Journal of Exposure Science and Environmental Epidemiology (2017) 27, 398–408

www.nature.com/jes

ORIGINAL ARTICLE

# Development of a source-exposure matrix for occupational exposure assessment of electromagnetic fields in the INTEROCC study

Javier Vila[1,2,3,7], Joseph D. Bowman[4,7], Jordi Figuerola[1], David Moriña[1], Laurel Kincl[5], Lesley Richardson[6], Elisabeth Cardis[1,2,3], On behalf of the INTEROCC Study Group[8]

To estimate occupational exposures to electromagnetic fields (EMF) for the INTEROCC study, a database of source-based measurements extracted from published and unpublished literature resources had been previously constructed. The aim of the current work was to summarize these measurements into a source-exposure matrix (SEM), accounting for their quality and relevance. A novel methodology for combining available measurements was developed, based on order statistics and log-normal distribution characteristics. Arithmetic and geometric means, and estimates of variability and maximum exposure were calculated by EMF source, frequency band and dosimetry type. The mean estimates were weighted by our confidence in the pooled measurements. The SEM contains confidence-weighted mean and maximum estimates for 312 EMF exposure sources (from 0 Hz to 300 GHz). Operator position geometric mean electric field levels for radiofrequency (RF) sources ranged between 0.8 V/m (plasma etcher) and 320 V/m (RF sealer), while magnetic fields ranged from 0.02 A/m (speed radar) to 0.6 A/m (microwave heating). For extremely low frequency sources, electric fields ranged between 0.2 V/m (electric forklift) and 11,700 V/m (high-voltage transmission line-hotsticks), whereas magnetic fields ranged between 0.14 μT (visual display terminals) and 17 μT (tungsten inert gas welding). The methodology developed allowed the construction of the first EMF–SEM and may be used to summarize similar exposure data for other physical or chemical agents.

*Journal of Exposure Science and Environmental Epidemiology* (2017) **27,** 398–408; doi:10.1038/jes.2016.60; published online 9 November 2016

**Keywords:** electromagnetic fields; log-normal distribution; occupational exposure assessment; semi-empirical exposure estimation; source-exposure matrix

## INTRODUCTION

Population-based case–control studies require the use of retrospective exposure assessment tools based on quality historical exposure data. However, the collection and analysis of these data is difficult, as measurements for some environmental and occupational agents, such as electromagnetic fields (EMF), are not systematically collected and, when available, are almost exclusively reported as aggregated and summarized results. Past efforts analysed and combined available exposure data in the literature for different agents.[1–8] They involved estimation of specific parameters from scarce measurements, using a limited number of equations based on the assumption of data log normality. Monte-Carlo simulations[1,2,7] were also used to recreate exposures when measurement data were sparse.

Measurements collected from the literature have been used in the construction of job-exposure matrices (JEMs), either alone or in combination with expert judgments. For EMF, JEMs have been created only for extremely-low frequency (ELF) magnetic fields[9–12] and electric shocks.[13,14] However, a worker's job title is insufficient to explain between-subject variability as exposure levels are influenced by other characteristics, such as industry, worker's tasks, specific equipment used or physical configuration of the workplace.[15,16] The JEM's mean of exposure measurements from a sample of workers with the same occupation introduces Berkson errors in the exposure estimate for a given individual, reducing the study's power to detect true hazards[17] and potentially biasing risk estimates.[18] Some authors[16] suggested the use of source-based measurements and questionnaires to improve EMF exposure assessment, allowing for a more individualized exposure estimation.

### The INTEROCC EMF Measurement Database

As part of the INTEROCC/INTERPHONE study of brain cancer, detailed information was collected for each job held by the study participants through a questionnaire on work organization (e.g., manual/automated), tasks (e.g., welding) and sources of exposure (e.g., type of equipment), divided into 12 occupational sections to take industrial activity into account. The aim was to combine the

[1]ISGlobal, Center for Research in Environmental Epidemiology, Parc de Recerca Biomèdica de Barcelona, Barcelona, Spain; [2]Universitat Pompeu Fabra, Barcelona, Spain; [3]CIBER Epidemiología y Salud Pública, Madrid, Spain; [4]National Institute for Occupational Safety and Health, Cincinnati, OH, USA; [5]Oregon State University, Corvallis, OR, USA and [6]University of Montreal Hospital Research Centre, Montreal, QC, Canada. Correspondence: Javier Vila, ISGlobal, Centre for Research in Environmental Epidemiology, Parc de Recerca Biomèdica de Barcelona, Doctor Aiguader, 88 1st Floor, Room 183.01.C, Barcelona 08003, Spain.
Tel.: +34 932 147 325. Fax: +34 932 147 302.
E-mail: javier.vila@isglobal.org
[7]These authors contributed equally to this work.
[8]Members of the INTEROCC Study Group are listed before the References.
Received 26 February 2016; accepted 18 August 2016; published online 9 November 2016

interview data and EMF exposure measurements from the literature for each source and/or task, to estimate individual cumulative exposures to electric fields ($E$) and magnetic fields ($B$ for lower frequencies and $H$ for higher frequencies) in four frequency bands: 0 Hz for static magnetic fields (SMF), 3–3000 Hz for extremely-low frequencies (ELF), 3 kHz–10 MHz for intermediate frequencies (IF) and 10 MHz–300 GHz for radio frequencies (RF).

Measurements for all the EMF sources identified through the study questionnaire were compiled into an occupational exposure measurement database (OEMD) with over 3000 records. The measurements collected were abstracted from published and unpublished resources, which were assessed based on their quality and relevance for our occupational brain cancer study. The OEMD was augmented with estimates of exposure range for 39 RF sources without available measurements in the literature, obtained from expert judgements. In total, exposures were compiled for 312 EMF sources commonly found in workplaces, covering the entire EMF frequency range. In this database, an EMF source refers to a specific piece of equipment and/or task that can lead to EMF exposure. Details of the construction and content of the OEMD were recently published,[19] and public access to this database is available at www.crealradiation.com/index.php/en/databases.

