Annals of Work Exposures and Health, 2026, **70**, wxaf069
https://doi.org/10.1093/annweh/wxaf069
Advance access publication date: 23 October 2025
Original Article

BOHS British Occupational Hygiene Society
The Chartered Society for Worker Health Protection
OXFORD

Using casual inference and machine learning with exposure determinant modeling to identify important workplace controls

Abas Shkembi[1*], Mohammed Abbas Virji[2], Jie He[1], Kowit Nambunmee[3], Pablo Ruiz-Rudolph[4], and Richard L. Neitzel[1]

[1]Department of Environmental Health Sciences, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, United States
[2]School of Public Health, West Virginia University, Morgantown, WV, United States
[3]Urban Safety Innovation Research Group (USIR), School of Health Science, Mae Fah Luang University, Chiang Rai, Thailand
[4]Epidemiology Program, Institute of Population Health, Faculty of Medicine, Universidad de Chile, Independencia, Santiago, Chile

*Corresponding author: Email: ashkembi@umich.edu

Abstract

Objectives: Exposure determinant modeling can help industrial hygienists understand where, when, and how to control occupational exposures for their particular work environment. Yet, in practice, the ability to evaluate exposure determinants is degraded by selection bias (where only a subset of all exposed workers is sampled) and the statistical issue of "small n, large p" (few samples but many exposure determinants). This study explored the application of the causal inference framework and machine learning algorithms in exposure determinant modeling using a "small n, large p" example of potential determinants of heavy metal concentrations among informal electronic waste recycling workers.

Methods: As a case study, we used a multivariable logistic regression model to construct inverse probability weights to account for selection bias into a video substudy of 41 of 226 possible workers monitored for exposures to heavy metals. Forty-four determinants of biomarkers (eg tool use, job tasks, and personal protective equipment use) were quantified through video monitoring. Concentrations of heavy metals in blood (Pb and Mn) and urine (Ni and Cu) were sampled. We identified the best-performing biomarker determinant model by comparing the leave-one-out cross-validation root-mean-squared error (LOOCV-RMSE) of 5 models: 2 traditional models (multivariate linear regression and forward selection), and 3 machine learning algorithms (LASSO, boosted regression trees, and random forests). Using the best-performing model, we estimated reductions in heavy metal concentrations through hypothetical workplace controls to identify the most important determinant of biomarker concentrations.

Results: The random forest model had the lowest LOOCV-RMSE and was used as the final biomarker determinant model. Stopping workers from bending their backs while dismantling e-waste was the most important determinant of heavy metal concentrations. Using blood Pb as an example, this translated to an estimated reduction of 0.81 μg/dL (95% confidence interval: 0.66, 0.98) in comparison with maintaining the status quo. Using a traditional regression model (forward selection without inverse probability weights), back bending was not identified as an important determinant of blood Pb.

Discussion: Our causal inference approach with machine learning algorithms overcomes the common limitations of exposure determinant modeling and produces easy-to-interpret estimates of biomarker concentration reductions from hypothetical workplace controls. This can aid industrial hygienists in choosing the most effective hazard controls that can be contextualized to their particular work setting.

Keywords: boosted regression tree; counterfactual; e-waste; forward selection; LASSO; LMIC; random forests.

Received: May 8, 2025, Accepted: September 21, 2025

What's Important About This Paper?

Identifying control strategies contextualized to a particular work environment is an important part of industrial hygiene practice, yet it can be more of an "art" than a "science." This study demonstrates how using exposure determinant modeling alongside machine learning algorithms and the causal inference framework can counterfactually estimate the reduction in worker exposures from various hypothetical workplace controls using informal e-waste workers in Thailand and Chile as an example. This approach can aid industrial hygienists' expert judgment by quantifying the effectiveness of hypothetical controls prior to implementing them.

Introduction

A primary goal for industrial hygienists is to identify suitable workplace controls that reduce workers' exposure to hazards, and subsequently dose, that are contextualized to their work environment. We can apply exposure determinant modeling to achieve this (Burstyn and Teschke 1999). This approach involves collecting exposure measurements (as the outcome) and detailed contextual information on the worker and the work environment (to quantify potential determinants of exposure). Using statistical modeling (eg linear regression), we can identify important factors that influence exposures to chemical and nonchemical hazards in the workplace. Theoretically, by implementing workplace controls on these important factors, workers' exposures could be reduced. Despite many studies highlighting the relevance of exposure determinant studies for occupational exposure assessment, such analyses are not commonly applied beyond research settings. In practice, there are many methodological and statistical issues that arise.

Potential issues with exposure determinant modeling in practice include, but are not limited to (i) the lack of systematically gathered and comprehensive contextual information on the worker and the work environment to quantify determinants of exposure, (ii) the large number of possible determinants to analyze, (iii) the infeasibility of monitoring all exposed workers; and as a result, and (iv) the small number of total samples usually collected during monitoring. Keeping these impediments in mind, we sought to present an illustrative example of exposure determinant modeling using heavy metal biomarkers among informal electronic waste (e-waste) recyclers in Thailand and Chile. We extend previous exposure determinant modeling approaches to demonstrate how using a causal inference framework in combination with machine learning models may overcome the commonly faced issues of exposure determinant modeling (Mooney et al. 2021). In doing so, we demonstrate how our extended exposure determinant modeling can be used to estimate potential reductions in workers' dose as a consequence of hypothetical workplace controls, following the hierarchy of controls. This approach can aid industrial hygienists' goal of identifying suitable workplace controls that are contextualized to their workers and work environment.

Going beyond linear regression

A common issue with industrial hygiene data is small sample sizes. Traditional exposure monitoring is often costly, both in time and in the resources and equipment needed. As a result, monitoring is often completed among a randomly selected subset of workers in similar exposure groups or, in the case of regulatory inspections, among those believed to be most highly exposed. These approaches often lead to small sample sizes, with the American Industrial Hygiene Association recommending a minimum of 6 samples for a similar exposure group (American Industrial Hygiene Association Exposure Assessment Strategies Committee 2015). Given that the number of possible determinants for a particular exposure can be quite large, we run into the issue of "small n, large p." "Small n" refers to the small number of measurements that are taken and "large p" refers to the large number of possible predictors, or exposure determinants. For example, to observe a medium effect size when we have 40 possible exposure determinants requires over 200 samples to be collected (Green 1991). In practice, fewer than 40 samples are likely to be collected during exposure monitoring, resulting in a "small n, large p" problem.

Traditional regression techniques, such as the commonly used multiple linear regression, perform poorly with a "small n, large p" problem. Linear regression implicitly assumes linearity between determinants and exposures, that there are enough samples relative to the number of determinants, and that the many determinants are independent and not highly correlated with one another (ie lack of multicollinearity). They are unable to provide unbiased estimates of the exposure determinants when these assumptions are violated because this can lead to model instability, resulting in imprecise regression coefficients. Such biased estimates can lead to incorrect conclusions that a particular determinant is the most important factor of a given exposure. This can result in erroneous decision-making when determining appropriate workplace controls from a statistical model.

