

## Original article

# Estimation of Lead Exposure Intensity by Industry Using Nationwide Exposure Databases in Korea



Dong-Hee Koh<sup>1,\*</sup>, Ju-Hyun Park<sup>2</sup>, Sang-Gil Lee<sup>3</sup>, Hwan-Cheol Kim<sup>4</sup>, Hyejung Jung<sup>1</sup>, Inah Kim<sup>5</sup>, Sangjun Choi<sup>6</sup>, Donguk Park<sup>7</sup>

<sup>1</sup> Department of Occupational and Environmental Medicine, International St. Mary's Hospital, Catholic Kwandong University, Incheon, Republic of Korea

<sup>2</sup> Department of Statistics, Dongguk University, Seoul, Republic of Korea

<sup>3</sup> Occupational Safety and Health Research Institute, Korea Occupational Safety and Health Agency, Ulsan, Republic of Korea

<sup>4</sup> Department of Occupational and Environmental Medicine, Inha University, Incheon, Republic of Korea

<sup>5</sup> Department of Occupational and Environmental Medicine, College of Medicine, Hanyang University, Seoul, Republic of Korea

<sup>6</sup> Department of Preventive Medicine, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

<sup>7</sup> Department of Environmental Health, Korea National Open University, Seoul, Republic of Korea

## ARTICLE INFO

## Article history:

Received 12 March 2021

Received in revised form

27 June 2021

Accepted 12 July 2021

Available online 17 July 2021

## Keywords:

Cancer

Carcinogen

Exposure

Occupational cancer

Occupational exposure

## ABSTRACT

**Background:** In a previous study, we estimated exposure prevalence and the number of workers exposed to carcinogens by industry in Korea. The present study aimed to evaluate the optimal exposure intensity indicators of airborne lead exposure by comparing to blood lead measurements for the future development of the carcinogen exposure intensity database.

**Methods:** Data concerning airborne lead measurements and blood lead levels were collected from nationwide occupational exposure databases, compiled between 2015 and 2016. Summary statistics, including the arithmetic mean (AM), geometric mean (GM), and 95th percentile level (X95) were calculated by industry both for airborne lead and blood lead measurements. Since many measurements were below the limits of detection (LODs), the simple replacement with half of the LOD and maximum likelihood estimation (MLE) methods were used for statistical analysis. For examining the optimal exposure indicator of airborne lead exposure, blood lead levels were used as reference data for subsequent rank correlation analyses.

**Results:** A total of 19,637 airborne lead measurements and 32,848 blood lead measurements were used. In general, simple replacement showed a higher correlation than MLE. The results showed that AM and X95 using simple replacement could be used as optimal exposure intensity indicators, while X95 showed better correlations than AM in industries with 20 or more measurements.

**Conclusion:** Our results showed that AM or X95 could be potential candidates for exposure intensity indicators in the Korean carcinogen exposure database. Especially, X95 is an optimal indicator where there are enough measurements to compute X95 values.

© 2021 Occupational Safety and Health Research Institute, Published by Elsevier Korea LLC. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cancers are among the leading causes of death worldwide and can lead to a heavy economic burden on patients and society [1]. Cancer is a multifactorial disease caused by genetic and environmental factors [2]. Among these factors, occupational exposure is estimated to contribute to 2–8% of all cancers [3–5]. Especially, occupational exposure is estimated to account for 14.5% of lung

cancer cases and 32.7% of sinonasal cancer cases [5]. The critical fact is that occupational cancers are largely avoidable.

In an effort to prevent occupational cancers, many countries have developed occupational carcinogen exposure surveillance systems, such as the Finnish Job-Exposure Matrix and CARcinogen EXposure (CAREX) [6–8]. Furthermore, the Korean CAREX (K-CAREX) was developed for 20 definite human carcinogens (group 1) for Korean working populations [9]; however, it focused on the

\* Corresponding author. Department of Occupational and Environmental Medicine, International St. Mary's Hospital, Catholic Kwandong University, 25, Simgok-ro 100 Beon-gil, Seo-gu, Incheon, 22711, Republic of Korea.