EMF data are usually reported using a variety of summary statistics, from arithmetic and geometric means (AM and GM), minimum (Min) and maximum (Max), only maximum, or values below or above the EMF meter's limits of detection, i.e., outside dynamic range (ODR). Several dosimetry types can be used when sampling EMF (i.e., personal, operator position or spot). Personal measurements are obtained with dosimeters by collecting exposures over an hour, a shift, or longer. Spot measurements are made at different distances from the source over shorter periods of time. Spot measurements performed at the usual worker's position are called "operator position" measurements.[20] The analysis and combination of these data entail several difficulties, as highlighted in similar efforts.[3,4,7] Since measurements are collected for different purposes and following different sampling strategies, quality and relevance for epidemiological studies also needs to be considered.

The aim of this article is to describe the methodology developed to combine the OEMD data into a source-exposure matrix (SEM), containing representative exposure estimates and their within-source variability for all EMF sources identified in the study.

## MATERIALS AND METHODS

The methodology developed has two main stages: (1) calculation of semi-empiric estimates of missing summary statistics in OEMD studies and (2) pooling of reported and/or estimated summary statistics. Pooled statistics were weighted by semiquantitative ratings from expert confidence evaluations on whether a study's measurement data are accurate and representative of long-term brain exposure.

### Semi-Empirical Methods For Estimating Missing Summary Statistics

Each OEMD record for a given EMF source may contain values for combinations of Min, Max, AM, GM, $N$ (sample size) and the minimum or maximum ODR limit for a specific frequency band and dosimetry type. To construct the SEM, we estimated AM, GM, SD and GSD for all EMF sources using this varied information. Our approach assumed that EMF exposure, like other environmental and occupational agents,[21–23] is log-normally distributed. The summary statistics from log-normal data obey several mathematical relationships (see Supplementary Appendix), including this equation for the standard normal quantile $z$ of the maximum data point:

$$z_{\text{Max}} = \frac{\ln \text{Max} - \ln \text{GM}}{\ln \text{GSD}} \tag{1}$$

and the analogous equation for $z_{\text{Min}}$. Our second assumption was that $z_{\text{Max}}$ and $z_{\text{Min}}$ are symmetric about zero:

$$z_{\text{Max}} = -z_{\text{Min}} \tag{2}$$

Equation 1 and other relationships[24] between the summary statistics AM, SD, GM and GSD, of a log-normal distribution, and parameters $z_{\text{Max}}$, Min and Max, were used to derive estimation formulae for missing statistics, depending on available values in the OEMD (Table 1). For OEMD records with values for $N$ (22% of the total), we further assumed that $z_{\text{Min}}$ and $z_{\text{Max}}$ were equal to their expected normal-order statistics,[25,26] which we call $E_N[z_{\text{Min}}]$ and $E_N[z_{\text{Max}}]$, as the expectation values of the extreme normal quantiles also have the symmetric quantile property.[22]

With values for $E_N[z_{\text{Max}}]$ obtained from a numerical algorithm,[25] these log-normal relationships could be solved exactly to obtain all summary statistics for OEMD records with three or more parameter values (estimation methods 1 and 2 in Table 1). When less information was available, solutions for the desired summary statistics were made possible by replacing the unknown GSD with its central tendency, $\overline{\text{GSD}}$, calculated from an OEMD subset with enough data for exact calculations using these two methods. This semi-empiric parameter plus the above approximations resulted in the formulae for estimation methods 3–5 in Table 1. For OEMD records without $N$, we replaced $E_N[z_{\text{Max}}]$ with a semi-empiric parameter, $\overline{z_{\text{Max}}}$, which equals the central tendency of $E_N[z_{\text{Max}}]$ from all OEMD records with values for $N$. With this substitution plus the symmetric quantile relationship (Eq. 2), formulae similar to those in Table 1 were derived (see Supplementary Appendix). OEMD records with $N = 1$ (i.e., single measurements) were considered to equal their AM and GM, whereas SD and GSD are undetermined.

When ODR measurements were reported, providing their corresponding limits of detection, they were entered into OEMD as ODRMin or ODRMax, with the corresponding Max or Min. For these entries, we estimated the desired statistics with models for the extreme exposures outside the dynamic range:

$$\text{Min} = \text{ODRMin} \times k_{\text{under}} \tag{3a}$$

$$\text{Max} = \text{ODRMax} \times k_{\text{over}} \tag{3b}$$

The correction factors $k_{\text{under}}$ and $k_{\text{over}}$ were estimated semi-empirically from a subset of ODR measurements that also reported the AM, so that:

$$k_{\text{under}} = \frac{\text{AM}^2}{\text{Max} \times \text{ODRMin} \times \sqrt{Q}} \tag{4a}$$

$$k_{\text{over}} = \frac{\text{AM}^2}{\text{Min} \times \text{ODRMax} \times \sqrt{Q}} \tag{4b}$$

where the parameter $Q = \overline{\text{GSD}}^{\ln \overline{\text{GSD}}}$ uses the central tendency $\overline{\text{GSD}}$, previously described. The central tendencies $\overline{k_{\text{over}}}$ and $\overline{k_{\text{under}}}$ were then used to obtain the desired statistics with the formulae in Table 2 (derived in the Supplementary Appendix).

The distributional characteristics of the data sets used to compute the semi-empiric parameters $\overline{\text{GSD}}$, $\overline{z_{\text{Max}}}$, $\overline{k_{\text{under}}}$ and $\overline{k_{\text{over}}}$ were examined to decide the best measure of their central tendency. Overall, data used for estimation of these semi-empiric statistics were not normally distributed; hence, the AM was never selected. When we confirmed that the data followed a log-normal distribution, the GM was used as the best measure of the central tendency. However, when the shape of the distribution was not clearly right-skewed, we chose the median value as it is considered the most appropriate metric for general skewed distributions.[27] Finally, we estimated mid point values for $k_{\text{over}}$ and $k_{\text{under}}$ using Eq. 4a and Eq. 4b. The median value was selected as the best estimate of the central tendency for these correction factors, complying the assumptions that $k_{\text{over}} > 1$ and $k_{\text{under}} < 1$.