To circumvent the issue of "small n, large p," and thus reduce potential erroneous decision-making, there are 2 simple approaches available: continue exposure monitoring to increase the number of measurements (often not feasible due to resource or logistical constraints, although sensor applications have become more low cost) and/or try to reduce the number of possible exposure determinants. Statistical methods for variable selection have been developed to achieve the latter. Stepwise selection, such as the forward selection of important determinants by their *P*-value, has traditionally been a popular method for reducing the number of predictors in a statistical model (Efroymson 1960). However, evidence has demonstrated that stepwise selection is likely to identify determinants that do not have any real relationship with the outcome (Smith 2018). This can similarly lead to false findings and erroneous decision-making.

Rather than trying to circumvent the "small n, large p" problem, alternative models to stepwise selection, such as shrinkage and machine learning methods, have been developed in the past few decades (Tibshirani 1996; Breiman 2001; Friedman 2001), but their utilization in exposure determinant modeling has been limited. This study will demonstrate their relevance to exposure determinant modeling. But before considering these statistical models, it is critical to understand the purpose of conducting exposure determinant modeling. Beyond simply identifying determinants of exposure, exposure determinant modeling can also assess whether controlling for these causes would reduce worker exposure in a particular work environment, and by how much. The causal inference framework can be used to achieve this goal, although it has not often been applied to such modeling.

Causal inference for the industrial hygienist

Returning to our central goal—implementing workplace controls that reduce exposure—an industrial hygienist's primary concern is cause and effect. Thus, industrial hygienists are interested in "causal inference." In observational epidemiology, causal inference seeks to assess the cause-and-effect relationships between independent risk factors and the development of adverse health outcomes (Vandenbroucke et al. 2016). Applying it to the context of industrial hygiene assessment, the causal inference framework is not necessarily concerned with traditionally "causal" determinants of exposure related to physical and chemical properties of hazards, like vapor pressure. Rather, it allows us to consider both causes of exposure (eg welding) and surrogates of exposure (eg body positioning when welding) as potential intervenable determinants of workplace exposures. While other mathematical and mechanistic models such as the Advanced REACH Tool can estimate likely occupational exposures under various scenarios (Tielemans et al. 2011), the advantage of the causal inference framework in observational studies is the ability to estimate potential exposure reductions using workplace-specific data and considering nonmechanistic causes of dose (eg human factors).

For example, the causal inference framework gathers observational information from a specific workplace to reasonably predict answers to hypothetical questions such as "How many fewer workers would be exposed to noise levels above the NIOSH Recommended Exposure Limit of 85 dBA if we provided all workers with hearing protection devices, in comparison with maintaining the status quo (ie doing nothing)?" or "How much, on average, can we reduce worker noise exposure if we highly control a few of the noisiest machines versus provide moderate control for all machines?"

In theory, with every worker properly trained and fit tested, we would expect no workers to be overexposed. In practice, however, this may not be the case. In some work environments with a robust safety culture, workers may be more likely to follow personal protective equipment (PPE) guidance. In others where safety culture is lacking, some workers use hearing protection only occasionally or not at all. The causal inference framework can estimate an answer to this question prior to implementing a new control by using observational information on workers in their own work environment. This contextualizes the potential effectiveness of controls to the workers' own work environment. This allows us to quantify the uncertainty around this hypothetical question by implicitly accounting for different worker behaviors. This is an important aspect of implementing new controls, as variability in how workers behave with the controls in their own work environment leads to uncertainty in how effective it will actually be. We can also assess many possible controls associated with the determinants of an exposure to try and predict which one would reduce exposures the most. However, in order to obtain reliable predictions of the effectiveness of possible controls, the assumptions of causal inference need to be satisfied.

A central concern that often limits causal inference in industrial hygiene data is the "selection bias" as a consequence of sampling from a subset of workers. Selection bias in exposure monitoring violates a principal assumption of causal inference—exchangeability (Hernán and Robins 2020). Ideally, we would want the subset of workers sampled to be randomly selected, as the assumption of exchangeability means that the workers sampled and those not sampled have the same set of characteristics that could influence exposure (eg age, job title, etc.). However, in practice, this assumption can be unrealistic if industrial hygienists prioritize monitoring the most likely exposed workers. This inherently makes the subset of those sampled "unexchangeable" with those not sampled. However, we can make these 2 groups "conditionally exchangeable" in observational studies through statistical methods, such as inverse probability weighting, prior to running the final exposure determinant model (Hernán and Robins 2020). While consideration of exchangeability for causal inference is most common in epidemiological settings, the concepts are transferable to workplace exposures, wherein the goal is not exposure–disease relationships but determinant–exposure relationships.

Determinants of heavy metal biomarkers among e-waste workers

With these statistical approaches (ie machine learning and causal inference) that can overcome common impediments with industrial hygiene data (ie small n, large p; lack of exchangeability), we sought to illustrate how these statistical approaches can be applied to "real-world" industrial hygiene data. We used an illustrative example of electronic waste (e-waste) recyclers in Thailand and Chile. The main workplace exposures of interest are toxic heavy metals, in particular lead (Pb), manganese (Mn), nickel (Ni), and copper (Cu). E-waste recyclers likely have exposures to these heavy metals, as they are commonly found in the e-waste they dismantle, recover, and sell (Heacock et al. 2016). However, recyclers—especially those working informally—typically have little to no training in the safe extraction of the valuable metals. They often dismantle e-waste without any PPE and with crude dismantling methods such as breaking lead-containing TV screen glass with blunt striking instruments, which can cause dust with heavy metals to be suspended in the air and settle on workers and their work environment, which is often also their home environment. As a result, workers experience multiple routes of exposure (inhalation, dermal, and ingestion) that can cause elevated levels of heavy metals in the blood and urine of e-waste recyclers (Okeme and Arrandale 2019).

Figure 1a shows a common work environment for e-waste recyclers in Thailand, which is typically their own homes. Note that while Chilean e-waste recycling is typically performed in a standard workplace, the informal nature of their work similarly leads to hazardous environments such as those in Thailand. There are many determinants of heavy metal exposure and feasible controls to reduce concentrations. For example, recyclers could eliminate the dismantling of specific e-waste that could contain high levels of heavy metals (Fig. 1b). However, this control may not be feasible for most workers, as e-waste with most heavy metals may also yield the most economic value upon recycling. Regardless, we can conceptually use the causal inference framework to ask, "how much, on average, can we reduce blood and urinary heavy metal concentrations if certain e-waste dismantling was eliminated, in comparison with doing nothing?". Recyclers could also substitute certain tools that elevate the risk of exposure while dismantling for tools with a lower risk of exposure. Given the informal nature of e-waste recycling, engineering controls are often not feasible and were not examined in this study. Administrative controls can be implemented to reduce heavy metal concentrations, such as training recyclers to not sit low to the ground or to reduce the time spent recycling. Note that exposures due to posture (eg sitting low to the ground) can also be reduced through engineering controls (eg workplace design), although for the purposes of our study, we simply classified these exposure determinants as administrative controls. Lastly, PPE, such as close-toed shoes, could be worn by recyclers to reduce contact with heavy metals.