E-mail address: [koh.donghee@gmail.com](mailto:koh.donghee@gmail.com) (D.-H. Koh).

exposure prevalence and the number of exposed workers, while the exposure intensity had not been established.

In the present study, we aimed to examine various exposure intensity indicators to be used in the K-CAREX. To do so, we estimated exposure intensity by industry using lead as an exemplary carcinogen. We selected lead because there were relatively abundant measurements both for airborne and blood lead. Also, blood lead is known to reflect airborne lead exposure well [10]. Lead is classified as a probable carcinogen (Group 2A) related to excess risk of stomach, kidney, and brain cancers [11]. The results can be applied to other carcinogens for the future development of the K-CAREX exposure intensity.

## 2. Methods

### 2.1. Data sources

There is an occupational exposure surveillance system administered by the government in Korea. Every workplace should undergo periodic workplace exposure monitoring for 190 chemical agents and two physical agents. Workplaces are requested to undergo environmental monitoring basically twice a year, and the schedules change according to the hazardous agent and previous compliance/violation conditions [12]. The measurement data are centrally collected by the Korea Occupational Safety and Health Agency (KOSHA) as an electronic database since 2002, which is referred to as the Work Environment Measurement Database (WEMD) [13]. The WEMD contains information such as industry, work process, sampling site, sampling time, and time-weighted average. We extracted airborne lead measurement data, including sampling time, industrial code, and concentration, from the WEMD between 2015 and 2016.

There is an occupational health surveillance system administered by the government in Korea. Every worker exposed to any one of 176 physical and chemical agents (including night-shift work) should undergo a periodic health examination for the agent(s). The results are centrally collected by the KOSHA and deposited into an electronic database since 2000, which is named the Special Health Examination Database (SHED) [14]. The SHED contains information such as blood tests and biological monitoring results. We extracted blood lead level data from the SHED between 2015 and 2016.

### 2.2. Industrial classification

In the WEMD and the SHED between 2015 and 2016, data were recorded using the Korean Standard Industrial Classification (KSIC-9), which is almost identical to the International Standard Industrial Classification (ISIC, 4th revision) because KSIC-9 was developed based on the ISIC. We used a 3-digit ISIC industrial group as a standard industrial classification (SIC) for subsequent analyses.

### 2.3. Data cleaning and treatment

In workplace monitoring in Korea, a personal sample is typically required to be measured for at least 6 hours. We excluded measurements taken for <4 hours or >12 hours. We also excluded measurements with incorrect SIC codes.

We restricted analyses to workplaces having data pertaining to both airborne and blood lead. Then, industries with at least two nonmissing, noncensored, distinct levels were included for analyses.

Each workplace monitoring occurrence might have its level of a limit of detection (LOD); however, that information is not available in the WEMD and SHED. Therefore, for airborne lead, we assumed a single LOD according to a standard sampling and analytical method

of the KOSHA (KOSHA GUIDE A-2-2012), which is largely based on the US National Institute of Occupational Safety and Health (NIOSH) method [15]. The LOD is estimated assuming 6 hours sampling time using a mixed cellulose filter, resulting in  $1.8 \mu\text{g}/\text{m}^3$ .

For blood lead level (BLL), similar to airborne lead measurements, a single LOD was assumed. For measuring BLL, a standard analytical method of the KOSHA (KOSHA GUIDE H-21-2011) was employed, which is largely based on the US NIOSH method using graphite furnace atomic absorption spectrophotometry [16]. The LOD of BLL was estimated at  $0.85 \mu\text{g}/\text{dL}$ .

### 2.4. Statistical analysis

Environmental measurements are known to be log-normally distributed [17]. Since the lead measurement data were left-censored below the LOD, we examined a log-normality of overall data by comparing an empirical cumulative distribution function (CDF) with a hypothesized fitted CDF using the “EnvStats” package [17] of R [18]. Also, we examined the ordered data (the empirical quantiles) vs. the corresponding quantiles from the theoretical probability distribution using a quantile-quantile plot (Q-Q plot).

