Safety and Health at Work 16 (2025) 361–370

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Original article

Development and Application of Standard Codes for the National Exposure Surveillance System Utilizing Korea's Nationwide Exposure Database

Sangjun Choi [1,2,*], Ju-Hyun Park [3], Dong-Hee Koh [4], Dae Sung Lim [5], Hwan-Cheol Kim [6], Jin-Ha Yoon [7], Dong-Uk Park [8], Su Min Oh [1], Hoekyeong Seo [9]

[1] *Graduate School of Public Health and Healthcare Management, The Catholic University of Korea, Seoul, Republic of Korea*
[2] *Catholic Institute for Public Health and Healthcare Management, Seoul, Republic of Korea*
[3] *Department of Statistics, Dongguk University, Seoul, 04620, Republic of Korea*
[4] *Occupational and Environmental Medicine, Severance Hospital, Yonsei University, Seoul, Republic of Korea*
[5] *Hansung Health and Safety Technology Co., Ltd., Daejeon, Republic of Korea*
[6] *Department of Occupational and Environmental Medicine, Inha University, Incheon, Republic of Korea*
[7] *Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea*
[8] *Department of Environmental Health, Korea National Open University, Seoul, Republic of Korea*
[9] *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 25 December 2024
Received in revised form
17 April 2025
Accepted 25 May 2025
Available online 29 May 2025

Keywords:
Exposure condition category
Job-exposure matrix
Standard occupation classification
Standard process classification

A B S T R A C T

Background: The Work Environment Measurement Database (WEMD) is a valuable system for occupational exposure surveillance. However, its use is limited due to the lack of proper standardization of exposure-related variables. Hence, we developed a new standard process classification (SPC), standard occupation classification (SOC), and exposure condition category (ECOC) codes to establish an exposure surveillance system using the WEMD. Additionally, we assessed the feasibility of constructing a job-exposure matrix (JEM) using standardized codes.
Methods: The SPC and SOC were reclassified based on similarity from an exposure perspective, using established codes refined through reviews by industrial hygiene experts. The ECOC codes were based on the conceptual exposure assessment model. Ten experts conducted a pilot project to evaluate the applicability of the newly reclassified SPC, SOC, and ECOC codes.
Results: We developed 77 SPC, 82 SOC, and 12 ECOC codes, which were assigned to over 98% of the data by experts, demonstrating their practical applicability. A JEM linking industry, occupation, process, and exposure condition was constructed into an interactive dashboard based on expert evaluations, demonstrating feasibility and enabling better interpretation of exposure levels through user-controlled variables. Exposure levels varied significantly across ECOC groups, showing a clear linear trend with higher exposures in conditions representing greater exposure potential, such as proximity to the source and lack of control measures.
Conclusion: The newly developed standardized codes are easily applicable by industrial hygienists and can be integrated into the WEMD, supporting its expected use as an exposure surveillance system.

Sangjun Choi: https://orcid.org/0000-0001-8787-7216; Ju-Hyun Park: https://orcid.org/0000-0001-9675-6475; Dong-Hee Koh: https://orcid.org/0000-0002-2868-4411; Dae Sung Lim: https://orcid.org/0000-0003-4190-0390; Hwan-Cheol Kim: https://orcid.org/0000-0002-3635-1297; Jin-Ha Yoon: https://orcid.org/0000-0003-4198-2955; Dong-Uk Park: https://orcid.org/0000-0003-3847-7392; Su Min Oh: https://orcid.org/0009-0002-1769-1998; Hoekyeong Seo: https://orcid.org/0000-0002-8069-3788

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

2093-7911/$ – see front matter
https://doi.org/10.1016/j.shaw.2025.05.006

1. Introduction

National-level health and exposure monitoring systems must be established to prevent occupational illness and develop effective health and safety policies [1,2]. Job-exposure matrices (JEMs) that utilize quantitative exposure assessment data are valuable tools for epidemiological research and national health and safety policy development. Hence, various countries have adopted such databases, including the Finnish Job-Exposure Matrix in Finland [3], MEGA in Germany [4], and COLCHIC [5] and SCOLA [6] in France.

Nationwide workplace exposure and workers' health surveillance systems have been established in the Republic of Korea under the Occupational Safety and Health Act (OSHAct). Annual monitoring of chemical and physical hazards designated by the Ministry of Employment and Labor (MOEL) is conducted by private Work Environment Monitoring Institutions (WEMIs). Since 2002, results, including exposure levels and workplace details, have been recorded electronically in the Work Environment Measurement Database (WEMD) managed by the Korea Occupational Safety and Health Agency (KOSHA) [7]. The Republic of Korea also conducts special health examinations for workers exposed to hazardous agents identified by MOEL. These examinations, carried out by occupational physicians at private Special Health Examination Institutions (SHEIs), include physical and biochemical tests. Since 2000, the results have been compiled into the Special Health Examination Database (SHED), which is also managed by KOSHA [8]. Data for both WEMD and SHED are collected through the "Kosha to Business" (K2B; https://k2b.kosha.or.kr/index.do) system, where WEMIs and SHEIs input relevant information. According to statistics from the MOEL, in 2021 alone, measurements were taken by 188 WEMIs for 640,451 processes at 75,377 workplaces. However, the MOEL has only conducted simple analyses of processes or workplaces exceeding occupational exposure limits (OELs) based on the WEMD and has not utilized it for a national exposure surveillance system [9].

To analyze the use of the WEMD for an exposure surveillance system, we must standardize the codes to classify variables, such as the industry, occupation, and process. This will help us understand the exposure characteristics of workplaces and workers. Information about the workplace's industry is reported in a standardized format, using 5-digit codes from the Korea Standard Industrial Classification (KSIC); however, no occupation data exist. Although JEMs have been developed for asbestos [10], benzene [11], lead [7], and the K-CAREX system [12] using industry codes from the WEMD, they have limitations when it comes to investigating specific exposure characteristics. To address this shortcoming, we tried to develop standard process classification (SPC) codes for the WEMD in 2021 and conducted a pilot study to construct a JEM using the SPC codes for lead [13].

As occupation information is included in the SHED but not the WEMD, researchers attempted to develop an occupation-based JEM by linking these two databases [14]. However, if workplaces do not provide job-related information, SHEIs face limitations in accurately entering the 5-digit Korean Standard Classification of Occupations (KSCO) code into the SHED. Thus, it is crucial to develop standard occupation classification (SOC) codes that can be directly applied to the WEMD.

In this article, we summarized the results of standard code development for processes and occupations and assessed their usefulness for experts responsible for entering standard codes. We additionally developed exposure condition category (ECOC) codes, allowing the identification of exposure circumstances, and assessed the potential for constructing a JEM that combines the standardized codes.