With the goal of identifying the most effective workplace controls for these particular informal e-waste workers, we constructed an exposure determinant model that accounted for sampling selection bias and could handle small sample sizes with a large number of exposure determinants. The first objective of this study was to identify the best statistical model for predicting heavy metal concentrations, comparing traditional methods (linear regression, forward selection) with machine learning algorithms. Second, using the best-performing statistical model, we applied the causal inference framework to estimate hypothetical reductions in workers' blood and urinary heavy metal concentrations, which identified the most important determinants of Pb, Mn, Ni, and Cu concentrations. Finally, we assessed how the choice of statistical model and failure to consider exchangeability can bias the estimated hypothetical reductions.

Materials and methods

Study population and design

We conducted a cross-sectional study of exposures and health effects among e-waste recycling workers in Thailand and Chile in 2017. Detailed descriptions of the e-waste sites that participated in this study and characterization of the Chilean and Thai participants have been published (Seith et al. 2019; Yohannessen et al. 2019; Shkembi et al. 2021). Briefly, participants were at least 18 yr of age from either country and recruited through convenience sampling (ie nonrandom sample of workers who were interested in participating at each visited worksite). In Thailand, workers from 4 villages from a rural, low-income community in the Northeastern region were sampled; in Chile, informal workers were recruited from Santiago and Temuco,

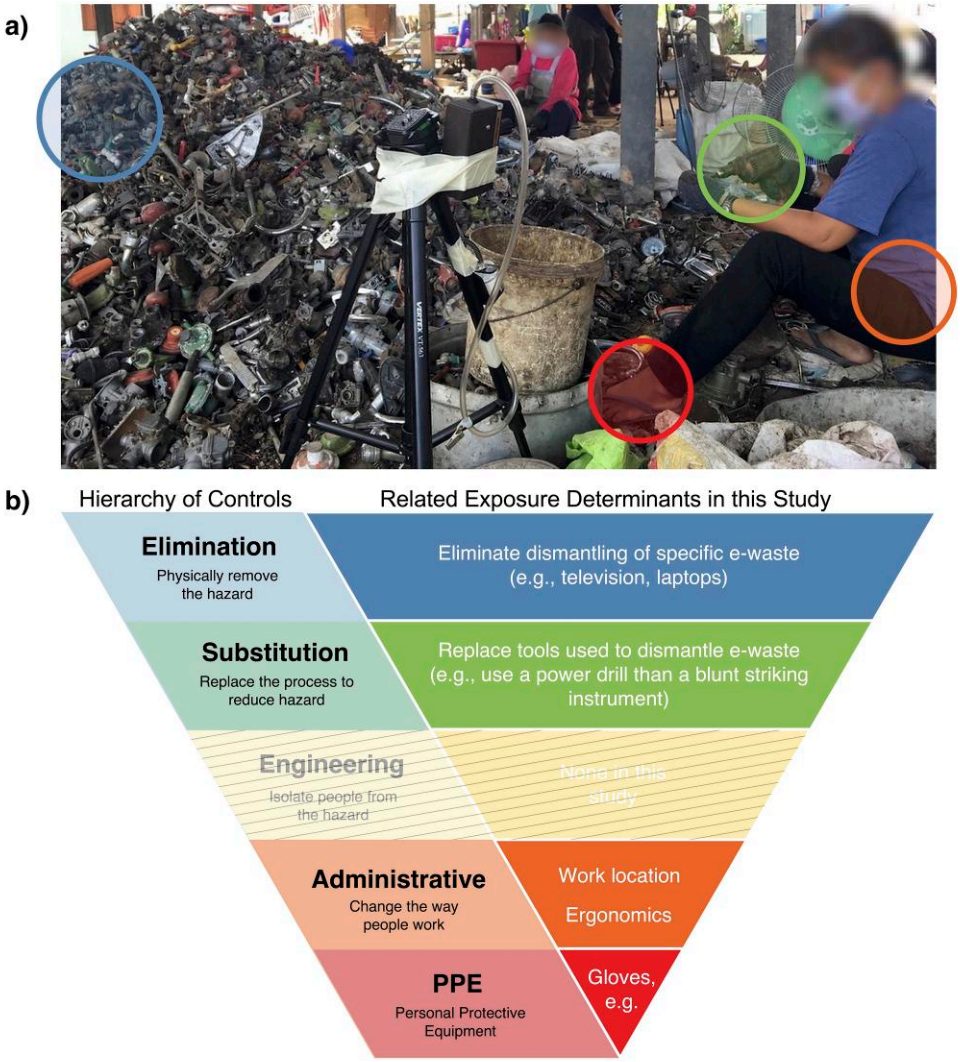

Fig. 1. Hierarchy of controls applied to informal e-waste recycling. (a) Common work environment for e-waste recyclers in Thailand. The colored circles correspond to possible hazards/hazardous work activities that could be mitigated through controls, examples of which are shown in (b) the hierarchy of controls for e-waste workers.

and formal workers were sampled from a recycling plant in Chillán. Research methods were approved by the University of Michigan Institutional Review Board (HUM0014562), Mae Fah Luang University (REH-59104), and the University of Chile-Santiago (Archive Project No. 101-2017; Act No 45). Informed consent was obtained from all participants before study enrollment.

Blood and urinary heavy metal measurement

Blood and urine samples were collected from workers. Whole blood was analyzed for Pb and Mn, and urine was analyzed for Ni and Cu in addition to creatinine, which was used for the adjustment of differences in urinary generation rates and hydration. All samples were frozen and transferred to the Thai Ministry of Public Health Central Laboratory for analysis, as previously described (Arain 2019; Seith et al. 2019). Blood samples were reported in µg/dL for Pb and µg/L for Mn. Urine samples are reported in µg/g creatinine.

Biomarker determinants and video substudy

A total of 44 potential determinants (relating to ergonomics, tool use, PPE use, and other job characteristics; see Appendix B for more details) of heavy metal concentrations were quantified through a video substudy of 41 e-waste workers out of the 226 enrolled in the study who consented to having their work shift recorded.