We calculated summary statistics for each industry, including arithmetic mean (AM), standard deviation (SD), geometric mean (GM), geometric standard deviation (GSD), and 95th percentile level (X95). Korean occupational exposure limit (OEL) of lead was  $50 \mu\text{g}/\text{m}^3$ . The biological exposure index (BEI) of BLL was  $30 \mu\text{g}/\text{dL}$  in line with the BEI of the American Conference of Governmental Industrial Hygienists (ACGIH), although the ACGIH BEI has been reduced from  $30 \mu\text{g}/\text{dL}$  to  $20 \mu\text{g}/\text{dL}$  since 2016 [19].

Summary statistics were calculated using two methods: simple replacement and the maximum likelihood estimation (MLE). Levels below the LOD were replaced with the value of LOD/2 for a simple replacement method [20]. For the MLE method, levels below the LOD were treated as censored, then used for estimating distribution parameters [17,21]. For MLE, GM and GSD were computed using a function of “elnormCensored” of “EnvStats” package of R. Estimated X95 was computed using a function of “eqlnormCensored” of “EnvStats” package. In cases where the industry had no censored value, “elnorm”, and “eqlnorm” functions were used instead.

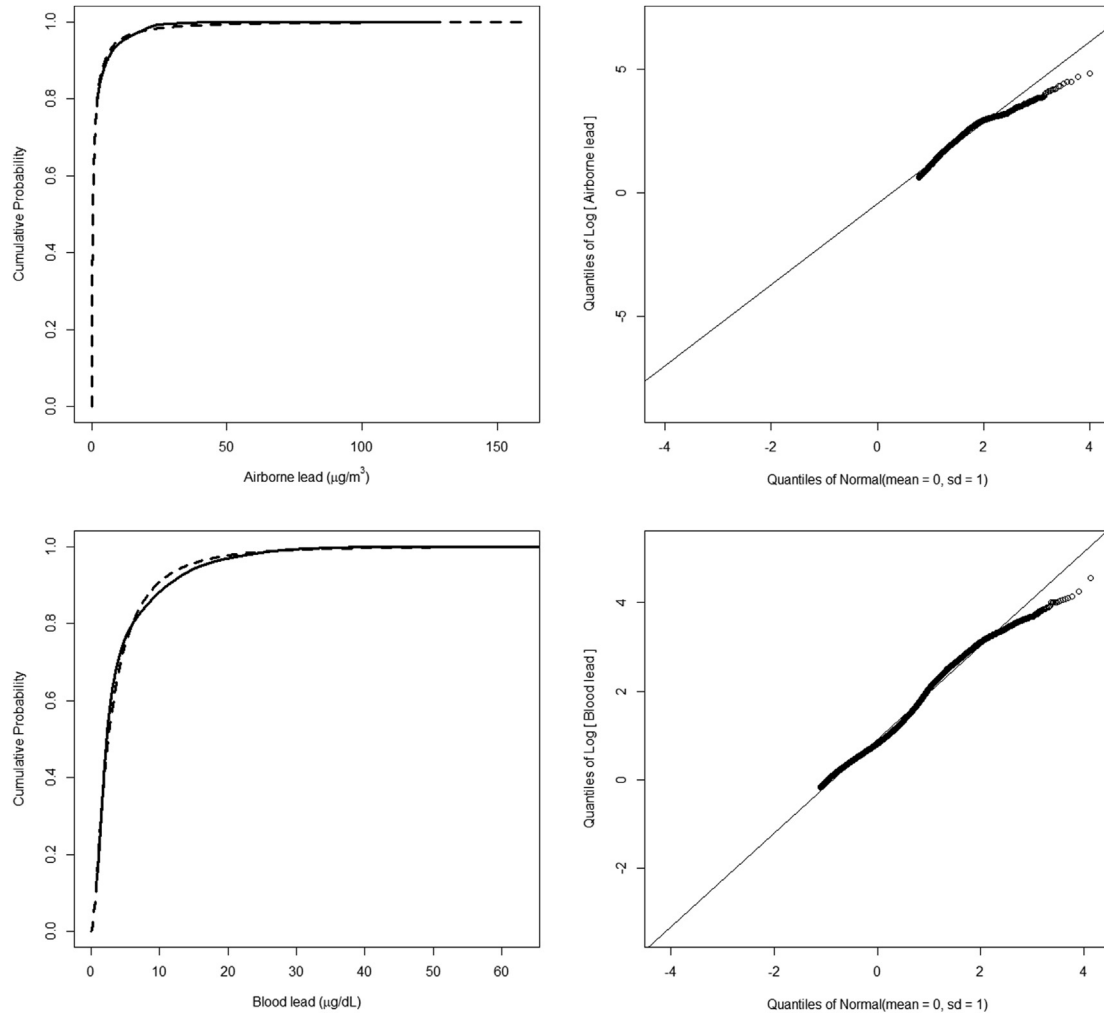
Spearman rank correlation analyses were performed across industries to compare rank orders of the summary statistics (AM, GM, and X95) of airborne lead with those of blood lead, which were treated as the reference. Furthermore, these analyses were repeated, restricting for industries with  $\geq 20$  and  $\geq 100$  measurements, respectively.