2. Materials and methods

2.1. Data sources

The WEMD is collected according to specific criteria mandated by the Republic of Korea OSHAct, including departments, processes, unit workplaces, hazardous agents measured, exposure levels, and OELs. KOSHA has required WEMIs performing these measurements to input their information into the K2B system using the KSIC codes, while processes have been input using standard codes independently developed by KOSHA. More specific measurement methods and related systems have been described in detail in our previous study [8,13]. We reviewed variables comprising the WEMD and the K2B input system data to develop suitable standard codes. Additionally, we constructed a JEM using selected WEMD data from 2024.

2.2. Development of the SPC codes

Up to 2019, 1390 SPC codes (SPC2019) were used in K2B. In 2020, the system was modified to enable users to generate codes autonomously, resulting in 2,807 standard codes (SPC2020) being used in September 2021 [13]. Standard code development proceeded in two stages in 2021 and 2023; the overall development process is summarized in Supplementary Fig. S1.

In Phase 1 in 2021, professional industrial hygienists (SC and DP) reviewed the 1,390 codes from SPC2019 and 2,807 codes from SPC2020 and reclassified these into 37 standard processes (SPC2021). Additionally, we extracted words, focusing on nouns, from the process names and explanations in SPC2019 and SPC2020 using functions from the R package KoNLP (e.g., extractNouns). We selected key index words that best explained the corresponding standard process from these while showing properties that excluded other processes. Additional index words were selected based on words not included in the key index words and words extracted from the process names and outlines of the 549 processes in the Occupational Health List [15].

In Phase 2, we reviewed the 37 standard process codes developed in 2021 to determine their applicability to data from several industries, including manufacturing, construction, and healthcare. The original codes were expanded, resulting in a final selection of 77 standard process codes.

2.3. Development of the SOC codes

The process for SOC code development is summarized in Supplementary Fig. S2. Although no occupation information is included in the WEMD, the SHED includes occupation codes from KSCO-7. Hence, the SOC codes applied to the WEMD were also based on the KSCO-7 codes. However, with 1,231 5-digit codes, SHEI experts may face difficulty selecting the correct occupation code. For example, the current 5-digit codes divide nurses into general nurses (KSCO = 24302) and nurse practitioners (KSCO = 24301). Selecting the appropriate code can be challenging if the person entering the data lacks detailed information. Since both codes are expected to have similar characteristics in terms of exposure, they can be combined into a single category labeled "nurse." This highlights the need to simplify SOC codes by grouping similar occupations.

To combine occupations, we referred to the 2018 revision of the Korean Employment Classification of Occupations (KECO). KECO was developed to collect data for suitable occupation units in the labor market and provide meaningful statistical data. Unlike KSCO, it is a division-based occupation classification system that prioritizes skill type over skill level. For example, KSCO divides the classes "sheet metal maker" (7422) and "sheet metal machine operator" (8417) into different sections (7 and 8) based on skill level and type, whereas KECO focuses on the skill type and classifies both occupations in the same group of "pipe and sheet metal makers" (822). Therefore, we grouped occupations that were deemed similar in terms of exposure, even if they belonged to different categories based on varying skill levels. In terms of exposure, we judged the KECO classification to be more appropriate, and so, with reference to KECO, the KSCO classes were reclassified as similar occupations. Hence, the 1,231 subclass codes from KSCO-7 were initially standardized into 47 codes. Then, after considering the properties of non-manufacturing industries, such as construction and human health activities, we generated 82 SOC codes.

2.4. Development of the ECOC codes

Industrial hygienists collect the information in the WEMD using the legal form prescribed under the Republic of Korea OSHAct. The form includes information about departments, processes, unit workplaces, OELs, and whether the exposure levels exceeded the OEL. However, no items enable the identification of working conditions at the time of measurement. For example, when the process information states "welding," welding in a place with a local ventilation system and welding in a confined, poorly ventilated space significantly differ regarding the exposure level of welding fumes. The legal form is limited because it cannot capture information about exposure conditions. Therefore, following a conceptual exposure model for assessing inhalation exposure [16], we developed ECOC codes for the three components judged to have the greatest influence on worker exposure level (isolation of workers from the source, operation of local ventilation equipment, and the frequency of work within $<$1 m from the source).

The first component was defined based on whether the workers were physically isolated from the source. The conceptual exposure model includes all separation performed to isolate the source from the recipient. "Source enclosed/isolated" referred to all situations in which source processes, devices, or equipment were located in a designated closed area within the workplace or where workers typically operated from a distinct space away from the source, such as a central control room. Cases where the source and worker shared the same workspace were classified as "source not enclosed/isolated".

The second component was the condition of control of hazardous factors generated from the source. The determinants were divided into "ventilation system on" and "ventilation system absent or off" depending on whether a local ventilation system was in operation.

The third component involved the worker's proximity to the source and the frequency of exposure. In the conceptual exposure model, near-field work was defined by the frequency during the last day with a distance of less than 1 m from the source. Specifically, three determinants were defined: "intermittent ($<$1 h)," "partial, (half-shift, 1–4 h)," and "most work time (full-shift, $>$ 4 h)."

Modifying factors (MFs) were assigned to compare the effects on relative exposure potential per determinant quantitatively. MFs are quantitative multipliers used in exposure modeling to adjust baseline exposure estimates based on specific conditions that influence exposure levels. Determinants refer to the qualitative or categorical characteristics that define each MF.

As in previous studies on retrospective exposure assessment [17] and modeling [16], the MFs in this study were assigned as dimensionless values based on the exposure control efficacy library [18]. This library was derived from intervention studies on exposure controls such as source enclosure, local exhaust ventilation, and worker separation, with effectiveness estimated by the reduction in measured exposure concentrations. For example, if exposure decreased from 10 mg/m^3 to 1 mg/m^3 after an intervention, the reduction efficiency was 90%, indicating a tenfold potential increase without the control. In this study, the reduction efficiencies were converted into dimensionless exposure potential values and assigned as MF scores. Reported reduction efficiencies were 86% (score = 7.1) for full enclosure, 82% (5.6) for local exhaust ventilation, 90% (10) for full worker separation (full-time near-field work in the absence of intervention), and 71% (5.3) for partial separation (partial near-field work). Considering variability, we simplified the MF scores to 1, 5, and 10, meaning that lack of controls could result in exposure levels up to 10 times higher than in fully controlled conditions.

The final ECOC was classified using 12 codes, corresponding to the combinations of determinants for each of the three components ($2 \times 2 \times 3 = 12$). The ECOC code numbers were assigned in order of increasing total MF, and cases with the same MF were assigned the same code number, differentiated by a lower-case letter (e.g., ECOC-2a, ECOC-2b). The total MF of each ECOC code was calculated as a multiplicative manner using the exposure MF of each determinant according to the previous study [16,17].

2.5. Development of a standardized code finder (SCF)

We developed two SCFs—standard code search engines—to enable accurate search of the SPC codes and SOC codes developed in this study.