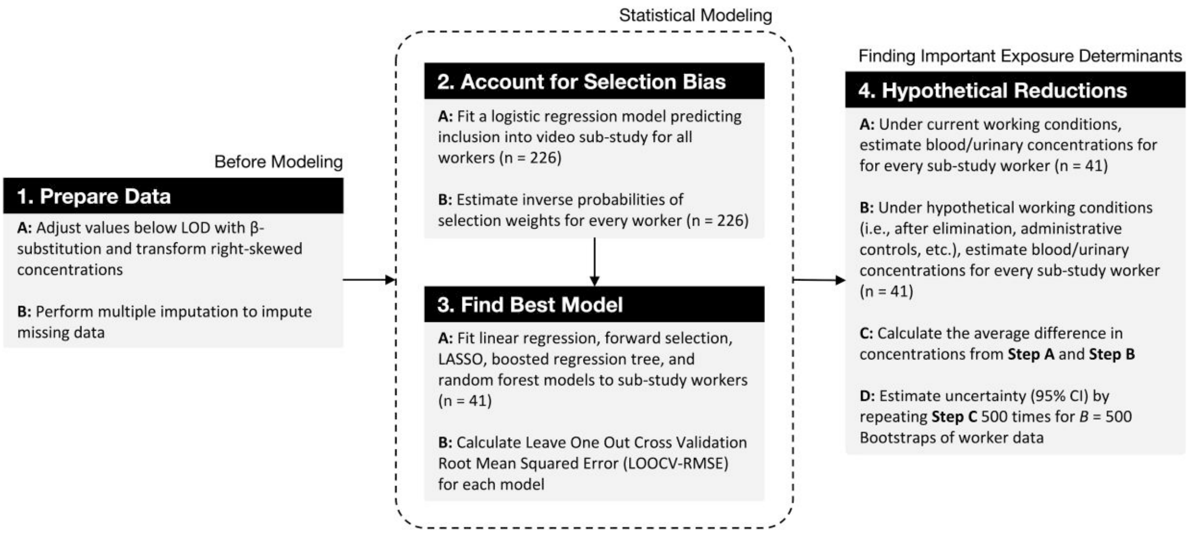

Fig. 2. A flowchart showing the analytical process to identify important determinants of heavy metal concentration among a subset of electronic waste workers who consented to video monitoring. LOD, limit of detection; LASSO, least absolute shrinkage and selection operator.

While some of the potential determinants may not be directly causally linked to increased heavy metal exposure, we included determinants that may be a proxy for exposure. For example, noise exposure does not increase Pb exposure; however, a noisy activity may generate excess dust that can increase Pb exposure. Thus, controlling noisy activities as a proxy for exposure could be an important workplace control.

Using a GoPro camera, videos were collected for workers performing e-waste recycling tasks at a fixed location within each worksite. A total of 10 min of footage per worker was selected by sampling at regular intervals of each video to yield 5- and 2-min increments. Each video recorded a worker's entire shift, except when they were not working (eg eating lunch). This video sampling approach was utilized because tasks across a whole shift of e-waste recycling typically remain the same; on a given day, a recycler dismantles similar e-waste parts using a typical set of tools for e-waste of different shapes and sizes. As such, the video sampling of each worker's shift is expected to adequately capture typical worker behaviors. A quantitative video assessment tool was created to rate the frequency of each potential determinant across the entire duration of each 2-min segment into 4 categories: 0 = never occurred or took zero seconds (s); 1 = occurred 1 to 3 times, or up to 20 s; 2 = occurred 4 to 10 times or 21 to 80 s; and 3 = occurred >10 times or >80 s. Five research assistants applied the quantitative tool to the entire video segment collection after reaching a kappa statistic >0.6 among a pilot set of videos, suggesting acceptable inter-rater agreement. The average rating (a continuous variable from 0 to 3) of each of the 44 potential determinants for every worker was calculated to represent a worker's "typical" job characteristics.

Worker characteristics

Personal characteristics of the workers were gathered with a 143-item survey on demographic characteristics, work history, and health status. Interviews were conducted face-to-face with a native language speaker, and the responses were entered directly into a Qualtrics electronic survey form (Qualtrics, Seattle, WA, USA). These characteristics were assessed to examine selection bias into the video substudy.

Statistical analysis

All analyses were completed using R v4.2.1. An overview of the analytical process is illustrated in Fig. 2. Briefly, we first prepare the data using standard data cleaning procedures (ie handling censored exposure values and imputing missing data). We then construct inverse probability weights (IPWs) to account for selection bias, making the subset of workers ($n = 41$) conditionally exchangeable with all workers ($n = 226$). We then find the best-performing statistical model, with regard to predictive performance, to finally estimate hypothetical reductions in workers' blood and urinary heavy metal concentrations.

Data preparation and descriptives

Blood Pb had 30% of samples below the analytical limit of detection (LOD). We adjusted these values using a β-substitution algorithm, which has been shown to have lower bias compared with common LOD/2 or

LOD/$\sqrt{2}$ (Ganser and Hewett 2010). None of the other heavy metal concentrations had samples below the LOD. Right-skewed concentrations for each heavy metal were log-transformed to achieve normality. Lastly, we performed multiple imputation, a statistical technique that predicts missing data using all other variables, where metal biomarkers were not measured, or the frequency of a determinant could not be estimated using the quantitative video assessment tool (Rubin 1996). A total of 10 data cells were missing and subsequently imputed (accounting for 0.5% of all data). After preparing the data, we performed simple statistical descriptive analyses to examine distributions of each heavy metal concentrations. Since unadjusted concentrations were right-skewed, we examined medians, interquartile ranges (IQRs), and maxima. We also examined how many workers were above Biological Exposure Index (BEIs) set by the American Conference of Government Industrial Hygienists (ACGIH) for each of the 4 heavy metals (ACGIH, 2023).

Inverse probability of selection weights

To address potential selection bias as a consequence of only having complete data for a subset of workers ($n = 41/226$), we created IPWs for the subset of workers (Cole and Hernán, 2008). We used a single logistic regression model to predict the probability of a worker being in the video substudy. We used *a priori* worker characteristics (age, number of family members supported, whether a worker earns above the minimum wage, and education level) as predictors of video substudy participation. The country of the worker was added as a random intercept in the model. Unstabilized weights were calculated as the inverse of the probability of each worker participating in the video substudy. These weights make the subset of workers sampled ($n = 41$) conditionally exchangeable with all participating workers ($n = 226$), reducing potential bias in statistical modeling due to selection bias.

Comparing potential exposure determinant models

To address our first study objective of comparing traditional statistical models with machine learning algorithms in order to identify the best exposure determinant model for predicting heavy metal concentrations, we compared the predictive performance of 5 separate regression models among the video substudy workers for the log-transformed concentrations of each of the 4 heavy metals: (i) traditional linear regression, (ii) forward selection linear regression, (iii) LASSO (least absolute shrinkage and selection operator, a type of shrinkage regression technique), (iv) gradient boosted regression trees, and (v) random forests. For each of the 4 outcomes, we fitted the 44 determinants as predictors. Each model type used IPWs for participation in the video substudy.

The predictive performance for each model varied based on the determinants that were chosen as important, except in the case of linear regression, which would have included all 44 determinants for prediction. We used the root-mean-squared error (RMSE) to compare each model's predictive performance. Given the small number of video substudy participants and the risk of a higher-than-expected influence for any individual participant, we conducted leave-one-out cross-validation (LOOCV) to examine RMSEs.

Briefly, for any given heavy metal, the procedure of estimating the LOOCV-RMSE of each model type involved iteratively removing 1 worker at a time from the model construction and predicting the log-transformed concentration for the worker that was not included in the model. We then calculated the RMSE of each worker by taking the difference in their observed log-transformed concentration with their model-predicted value, averaging each difference across all workers, and taking the square root of this value. A value closer to 0 indicates a better-performing model, and careful attention was paid to ensure the most suitable model did not have an excessively large LOOCV-RMSE given the small sample size. For this study, as a result, 41 separate models were constructed for every model type and for every heavy metal (41 participants × 5 model types × 4 heavy metals = 820 total models constructed).

