Safety and Health at Work 13 (2022) 493–499

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Original article

A Pilot Establishment of the Job-Exposure Matrix of Lead Using the Standard Process Code of Nationwide Exposure Databases in Korea

Ju-Hyun Park [1], Sangjun Choi [2,*], Dong-Hee Koh [3], Dae Sung Lim [4], Donguk Park [5], Hwan-Cheol Kim [6], Sang-Gil Lee [7], Jihye Lee [7], Ji Seon Lim [1], Yeji Sung [2], Kyoung Yoon Ko [2]

[1] *Department of Statistics, Dongguk University, Seoul, Republic of Korea*
[2] *Graduate School of Public Health and Healthcare Management, The Catholic University of Korea, Seoul, Republic of Korea*
[3] *Department of Occupational and Environmental Medicine, International St. Mary's Hospital, Catholic Kwandong University, Incheon, Republic of Korea*
[4] *Hansung Health and Safety Technology Co., Ltd., Daejeon, Republic of Korea*
[5] *Department of Environmental Health, Korea National Open University, Seoul, Republic of Korea*
[6] *Department of Occupational and Environmental Medicine, Inha University, Incheon, Republic of Korea*
[7] *Occupational Safety and Health Research Institute, Korea Occupational Safety and Health Agency, Ulsan, Republic of Korea*

A R T I C L E I N F O

Article history:
Received 4 January 2022
Received in revised form
30 August 2022
Accepted 1 September 2022
Available online 9 September 2022

Keywords:
Contrast
Job-exposure matrix
Lead
Standard process

A B S T R A C T

Background: The purpose of this study is to construct a job-exposure matrix for lead that accounts for industry and work processes within industries using a nationwide exposure database.
Methods: We used the work environment measurement data (WEMD) of lead monitored nationwide from 2015 to 2016. Industrial hygienists standardized the work process codes in the database to 37 standard process and extracted key index words for each process. A total of 37 standardized process codes were allocated to each measurement based on an automated key word search based on the degree of agreement between the measurement information and the standard process index. Summary statistics, including the arithmetic mean, geometric mean, and 95th percentile level (X95), was calculated according to industry, process, and industry process. Using statistical parameters of contrast and precision, we compared the similarity of exposure groups by industry, process, and industry process.
Results: The exposure intensity of lead was estimated for 583 exposure groups combined with 128 industry and 35 process. The X95 value of the "casting" process of the "manufacture of basic precious and non-ferrous metals" industry was 53.29 $\mu g/m^3$, exceeding the occupational exposure limit of 50 $\mu g/m^3$. Regardless of the limitation of the minimum number of samples in the exposure group, higher contrast was observed when the exposure groups were by industry process than by industry or process.
Conclusion: We evaluated the exposure intensities of lead by combination of industry and process. The results will be helpful in determining more accurate information regarding exposure in lead-related epidemiological studies.

1. Introduction

To prevent occupational diseases and establish appropriate safety and health policies, a national exposure surveillance system is required. Similar to Germany's MEGA [1,2] and France's COLCHIC [3], Republic of Korea has the Workplace Environment Measurement Database (WEMD), a nationwide exposure database collected by the work environment monitoring (WEM) system under the Korean Occupational Safety and Health Act (OSHAct). All employers are obligated to perform a regular evaluation of the exposure level of workers to harmful factors twice a year [4]. Measurement of the working environment is performed by

Ju-Hyun Park: https://orcid.org/0000-0001-9675-6475; Sangjun Choi: https://orcid.org/0000-0001-8787-7216; Dong-Hee Koh: https://orcid.org/0000-0002-2868-4411; Dae Sung Lim: https://orcid.org/0000-0003-4190-0390; Donguk Park: https://orcid.org/0000-0003-3847-7392; Hwan-Cheol Kim: https://orcid.org/0000-0002-3635-1297; Sang-Gil Lee: https://orcid.org/0000-0001-8173-3940; Jihye Lee: https://orcid.org/0000-0003-0484-6698; Ji Seon Lim: https://orcid.org/0000-0002-6642-5031; Yeji Sung: https://orcid.org/0000-0003-2267-5490; Kyoung Yoon Ko: https://orcid.org/0000-0001-6616-5164

* Corresponding author. Graduate School of Public Health and Healthcare Management, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul, 06591, Republic of Korea.
E-mail address: junilane@gmail.com (S. Choi).

https://doi.org/10.1016/j.shaw.2022.09.001

industrial hygienists from monitoring institutions designated by the Ministry of Employment and Labor (MOEL). Since 2002, the results of measurements have been sent to the Republic of Korea Occupational Safety and Health Agency (KOSHA) through the "KOSHA to Business (K2B)" computer system (https://k2b.kosha.or.kr/index.do) and stored in WEMD, which contains data on more than 10 million quantitative exposures for 190 types of harmful chemical factors.