SCF-1 was a keyword-based search tool constructed using R Shiny [19]. First, the search text was pre-processed to remove non-words, such as grammatical markers. The SCF was developed to show the standard process and occupation codes in descending order, with the highest total matching score with the standard process and occupation databases (2 points for each match with the name, 1 point for each key index word, and 0.5 points for each explanation text and additional index word).

SFC-2 was designed to support existing keyword-based search methods by providing a text similarity-based search function using the Bidirectional Encoder Representations from Transformers (BERT) model from Google [20] and the Facebook AI Similarity Search (FAISS) language model library from Meta [21]. The keyword-based search method was advantageous because it enables the user to retrieve the components of a keyword. Nevertheless, there were difficulties in effectively searching for expressions with a similar meaning. To overcome these limitations, applying sentence embedding and similarity search methods using language models is essential. Language models (e.g., BERT) can understand the context and analyze the relationships between words, providing more accurate and meaningful search results.

2.6. Application of standard codes

To assess whether industrial hygienists who had previously collected actual measurement data could appropriately allocate the developed SPC, SOC, and ECOC codes, we selected 11,781 data points collected from 209 workplaces by 10 industrial hygienists in the first half of 2024. The variables included classification of the industry, main product, department name, process name, unit workplace, hazard name, measurement location (or worker name), measured level, and OELs at the time of measurement. The industry classification was determined using the 10th revision of KSIC, which is based on the International Standard Industrial Classification rev. 4.

The experience of industrial hygienists who participated in the assessment was diverse, ranging from 1 to 32 years. There was a junior group of five persons with $\leq$ seven years of experience and a senior group of five persons with > seven years of experience. Each expert was asked to use SCFs to allocate SPC and SOC codes for measurements that they had taken directly. In cases where it was difficult to select the SPC or SOC code, the experts were asked to select "Other." Further, the experts were asked to assess the utility of the two SCF types and identify which was more helpful. Moreover, they were asked to assess the three categories for the ECOC and to select "not sure" in cases that were difficult to assess.

2.7. Data analysis

Of the 11,781 data points, we excluded data with missing measurements due to reasons such as "process closure." We constructed a JEM with the remaining 11,607 data points using Tableau Desktop Professional Edition (2024.2) to compare exposure levels depending on combinations of standard industry, occupation, process, and exposure conditions. To compare exposure levels of hazardous agents with different units (e.g., mg/m^3, ppm) and varying OELs, we calculated an exposure index (EI) by dividing the measured concentration of each agent by its corresponding OEL. For noise, the EI was calculated by dividing the actual exposure time by the permissible exposure duration corresponding to the measured sound level. The permissible exposure duration is calculated based on the OEL set by the MOEL of the Republic of Korea, using the following formula.

$$PD = CT \times 2^{[(CL-SL)/ER]} = 8 \times 2^{[(90-SL)/5]}$$

- PD: Permissible exposure duration,
- CT: Criterion time (permissible exposure duration for the criterion level: 8 h),
- CL: Criterion level (8-hour permissible exposure limit for noise: 90 dBA),
- SL: Measured sound level (dBA),
- ER: Exchange rate (5 dBA)

JEM was presented as a dashboard to compare the EI of each hazardous agent by showing the mean and 95th percentile across standard industry, occupation, process, and exposure condition codes.

We compared the average EI across ECOC groups to evaluate their association. ANOVA was used to test for overall differences among groups, and an orthogonal linear trend test was conducted to assess whether there was a statistically significant linear trend in the average EI across the ordered ECOC groups. Statistical analyses, including making box plots, were conducted with the statistical software R version 4.4.1 [22].

Table 1
Assignment of SPC, SOC, and ECOC based on industrial hygiene expert assessment

Category	Assessment outcome	Exposure assessment career		
		Junior (n = 5, ≤7 years)	Senior (n = 5, >7 years)	Total (%)
SPC	Assigned	4,569	7,029	11,598 (98.4)
	Unassigned	183	0	183 (1.6)
SOC	Assigned	4,746	7,027	11,773 (99.9)
	Unassigned	6	2	8 (0.1)
ECOC	Assigned	4,668	7,029	11,697 (99.3)
	Unassigned	84	0	84 (0.7)
Total		4,752	7,029	11,781 (100.0)

SPC: standard process classification, SOC: standard occupation classification, ECOC: exposure condition category.

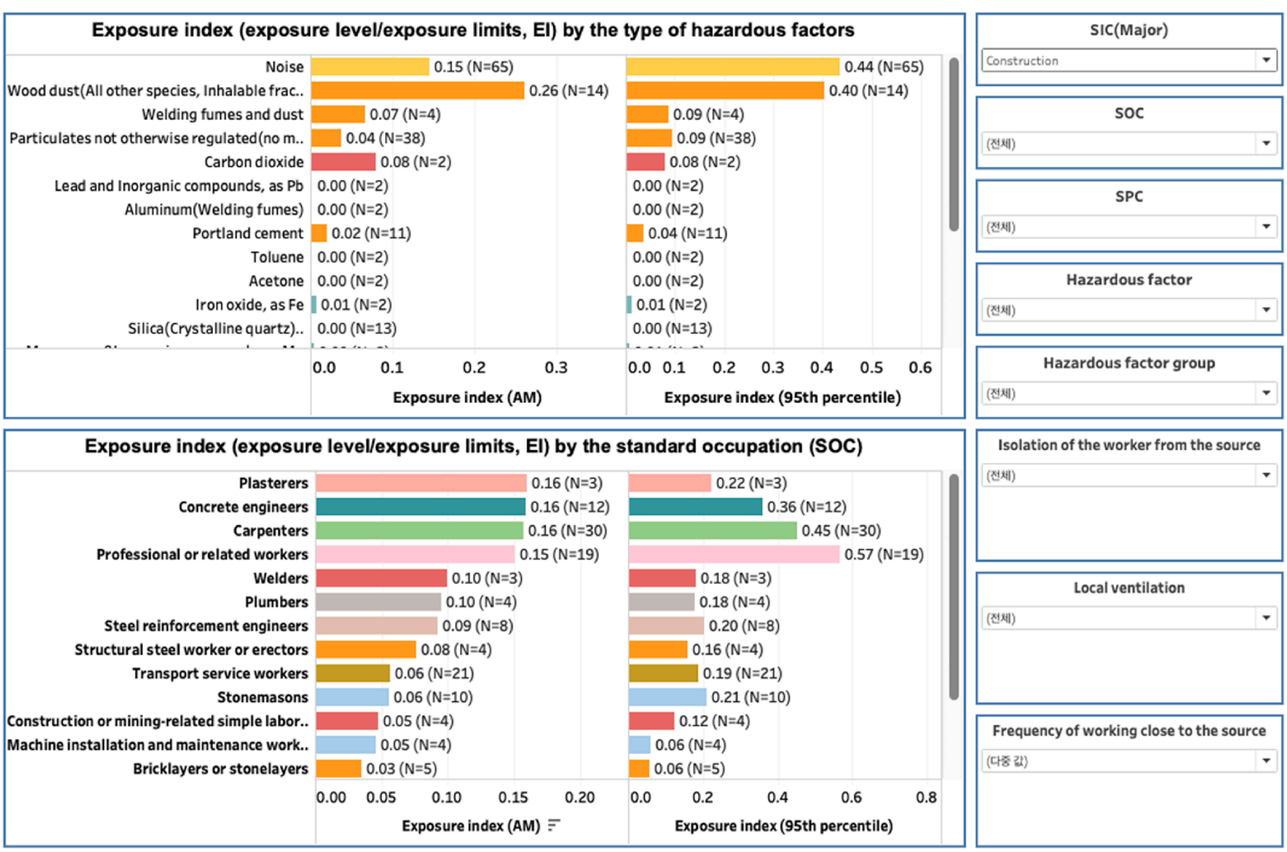


Fig. 1. Job-exposure matrix dashboard divided by industry, occupation, process, and exposure conditions (EI per hazardous factor, EI per occupation).