Identifying important determinants with hypothetical reductions in heavy metals

To address the second study objective of identifying the most important determinants of Pb, Mn, Ni, and Cu concentration, we applied the causal inference framework to estimate hypothetical reductions in workers' blood and urinary heavy metal concentrations. Many machine learning algorithms, such as random forests, identify important variables typically as a combination of how often a variable is used and the amount of error it reduces in the algorithm. Since our primary goal is not the predictive capability of exposure determinants, but estimating potential reductions in exposure, we ignored traditional "variable importance" outputs of machine learning algorithms.

Instead, using the best-performing model (ie lowest LOOCV-RMSE), we estimated average hypothetical reductions in heavy metal concentrations, including uncertainty. First, we estimated the concentration of a given heavy metal for every worker under observed working conditions (denoted as $\hat{C}_{obs}$). Using our final model, we predicted the heavy metal concentrations using the observed determinants as model covariates. Next, we estimated the concentration of a given heavy metal for every worker under hypothetical working conditions (denoted as $\hat{C}_{hyp}$). Estimating $\hat{C}_{hyp}$ involved altering the observed values of each determinant, 1 determinant at a time. Thus, using the same final model, we predicted the heavy metal concentrations using the altered determinants as model covariates. For non-PPE-related determinants, each observed value for a given determinant was set to 0 to reflect workers stopping using/performing certain work tasks (eg what if workers stopped dismantling TVs or what if workers stopped sitting low to the ground when dismantling e-waste?). For PPE-related determinants, each observed value for a given determinant was set to 3 to reflect workers always using certain PPE (eg what if workers always wore close-toed shoes while dismantling e-waste?).

Finally, we calculated $\hat{C}_{obs} - \hat{C}_{hyp}$ to estimate the hypothetical reduction in heavy metal concentrations after controlling for a particular concentration determinant. We generated 500 bootstrap samples of the data and recalculated $\hat{C}_{obs} - \hat{C}_{hyp}$ to generate an overall reduction (by taking the median of the bootstrap estimates) and a 95% confidence interval (CI; by taking the 2.5th and 97.5th percentiles of the bootstrap estimates) to capture the uncertainty in our estimated reductions. We conducted this process for each of the 44 determinants and each heavy metal.

Sensitivity analysis

To illustrate how ignoring selection bias (IPW) and identification of best-performing models could influence the findings of concentration determinant modeling (Study objective 3), we re-estimated the average hypothetical reduction in blood Pb concentrations and reassessed the sets of proposed reductions under 4 scenarios: (i) using the best-performing model along with IPW, (ii) using the best-performing model but not IPW, (iii) using a more traditionally utilized, forward selection linear regression model along with IPW, and (iv) using a forward selection linear regression model without IPW.

Results

Among 226 informal e-waste workers from Thailand and Chile, no workers had blood Pb concentrations above the ACGIH BEI of 20 µg/dL for blood Pb (Table 1). The median (IQR) level of blood Pb was 2.1 µg/dL (1.9, 4.0). Workers had a median of 11.0 µg/L (6.6, 14.5) in blood Mn. Among urinary biomarkers, workers had concentrations of 2.8 µg/g creatinine (1.3, 4.9) and 10.6 µg/g creatinine (7.2, 18.6) of Ni and Cu, respectively. Workers in the video substudy ($n = 41$, or 18% of workers) had slightly lower blood Pb, blood Mn, and urinary Cu (Table S1) than those not in the video substudy (Table S2).

Table 1. Blood and urinary heavy metal concentrations among 226 informal e-waste workers in Thailand and Chile.

	Median (IQR)	Max	ACGIH BEI[a]
Pb, blood (µg/dL)	2.1 (1.9, 4.0)	12.4	20
Mn, blood (µg/L)	11.0 (6.6, 14.5)	86.2	…
Ni, urine (µg/g creatinine)	2.8 (1.3, 4.9)	38.6	…
Cu, urine (µg/g creatinine)	10.6 (7.2, 18.6)	370	…

[a]BEIs for Mn, Ni, and Cu do not exist. BEIs according to 2023 limits. BEI, Biological Exposure Indices; ACGIH, American Conference of Governmental Industrial Hygienists.

Comparing the 5 potential models to be used in the main analysis (Table S3; each including IPW), the random forest model had the lowest LOOCV-RMSE for all heavy metals except blood Mn ($\text{LOOCV-RMSE}_{\text{LASSO}} = 0.40$ versus $\text{LOOCV-RMSE}_{\text{Random Forest}} = 0.41$). This suggests that the random forest models were the best-performing models and hence were used in all subsequent analyses.

Figure 3 (and Table S4) displays the most important determinants for each heavy metal concentration, along with hypothetical reductions in blood and urinary concentrations. The most important determinants were identified as those that estimated a statistically significant 1% increase or reduction in blood/urinary concentrations. The results suggested that having workers stop bending their back while dismantling e-waste may be the most important determinant of blood Pb, urinary Ni, and urinary Cu. If all workers stopped bending their back while dismantling e-waste, we would predict an average reduction of 0.81 µg/dL (95% CI: 0.66, 0.98) in blood Pb, 1.31 µg/g creatinine (0.99, 1.65) in urinary Ni, and 1.79 µg/g creatinine (95% CI: 0.88, 3.29) in urinary Cu. The next most important determinant was the use of blunt striking instruments. If workers were to stop using blunt striking instruments to dismantle e-waste, we would predict an average reduction of 0.26 µg/dL (0.14, 0.42) in blood Pb, 2.25 µg/L (1.48, 3.08) in blood Mn, and 0.93 µg/g creatinine (0.60, 1.33) in urinary Ni.

Among other determinants of blood Pb, eliminating neck bending and repetitive hand motion when dismantling e-waste were factors that could significantly reduce concentrations (Fig. 3). Eliminating the dismantling of random e-waste parts, not using a chisel to dismantle e-waste, not sitting close to the ground, and not bending the neck were also important factors that could significantly reduce blood Mn concentrations. Not using chisels, not working in noisy