Monitoring institutions are required to enter standardized industry and process information for the workplace where the measurement was conducted. For example, if the noise was measured in "Assembly Part 1" and "Assembly Part 2" at an automobile manufacturing site, the standardized process name 'Assembly' must be entered additionally. WEMD has been used in previous studies to estimate the exposure prevalence and exposure intensity to asbestos [5], benzene [6], lead [7–9], and 20 human carcinogens [10]. However, all previous studies evaluated exposure characteristics according to industry, thus there are limitations in identifying more specific exposure characteristics related to jobs and processes performed by employees within each industry.

Until 2019, a total of 1,390 process standard codes (SPC_{2019}) were used in WEMD. However, since only a small number of process codes were explained with definition, the K2B users assigned a code of "not otherwise classified (NOC)" to many processes. Therefore, the K2B system was changed in 2020 to lower the proportion of NOC code input by allowing the monitoring institutions to generate a new process code with a brief description if it is not on the list. Although the proportion of the NOC code input was lowered, the number of newly generated codes was increased to 2,807 in 2020 (SPC_{2020}), making it difficult to identify exposure characteristics using standard process information.

In this article, we propose a new method of standardization for the K2B process code and aims to establish a job-exposure matrix (JEM) for each industry process by applying it on a trial basis to the WEMD of lead. We selected lead because it was evaluated for appropriate exposure intensity indicators for WEMD in previous study [9]. The results of this study could be applied in the future to other hazardous agents in WEMD for the development of JEM.

2. Methods

2.1. Data collection

A total of 47,575 WEMD measurements of airborne lead reported to KOSHA from 2015 to 2016 were used to construct the industry process JEM. For comparison purposes, we extracted the same lead data from WEMD as in the previous study [9] that conducted exposure intensity evaluation according to industry only. The WEMD of the lead includes information such as the industry code and process code of the measurement and the process name, department, unit workplace, measurement time, and measured concentration entered directly by the monitoring institutions. A total of 47,280 WEMD datasets of lead, excluding those with less than 4 hours of measurement time and inaccurate industry code information, were analyzed.

2.2. Standard industry classification

The industrial classification of WEMD is input as a subclassification (5 digits) of the Republic of Korea Standard Industrial Classification, 9th revision (KSIC-9). KSIC-9 is based on the International Standard Industrial Classification (ISIC), 4th revision, and they are interconnected. The three-digit subclassification code of the Standard Industrial Classification (SIC) was used in this study.

2.3. Standard process classification

Development of the new standard process classification codes, SPC_{2021}, proceeded as follows and is summarized in Fig. 1. First, 1,390 codes of SPC_{2019} and 2,807 codes of SPC_{2020} were reclassified into 37 new standard processes codes based on the descriptions of the processes and the exposure similarities reviewed by industrial hygiene experts (SC and DP). For a new standard process code of SPC_{2021}, nouns were extracted from the process names and descriptions of corresponding SPC_{2019} and SPC_{2020} using functions of an R package KoNLP (e.g., extractNoun) and were configured as key index words that best described the standard process and had characteristics that were exclusive with other processes.