## 3. Results

The overall number of measurements for airborne lead and blood lead were 47,575 and 56,121, respectively. Finally, excluding measurements not eligible for analysis, a total of 19,637 airborne lead (41%) and 32,848 blood lead measurements (59%) were used for analysis. Censoring rates of airborne lead and blood lead measurements were 78.0% and 13.5%, respectively.

The CDF and the Q-Q plot of overall measurements show that lead data have a distribution that is close to a log-normal one (Fig. 1).

Table 1 shows the results of Spearman rank correlation between airborne lead and blood lead measurements. In general, simple replacement showed a better correlation than MLE. In terms of exposure intensity indicator, AM showed the highest correlation with all data, while X95 showed the highest correlation when restricting the analysis to industries with 20 or more measurements. The correlation coefficients further increased when restricting the analysis to industries with 100 or more



**Fig. 1.** Empirical CDF for lead (censored: solid line) with fitted log-normal CDF (dash line) (Left). Normal Q-Q plot for left-censored log-transformed lead measurements (Right).

**Table 1**  
Results of Spearman rank correlation between airborne lead and blood lead measurements across industries

	All industries (64 industries, 19,637 airborne and 32,848 blood lead measurements)				Industries with $\geq 20$ measurements (49 industries, 19,329 airborne and 32,560 blood lead measurements)				Industries with $\geq 100$ measurements (26 industries, 17,283 airborne and 28,793 blood lead measurements)			
	Simple replacement		MLE		Simple replacement		MLE		Simple replacement		MLE	
	$\gamma$	$p$	$\gamma$	$p$	$\gamma$	$p$	$\gamma$	$p$	$\gamma$	$p$	$\gamma$	$p$
AM	0.470	<0.001			0.553	<0.001			0.653	<0.001		
GM	0.426	<0.001	0.227	0.071	0.500	<0.001	0.194	0.181	0.635	0.001	0.411	0.038
X95	0.443	<0.001	0.448	<0.001	0.581	<0.001	0.578	<0.001	0.722	<0.001	0.649	<0.001

AM, arithmetic mean; GM, geometric mean; MLE, maximum likelihood estimation;  $\gamma$ , Spearman rank correlation coefficient; X95, 95th percentile level.

measurements, although the number of industries eligible for the analysis decreased. In general, AM with simple replacement or X95 with simple replacement can be candidates for exposure intensity indicators in the WEMD database. However, considering that a small number of measurements render the intensity estimates unstable, it is reasonable to restrict analyses to industries with 20 or more measurements and employ X95 as an optimal exposure intensity indicator.