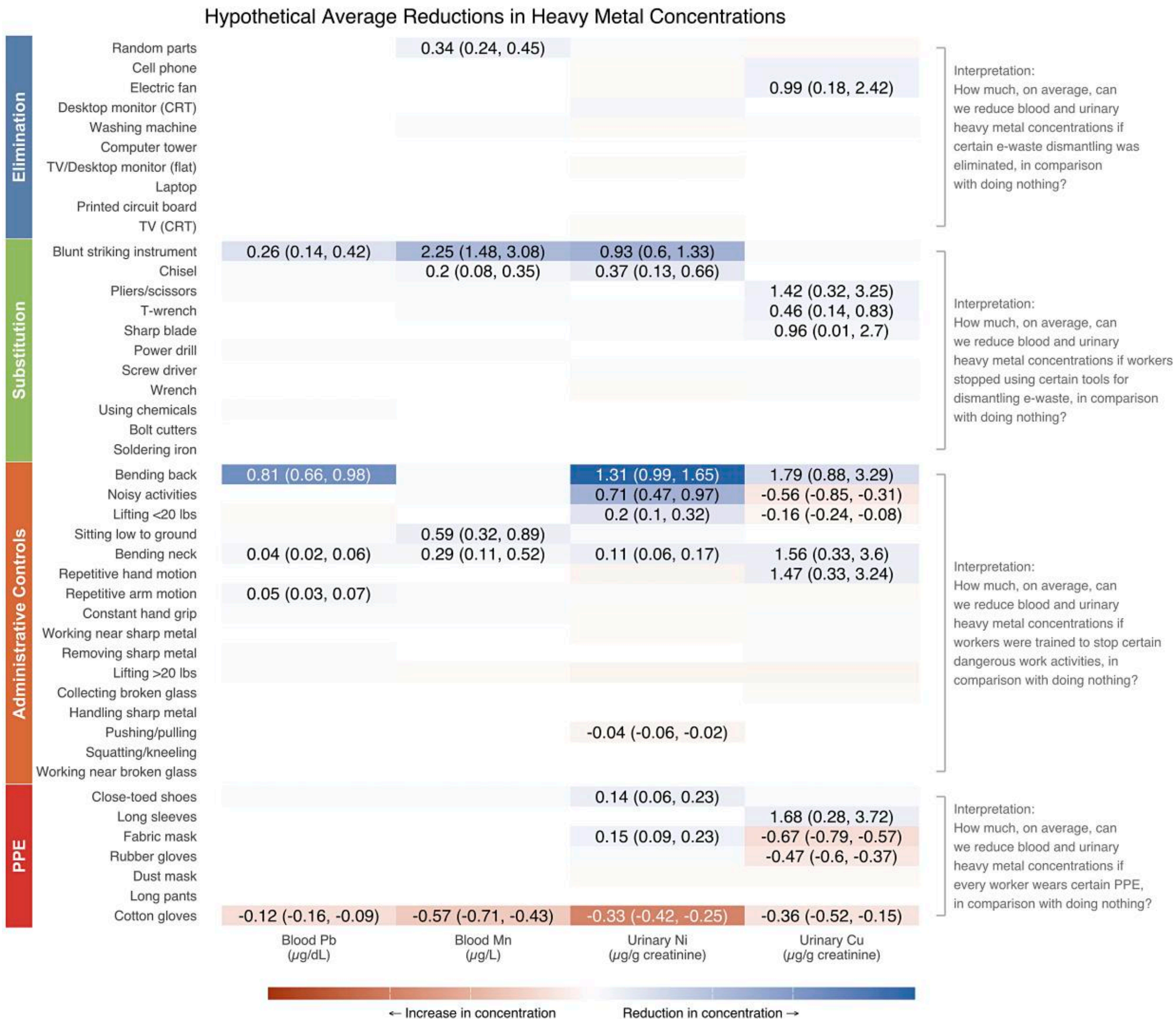

Fig. 3. Estimated average reductions in 4 heavy metal concentrations from hypothetical workplace controls. Hypothetical reductions were estimated from 4 random forest models (1 for each heavy metal) using the 44 workplace controls as predictors after accounting for selection bias using inverse probability weighting. Only reductions/increases that were statistically significant and had >1% difference are presented in the figure.

environmental, not lifting <20 lbs, not bending the neck, wearing close-toed shoes, and wearing fabric masks could significantly reduce urinary Ni concentrations. Eliminating the dismantling of electric fans was estimated to reduce urinary Cu concentrations by 0.99 µg/g creatinine (95% CI: 0.18, 2.42). Other important determinants were using pliers/scissors, a T-wrench, a sharp blade, neck bending, repetitive hand motions, and not wearing long sleeves. On the other hand, requiring all workers to use cotton gloves while dismantling e-waste was estimated to significantly increase concentrations of all heavy metals.

In a sensitivity analysis to illustrate the importance of using the best-performing statistical model (in this case, random forest) and using IPW to weight for participation in the video substudy (analogous to sampling only a subset of the most likely exposed workers), we compared the 5 most important determinants of blood Pb concentrations using 4 different models (Table 2). When using a random forest model with IPW (as in Fig. 3), back bending, using blunt striking instruments, lifting >20 lbs, using chisels, and wearing close-toed shoes were the 5 most important determinants of blood Pb concentrations among a total of 44 possible determinants. Not using IPW would have resulted in the 3rd, 4th and 5th most important determinants becoming the 7th, 8th, and 10th, respectively (Table 2).

A forward selection linear regression model with no IPW would have identified using blunt striking

Table 2. Comparison of most important biomarker determinants of blood Pb and their hypothetical average reductions (in μg/dL) based on model choice (random forest versus forward selection) and whether IPWs are included.

Most important determinants	Random forest & IPW		Random forest & no IPW		Forward selection & IPW		Forward selection & no IPW	
	Rank	Reduction (95% CI)	Rank	Reduction (95% CI)	Rank	Reduction (95% CI)	Rank	Reduction (95% CI)
Bending back	1	0.8 (0.7, 1.0)	1	0.8 (0.7, 0.9)	N/A	N/A	N/A	N/A
Blunt striking instrument	2	0.3 (0.1, 0.4)	2	0.2 (0.1, 0.3)	1	0.8 (0.5, 1.2)	1	0.8 (0.5, 1.2)
Lift >20 lbs	3	0.1 (0.0, 0.2)	7	0.0 (0.0, 0.1)	3	0.3 (0.2, 0.4)	2	0.3 (0.2, 0.5)
Chisel	4	0.1 (0.0, 0.1)	8	0.0 (0.0, 0.1)	N/A	N/A	N/A	N/A
Close-toed shoes	5	0.1 (0.0, 0.1)	10	0.0 (0.0, 0.1)	2	0.3 (0.2, 0.5)	3	0.3 (0.2, 0.5)

Ranking determined in order of largest, hypothetical average reduction in blood Pb, out of 44 possible determinants. Estimates with "N/A" could not be estimated as they were not selected by the forward selection model. Reference determinants are shaded in gray. IPW, inverse probability weights.

instruments, lifting >20 lbs, wearing close-toed shoes, dismantling electric fans, and using chemicals as the 5 most important determinants of blood Pb (Table 2). While the not bending the back was the most important determinant identified in the random forest model with IPW, it would not have identified back bending or chisel use as important determinants at all. A similar conclusion would have been made even after accounting for IPW.

Discussion

While many studies have demonstrated the usefulness of statistical modeling to identify important determinants of worker exposure, the applicability of such modeling in practice is impeded by a few common statistical and methodological issues. This study in particular focused on the issues of (i) selection bias, in which only a subset of all exposed workers is typically sampled and (ii) "small n, large p," where a small number of samples are taken despite a large number of possible determinants, making traditionally used statistical approaches such as multiple linear regression and stepwise selection infeasible.

Using an illustrative example of 44 potential determinants of heavy metal concentrations among 41 informal e-waste workers in Thailand and Chile, we used (i) inverse probability of selection weights and (ii) machine learning algorithms, respectively, to overcome these issues. This allowed the estimation of reductions in worker dose based on hypothetical controls in the informal e-waste workplace environment. Using this approach, our final random forests model with inverse probability weighting found that stopping workers from bending their backs and stopping the use of blunt striking instruments while dismantling e-waste were the most important determinants that could significantly reduce exposure to heavy metals.