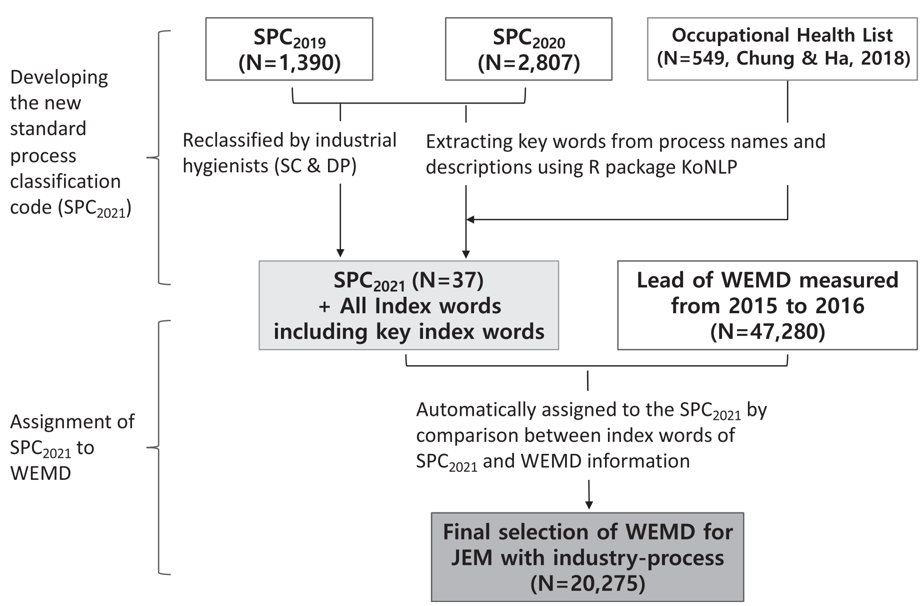

Fig. 1. Flow chart for the development of new standard process codes and automatic allocation (SPC, standard process classification; WEMD, workplace environment measurement database).

All index words for the standard process were then constructed by combining all the extracted words including the key index words with the nouns further extracted using the same method from the process names and outlines of 549 processes on the Occupational Health List [11], which contains process-related information that the key index words might miss. For example, in the case of "Injection (SPC2)," five words (injection, insertion, addition, charging, and input) are the key indexes, but 75 words, including other variants of the key index "input," such as 'container input" and "catalyst input," were composed of all index words. Supplementary Table 1 shows the process names, process definitions, key index words, and all index words of the 37 final SPC_{2021}.

An allocation algorithm was created to automatically assign the standard processes to WEMD of airborne lead monitored from 2015 to 2016 using the key index words and all index words of the newly created SPC_{2021} (Fig. 1). The algorithm first compared the name of the existing process and unit workplace of each WEM data with the key index words of SPC_{2021} and then allocated the standard process that showed the highest matching proportion. The algorithm also performed an additional process allocation in the same way but with all index words of SPC_{2021}, separately. Such an algorithm with two sets of index words with different priorities was intended to increase the specificity of the process allocation, and in a case where the existing process name and the unit workplace were not specific enough, multiple standard process allocations might result with the same highest matching proportion. Note that no allocation can occur if neither the name of the existing process nor the unit workplace has information sensitive enough to be matched with either of two sets of index words.

The allocation results of WEMD of airborne lead depending on whether no, single, or two allocations occurred with two sets of index words, resulting in a total of 11 combinations, are shown in Supplementary Table 2. Case 1 in the table, which matched both key index words and all index words equally with a single process code of SPC_{2021} compared with measurement information, was considered a highly reliable case. A total of 20,275 airborne lead measurements belonging to Case 1, the most reliable case, were used for construction of JEM.

2.4. Evaluation of similar exposure level of job-exposure matrix for each industry and process

Similar exposure levels were assessed using the contrast between exposure groups and the precision within each exposure group, which are the statistics commonly used to evaluate the classification suitability of exposure groups in epidemiologic studies [12,13]. Three different grouping schemes were considered: 1) industry only, 2) process only, and 3) a combination of industry process and the contrast (ε) and precision (π) for each grouping scheme were calculated with the sample variance between groups of lead exposure concentrations (S^2_{BG}), the sample variance within the group (S^2_{WG}), and the sample variance within the individual company (S^2_{WC}) using the following equations:

$$\varepsilon = \frac{S^2_{BG}}{S^2_{BG} + S^2_{WG}} \tag{1}$$

$$\pi = \frac{1}{\sqrt{S^2_{WG}\big/n_1 + S^2_{WC}\big/n_2}} \tag{2}$$

where n_1 is the number of companies for each exposure group, and n_2 is the number of measurements for each exposure group. As shown in Equation (1), the contrast measures the exposure similarity between groups and can take a value close to 1 if a grouping scheme separates distinct groups well with each being homogeneous. On the other hand, a large precision value for a group means that the similarity between the company-specific measurements within the exposure group and/or in a company's repetitive measurements is high. Because the precision was calculated for each group, the median precision of all exposure groups for each grouping scheme was reported as a representative summary statistic. In summary, the contrast and precision for a grouping scheme increase, as the variance within the group decreases, which is the case when groups are classified as well by the grouping scheme.