Summary statistics of measurements of the top-ten upper and lower industries according to X95 with the simple replacement for airborne lead are presented in Table 2. The “Cast of Metals” industrial group (SIC 243) showed the highest X95 level of 35.54  $\mu\text{g}/$

$\text{m}^3$ . In general, industries involving high exposures showed low censoring rates, as well as high AM and GM values. Detailed results of airborne and blood lead by the 3-digit industrial group are available online in Supplemental Table 1.

#### 4. Discussion

Using nationwide occupational exposure databases, in the present study, we estimated lead exposure intensity by industry. We examined AM, GM, and X95 as indicators of exposure intensity by comparing airborne lead to blood lead as a reference. Our approach is reasonable to determine an optimal exposure intensity indicator



across industries, which will be used to develop exposure intensity to other carcinogens in the K-CAREX.

There is currently no single optimal estimation method that is used for the values less than the LOD, and the accuracy of any estimation is dependent on the censoring rate, sample size, and LOD type (for instance, single or multiple). However, for AM and the 95th percentile, it is reported that the MLE-based method is largely recommended regardless of censoring rate and sample size [21]. Furthermore, for estimating upper extreme quantiles based on a small sample of less than 30, reporting the MLE of quantiles under the log-normal assumption is suggested [22]. However, in the present study, simple replacement showed a better correlation with blood lead compared to MLE, which is contrary to the aforementioned reports, and we could not establish the cause of this discrepancy. In terms of statistical distribution, the MLE method is based on the log-normality assumption; however, in reality, the assumption did not hold in every industry, which might have led to inconsistent results. In terms of study design, previous studies had largely employed a simulation method to compare incomplete data with values less than the LOD and complete dataset. However, our study compared airborne measurements with biomonitoring data.

AM with simple replacement and X95 with simple replacement are selected as candidates as exposure intensity indicators in the Korean CAREX. Regarding AM, the percentage bias (%bias) is strongly dependent on the percentage of censored observations, which seems to be independent of sample size. When the percentage of a censored value is greater than 50%, a substantial negative bias with a %bias lower than  $-20\%$  is reported. Regarding the 95th percentile, the %bias is lower than that of AM [23]. Therefore, considering that there is a high proportion of values below the LOD in the WEMD, we prefer X95 with a simple replacement as a default exposure intensity indicator.

It is suggested that monitoring initial exposure should consist of at least six to ten measurements taken from randomly chosen workers [24]. In general, 10–20 measurements per observational group (for instance, two measurements from 5–10 randomly selected persons) are regarded to be sufficient for an initial exposure assessment [25]. Although airborne samples were non-randomly selected, we conclude that it is reasonable to assign X95 only to industries with 20 or more measurements. In fact, restricting to industries with 20 or more measurements seems to be intuitively appropriate because it requires at least 20 samples for calculating the 95th percentile value.

Strength of correlation increases when we restrict our analysis to industries that have more measurements, while information is lost by decreasing the number of assessed industries. There is a trade-off between the strength of correlation and the amount of information. We, therefore, prefer restricting to industries with 20 or more measurements because this increases the strength of correlation while not losing many industries.

Although the strength of the correlation seems fair [26], in some industries, the airborne lead level was not significantly correlated with BLL. For instance, “Manufacture of Basic Iron and Steel” showed a high rank of airborne X95 lead level but showed a low blood X95 lead level (Table 2). There would be several factors associated with the discrepancy, such as ventilation and respirator use. However, in this case, we suspect that health surveillance might be overly performed where only a small proportion of workers are exposed. This is represented by a high censoring rate of blood lead levels, which resulted in low X95 blood lead levels. Therefore, it is important to recognize that our exposure intensity indicators only applied to exposed workers in an industry, but not all workers in that industry.

The strength of our study lies primarily in using large measurement databases. We used nationwide blood lead measurement

data as a reference to be compared with nationwide airborne lead data. Besides, we examined various exposure indicators such as AM, GM, and X95, which may be used for evaluating exposure in various ways.