Ignoring inverse probability weighting and employing stepwise selection regression methods instead (a more "traditional" approach of exposure determinant modeling) failed to identify back bending as the most important determinant of Pb exposure. This discrepancy is likely due to the inability of stepwise regression to handle dependence and interactions among determinants. Tree-based machine learning algorithms like random forests, on the other hand, have the particular advantage of allowing for dependence between all possible determinants. This better mimics how determinants in the workplace interact with one another and could explain why the random forests model displayed the best performance with regard to LOOCV-RMSE. While inclusion of the IPWs did not alter the rankings of the determinants as greatly as model choice, slight differences were still observed. This is likely due to the fact that those sampled in the video substudy had lower heavy metal concentrations than those not in the substudy, suggesting the 2 groups were not exchangeable and should have been made conditionally exchangeable through inverse probability weighting to reduce bias in the hypothetical reduction estimates. The difference in the observed concentrations could have been due to random chance or because the workers knew they were being monitored and altered their work practices.

How causal is our causal inference modeling?

Despite the advantages of the tree-based machine learning algorithm over traditional regression methods, our hypothetical reductions can only be interpreted causally if all of the assumptions of the causal inference framework are met. The main assumption this paper focused on that is typically violated in exposure monitoring assessment is exchangeability. We demonstrated how inverse probability weighting can account for the selection bias of the subset of workers who were sampled in the video substudy ($n = 41$ of 226 participating workers). It is important to note that an improperly specified model that does not properly account for confounding of the selection bias will not weight the 2 groups (those in the subset and those not) correctly. As such, carefully considering and accounting for factors that would impact the selection bias is crucial to not introduce unmeasured confounding that violates the exchangeability assumption.

The 2 other assumptions of causal inference are consistency and positivity (Hernán and Robins 2020). These assumptions were met in this specific study, but violations of these assumptions in other concentration determinant models would limit the causal interpretation of hypothetical reductions. Consistency implies that potential determinants are well-defined. In this study, we defined each and every one of the determinants as a combination of frequency and duration (see Materials and methods section). Positivity implies that there is variability in determinants. For example, we excluded the determinant of wearing a respirator from our modeling because we observed that none of the 41 substudy workers ever wore a respirator. Thus, estimating the hypothetical reductions in heavy metal concentrations from having all workers wear a respirator would be impossible to estimate given the information we have observed from this set of workers. This reiterates the importance of defining exposure or concentration determinants well.

Causal inference versus Bradford-Hill criteria of causality

The most common limitation of cross-sectional, observational studies is that causality cannot be inferred. This is in contradiction to the premise of this study and the causal inference framework. Indeed, while causality in cross-sectional studies is more difficult to infer, this does not make causal inference impossible. Beyond ensuring that the assumptions (exchangeability, consistency, and positivity) of the causal inference framework are met, we can also assess the Bradford-Hill criteria for establishing causality (strength, consistency, temporality, gradient, specificity, plausibility, coherence, experiment, and analogy; Hill 1965). While the Bradford-Hill criteria are typically applied to a body of literature, they can also be useful for considering the causal limitations of 1 particular study.

Generally, the criteria of specificity, plausibility, coherence, consistency, and experiment are satisfied in a well-designed exposure/biomarker determinant study. This is because determinants are typically included due to their established impacts on exposure/dose (eg it is a fact that appropriate and properly fit PPE reduces worker dose). For example, the modeling estimated that stopping workers from using blunt striking instruments would substantially reduce Pb, Mn, and Ni concentrations. This seems plausible because blunt striking instruments can generate dust. Similarly, the estimated reductions from stopping workers from bending their backs are plausible because back bending may be putting workers closer to the source of the hazard, increasing their exposure.

The criteria of gradient and temporality are likely to hold in this study but are the 2 criteria with the most limitations in exposure determinant modeling. In this study, we estimated hypothetical reductions through binary workplace controls (eg eliminating certain e-waste dismantling), making a gradient difficult to ascertain. Lastly, temporality is difficult to satisfy in any cross-sectional study, as the determinants were measured at the same time as the heavy metal concentrations. However, the central concern with temporality is reverse causation, where the predictors do not cause the outcome, but the outcome causes the predictors. This assumption is not likely to be violated in exposure determinant modeling. For example, Pb concentrations do not likely influence the use of blunt striking instruments to dismantle e-waste. Further, concentrations were low enough that they were unlikely to produce any health effect that might change exposures (eg change of work). Another central concern with temporality is that the determinants occurred before the exposures were measured. This may be particularly concerning for some heavy metal biomarkers, such as blood Pb, which have a long half-life and a cross-sectional assessment of determinants and biomarkers may be tenuous due to previous exposure. Thus, the results of an exposure determinant model must be carefully interpreted in the context of temporality. Were there previous exposures? Are there exposures occurring outside of the workplace? Are the potential measured determinants of exposure generally constant over time? Randomly surveying workers to measure potential determinants of exposure/concentration also helps strengthen this assumption but can also unintentionally alter worker behaviors, as knowing that their behaviors may be observed could cause some workers to uncharacteristically behave in a safer manner.

Despite this limitation, the satisfaction of the other criteria for causality, in combination with utilizing the causal inference framework, makes the approach outlined in this study robust for identifying important determinants of workplace exposures. Beyond simply estimating average reductions in exposures/concentrations, the causal inference framework can also be used to estimate reductions in the number of workers who may be exposed above an occupational exposure limit. We did not utilize this approach in this study because 3 of 4 heavy metals examined do not have ACGIH BEIs and none of the workers were overexposed above the blood Pb BEI of 20 μg/dL. However, this approach can also be extended to estimating reductions in how many workers are above 50% or 10% of an OEL (American Industrial Hygiene Association Exposure Assessment Strategies Committee 2015). Future approaches should also consider whether nonoccupational exposures could be influencing workplace exposures to account for further, unmeasured confounding.

Statistical limitations

With all statistical modeling, there are limitations that must be carefully considered. First, we relied on several statistical analyses (eg β-substitution, multiple imputation, inverse probability weighting, and machine learning models) to minimize the bias in this small sample population of workers. This approach was motivated by "real-world," small, and biased industrial hygiene data. However, sophisticated statistical techniques should not be a substitute for improved study design (such as increased sample size and random sampling of workers), as weak study design can fundamentally bias results beyond what statistical techniques can minimize.

Similarly, it's important to note that the "right" statistical model can never be selected. Thus, while the random forest model with inverse probability weighting performed the best in this particular study, the estimated hypothetical reductions may still be biased. However, given the theoretical frameworks that underpin the selection of this model, we would expect the hypothetical reductions estimated from this model to be the least biased among the set of models we compared here. Specifically, given that the final model had the lowest LOOCV-RMSE, we would expect it to be the least prone to overlearning and thus likely to estimate the least biased hypothetical reductions. Additionally, given that the final model accounted for selection bias with inverse probability weighting, we expect it to estimate less biased hypothetical reductions than if IPWs were not incorporated into the model.