3. Results

3.1. Standard code and SCF

The detailed SPC and SOC code lists are shown in Supplementary Tables S1 and S2. The SOC codes are displayed alongside the corresponding KSCO-7 codes to which they can be linked. The Korean descriptions and keywords for the standard codes can be found in SFC-1 (https://kscf.shinyapps.io/scf_app/). SFC-2 is available at https://kosha.pro/spc/. For the ECOC codes, descriptions and corresponding MF values for each determinant are summarized in Supplementary Table S3, while the final MF values assigned to each ECOC code are presented in Supplementary Table S4.

3.2. Basic properties of the data for standard code application by expert assessment

Supplementary Table S5 summarizes each industry and the types of workplace hazardous substances. The 209 workplaces were distributed across industries in 13 sections in KSIC. The most common industry was manufacturing, accounting for 131 workplaces, followed by professional, scientific, and technical activities (15 workplaces) and construction (14 workplaces). Of the 11,781 data points, the majority were from the manufacturing industry (n = 9,168), followed by transportation and storage (n = 598), human health and social work activities (n = 524), and professional, scientific, and technical activities (n = 436).

Among hazardous factors, chemical factors were categorized depending on physicochemical characteristics as gaseous substances, metalworking fluids, metals, dust, acids and alkalis, and organic compounds. Heat, noise, and illuminance were categorized as physical factors. Measurements were taken for three types of physical factors and 170 types of chemical factors. Of these, organic compounds were the most common (n = 5,867), followed by metals (n = 1,765), acids and alkalis (n = 1,338), noise (n = 1,217), and dust (n = 844).

3.3. Results of applying standard codes through expert assessment

Table 1 shows that SPC, SOC, and ECOC codes were all assigned in over 98% of cases. As shown in Supplementary Table S6, 52 SPC codes were assigned, with the most common being testing, followed by maintenance and welding. Only 1.6% (n = 183) of all data were not assigned and classified as “Other”; all of these cases were in the Junior group.

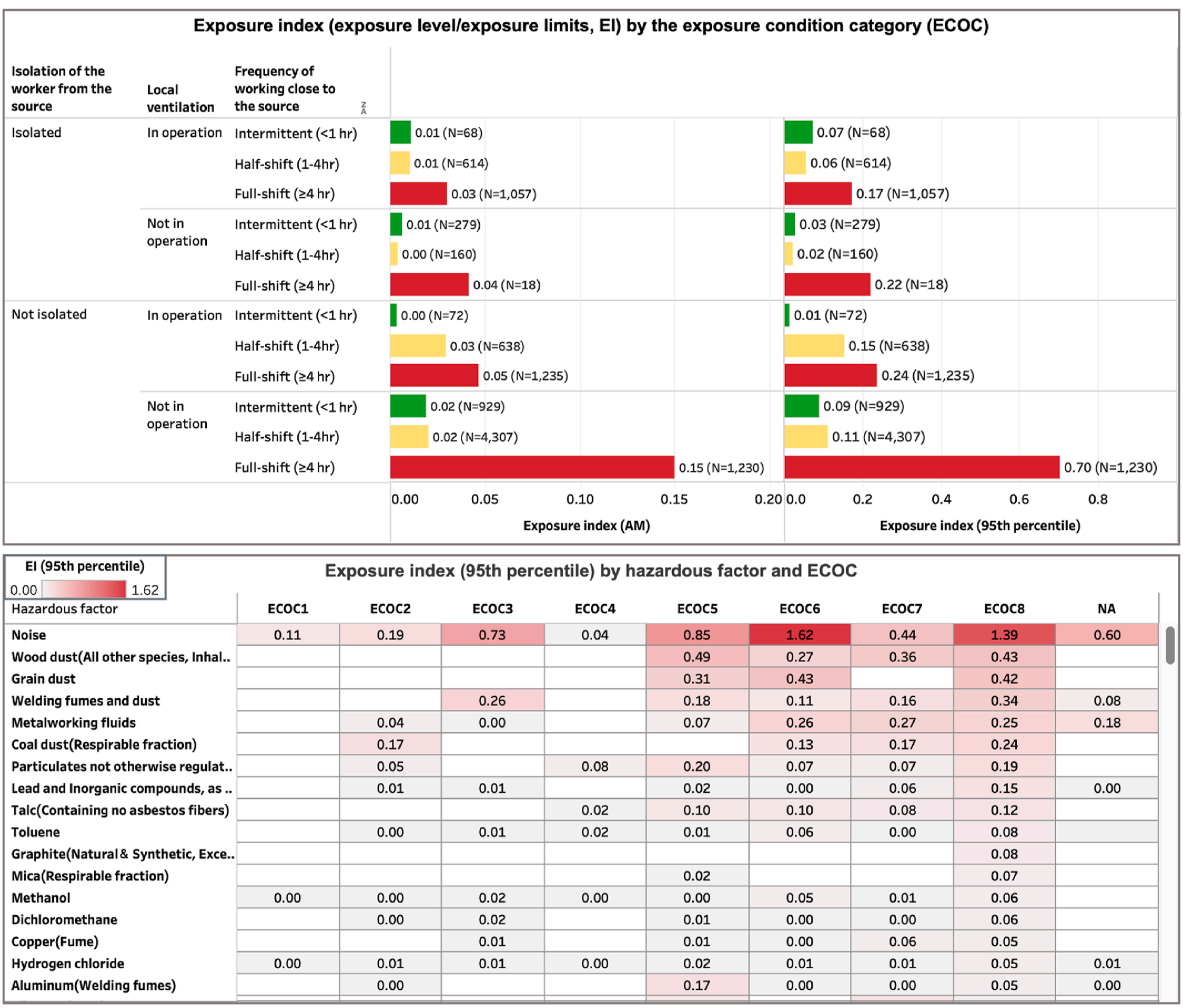


Fig. 2. Job-exposure matrix dashboard divided by industry, occupation, process, and exposure conditions (EI per exposure conditions).