Our study has several limitations. First, work environment monitoring is conducted as a compliance measurement. Industrial hygienists select workers to be measured according to their own judgment, which is based on maximum potential exposure rather than random sampling. Using judgment sampling to select workers can result in an inability to compute precision, causing bias in estimates that are computed from these measurements [17]. Judgment sampling based on maximum potential exposure may lead to potential overestimation; on the other hand, because the cost of work environmental monitoring is paid by businesses, there is an inherent potential underestimation of exposure. We could not figure out the net effects of these factors, but we suspect that there is an underestimation of exposure. To cope with the potential underestimation, we considered analyzing data that only includes measurements that are above the LOD; however, this may only highlight the high exposure levels and poses a potential risk of overestimation. Therefore, we included all measurements in our analyses, and the results should be interpreted with caution considering this aspect.

Second, while workers to be measured are selected based on maximum potential exposure in workplace exposure monitoring, health examinees are selected based on minimum exposure, which means all workers exposed to the agent(s) should undergo special health examination regardless of exposure level. Considering the different sampling strategies, we used the Spearman rank correlation test rather than the Pearson correlation test.

Third, work environment measurement agencies are required to be equipped to perform atomic absorption spectrometry (AAS) by the Occupational Safety and Health Act in Korea. Therefore, airborne lead samples were quantified mainly by the AAS, while a small proportion of samples were analyzed by inductively coupled plasma spectrometry (ICP). The LOD of ICP is far lower than that of AAS [27]; however, we could not obtain those LODs, and therefore applied a single LOD based on the AAS method.

Fourth, we analyzed measurements from workplaces conducted both airborne lead and blood lead monitoring. However, workplaces that conducted only airborne lead monitoring showed lower summary statistics than workplaces that conducted both airborne lead and blood lead monitoring (data not shown). It may suggest that workplaces with high exposure are more likely to conduct blood lead monitoring. Therefore, our results may have a potential limitation in applying to all workplaces.

Fifth, there are potential errors in the coding of industries in the database. For instance, “Spinning of Textiles and Processing of Threads and Yarns” (SIC: 131) showed a high level of exposure; however, when further investigating the work process and sampling sites, we found that the industry involved the manufacturing of steel wires (rolling/heat treatment/chamfer) and was incorrectly coded. Therefore, we re-assigned these measurements to “Manufacture of Basic Iron and Steel” (SIC: 241). However, we could not check every industrial code and, therefore, the possibility of unchecked coding errors should be considered, especially in industries with small sample sizes.

The WEMD consists of a large proportion of measurements below LODs, which leads to difficulty in assessing and ranking exposure intensity. In the present study, we computed summary statistics of airborne lead exposure and tested optimal intensity indicators of airborne lead exposure by comparing to blood lead data across industries. This result may aid in establishing the K-CAREX exposure intensity, which would assess exposures to other carcinogens.

## Ethics

The study protocol was reviewed and approved by the Institutional Review Board of the Catholic Kwandong University, International St. Mary's Hospital, Incheon, Korea (IS17QIMI0035).

## Conflicts of interest

The authors declare that they have no conflict of interests.

## Acknowledgment

This study was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (grant number 2017R1D1A1B04032379). This study was also supported by the Occupational Safety and Health Research Institute grant (2020-OSHRI-845) funded by Korea Occupational Safety and Health Agency.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.shaw.2021.07.008>.