2.5. Statistical analysis

While there may be non-detected data in WEMD, checking information on the accurate limit of detection (LOD) is not feasible because multiple monitoring institutions did not input such information into the K2B system. However, participation in quality control programs operated by KOSHA [4,14] is required for all measuring institutions and performance of measurement analysis according to KOSHA guides similar to those of the US National Institute of Occupational Safety and Health [15] is required. Therefore, using the analytical LOD value suggested in KOSHA Guide A-2-2012, the standard method for measurement and analysis of airborne lead, and the 6-hour measurement time, the LOD value of the airborne lead concentration was estimated as 1.8 μg/m^3. According to the results of a recently published study [9] reporting that the simple replacement was better than the maximum likelihood estimation (MLE) method for correcting the value below LOD, left-censored data, using the airborne lead WEMD data [9], this study calculated the summary statistics of censoring rate (in other words of proportion of data not being detected), the arithmetic mean (AM), standard deviation (SD), geometric mean (GM), geometric standard deviation (GSD), and summary statistics of the 95th percentile (X95) of each exposure group (by industry, process, and industry process) by replacing the concentration of airborne lead below LOD to LOD/2. When the censoring rate is very high, for example, >0.8, it is highly likely that all the summary statistics are unstable with the simple replacement method for LOD and some values of AM, GM, and X95 are less than the LOD of 1.8 μg/m^3. In such a case, these statistics were presented as "<1.8 μg/m^3" and the variability statistics, SD and GSD, were not calculated and presented as NA (not available). In an extremely high censoring rate>0.95, all the summary statistics are meaningless and thus treated as NA.

The sample variances between exposure groups (S^2_{BG}), within a group (S^2_{WG}), and within a company (S^2_{WC}) needed to calculate the contrast and precision were obtained through the following analysis of variance model with two-way random effects, one for the exposure group and the other for individual companies within each exposure group.

$$Y_{ijk} = \ln(X_{ijk}) = \mu + \alpha_i + \beta_{ij} + \in_{ijk'} \tag{3}$$

where Y_{ijk} is the log-transformed value of the k-th airborne lead concentration measurement in the j-th company of the i-th exposure group, μ is the total average value, $\alpha_i = \sim N(0, \sigma^2_{BG})$ is the random effect of the i-th exposure group, μ is the random effect of the j-th company of the i-th exposure group, and $\beta_{ii} = \sim N(0, \sigma^2_{WG})$ is the random variation of the k-th log-transformed airborne lead concentration in the j-th company of the i-th exposure group. The contrast ($\varepsilon_{ii} = \sim N(0, \sigma^2_{WC})$) and precision ($\pi$) of each exposure group were calculated using the estimates of variance terms, that is, $S^2_{BG} = \hat{\sigma}^2_{BG}$, $S^2_{WG} = \hat{\sigma}^2_{WG}$, and $S^2_{BC} = \hat{\sigma}^2_{BC}$, obtained by fitting the model in Equation (3) to the data. All statistical analyses were performed using R version 4.1.0 [16].

3. Results

JEM for each minor industry group (three-digit) was constructed using the 47,280 results of lead measurement. Among 66 industries with 20 or more measurements and estimated 95 percentile (X95) values, the lead exposure level in the top 10 exposure industries and the bottom 10 exposure industries according to X95 of lead level is shown in Table 1. The X95 value of "manufacture of basic precious and non-ferrous metals (SIC 242)" was the highest at 23.68 μg/m^3, and a censoring rate of 85% or higher was observed for all 10 low-exposure industries. The statistics of airborne lead for all 133 industries are summarized in Supplementary Table 3.

Reclassification of 37 SPC$_{2021}$ through the proposed allocation algorithm using the two sets of index words created JEMs of 583 industry process groups for 20,275 lead measurement data measured in 125 industries. Detailed results of airborne lead by industry process groups and by process groups can be found in Supplementary Tables 4 and 5, respectively. The largest number of measured samples was observed for the "welding" process, with 5,433, and the "inspection" process was the most common standard process identified in 92 industries.

Because there was considerable variation in the number of measurements between the industry process groups and the overall censoring rate was too high (80%), further analysis was restricted to industry process groups with 20 or more measurements and a censoring rate less than 50%, resulting in 18 groups in seven industries; the results for the standard process with the highest GM of exposure level in the seven industries are shown in Table 2. The X95 value of the "casting" (SPC 4) process of the "manufacture of basic precious and non-ferrous metals" (SIC 242) industry was 53.29 μg/m^3, exceeding the occupational exposure limit (OEL) of 50 μg/m^3. The "molding" (SPC 9) process of the "cast of metals" (SIC 243) industry and the "mixing" (SPC 19) process of the "manufacture of primary cells and batteries and accumulators" (SIC 282) industry exceeded 50% of the OEL with 45.26 μg/m^3 and 33.88 μg/m^3, respectively.

The contrast and precision for three grouping schemes were calculated and are shown with other variant characteristics such as the number of groups and the sample variance within-group in Table 3. Higher contrast values were observed for the grouping schemes using the new SPC$_{2021}$ compared with grouping by industry only (the reference group) with the highest value for the grouping by industry and process together. The highest precision value was observed for grouping by industry only. This is because the grouping scheme resulted in significantly fewer groups than the other two grouping schemes, whereas the number of companies per exposure group and the three GSDs were comparable. The same results hold regardless of the limitation of the minimum number of measurements in the exposure group.