### Confidence-Weighting Of Pooled Estimates

The lack of information on sample size and/or variance for many OEMD measurements ruled out inverse variance and other traditional measurement quality-weighting procedures.[28] Therefore, a methodology was developed to weight our pooled measurements based on their quality and relevance for epidemiological studies, in particular for INTEROCC. The weighting approach was based on the use of expert confidence ratings as weights. These ratings had been initially used to include/exclude measurements from the OEMD. INTEROCC experts, with experience in occupational EMF measurements, used a semiquantitative approach to derive an average rating for each set of measurements extracted from a study. Using a confidence evaluation form published with the OEMD

**Table 1.** Formulae for estimating AM, GM, SD and GSD from the OEMD data, not including values outside the dynamic range (ODR).

| *Method #* | *OEMD data* | *Estimated statistic* | *Formula* | *Assumptions*[a] |
|---|---|---|---|---|
| 1 | $N$, Min, Max and AM[b] | $\widehat{\mathrm{AM}} =$ | $\mathrm{AM}$ | — |
| | | $\widehat{\mathrm{GSD}} =$ | $\left(\sqrt{\mathrm{Max}/\mathrm{Min}}\right)^{1/E_N[z_{\mathrm{Max}}]}$ | A or B[b] |
| | | $\widehat{\mathrm{GM}} =$ | $\mathrm{AM}/\sqrt{\widehat{GSD}^{\ln\widehat{GSD}}}$ | — |
| | | $\widehat{\mathrm{SD}} =$ | $\frac{\mathrm{AM}}{\widehat{\mathrm{GM}}}\sqrt{\mathrm{AM}^2-\widehat{\mathrm{GM}}^2}$ | — |
| 2 | $N$, Min and Max | $\widehat{\mathrm{GM}} =$ | $\sqrt{\mathrm{Min}\times\mathrm{Max}}$ | — |
| | | $\widehat{\mathrm{GSD}} =$ | $\left(\sqrt{\mathrm{Max}/\mathrm{Min}}\right)^{1/\overline{z_{\mathrm{Max}}}}$ | B |
| | | $\widehat{\mathrm{AM}} =$ | $\widehat{GM}\sqrt{\widehat{GSD}^{\widehat{GSD}}}$ | — |
| | | $\widehat{\mathrm{SD}} =$ | $\frac{\widehat{\mathrm{AM}}}{\widehat{\mathrm{GM}}}\sqrt{\widehat{\mathrm{AM}}^2-\widehat{\mathrm{GM}}^2}$ | — |
| 3 | $N$[b] and [c]Max | $\widehat{\mathrm{GM}} =$ | $\mathrm{Max}/\overline{\mathrm{GSD}}^{E_N[z_{\mathrm{Max}}]}$ | A |
| | | $\widehat{\mathrm{AM}} =$ | $\mathrm{Max}\sqrt{Q^{-2E_N[z_{\mathrm{Max}}]}},\ \text{where}\ \ Q=\overline{\mathrm{GSD}}^{\ln\overline{\mathrm{GSD}}}$ | A or B[b], C |
| | | [d]$\widehat{\mathrm{GSD}} =$ | $\overline{\mathrm{GSD}}$ | C |
| | | $\widehat{\mathrm{SD}} =$ | $\frac{\widehat{\mathrm{AM}}}{\widehat{\mathrm{GM}}}\sqrt{\widehat{\mathrm{AM}}^2-\widehat{\mathrm{GM}}^2}$ | — |
| 4 | AM | $\widehat{\mathrm{AM}} =$ | $\mathrm{AM}$ | — |
| | | $\widehat{\mathrm{GM}} =$ | $\mathrm{AM}/\sqrt{Q}$ | C |
| | | $\widehat{\mathrm{GSD}} =$ | $\overline{\mathrm{GSD}}$ | C |
| | | $\widehat{\mathrm{SD}} =$ | $\frac{\mathrm{AM}}{\widehat{\mathrm{GM}}}\sqrt{\mathrm{AM}^2-\widehat{GM}^2}$ | — |
| 5 | GM | $\widehat{\mathrm{GM}} =$ | $\mathrm{GM}$ | — |
| | | $\widehat{\mathrm{AM}} =$ | $\mathrm{GM}\sqrt{Q}$ | C |
| | | $\widehat{\mathrm{GSD}} =$ | $\overline{\mathrm{GSD}}$ | C |
| | | $\widehat{\mathrm{SD}} =$ | $\frac{\widehat{AM}}{\mathrm{GM}}\sqrt{\widehat{\mathrm{AM}}^2-\mathrm{GM}^2}$ | — |

Notation: Hats denote estimates; bars denote semi-empiric parameters; other variables are input values. [a]In addition to the log-normality assumption and the symmetric quantile approximation (Eq. 2), additional assumptions were needed to derive some formulae: (A) Expected normal order statistic approximation: $z_{\mathrm{Max}}=E_N[z_{\mathrm{Max}}]$; (B) Semi-empiric value for $z_{\mathrm{Max}}$; (C) Semi-empiric value for GSD. [b]Where $N$ is not available, $E_N\,[z_{\mathrm{Max}}]$ is replaced with $\overline{z_{\mathrm{Max}}}$, which equals the median $E_N[z_{\mathrm{Max}}]$ from all available $N$ values (except for method #2, see main text). [c]Formulae when Min is the only input data are omitted because this case does not occur in the OEMD. [d]The semi-empiric parameter $\overline{\mathrm{GSD}}$ is calculated from OEMD records with data for methods #1 and #2.

paper,[19] each EMF expert first assigned a rating between 0 and 3 (0–1: low confidence; ≥ 1–2: moderate confidence; ≥ 2–3: high confidence) to eight specific factors of interest: sampling strategy, sample size, type of statistic reported, duty factor, dosimetry type, anatomical location, nature of exposure scenario and overall quality and reliability. Each set of measurements was rated by at least two experts and an average rating was assigned. We now used these ratings to adjust the pooled estimates to our confidence in the quality and relevance of the measurements.