The performance of statistical models may also be influenced by the model specifications. The choice of the regularization parameter in LASSO, the number of trees grown in random forests, and the tree depth of a gradient boosted regression tree are all examples of model specifications that should be optimized in future applications of these methods in order to improve the performance of the model.

An additional source of bias that can influence the statistical models is potential misclassification in the determinants. Due to logistical constraints, our study team was unable to watch each worker's shift live to quantify potential determinants; rather, we video recorded these shifts. While this is advantageous to increase the sample size, this video recording may have impeded the ability to interpret the frequency of the determinants and lead to misclassification in our predictors. Similarly, we used multiple imputation to not reduce the sample size of our analysis further due to missing values during data collection. While our assumption was that our data were missing randomly, it is important to note that multiple imputation can introduce additional bias into statistical analyses if the data are not missing randomly (ie due to some systematic error in data collection).

In this illustrative example, the sample size ($n = 41$) was small, and the number of determinants was large ($p = 44$). While this was intentional to mimic practical industrial hygiene data, attention needs to be paid toward whether an acceptable sample size has been reached. Power calculations to derive suitable sample sizes are common for traditional statistical models like linear regression, but not for machine learning algorithms. For this reason, we determined that we have an acceptable sample size because the predictive error of the random forest model was minimal, and we did not estimate very wide 95% CIs (which would indicate high uncertainty in our estimates). For future applications of these methods, it is important to note that there are not simple rules of thumb to follow and that one must carefully examine whether the predictive performance of the modeling is low and whether the uncertainty in the hypothetical reductions is small.

Caution must also be taken regarding interpretation of the possible workplace controls estimated here. While it is important to incorporate as many reasonable, potential determinants of exposure in the statistical modeling (as we did here) so that potential bias due to unmeasured confounding is minimized, not all controls may be feasible in a particular work environment. For example, in our illustrative example, eliminating electric fans could hypothetically reduce urinary Cu concentrations. However, this potential reduction must be weighed against the economic loss of no longer dismantling that type of e-waste. This is where the judgment of an industrial hygienist is critical to the modeling approach so that potential controls are carefully balanced between estimated exposure reductions and economic feasibility.

Conclusion

Our causal inference approach with machine learning algorithms overcomes the common limitations of cross-sectional, exposure determinant modeling and produces easy-to-interpret estimates of concentration reductions from hypothetical workplace controls. This can aid industrial hygienists in choosing the most effective hazard controls that are contextualized to their particular work environment.

Acknowledgments

The authors acknowledge all of the e-waste workers in Chile and Thailand who graciously participated in the study, without whom none of this research would have been possible. They also thank the Graham Sustainability Institute at the University of Michigan for providing funding for this research.

Supplementary material

Supplementary material is available at *Annals of Work Exposures and Health* online.

Funding

This research was funded by a grant from the Graham Sustainability Institute of the University of Michigan. AS was additionally supported by a National Institute of Occupational Safety and Health Targeted Research Training Scholarship (2T42OH008455-19).

Conflict of interest

None declared.

Data availability

The data underlying these analyses are available from the authors upon reasonable request. The code used for statistical analyses is stored in a publicly available repository at https://github.com/abasshkembi/causal-inference-IH

References

ACGIH. 2023. 2023 TLVs and BEIs based on the documentation of the threshold limit values for chemical substances and physical agents. American Conference of Governmental Industrial Hygienists.

American Industrial Hygiene Association Exposure Assessment Strategies Committee. 2015. A strategy for assessing and managing occupational exposures. 4th edn. AIHA.

Arain AL. 2019. Environmental Health Impacts of Informal Electronic Waste Recycling (Doctoral Dissertation).

Breiman L. 2001. Random forests. Mach Learn. 45:5–32. https://doi.org/10.1023/A:1010933404324.

Burstyn I, Teschke K. 1999. Studying the determinants of exposure: a review of methods. Am Ind Hyg Assoc J. 60:57–72. https://doi.org/10.1080/00028899908984423.

Cole SR, Hernán MA. 2008. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 168:656–664. https://doi.org/10.1093/aje/kwn164.

Efroymson MA. 1960. Multiple regression analysis. In: Ralston A, Wilf H, editors. Mathematical methods for digital computers. Wiley. p. 191–203.

Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. Ann Stat. 29:1189–1232. https://doi.org/10.1214/aos/1013203451.

Ganser GH, Hewett P. 2010. An accurate substitution method for analyzing censored data. J Occup Environ Hyg. 7: 233–244. https://doi.org/10.1080/15459621003609713.

Green SB. 1991. How many subjects does it take to do a regression analysis. Multivariate Behav Res. 26:499–510. https://doi.org/10.1207/s15327906mbr2603_7.

Heacock M et al. 2016. E-waste and harm to vulnerable populations: a growing global problem. Environ Health Perspect. 124:550–555. https://doi.org/10.1289/ehp.1509699.

Hernán MA, Robins JM. 2020. Causal inference: what if. Chapman & Hall/CRC.

Hill AB. 1965. The environment and disease: association or causation? Proc R Soc Med. 58:295–300. https://doi.org/10.1177/003591576505800503.

Mooney SJ, Keil AP, Westreich DJ. 2021. Thirteen questions about using machine learning in causal research (you won't believe the answer to number 10!). Am J Epidemiol. 190:1476–1482. https://doi.org/10.1093/aje/kwab047.

Okeme JO, Arrandale VH. 2019. Electronic waste recycling: occupational exposures and work-related health effects. Curr Environ Health Rep. 6:256–268. https://doi.org/10.1007/s40572-019-00255-3.

Rubin DB. 1996. Multiple imputation after 18+ years. J Am Stat Assoc. 91:473–489. https://doi.org/10.1080/01621459.1996.10476908.

Seith R et al. 2019. Self-reported health and metal body burden in an electronic waste recycling community in Northeastern Thailand. J Occup Environ Med. 61:905–909. https://doi.org/10.1097/JOM.0000000000001697.

Shkembi A et al. 2021. Work task association with lead urine and blood concentrations in informal electronic waste recyclers in Thailand and Chile. Int J Environ Res Public Health. 18:10580. https://doi.org/10.3390/ijerph182010580.

Smith G. 2018. Step away from stepwise. J Big Data. 5:32. https://doi.org/10.1186/s40537-018-0143-6.

Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol. 58:267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Tielemans E et al. 2011. Advanced REACH Tool (ART): Overview of Version 1.0 and Research Needs. Ann Occup Hyg. 5:949–956. https://doi.org/10.1093/annhyg/mer094.

Vandenbroucke JP, Broadbent A, Pearce N. 2016. Causality and causal inference in epidemiology: the need for a pluralistic approach. Int J Epidemiol. 45:1776–1786. https://doi.org/10.1093/ije/dyv341.

Yohannessen K et al. 2019. Health assessment of electronic waste workers in Chile: participant characterization. Int J Environ Res Public Health. 16:386. https://doi.org/10.3390/ijerph16030386.