4. Discussion

The population-based JEM is useful for estimating exposure levels in epidemiologic studies of occupational factors and diseases. JEM was also used in a number of epidemiological studies on lead. These studies estimated the exposure intensity in each industry [9,17], occupation [18], or industry occupation [19–22] but did not use standardized processes. In this study, we were able to estimate the lead exposure intensity by each of 133 industries for 47,280

Table 1
Summary statistics of the top 10 industries with high and low exposure to airborne lead based on the 95th percentile level (number of measurements per industry≥20, 95th percentile level>1.8)

SIC	Industry	Airborne lead, μg/m^3							
		Censoring rate			Simple replacement				
		Censored	Total	Rate (%)	AM	SD	GM	GSD	X95
High-exposure industry									
242	Manufacture of basic precious and non-ferrous metals	535	1193	44.8	6.80	17.78	2.85	3.42	23.68
251	Manufacture of structural metal products, tanks, reservoirs, and steam generators	333	1048	31.8	9.48	9.04	4.88	3.70	23.20
243	Cast of metals	447	758	59.0	5.31	9.31	2.15	3.34	22.33
239	Manufacture of other non-metallic mineral products	71	101	70.3	5.07	14.43	1.72	3.14	19.60
282	Manufacture of primary cells and batteries and accumulators	442	1412	31.3	6.53	7.15	3.76	3.04	19.10
383	Recovery of metal and non-metal waste and scrap	33	52	63.5	4.05	7.97	1.78	2.92	17.40
201	Manufacture of basic chemicals	229	391	58.6	3.73	4.98	1.97	2.84	16.00
222	Manufacture of plastic products	808	1084	74.5	3.05	5.76	1.49	2.61	15.41
181	Printing and service activities related to printing	28	33	84.8	2.35	4.33	1.24	2.35	14.50
259	Manufacture of other metal products; metal working service activities	898	1297	69.2	2.72	4.18	1.54	2.46	12.30
Low-exposure industry									
313	Manufacture of semiconductor	579	624	92.8	<1.8	NA	<1.8	NA	2.47
332	Software development and supply	65	74	87.8	<1.8	NA	<1.8	NA	2.40
351	Manufacture of medical appliances and instruments	199	218	91.3	<1.8	NA	<1.8	NA	2.33
370	Manufacture of cement, lime, and plaster and its products	286	336	85.1	<1.8	NA	<1.8	NA	2.30
478	Manufacture of pharmaceutical goods other than medicaments	55	62	88.7	<1.8	NA	<1.8	NA	2.29
511	Business facilities support management services	32	34	94.1	<1.8	NA	<1.8	NA	2.26
715	Manufacture of computers and peripheral equipment	130	139	93.5	<1.8	NA	<1.8	NA	2.10
759	Waste collection	82	88	93.2	<1.8	NA	<1.8	NA	2.00
861	Manufacture of rubber products	147	157	93.6	<1.8	NA	<1.8	NA	1.98
869	Transit and ground passenger transportation	75	80	93.8	<1.8	NA	<1.8	NA	1.81

SIC, Standard Industrial Classification (KSIC 9th revision); AM, arithmetic mean; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation; X95, 95th percentile level; NA, not available.

Table 2
Lead concentration level of the process with the highest exposure to lead in each industry (number of measurements per process≥20, censoring rate<50%)

Standard industry		Standard process		Censoring rate			Airborne lead, $\mu g/m^3$				
									Simple replacement		
SIC	Industry	SPC	Process	Censored	Total	Rate (%)	AM	SD	GM	GSD	X95
242	Manufacture of basic precious and non-ferrous metals	SPC4	Casting	25	94	26.6	11.67	16.57	5.29	3.75	53.29
243	Cast of metals	SPC9	Molding	6	22	27.3	14.35	15.48	6.28	4.47	45.26
282	Manufacture of primary cells and batteries and accumulators	SPC19	Mixing	4	50	8.0	11.66	13.88	7.60	2.55	33.88
251	Manufacture of structural metal products, tanks, reservoirs, and steam generators	SPC3	Melting	8	64	12.5	13.10	8.23	9.34	2.80	23.27
259	Manufacture of other metal products; metal working service activities	SPC8	Polishing	10	29	34.5	6.37	6.87	3.52	3.20	22.36
201	Manufacture of basic chemicals	SPC20	Chemical reaction	0	22	0.0	7.88	3.44	7.26	1.52	13.02
264	Manufacture of telecommunication and broadcasting apparatuses	SPC9	Molding	7	22	31.8	3.74	2.37	2.82	2.34	6.90