As shown in Supplementary Table S7, 51 SOC codes were assigned, with the most common being chemical-related machine operators followed by expert or related workers, machinery installation or maintenance workers, and welders, with very few cases of non-assignment (n = 8).

Supplementary Table S8 shows the assessment results for each item in the ECOC. For isolation of the workers from the source, eight cases were categorized as “not sure.” In 78 cases, it could not be determined if a local ventilation device was used, all of which were in the Junior group.

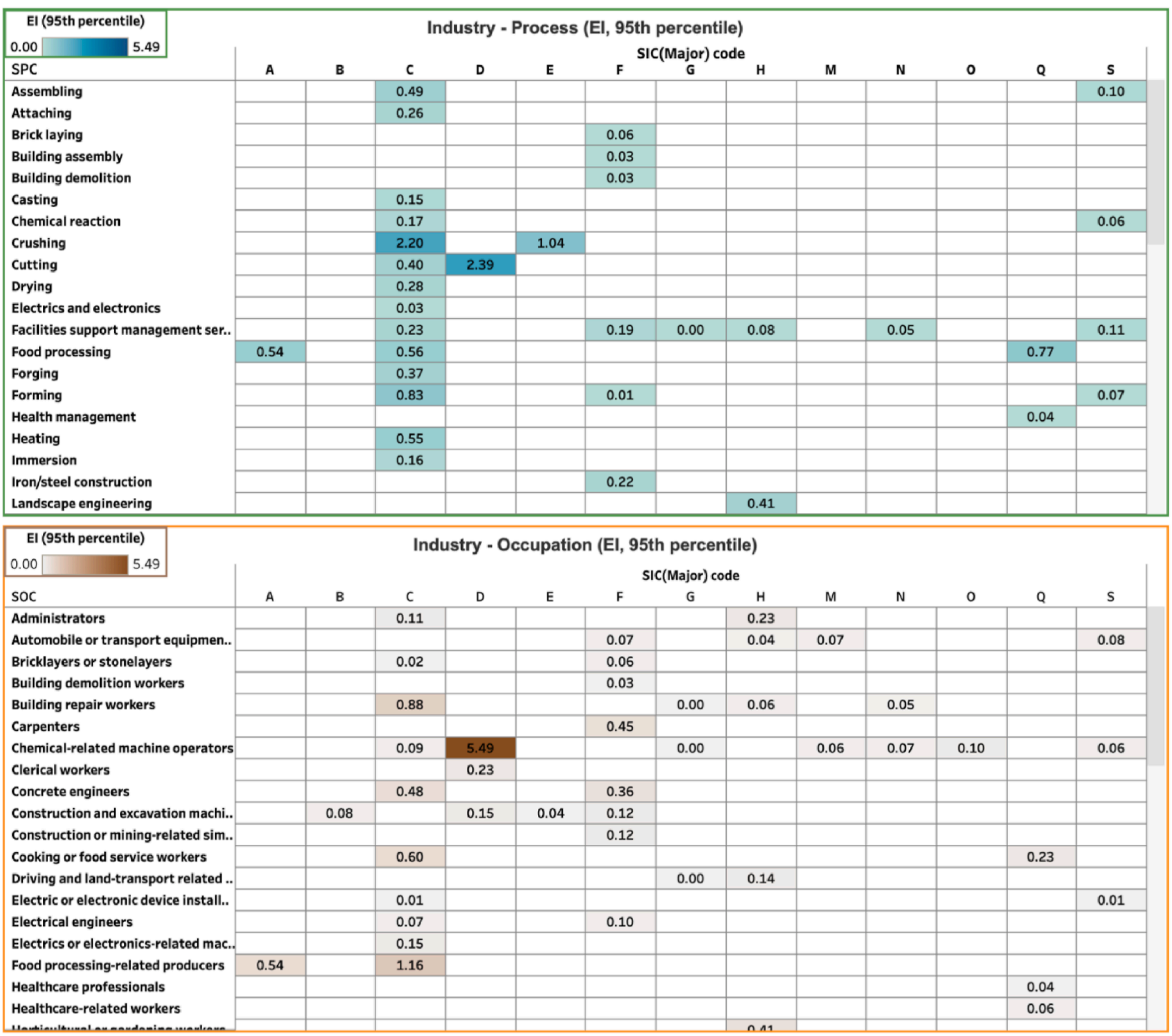
3.4. JEM using the industry, occupation, process, and ECOC codes

A web-based dashboard was developed for the JEM, allowing users to explore exposure levels by selecting variables such as hazardous agents, standardized industry, process, occupation, and ECOC code. The dashboard is accessible via the provided link (https://public.tableau.com/views/WEMD_JEM2024_eng/WEMDJEM2024_eng) and consists of four main sections, each illustrated from Figs. 1–4.

The first section allows the user to view the distribution of EI values for each type of hazardous factor and occupation (Fig. 1). The second section allows users to compare the EI values between ECOCs for a chosen combination of industries, occupations, processes, and hazardous factors (Fig. 2). The third section allows the user to compare the 95th percentiles of EI values for industry-process and industry-occupation pairs in the form of a heat map (Fig. 3). The final section allows the user, depending on whether only a single factor was measured by the same investigator (single exposure) or multiple hazardous factors were measured simultaneously (multiple exposure), to verify which hazardous factor was measured most frequently and to inspect the mean EI values per occupation (Fig. 4).

3.5. Comparison of exposure levels by ECOC

As shown in Fig. 5, the distribution of EI values for 2,918 samples of chemical agents with quantifiable concentrations was compared across the ordered ECOC groups. The analysis revealed statistically significant differences in mean exposure levels among



SPC	A	B	C	D	E	F	G	H	M	N	O	Q	S
Assembling			0.49										0.10
Attaching			0.26										
Brick laying						0.06							
Building assembly						0.03							
Building demolition						0.03							
Casting			0.15										
Chemical reaction			0.17										0.06
Crushing			2.20		1.04								
Cutting			0.40	2.39									
Drying			0.28										
Electrics and electronics			0.03										
Facilities support management ser..			0.23			0.19	0.00	0.08		0.05			0.11
Food processing	0.54		0.56									0.77	
Forging			0.37										
Forming			0.83			0.01							0.07
Health management												0.04	
Heating			0.55										
Immersion			0.16										
Iron/steel construction						0.22							
Landscape engineering								0.41					

SOC	A	B	C	D	E	F	G	H	M	N	O	Q	S
Administrators			0.11					0.23					
Automobile or transport equipmen..						0.07		0.04	0.07				0.08
Bricklayers or stonelayers			0.02			0.06							
Building demolition workers						0.03							
Building repair workers			0.88				0.00	0.06		0.05			
Carpenters						0.45							
Chemical-related machine operators			0.09	5.49			0.00		0.06	0.07	0.10		0.06
Clerical workers				0.23									
Concrete engineers			0.48			0.36							
Construction and excavation machi..		0.08		0.15	0.04	0.12							
Construction or mining-related sim..						0.12							
Cooking or food service workers			0.60									0.23	
Driving and land-transport related ..							0.00	0.14					
Electric or electronic device install..			0.01										0.01
Electrical engineers			0.07			0.10							
Electrics or electronics-related mac..			0.15										
Food processing-related producers	0.54		1.16										
Healthcare professionals												0.04	
Healthcare-related workers												0.06	

Fig. 3. JEM matrix dashboard divided by industry, occupation, process, and exposure conditions (EI heatmaps per industry-process and industry-occupation pairs).