## Data Pooling and Calculation of Confidence-Weighted Statistics

Finally, the EMF exposure statistics ($\mathrm{AM}_i$ and $\mathrm{GM}_i$), for each OEMD record $i$, were pooled to obtain mean exposure statistics by EMF source, frequency band and dosimetry type, using the expert ratings as confidence weights ($C_i$). Thus, confidence-weighted means (${}_{cw}\mathrm{AM}$ and ${}_{cw}\mathrm{GM}$) and SDs (${}_{cw}\mathrm{SD}$ and ${}_{cw}\mathrm{GSD}$) were calculated for each electric or magnetic field with these formulae derived in the Supplementary Appendix:

$$ {}_{cw}\mathrm{AM}=\sum_{i=1}^{N}C_iN_i\mathrm{AM}_i\Big/\sum_i C_iN_i \tag{5}$$

$$ \ln_{\mathrm{cw}}\mathrm{GM}=\sum_{i=1}^{N}C_iN_i\ln\mathrm{GM}_i\Big/\sum_i C_iN_i \tag{6}$$

$$ {}_{cw}\mathrm{SD}^2=\frac{\sum_i C_i\left[(N_i-1)\mathrm{SD}_i^2+N_i\left(\mathrm{AM}_i^2-{}_{cw}\mathrm{AM}^2\right)\right]}{\sum_i C_iN_i-\left(\sum_i C_i^2N_i/\sum_i C_iN_i\right)} \tag{7}$$

$$ \ln^2{}_{cw}\mathrm{GSD}=\frac{\sum_i C_i\left[(N_i-1)\ln^2\mathrm{GSD}_i+N_i\left(\ln^2\mathrm{GM}_i-\ln^2{}_{cw}\mathrm{GM}\right)\right]}{\sum_i C_iN_i-\left(\sum_i C_i^2N_i/\sum_i C_iN_i\right)} \tag{8}$$

**Table 2.** Formulae for calculating AM, GM, SD and GSD from OEMD data, including values outside the dynamic range (ODR).

| *Method #* | *OEMD data* | *Estimated statistic* | *Formula* | *Assumptions*[a] |
|---|---|---|---|---|
| 6 | *N,* Min, ODRMax and AM[c] | $^{b}\widehat{k_{over}}=$ | $\frac{\mathrm{AM}^2}{\mathrm{Min}\times \mathrm{ODRMax}\times \sqrt{Q}}$ | C and D |
| | | $^{b}\widehat{\mathrm{AM}}=$ | AM | — |
| | | $\widehat{\mathrm{GSD}}=$ | $\left(\mathrm{ODRMax}\times \overline{k_{\mathrm{over}}}/\mathrm{Min}\right)^{1/2E_N[z_{\mathrm{Max}}]}$ | A or B[c] and D |
| | | $\widehat{\mathrm{GM}}=$ | $\sqrt{\mathrm{Min}\times \overline{k_{\mathrm{over}}}\times \mathrm{ODRMax}}$ | D |
| | | $\widehat{\mathrm{SD}}$ | $\frac{\mathrm{AM}}{\mathrm{GM}}\sqrt{\mathrm{AM}^2-\widehat{\mathrm{GM}}^2}$ | — |
| 7 | *N,* ODRMin, Max and AM[c] | $^{b}\widehat{k_{under}}=$ | $\frac{\mathrm{AM}^2}{\mathrm{ODRMin}\times \mathrm{Max}\times \sqrt{Q}}$ | C and E |
| | | $\widehat{\mathrm{AM}}=$ | AM | — |
| | | $\widehat{\mathrm{GSD}}=$ | $\left(\mathrm{Max}/\mathrm{ODRMin}\times \overline{k_{\mathrm{under}}}\right)^{1/2E_N[z_{\mathrm{Max}}]}$ | A or B[c] and E |
| | | $\widehat{\mathrm{GM}}=$ | $\sqrt{\mathrm{ODRMin}\times \overline{k_{\mathrm{under}}}\times \mathrm{Max}}$ | E |
| | | $\widehat{\mathrm{SD}}=$ | $\frac{\mathrm{AM}}{\mathrm{GM}}\sqrt{\mathrm{AM}^2-\widehat{\mathrm{GM}}^2}$ | — |
| 8 | *N,* Min and ODRMax[c] | $\widehat{\mathrm{GM}}=$ | $\sqrt{\mathrm{Min}\times \overline{k_{\mathrm{over}}}\times \mathrm{ODRMax}}$ | D |
| | | $\widehat{\mathrm{GSD}}=$ | $\left(\mathrm{ODRMax}\times \overline{k_{\mathrm{over}}}/\mathrm{Min}\right)^{1/2E_N[z_{\mathrm{Max}}]}$ | A or B[c] and D |
| | | $\widehat{\mathrm{AM}}=$ | $\widehat{GM}\sqrt{\widehat{GSD}^{\widehat{GSD}}}$ | — |
| | | $\widehat{\mathrm{SD}}=$ | $\frac{\widehat{\mathrm{AM}}}{\mathrm{GM}}\sqrt{\widehat{\mathrm{AM}}^2-\widehat{\mathrm{GM}}^2}$ | — |
| 9 | *N, ODRMin and Max*[c] | $\widehat{\mathrm{GM}}=$ | $\sqrt{\mathrm{Max}\times \overline{k_{\mathrm{over}}}\times \mathrm{ODRMin}}$ | E |
| | | $\widehat{\mathrm{GSD}}=$ | $\left(\mathrm{Max}/\mathrm{ODRMin}\times \overline{k_{\mathrm{under}}}\right)^{1/2E_N[z_{\mathrm{Max}}]}$ | A or B[c] and E |
| | | $\widehat{\mathrm{AM}}=$ | $\widehat{GM}\sqrt{\widehat{GSD}^{\widehat{GSD}}}$ | — |
| | | $\widehat{\mathrm{SD}}=$ | $\frac{\widehat{\mathrm{AM}}}{\mathrm{GM}}\sqrt{\widehat{\mathrm{AM}}^2-\widehat{\mathrm{GM}}^2}$ | — |

Notation: Hats denote estimates; bars denote semi-empiric parameters; other symbols are input values. [a]In addition to the log-normality assumption and the symmetric quantile approximation (Eq. 2), additional assumptions were used to derive some formulae: assumptions A–C from Table 1, (D) semi-empiric value for $k_{\mathrm{over}}$; (E) semi-empiric value for $k_{\mathrm{under}}$. [b]Where $N$ is not available, $E_N$ [$z_{\mathrm{Max}}$] is replaced with $\overline{z_{\mathrm{Max}}}$, which equals the median of $E_N[z_{\mathrm{Max}}]$ from all available $N$ values in OEMD. [c]OEMD records with the data in methods #6 and #7 are used to calculate the semi-empiric parameters $\overline{k_{\mathrm{over}}}$ and $\overline{k_{\mathrm{under}}}$.

where $N_i$ is the number of individual measurements $i$ used to calculate the pooled summary statistics for each record $i$ in the OEMD. When $N_i$ was not available, the median $N=10$ from the OEMD records was used. Equations 7 and 8 were derived from the general formula for the unbiased weighted sample variance with non-random (namely, reliability) weights,[29] and become the classic formulae for the unweighted SD and GSD when $C_i=1$.