SIC, Standard Industrial Classification (KSIC 9th revision); SPC, Standard Process Classification; AM, arithmetic mean; SD, standard deviation; GM, geometric mean; GSD, geometric standard deviation; X95, 95th percentile level.

measurements of airborne lead in the national exposure database WEMD from 2015 to 2016 and the lead concentration of each of 583 exposure groups for 37 standard processes in 125 industries. In the case of previous study [9] that established JEM by industry with lead data from the same period, the ‘cast of metals” industry showed the highest exposure intensity, but it is limited to understanding the exposure characteristics because there is no detailed process information. In this study, exposure levels for each of the 10 processes could be additionally identified, including the molding process (X95 = 45.26 $\mu g/m^3$), which has the highest exposure level in the “cast of metals” industry.

Although establishment of a JEM for occupation is difficult because WEMD does not include information regarding occupation, more specific exposure characteristics can be identified by linking industries with standard processes. As shown in the evaluation results, the contrast is approximately twice as high when an exposure matrix is created in connection with industry and process than the exposure matrix for the standard industry (Table 3).

WEMD is a nationwide quantitative exposure database created by the WEM system. The WEM system is implemented as prescribed by the Korean OSHAct to ensure that employers improve their working environment according to the level of exposure to workers by providing the results of the measurements by the occupational hygiene professionals of the monitoring institution [4]. Therefore, the names of the department and process actually used in individual workplaces instead of standardized occupations or processes were used in the initial measurement report. The database has accumulated, and standardized process codes have been used since 2002 as the occupational hygiene experts have sent measurement results directly to KOSHA using the computer system (K2B). However, because the standard process code was too subdivided and investigation was not legally mandatory, the reliability of the entered standard process code information was lower than the company's actual department name or process name information. Therefore, this study allocated a new standard process code according to the degree of agreement between the process name, unit workplace information, highly reliable information, and the index words of SPC_{2021}. As shown in Supplementary Table 2, 32.7% of the total data could not be allocated to standard processes, and 42.9% could be allocated to a highly reliable single standard process. Because a single process matches the results of the use of key indexes and all indexes, even in Cases 2 through 4, the standard process can be allocated through review of data by experts. However, too much time is required. Therefore, to reliably increase the standard process allocation rate, supplementing the constructed index word lists of the standard processes is necessary by analyzing the input content characteristics of unallocated data without NOCs. Nevertheless, since the total number of WEMD data is more than 10 million, its use for JEM construction should be possible by industry process through automatic allocation using the current standard process index word lists.

There are several limitations to this study. First, because WEM is a compliance measurement performed to determine whether employers are managing exposure levels below the legal exposure limit, selection, and measurement of the worst case among workers working in a process by industrial hygiene experts is required. Therefore, the exposure level of WEMD can be overestimated. On the other hand, there is also an institutional vulnerability that can underestimate the exposure level because the employer pays the measurement institution, and the measurement is performed according to the schedule preferred by the employer. We estimate that the measurement levels can be underestimated because of the high censoring rate. Therefore, use of X95 of the exposure group with a low censoring rate and at least 20 measured samples as a more reliable indicator of exposure [9] and careful interpretation of

Table 3
Comparison of contrast and precision for three grouping schemes

Grouping scheme	Restriction	Number			Sample variance			Contrast	Precision
		Group	Company	Total	Between groups (S^2_{BG})	Within a group (S^2_{WG})	Within a company (S^2_{WC})		
SIC	N > 2	122	2247	19,929	0.23	0.05	0.33	0.17	4.0
SPC	N > 2	34	2664	19,929	0.29	0.10	0.31	0.24	7.2
SIC + SPC	N > 2	546	2767	19,929	0.23	0.15	0.31	0.39	2.1
SIC	N > 5	98	2142	19,448	0.23	0.05	0.33	0.19	4.6
SPC	N > 5	32	2520	19,448	0.29	0.11	0.31	0.28	6.8
SIC + SPC	N > 5	393	2605	19,448	0.23	0.16	0.31	0.41	2.7
SIC	N > 10	75	2055	18,777	0.23	0.06	0.33	0.20	6.0
SPC	N > 10	30	2405	18,777	0.29	0.07	0.31	0.19	7.6
SIC + SPC	N > 10	299	2479	18,777	0.23	0.15	0.31	0.40	3.4
SIC	N > 20	58	1932	17,545	0.24	0.07	0.32	0.22	6.8
SPC	N > 20	27	2209	17,545	0.28	0.08	0.31	0.23	8.1
SIC + SPC	N > 20	191	2273	17,545	0.23	0.15	0.30	0.40	4.5