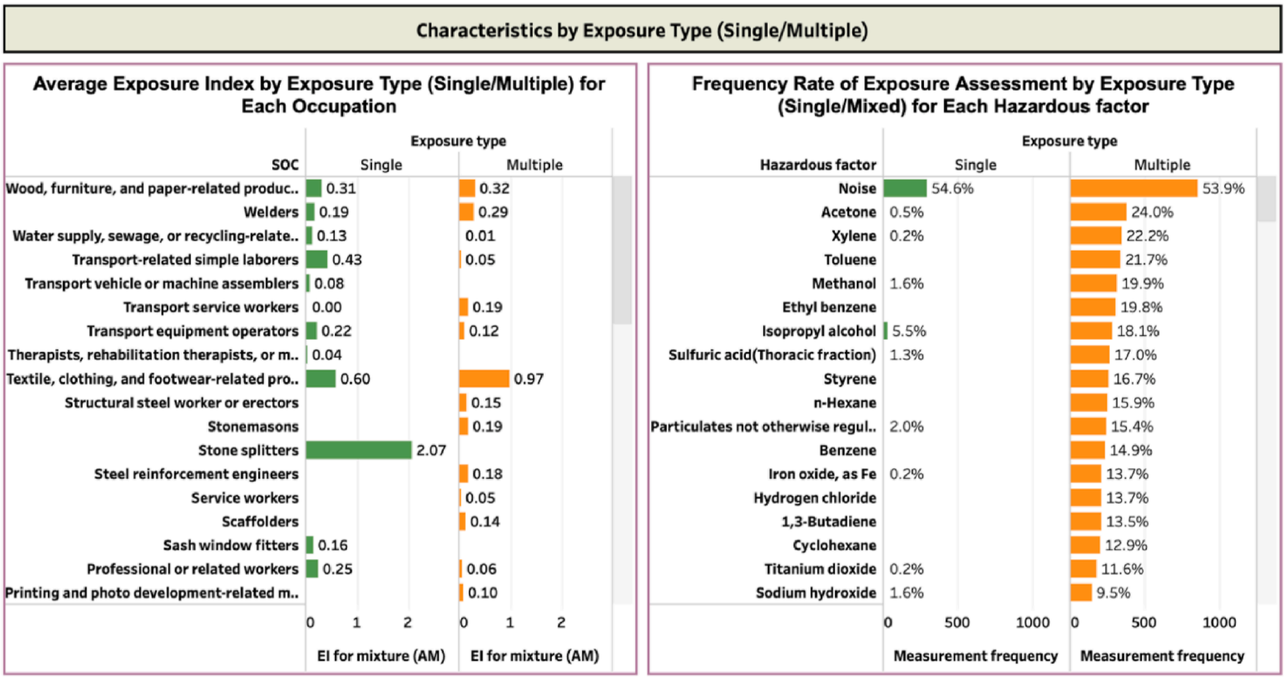


Fig. 4. JEM matrix dashboard divided by industry, occupation, process, and exposure conditions (analysis of characteristics, by type, of single/multiple exposure).

the ECOC groups ($P < 0.001$). Furthermore, linear trend analysis indicated a statistically significant positive trend across the categories ($P < 0.001$).

A comparison of the mean exposure level and 95th percentiles across exposure condition categories, based on personal sampling results and excluding heat stress and illuminance across all industries, is presented in Supplementary Fig. S3. The arithmetic means and 95th percentiles were both highest in cases where the worker and source were not isolated, local ventilation was not used, and near-field work was full-time, whereas these were lowest in cases where the worker and source were isolated and local ventilation was used. Moreover, increasing the frequency of near-field work was associated with a trend for increasing EI, suggesting that the ECOC could be useful for interpreting exposure levels. To analyze a group with similar exposure properties, we focused on the process of testing in manufacturing with the most measurements. We also found that EI was highest in cases with no isolation of the source, no local ventilation equipment, and full-time near-field work (Supplementary Fig. S4).

4. Discussion

This study outlines the standardization process of processes, occupations, and exposure conditions from 2021 to 2024 to support the use of WEMD, the Republic of Korea's largest quantitative exposure database, as a surveillance system. The standardized codes were designed to help users easily and accurately input data, and to allow analysis of exposure levels by assigning codes to previously unstandardized entries. The code development focused on two key principles: simplicity—ensuring categories are not overly technical but still distinct and relevant from an exposure perspective—and comprehensiveness—making the codes applicable across a wide range of industries and occupations.

In Phase 1 of the process standardization, the overly detailed K2B process codes—over 4,000 in total—were simplified into 37 codes, focusing on simplicity. However, when evaluating their applicability across all industries with measurements conducted in 2022, it was found that the construction industry, which had the second-highest number of measurements after manufacturing, lacked appropriate codes. This limitation arose because work environment measurements were historically focused on manufacturing, and the K2B process codes were based on manufacturing processes. Since the 2014 revision of the the Republic of Korea OSHAct, which required health managers in construction [23], measurements in the construction sector have increased, highlighting the need for more inclusive codes. In Phase 2, an additional 31 process codes related to the construction industry (SPC038–SPC068) were added, along with codes for other sectors such as healthcare and the electronics industry, resulting in a total of 77 SPC codes (Supplementary Table S1). The development of SOC codes followed a similar two-phase approach. In Phase 1, the 1,231 KSCO-7 codes were simplified into 47 SOC codes. In Phase 2, the broadly defined category "Construction and Mining-Related Trades (KSCO = 78)" was further subdivided into 30 codes (SOC039–SOC068) based on job characteristics, leading to a total of 82 SOC codes (Supplementary Table S2).

In addition to the SPC and SOC codes, we developed a set of standardized codes for exposure conditions to enhance the interpretation of exposure levels. For this, we referred to the components of the conceptual inhalation exposure model proposed by Tielemans et al. (2008) [16], as well as the associated concepts of MFs, determinants, and exposure prediction methods. While Tielemans et al. (2008) suggested nine MFs across eight model components, our study simplified these into three MFs including emission source, local ventilation, and near-field work frequency. This simplification was intended to enable industrial hygienists to quickly and easily identify key variables that may influence exposure levels under actual working conditions during measurement.