Since measurement data pooling was performed by dosimetry type, pooled exposure estimates obtained from spot measurements comprise several distances while those obtained from personal or operator position involve several anatomical locations (e.g., head, chest). Owing to the different availability of measurements, some sources in the SEM may have estimates for just one dosimetry type while others may have estimates for two or more. Maximum values, by source, frequency and dosimetry, were also included in the SEM, as well as information on the exact number of measurements pooled for each estimate.

To compare the values between pooled estimates for different dosimetries, we analyzed the overall difference between estimates for the same source by comparing different possible combinations (i.e., operator position *versus* spot; personal *versus* operator position and personal *versus* spot). For this analysis, we used the intraclass correlation coefficient (ICC) which, similarly to a one-way analysis of variance (ANOVA), allows comparing continuous values between groups.[30]

## Quality Control

To check the quality of the estimation process and ensure that the assumptions in our semi-empiric methods were appropriate for OEMD data, we performed tests based on fundamental statistical characteristics of log-normal distributions (such as Min < GM < AM < Max) as well as more specific checks based on EMF's physical properties.[19] Manual calculations were also performed, comparing the results with those from the programmed algorithms. Identified errors were corrected, ensuring that both statistical characteristics and physical laws were not breached.

## ANOVA

To test the ability of the SEM to assign different exposures to subjects in an epidemiological study, we performed a one-way ANOVA with EMF source as the independent variable and the (reported or estimated) $\mathrm{AM}_i$ from OEMD as the response variable. Because of the large number of sources in the matrix and the diversity of frequencies and EMF magnitudes, as an example, we compared the values for the mean electric fields for RF sources with three or more measurements at the operator position. As ANOVA requires normal residuals and equal variances, data were log-transformed for this analysis. Furthermore, after confirming heterocedasticity (unequal variances between groups) using Levene's test and assuming log-normality, we also used the non-parametric Welch's test with untransformed data.

### Validation

To test the validity of our methods to estimate parameters from limited summary statistics, Monte Carlo simulations were performed using the formulae in Table 1 on 10,000 random samples from a log-normal distribution with parameters similar to those found with EMF measurements (i.e., GM = 20 and GSD = 2.5). To make a realistic simulation, the sample size $N$ for each iteration was drawn randomly from the values in the OEMD and the semi-empiric parameters were derived from the simulated data, using the methods described above for obtaining $\overline{z_{\mathrm{Max}}}$ and $\overline{\mathrm{GSD}}$. For each simulation, the relative errors in the summary statistic estimates for all methods were calculated relative to the sample statistics (GM, AM, GSD, SD) calculated from the $N$ random draws:

$$\text{Relative error (RE)} = 100\%\frac{\text{Statistic estimate} - \text{Sample statistic}}{\text{Sample statistic}} \quad (9)$$

The mean of the RE over all simulations is, therefore, a measure of the bias, and its SD equals the relative SD (RSD), a measure of the precision. The overall uncertainty, which is considered an approximation to the accuracy,[31,32] can be estimated from these two values:

$$\text{Overall uncertainty} \equiv |\text{bias}| + 2 \times \text{RSD} \approx \text{Accuracy} \quad (10)$$

In addition, a split data set validation was performed for an RF source (dielectric heater), for which the mean pooled estimates were obtained from 84 E-field measurements. The confidence-weighted arithmetic mean was computed using both a test subset (i.e., random 50% samples) and the entire data set, repeating these computations 1000 times. To shed some light regarding possible changes over time for both exposure levels and measurements quality, we analyzed the available data for operator position measurements and confidence ratings — averaged by year — for two RF sources (aircraft radar, $n = 71$, years = 1974–1997; dielectric heater, $n = 84$, years = 1986–2004). Finally, to test our hypothesis that the measurements used in the SEM follow a log-normal distribution, we used the Shapiro–Wilk test on log-transformed data from EMF sources with three or more records. All statistical analyses and graphics were performed using R, version 3.2.3.[33]

## RESULTS

### Semi-Empirical Parameters $\overline{\mathrm{GSD}}, \overline{z_{\mathrm{Max}}}, \overline{k_{\mathrm{under}}}$ and $\overline{k_{\mathrm{over}}}$

Univariate statistics obtained for the parameters, $E_N[z_{\mathrm{Max}}]$, GSD, $k_{\mathrm{over}}$ and $k_{\mathrm{under}}$, are presented in Table 3. With $E_N[z_{\mathrm{Max}}]$, its distribution over all $N_i$ in OEMD was *a priori* unknown; therefore, we chose its median as the central tendency ($\overline{z_{\mathrm{Max}}} = 1.54$). GSD values tend to be log-normally distributed; therefore, we chose its GM as the semi-empiric parameter ($\overline{\mathrm{GSD}} = 2.31$). Following the logic with GSD's central tendency that the models are linear in the parameter logarithms, the GM was selected as the central tendency measure for the correction factors ($\overline{k_{\mathrm{over}}} = 1.47$) and ($\overline{k_{\mathrm{under}}} = 0.48$).

