated 2020 Dec 20; accessed 2025 Jun 30]. <https://www.nytimes.com/2020/12/20/health/covid-ppe-shortages.html>
- Jing X et al. 2023. BERT for aviation text classification. AIAA AVIATION 2023 Forum. <https://doi.org/10.2514/6.2023-3438>
- Justie B, Koonse T, Macias M, Ray J, Waheed S. 2022. Fast-food frontline: COVID-19 and working conditions in Los Angeles. UCLA Labor Center.
- LibreTranslate Authors. 2025. LibreTranslate Version 1.6.4. PyPI (Python Package Index).
- Lincoln AE et al. 2004. Using narrative text and coded data to develop hazard scenarios for occupational injury interventions. *Inj Prev*. 10(4):249–254. <https://doi.org/10.1136/ip.2004.005181>
- Meyersohn N. 2021. Store mask policies are a mess and nearly impossible to enforce. CNN; [updated 2021 May 20; accessed 2025 Jun 30]. <https://www.cnn.com/2021/05/20/business/masks-stores-walmart-starbucks>
- Murphy C. 2006. The 2003 SARS outbreak: global challenges and innovative infection control measures. *Online J Issues Nurs*. 11(1):6. <https://doi.org/10.3912/OJIN.Vol11No01Man05>
- Naseem U, Lee BC, Khushi M, Kim J, Dunn A. 2022. Benchmarking for public health surveillance tasks on social media with a domain-specific pretrained language model. Dublin, Ireland: Association for Computational Linguistics.
- National Institute for Occupational Safety and Health. 2019. A smarter national surveillance system for occupational safety and health in the 21st century. In: The NIOSH plan to implement the National Academies' Program evaluation recommendations. National Institute for Occupational Safety and Health.
- Northam J, Inskeep S, Martin R. 2014. Ebola protective suits are in short supply. National Public Radio; [updated 2014 Oct 7; accessed]. <https://www.npr.org/2014/10/07/354230895/ebola-protective-suits-are-in-short-supply>
- Occupational Safety and Health Administration. File a complaint. [accessed 2025 Feb 28]. <https://www.osha.gov/workers/file-complaint>

- Plana D et al. 2021. Assessing the filtration efficiency and regulatory status of N95s and nontraditional filtering face-piece respirators available during the COVID-19 pandemic. *BMC Infect Dis.* 21(1):712. <https://doi.org/10.1186/s12879-021-06008-8>
- Rebmann T, Wang J, Swick Z, Reddick D, delRosario JL. 2013. Business continuity and pandemic preparedness: US health care versus non-health care agencies. *Am J Infect Control.* 41(4):e27–e33. <https://doi.org/10.1016/j.ajic.2012.09.010>
- Sanh V, Debut L, Chaumond J, Wolf T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv.* abs/1910.01108.
- Scott E, Hirabayashi L, Levenstein A, Krupa N, Jenkins P. 2021. The development of a machine learning algorithm to identify occupational injuries in agriculture using pre-hospital care reports. *Health Inf Sci Syst.* 9(1):31. <https://doi.org/10.1007/s13755-021-00161-9>
- Sfeir MM. 2021. Frontline workers sound the alarm: be always sure you're right, then go ahead. *J Public Health (United Kingdom).* 43(4):899–901. <https://doi.org/10.1093/pubmed/fdaa066>
- Shaw K. 2006. The 2003 SARS outbreak and its impact on infection control practices. *Public Health.* 120(1):8–14. <https://doi.org/10.1016/j.puhe.2005.10.002>
- Tamers SL et al. 2021. The NIOSH future of work initiative research agenda. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. 2022-105.
- Vallmuur K. 2015. Machine learning approaches to analysing textual injury surveillance data: a systematic review. *Accid Anal Prev.* 79:41–49. <https://doi.org/10.1016/j.aap.2015.03.018>
- Vallmuur K et al. 2016. Harnessing information from injury narratives in the 'big data' era: understanding and applying machine learning for injury surveillance. *Inj Prev.* 22 Suppl 1(Suppl 1):i34–i42. <https://doi.org/10.1136/injury-prev-2015-041813>
- van Kampen V et al. 2023. Influence of face masks on the subjective impairment at different physical workloads. *Sci Rep.* 13(1):8133. <https://doi.org/10.1038/s41598-023-34319-0>
- World Health Organization. 2016. Personal protective equipment for use in a filovirus disease outbreak: rapid advice guideline. World Health Organization.
- Zaroli J. 2020. The customer is always right. Except when they won't wear a mask. *National Public Radio*; [updated 2020 Jul 14; accessed 2025]. <https://www.npr.org/2020/07/14/889721147/the-customer-is-always-right-except-when-they-wont-wear-a-mask>