To assess the applicability of the standardized codes developed in this study, ten industrial hygiene experts assigned SPC, SOC, and ECOC codes to 11,781 measurement records. The SPC, SOC, and ECOC codes were assigned to 98.4%, 99.9%, and 99.3% of the data, respectively, showing that the codes are practical and easy to apply

(Table 1). Although we could not independently verify the accuracy of each assignment through on-site confirmation, experts were instructed to leave comments rather than assign codes when uncertain. Therefore, the assigned codes are considered to be reasonably reliable. However, future studies should include validity testing, such as assessing inter- and intra-evaluator agreement, with a larger group of evaluators to further ensure the reliability of the coding system.

Of the 11,781 measurement records with assigned standard codes, 11,607 records with calculable EI values were used to develop the JEM. These data, collected from 209 workplaces across 13 major industries, include measurements of 170 chemical agents and 3 physical agents (Supplementary Table S5). Due to the complexity of presenting exposure data across multiple variables—industry, occupation, process, and exposure condition—a static table format would be difficult to interpret. Therefore, we developed a web-based dashboard that allows users to explore exposure levels interactively by selecting variables of interest. Fig. 1 shows the first section of the dashboard, displaying EI distributions by hazardous agent and occupation within the construction industry. The dashboard is interactive, and selections are linked across the sections shown in Fig. 2 through Fig. 4.

While the current dataset lacks national representativeness and is not intended for epidemiological use, the goal of this study was to assess the feasibility of applying standardized codes to WEMD and developing a practical JEM. Compared to previous JEMs based solely on industry [7], this approach—incorporating occupation, process, and exposure conditions—offers a more detailed understanding of exposure. Future work should apply this method to the full WEMD dataset to build a nationally representative JEM.

When comparing exposure levels by ECOC, a statistically significant positive linear trend was also observed, with higher total MF values associated with higher mean EI (Fig. 5). We also identified a trend for increasing exposure levels in the cases where the source and worker were not isolated, local ventilation was not used, and the frequency of near-field work was high (Supplementary Fig. 3, 4). This highlights the need for developing strategies to utilize ECOCs in the future assessment of work environment measurements and data transfer via K2B.

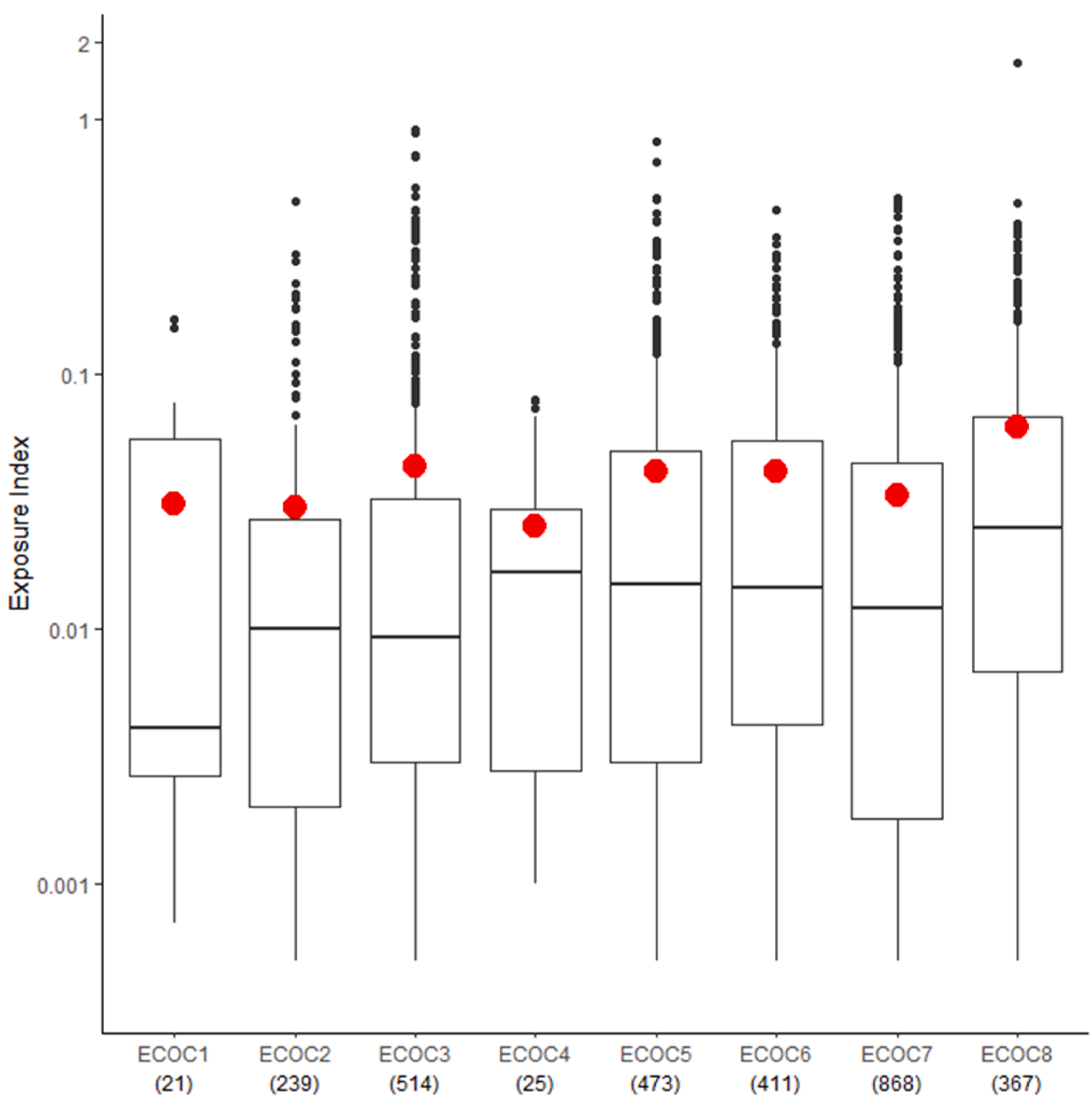


Fig. 5. Comparison of exposure index distributions for chemical agents by exposure condition category (ECOC) codes. Red dots indicate the arithmetic mean, and the number of samples for each ECOC code is shown in parentheses on the x-axis.

Despite these findings, the large variability observed within each ECOC group in Fig. 5 may be attributed to the heterogeneity of the samples. This is likely because the analysis was conducted using aggregated chemical agents, rather than evaluating exposure conditions for a single agent, due to limited sample size per agent. We believe that future research focusing on a single hazardous agent, within similar industries and processes, would allow for a more precise evaluation of the validity of the MF scoring system proposed in this study.

Several limitations should also be considered when constructing a JEM using the SPC, SOC, and ECOC codes developed in this study.

First, while this technique could be used in population-level epidemiological studies, uncertainty must be considered in individual epidemiological investigations. Even within the same industry, occupation, or process, there may be low homogeneity for exposure to certain hazardous factors [24], limiting the applicability of this technique in assessing previous exposure in individuals who have developed certain diseases.