**Table 3.** Descriptive statistics for the estimated parameters calculated from subsets of the OEMD data.

| *Statistics* | $E_N[z_{\mathrm{Max}}]$ | *GSD* | $^{a}k_{\mathrm{over}}$ | $^{b}k_{\mathrm{under}}$ | $^{c}N$ |
|---|---|---|---|---|---|
| #Records | 372 | 100 | 7 | 22 | 372 |
| Min | 0.56 | 1.1 | 0.09 | 0.01 | 2 |
| Max | 3.26 | 6.4 | 18.44 | 44.54 | 1075 |
| AM | 1.45 | 2.59 | 5.69 | 5.81 | 19 |
| Median | **1.54** | 2.09 | 0.26 | 0.58 | **10** |
| GM | 1.33 | **2.31** | **1.47** | **0.48** | 9 |
| SD | 0.56 | 1.3 | 8.00 | 13.70 | 59 |
| CV% | 39% | 67% | 141% | 236% | 3% |

The statistics selected as a central tendency for use as semi-empiric parameters in Tables 1 and 2 are marked in bold. [a]$k_{\mathrm{over}}$ values less than 1 were not included in the calculations. [b]$k_{\mathrm{under}}$ values greater than 1 were not included in the calculations. [c]$N$=1 records were not considered in this table nor in the simulations.

**Table 4.** Description of estimation methods and measurements used in the SEM.

| *Estimation method # (data available)* | *Number of OEMD measurements* | | | | | |
|---|---|---|---|---|---|---|
| | *E-field* | *H-field* | *B-field* | | *PD*[d] | *Total* |
| *%* | | | | | | |
| 1 AM, Min and Max | 13 | 2 | 156 | 4 | 175 | 8 |
| 2 Min and Max | 226 | 133 | 115 | 40 | 514 | 23 |
| 3 Max | 269 | 163 | 134 | 30 | 596 | 27 |
| 4 AM | 71 | 34 | 317 | 17 | 439 | 20 |
| 5 GM | 12 | 18 | 19 | 0 | 49 | 2 |
| 6 AM, [a]ODRMax and Min | 0 | 0 | 1 | 0 | 1 | 0.1 |
| 8 Min and ODRMax[b] | 4 | 1 | 0 | 2 | 7 | 0.3 |
| 9 Max and ODRMin | 0 | 0 | 1 | 0 | 1 | 0.1 |
| 10 Single measurement[c] | 218 | 35 | 95 | 88 | 436 | 20 |
| *# Measurements per estimate* | *Number of SEM estimates* | | | | | |
| One measurement | 98 | 152 | 142 | — | 392 | 49 |
| Two measurements | 23 | 51 | 31 | — | 105 | 13 |
| 3-5 measurements | 31 | 76 | 59 | — | 166 | 21 |
| 6–10 Measurements | 23 | 35 | 31 | — | 89 | 11 |
| > 10 Measurements | 14 | 21 | 11 | — | 46 | 6 |

[a]ODR: Outside Dynamic Range. [b]Estimation method 7 (AM, ODRMin and Max) was not used as the OEMD did not have any cases with these values. [c]OEMD records with $N$=1 (i.e.. single measurements) were considered as AM=GM; hence, method 10 is not considered for Tables 1 and 2. [d]PD: power density.

### Exposure Estimates in the SEM

The SEM contains AM, GM and maximum exposure estimates for 312 occupational sources of EMF exposure by frequency band, and estimates of their associated variability (SD and GSD). The maximum values for each source are the maxima of both the Max and AM values from the input OEMD records. Exposure estimates are provided for various types of dosimetry (i.e., personal, spot and operator position) as well as for literature reviews and expert judgments. In total, there are 401 combinations of EMF source, frequency band and dosimetry type. Table 4 summarizes the records used to obtain the different mean estimates. In total, over 3000 measurements were compiled to create the SEM.

This table also outlines the different estimation methods used. While methods 2–5 were more frequently used, methods 1, 6, 8 and 9 were used less often. More than 400 single measurements were used in the calculations; hence, method 10 was also common. More than 50% of the estimates were obtained from two or more measurements, whereas the remaining values used only one measurement. As an example of the SEM results, Figure 1 shows the EMF sources with the minimum and maximum confidence-weighted geometric means (operator position) in the RF, IF and ELF frequency bands. Figure 2 shows the evolution of exposure levels and measurement quality for two RF sources over time. A considerable decrease of exposure levels and a slight increase in data quality are appreciable.

### The Confidence Evaluation Process

A total of 268 quantitative ratings were used as weights, as the same rating was assigned to two or more measurements if they shared the same characteristics. Of these, 135 (~50%) are above 2 (high confidence), 120 (~45%) are between 1 and 2 (moderate confidence), whereas only 13 ratings (~5%) are below 1 (low confidence). To illustrate the impact of the weighting process in