SIC, Standard Industrial Classification (KSIC 9th revision); SPC, Standard Process Classification (SPC_{2021}); N, the number of measurements in an exposure group.

the results if the censoring rate is high and the number of measured samples is small is recommended. Koh et al. (2021) analyzed lead data with a high censoring rate by the simple replacement method (LOD/2) and the MLE and suggested that AM and X95 analyzed by the simple replacement method are appropriate exposure intensity indicators [9]. Park et al. (2022) compared five analysis methods including simple replacement method (LOD/2), β-substitution method, MLE method, Bayesian method, and regression on order statistics, by creating a data set with left-censored data with various combinations of censoring rate (1% to 90%) and sample size (30 to 300) [23]. They concluded the simple replacement method was inappropriate when the censoring rate was high, and the β-substitution method, MLE method, and Bayesian method can be widely applied. However, even when the censoring rate was large at 90%, predicting X95 with a simple replacement method was applicable at a level of relative bias less than 5%. To accurately predict other exposure intensity indicators such as AM and GM using WEMD with a high censoring rate in the future, it is necessary to use the β-substation method, the MLE method, and the Bayesian method.

Second, 42.9% of the total analysis target data are reliably allocated to standard processes, and 32.7% are unable to allocate standard processes through the automatic algorithm that we proposed. Therefore, the lead exposure level of each industry process has a limitation in that it does not encompass all workplaces that can be exposed to lead. In particular, there are limitations in that the index words of the standard process fit well with the manufacturing industry but not as well the construction and healthcare industries. The index words for the “inspection” (SPC 18) process were expanded to cover patient diagnosis, treatment, and pathological tests performed in the healthcare industry to compensate for it. The number of industries allocated for the “inspection” process was the highest with 92. In the future, supplementation of the standard processes and index word lists which can better reflect the characteristics of industrial processes other than manufacturing will be necessary.

Third, the industry process JEM of lead constructed in this study was attempted on a trial basis for a short period of time, so it cannot be used as the final JEM for lead. In the future, we will try to build a final lead JEM by expanding the exposure period and supplementing it in a way that can increase the standard process allocation rate of more industries.

In summary, we evaluated exposure intensity of lead for 583 exposure groups combined with 128 industry and 35 process using a nationwide workplace monitoring database. The results of this study will be helpful in determining more accurate information regarding exposure in lead-related epidemiological studies. In addition, building JEMs for each industry process will be possible by application of the standard process automatic allocation algorithm to large data regarding other hazardous chemical agents accumulated in WEMD.

Conflict of interest

Conflict of interest relevant to this article was not reported.

Acknowledgment

This study was supported by the Occupational Safety and Health Research Institute grant (2021-OSHRI-795) funded by the Republic of Korea Occupational Safety and Health Agency and by the research program of Dongguk University, 2020 (JHP).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.shaw.2022.09.001.

References

[1] Stamm R. MEGA-database: one million data since 1972. Appl Occup Environ Hyg 2001;16(2):159–63. https://doi.org/10.1080/104732201460262.
[2] Gabriel S. The BG measurement system for hazardous substances (BGMG) and the exposure database of hazardous substances (MEGA). Int J Occup Saf Ergon 2006;12(1):101–4. https://doi.org/10.1080/10803548.2006.11076673.
[3] Vincent R, Jeandel B. COLCHIC-occupational exposure to chemical agents database: current content and development perspectives. Appl Occup Environ Hyg 2001 Feb;16(2):115–21. https://doi.org/10.1080/104732201460190.
[4] Park HD, Chung EK, Kim K. The quality control program for industrial hygiene laboratories in Korea. Saf Health Work 2017;8(3):322–6. https://doi.org/10.1016/j.shaw.2017.08.003.
[5] Choi S, Kang D, Park D, Lee H, Choi B. Developing asbestos job exposure matrix using occupation and industry specific exposure data (1984-2008) in Republic of Korea. Saf Health Work 2017;8(1):105–15. https://doi.org/10.1016/j.shaw.2016.09.002.
[6] Koh DH, Jeon HK, Lee SG, Ryu HW. The relationship between low-level benzene exposure and blood cell counts in Korean workers. Occup Environ Med 2015;72(6):421–7. https://doi.org/10.1136/oemed-2014-102227.