Second, several aspects need to be considered when changing the method of standard code input. In particular, users should be able to input two or more codes simultaneously for standard processes. During the pilot test for the application of the SPC codes, one problem was that workers sometimes performed two or more processes at once, but the data input system in K2B only allowed for the selection of a single standard process. Hence, the K2B input system should be modified to enable the selection of multiple SPC codes.

Third, ECOC codes need to be applied to reduce uncertainty in the JEM using standardized variables, such as industry and occupation. Even within the same industry or occupation group, the exposure level is significantly affected by various factors. However, the previously measured and collected data in the WEMD do not include information about the exposure conditions, making it difficult to use. The items included in the WEMD are currently defined in a legal form prescribed under the Republic of Korea OSHAct. However, based on our study's assessment of the industrial hygienists, the addition of ECOCs is feasible. Even without modifying the legal reporting form, if the industrial hygienists were motivated and consented, it would be possible to input ECOC codes by adjusting the digital input method.

The results demonstrate the feasibility of building a national exposure surveillance system that actively uses SPC, SOC, and ECOC codes. Future studies should identify the hazardous factors that need to be prioritized for an exposure surveillance system.

CRediT authorship contribution statement

Sangjun Choi: Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization. **Ju-Hyun Park:** Software, Formal analysis, Data curation. **Dong-Hee Koh:** Methodology, Conceptualization. **Dae Sung Lim:** Methodology, Investigation. **Hwan-Cheol Kim:** Methodology, Investigation. **Jin-Ha Yoon:** Software. **Dong-Uk Park:** Supervision, Conceptualization. **Su Min Oh:** Visualization, Data curation. **Hoekyeong Seo:** Resources, Project administration.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Funding

This study was supported by a grant from the Occupational Safety and Health Research Institute (2024-OSHRI-527), funded by the Korea Occupational Safety and Health Agency.

Conflicts of interest

We have nothing to declare.

Acknowledgments

We would like to thank the experts who evaluated the applicability of the standard codes and Editage (www.editage.co.kr) for English language editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.shaw.2025.05.006.

References

[1] Kauppinen T, Rantanen J. Work and health country profiles and national surveillance indicators in occupational health and safety. Appl Occup Environ Hyg 2002;17:603–5.
[2] Brooke D, Cowley S, Else D, Leggett S. International review of surveillance and control of workplace exposures. Ballarat: VIOSH Australia; 2005.
[3] Kauppinen T, Uuksulainen S, Saalo A, Mäkinen I, Pukkala E. Use of the Finnish information system on occupational exposure (FINJEM) in epidemiologic, surveillance, and other applications. Ann Occup Hyg 2014;58(3):380–96. https://doi.org/10.1093/annhyg/met074. PMID 24401793.
[4] Stamm R. Mega-database: one million data since 1972. Appl Occup Environ Hyg 2001;16(2):159–63. https://doi.org/10.1080/104732201460262. PMID 11217704.
[5] Vincent R, Jeandel B. COLCHIC-occupational exposure to chemical agents database: current content and development perspectives. Appl Occup Environ Hyg 2001;16(2):115–21. https://doi.org/10.1080/104732201460190. PMID 11217697.
[6] Mater G, Paris C, Lavoué J. Descriptive analysis and comparison of two French occupational exposure databases: COLCHIC and SCOLA. Am J Ind Med 2016;59(5):379–91. https://doi.org/10.1002/ajim.22569. PMID 26901238.
[7] 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. PMID 34900362.
[8] Koh DH, Choi S, Park JH, Lee SG, Kim HC, Kim I, Min J, Kim Y, Park D. Development of Korean CARcinogen EXposure: occupation-based exposure matrix. Ann Work Exposures Health 2025;69(2):191–200. https://doi.org/10.1093/annweh/wxae090.
[9] MOEL. 2021 work environment measurement results. City. 2022;2021 work environment measurement results.
[10] 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. PMID 28344849.
[11] 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. PMID 25575529.
[12] 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 Exposures Health 2021;65(5):528–38. https://doi.org/10.1093/annweh/wxaa135. PMID 33528514.
[13] Park JH, Choi S, Koh DH, Lim DS, Park D, Kim HC, Lee SG, Lee J, Lim JS, Sung Y, Ko KY. A pilot establishment of the job-exposure matrix of lead using the standard process code of nationwide exposure databases in Korea. Saf Health Work 2022;13(4):493–9. https://doi.org/10.1016/j.shaw.2022.09.001. PMID 36579012.
[14] Koh DH, Choi S, Park JH, Lee SG, Kim HC, Kim I, Lim D, Seo H, Park D. Development of occupation-based exposure matrix of lead for Korean workers: challenges and opportunities. Ann Work Exposures Health 2024;68(3):256–68. https://doi.org/10.1093/annweh/wxae006. PMID 38377017.
[15] 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.
[16] Tielemans E, Schneider T, Goede H, Tischer M, Warren N, Kromhout H, Van Tongeren M, Van Hemmen J, Cherrie JW. Conceptual model for assessment of

inhalation exposure: defining modifying factors. Ann Occup Hyg 2008;52(7): 577–86. https://doi.org/10.1093/annhyg/men059. PMID 18787181.

[17] Cherrie JW, Schneider T. Validation of a new method for structured subjective assessment of past concentrations. Ann Occup Hyg 1999;43:235–45.

[18] Fransman W, Schinkel J, Meijster T, Van Hemmen J, Tielemans E, Goede H. Development and evaluation of an exposure control efficacy library (ECEL). Ann Occup Hyg 2008;52(7):567–75. https://doi.org/10.1093/annhyg/men054.

[19] Sievert C. Interactive web-based data visualization with R, plotly, and shiny. Chapman & Hall/CRC; 2020.

[20] Devlin J, Chang M, Lee K, Toutanova Kristina. Bert: pretraining of deep bidirectional transformers for language understanding. In proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2019;1:4171–86.

[21] Douze M, Guzhva A, Deng C, Johnson J, Szilvasy G, Mazaré PE, Lomeli M, Hosseini L, Jégou H. The FAISS library. arXiv Preprint arXiv:2401.08281 2024. Retrieved from, https://arxiv.org/abs/2401.08281; 2024.

[22] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2024. https://www.R-project.org/.

[23] Park J, Choi S, Sung Y, Chung J, Choi S. Workplace violence against female health managers in the male-dominated construction industry. Ann Work Expo Health 2022;66(9):1224–30. https://doi.org/10.1093/annweh/wxac025.

[24] Petit P, Bicout DJ, Persoons R, Bonneterre V, Barbeau D, Maître A. Constructing a database of similar exposure groups: the application of the Exporisq-HAP Database from 1995 to 2015. Ann Work Expo Health 2017;61(4):440–56. https://doi.org/10.1093/annweh/wxx017.