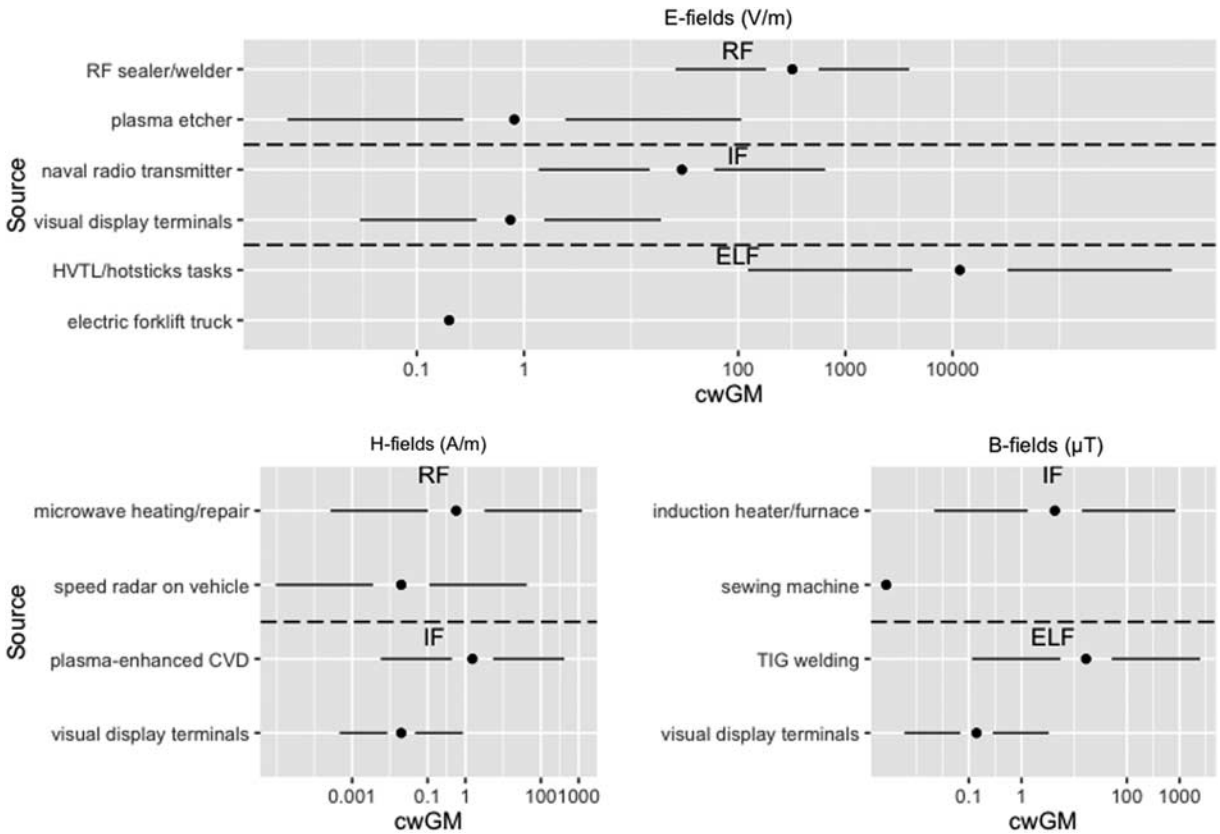


**Figure 1.** Quartile plots (25th and 75th percentiles) for EMF sources in the SEM with the highest and lowest *cw*GM for E-, H- and B-fields for operator position by frequency band. Estimates without whiskers (i.e., "electric forklift truck" and "sewing machine") were obtained from only one measurement. CVD, chemical vapor deposition; EMF, electromagnetic field; HVTL: high-voltage transmission lines; SEM, source-exposure matrix; TIG, tungsten inert gas.

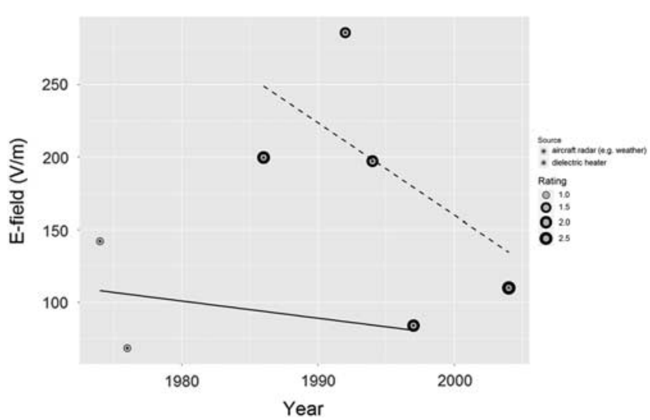


**Figure 2.** Operator position E-field measurements for two RF sources (i.e., aircraft radar and dielectric heater) collected from documents covering the time span 1974–2004. Data points and corresponding confidence ratings (i.e., the size of the point) were obtained by averaging the available data by year. The lines represent modeled linear regressions based on the averaged data. RF, radiofrequency.

the SEM calculations, Figure 3 shows the distribution of the E-field measurements used to calculate the mean (spot) estimate for an RF source, "continuous shortwave diathermy". These plots show weighted and unweighted regression lines over distance, highlighting the impact of the ratings on the weighted line (dashed). Measurements rated as low confidence are downplayed while moderate and high confidence values have a stronger influence on the final estimate.

ANOVA

In the ANOVA analysis to assess the ability of the SEM estimates to assign exposure variation for epidemiological analysis, the RF source explained almost 60% of the variability of the E-field and these differences were significant ($P < 0.0001$). The Welch's test ($P < 0.0001$) also confirmed these results.

Validation

The simulations based on the estimation formulae in Table 1 yielded overall uncertainties (i.e., accuracy) for GM and AM estimates between 47–143% (Table 5). For variability statistics, GSD estimates were obtained with accuracies between 33% and 78% while SD estimates yielded extreme overall uncertainties. An additional simulation using different $N$ values showed a clear pattern of better performance with larger sample sizes (data not shown). Furthermore, these simulations showed that some estimation methods have less overall uncertainty when $\overline{z_{\mathrm{Max}}}$ is used instead of $E_N[z_{\mathrm{Max}}]$ (see Table SIII in the Supplementary Appendix). Hence, our SEM calculations used the $z_{\mathrm{Max}}$ parameter that gave the best accuracy in the simulations of each statistic/method combination in Table 5. The split data set validation yielded a median relative error of −18%.

The Shapiro–Wilk test confirmed the log-normal hypothesis ($P$-value $> 0.05$) in ~85% of the analyzed sources. The ICC analysis showed moderate to substantial agreement for the compared dosimetries (i.e., ICC = 0.80 for spot *versus* operator position, $n = 18$;