[7] Koh DH, Park JH, Lee SG, Kim HC, Choi S, Jung H, Park DU. Combining lead exposure measurements and experts' judgment through a bayesian framework. Ann Work Expo Health 2017;61(9):1054–75. https://doi.org/10.1093/annweh/wxx072.

[8] Koh DH, Park JH, Lee SG, Kim HC, Choi S, Jung H, Park JO, Park DU. Estimation of lead exposure prevalence in Korean population through combining multiple experts' judgment based on objective data sources. Ann Work Expo Health 2018;62(2):210–20. https://doi.org/10.1093/annweh/wxx106.

[9] Koh DH, Park JH, Lee SG, Kim HC, Jung H, Kim I, Choi S, Park D. Estimation of lead exposure intensity by industry using nationwide exposure databases in Korea. Saf Health Work 2021;12(4):439–44. https://doi.org/10.1016/j.shaw.2021.07.008.

[10] Koh DH, Park JH, Lee SG, Kim HC, Choi S, Jung H, Kim I, Park D. Development of Korean CARcinogen EXposure: an initiative of the occupational carcinogen surveillance system in Korea. Ann Work Expo Health 2021;65(5):528–38. https://doi.org/10.1093/annweh/wxaa135.

[11] Chung E, Ha K. Development of occupational and industrial health guide and occupational health summary for manufacturing plant. Occupational Safety and Health Research Institute, Korea Occupational Safety and Health Agency; 2018.

[12] Kromhout H, Heederik D. Occupational epidemiology in the rubber industry: implications of exposure variability. Am J Ind Med 1995;27(2):171–85. https://doi.org/10.1002/ajim.4700270203.

[13] Park JY, Ramachandran G, Raynor PC, Eberly LE, Olson Jr G. Comparing exposure zones by different exposure metrics using statistical parameters: contrast and precision. Ann Occup Hyg 2010;54(7):799–812. https://doi.org/10.1093/annhyg/meq043.

[14] Paik NW, Levine SP, Schork A. Development and application of a quality control program for industrial hygiene laboratories in Korea. Appl Occup Environ Hyg 1997;12:46e53.

[15] NIOSH. LEAD by flame AAS METHOD 7082. In: NIOSH manual of analytical methods. 4th ed. Washington, D.C: National Institute of Occupational Safety and Helath; 2003. U.S. Government Printing Office.

[16] R Core Team. R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2021. URL, https://www.R-project.org/.

[17] Koh DH, Locke SJ, Chen YC, Purdue MP, Friesen MC. Lead exposure in US worksites: a literature review and development of an occupational lead exposure database from the published literature. Am J Ind Med 2015;58(6):605–16. https://doi.org/10.1002/ajim.22448.

[18] Dickerson AS, Hansen J, Specht AJ, Gredal O, Weisskopf MG. Population-based study of amyotrophic lateral sclerosis and occupational lead exposure in Denmark. Occup Environ Med 2019;76(4):208–14. https://doi.org/10.1136/oemed-2018-105469.

[19] Cocco P, Dosemeci M, Heineman EF. Brain cancer and occupational exposure to lead. J Occup Environ Med 1998;40(11):937–42. https://doi.org/10.1097/00043764-199811000-00001.

[20] van Wijngaarden E, Dosemeci M. Brain cancer mortality and potential occupational exposure to lead: findings from the national longitudinal mortality study, 1979-1989. Int J Cancer 2006;119(5):1136–44. https://doi.org/10.1002/ijc.21947. Erratum in: Int J Cancer. 2007 Aug 1;121(3):698.

[21] Koh DH, Bhatti P, Coble JB, Stewart PA, Lu W, Shu XO, Ji BT, Xue S, Locke SJ, Portengen L, Yang G, Chow WH, Gao YT, Rothman N, Vermeulen R, Friesen MC. Calibrating a population-based job-exposure matrix using inspection measurements to estimate historical occupational exposure to lead for a population-based cohort in Shanghai, China. J Expo Sci Environ Epidemiol 2014;24(1):9–16. https://doi.org/10.1038/jes.2012.86.

[22] Siemiatycki J, Lavoué J. Availability of a new job-exposure matrix (CANJEM) for epidemiologic and occupational medicine purposes. J Occup Environ Med 2018;60(7):e324–8. https://doi.org/10.1097/JOM.0000000000001335.

[23] Park JH, Choi S, Koh DH, Park DU, Sung Y. A comparison of analysis methods for work environment measurement databases including left-censored data. J Korean Soc Occup Environ Hyg 2022;32(1):21–30. https://doi.org/10.15269/JKSOEH.2022.32.1.21.