## References

- [1] Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a Systematic Analysis for the Global Burden of Disease Study 2016. *Lancet* 2017;390:1151–210.
- [2] Anttila SL, Boffetta P. Occupational cancers. 2nd ed. Cham: Springer; 2020.
- [3] Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Canc Inst* 1981;66:1191–308.
- [4] ACGIH. Lead and inorganic compounds BEL. In: ACGIH, editor. Documentation of threshold limit values and biological exposure indices. 7th ed. Cincinnati, OH: ACGIH; 2001.
- [5] Rushton L, Hutchings SJ, Fortunato L, Young C, Evans GS, Brown T, et al. Occupational cancer burden in Great Britain. *Br J Canc* 2012;107(Suppl. 1):S3–7.
- [6] Kauppinen T, Toikkanen J, Pukkala E. From cross-tabulations to multipurpose exposure information systems: a new job-exposure matrix. *Am J Ind Med* 1998;33:409–17.
- [7] Kauppinen T, Toikkanen J, Pedersen D, Young R, Ahrens W, Boffetta P, et al. Occupational exposure to carcinogens in the European Union. *Occup Environ Med* 2000;57:10–8.
- [8] Peters CE, Ge CB, Hall AL, Davies HW, Demers PA. CAREX Canada: an enhanced model for assessing occupational carcinogen exposure. *Occup Environ Med* 2015;72:64–71.
- [9] Koh D-H, Park J-H, Lee S-G, Kim H-C, Choi S, Jung H, et al. Development of Korean CARcinogen EXposure: an initiative of the occupational carcinogen surveillance system in Korea. *Ann Work Expos Health* 2021;65:528–38.
- [10] Symanski E, Greeson NMH, Chan W. Evaluating measurement error in estimates of worker exposure assessed in parallel by personal and biological monitoring. *Am J Ind Med* 2007;50:112–21.
- [11] IARC. IARC monographs on the evaluation of the carcinogenic risk to humans. In: Inorganic and organic lead compounds. Lyon: international agency for Research on cancer, vol. 87; 2006.
- [12] 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;12:46–53.
- [13] Koh DH, Park JH, Lee SG, Kim HC, Choi S, Jung H, et al. Combining lead exposure measurements and experts' judgment through a Bayesian framework. *Ann Work Expos Health* 2017;61:1054–75.
- [14] Koh DH, Park JH, Lee SG, Kim HC, Choi S, Jung H, et al. Estimation of lead exposure prevalence in Korean population through combining multiple experts' judgment based on objective data sources. *Ann Work Expos Health* 2018;62:210–20.
- [15] NIOSH. LEAD by flame AAS METHOD 7082. In: NIOSH manual of analytical methods. 4th ed. National Institute of Occupational Safety and Health. Washington, D.C: U.S. Government Printing Office; 2003.
- [16] NIOSH. LEAD in blood and urine 8003. In: NIOSH manual of analytical methods. 4th ed. National Institute of Occupational Safety and Health. Washington, D.C: U.S. Government Printing Office; 2003.
- [17] Millard SP. EnvStats, an R package for environmental statistics. Wiley StatsRef: Statistics Reference Online; 2014.
- [18] R Core Team. R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2020. <http://www.R-project.org/>.
- [19] ACGIH. 2019 TLVs and BEIs. Cincinnati, OH: American Conference of Governmental Industrial Hygienists; 2019.
- [20] Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg* 1990;5:46–51.
- [21] Hewett P, Ganser GH. A comparison of several methods for analyzing censored data. *Ann Occup Hyg* 2007;51:611–32.
- [22] GBD 2016 Occupational Carcinogens Collaborators. Global and regional burden of cancer in 2016 arising from occupational exposure to selected carcinogens: a systematic analysis for the Global Burden of Disease Study 2016. *Occup Environ Med* 2020;77:151.
- [23] European Food Safety Authority. Management of left-censored data in dietary exposure assessment of chemical substances. *EFSA Journal* 2010;8:1557.
- [24] Jahn SD, Bullock W, Ignacio JS. A strategy for assessing and managing occupational exposures. 4th ed. AIHA; 2015.
- [25] Rappaport SM, Kupper LL. Quantitative exposure assessment. El Cerrito, CA: Stephen Rappaport; 2008.
- [26] Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med* 2018;18:91–3.
- [27] NIOSH. Elements by ICP METHOD 7300. In: NIOSH manual of analytical methods. 4th ed. National Institute of Occupational Safety and Health. Washington, D.C: U.S. Government Printing Office; 2003.