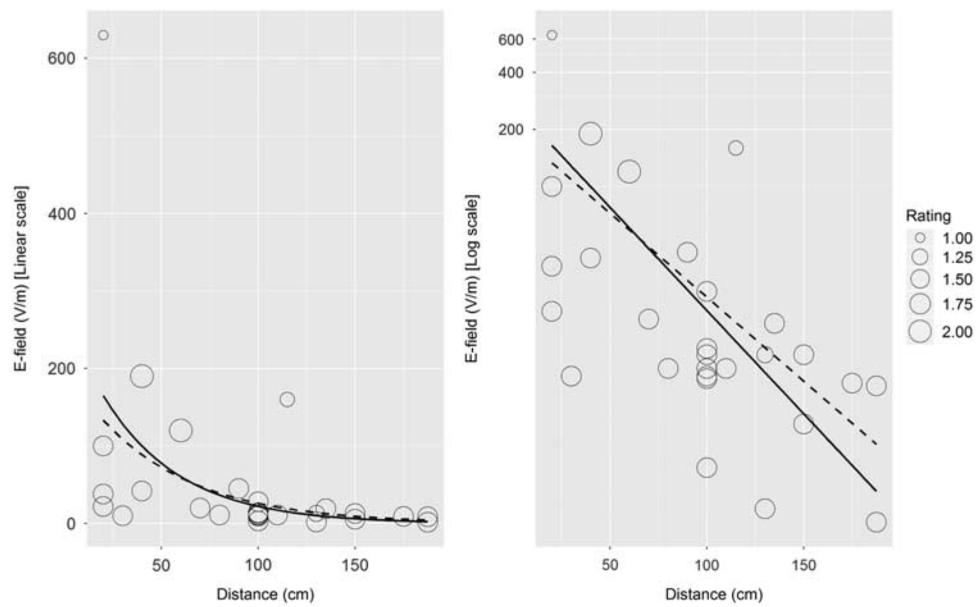


**Figure 3.** E-field measurements *versus* distance for OEMD data used to estimate the confidence-weighted mean exposure for the source "continuous shortwave diathermy" in the SEM. The bubbles represent data points with size proportional to the assigned rating level. The lines represent modeled exponential regression lines (dashed line, weighted) with the *y* axis in the linear (left graph) and logarithmic (right graph) scales. No ratings were assigned to these measurements below 1 or above 2. Thus, the "Rating" legend only includes a scale of sizes between these levels. OEMD, occupational exposure measurement database; SEM, source-exposure matrix.

**Table 5.** Uncertainties of the methods in Table 1 for estimating unknown summary statistics.

| *Estimated statistic* | *Measures of the relative error (RE)* | *Estimation method[a] (statistics with values from simulated sample[b])* | | | | |
|---|---|---|---|---|---|---|
| | | *1 (AM, Min and Max; %)* | *2 (Min and Max; %)* | *3 (Max; %)* | *4 (AM; %)* | *5 (GM; %)* |
| $\widehat{GM}$ | Mean RE (bias) | −9 | 3 | 14 | −5 | |
| | SD of RE (precision) | 51 | 25 | 64 | 21 | |
| | Accuracy | 51 | 53 | 143 | 47 | |
| $\widehat{AM}$ | Bias | | 21 | 16 | | 9 |
| | Precision | | 52 | 36 | | 20 |
| | Accuracy | | 125 | 88 | | 50 |
| ${}^{c}\widehat{GM}$ | Bias | 6 | 6 | 9 | 9 | 9 |
| | Precision | 13 | 13 | 35 | 35 | 35 |
| | Accuracy | 33 | 33 | 78 | 78 | 78 |
| $\widehat{SD}$ | Bias | 37 | 85 | 72 | 113 | 170 |
| | Precision | 74 | 254 | 411 | 1101 | 1651 |
| | Accuracy | 185 | 593 | 894 | 2316 | 3472 |

The uncertainty measures are calculated with Monte Carlo methods with 10,000 simulations of a random sample from a log-normal distribution with GM=20, GSD=2.5 and sample size [a]$N$ randomly drawn from those in OEMD. [a]Estimation methods #1, 3, 4 and 5 use $E_N[z_{\mathrm{Max}}]$ while method #2 uses $\overline{z_{\mathrm{Max}}}$ (i.e., median of $E_N[z_{\mathrm{Max}}]$). [b]The median of the $N$ values in OEMD=10. [c]Methods #3–5 for $\widehat{GSD}$ have the same results as they all use $\widehat{GSD}$.

ICC = 0.69 for personal *versus* operator position, $n = 9$; ICC = 0.53 for spot *versus* personal dosimetries, $n = 20$).

## DISCUSSION

This work allowed the construction of a SEM containing estimated exposure statistics for the most common occupational sources of EMF exposure, identified through the INTEROCC study questionnaire. This database represents a new approach for occupational exposure assessment, based on EMF sources independent of occupation. The SEM will be available online as a free-access tool at http://www.crealradiation.com/index.php/es/databases. Although the current version does not include all possible EMF sources, it can be updated with new or newly identified measurements and sources.

One advantage of the source-based approach is that personal determinants of exposure obtained from questionnaires should reduce Berkson errors, increasing the validity and reliability of both exposure and risk estimates.[15] However, the SEM mean exposures will still leave residual Berkson errors because of the combination of measurements from different studies and locations (i.e., distances or anatomical positions). Another advantage is the SEM's ability to evaluate occupational exposures to RF and IF fields. As no JEM yet exists for these higher frequencies, only a source-based approach can provide quantitative estimates of exposure for INTEROCC and other studies. The results of the ANOVA and the non-parametric test confirmed the existence of significant between-source variability, which allows the assignment of different exposures to study subjects, necessary for identifying exposure–response relationships in risk analysis. Previous efforts to reduce exposure misclassification included the development of task-exposure matrices for other agents.[34–37] However, earlier advocates of a source-based approach for EMF exposure assessment[38–41] recommended the use of combined estimates from a JEM together with information such as duration and location related to specific sources of exposure. To our knowledge, this is the first time that a full source-based approach, independent of the occupation, has been attempted.

The mean exposure (i.e., AM or GM) was selected as the central metric in the SEM because it best represents measurements taken in diverse settings. There has been considerable discussion whether the AM or GM from JEMs best reduces Berkson errors in an epidemiological analysis,[42–45] and these same considerations apply to the SEM. Although the GM is the best estimate of the central tendency for log-normally distributed data, the AM has been considered the best summary measure for linear and convex dose–response relationships, while the GM would be a better metric when the proposed mechanism is log-linear (i.e., the response is proportional to the logarithm of the exposure/dose).[45–49] The availability of both AM and GM in the SEM allows for the selection of the more appropriate metric for the study hypothesis. The provision of within-source variability statistics (i.e., $_{cw}SD$ and $_{cw}GSD$) also allows for the correction of risk estimates for bias attributable to Berkson error as well as for uncertainty propagation analysis.[18,50–52] Moreover, although bias estimates were provided for only half of the methods, the use of this information as weights for the pooled statistics should be explored in the future.

Several methods were developed for estimating parameters based on scarce measurement data. Methods 1 and 2 require enough available variables but allow estimating AM, GM, GSD and SD based on exact relationships between the true statistics of a log-normal distribution.[24] Method 2, in particular, was based on an estimation formula, $\widehat{\mathrm{GM}} = \sqrt{\mathrm{Min} \times \mathrm{Max}}$, which has recently been popularized by physicists for "guesstimation"[53,54] and variants have been used in exposure assessment efforts.[3,4,7] To extend this estimation technique to the other combinations of statistics, we introduced several semi-empirical methods to derive equations where the literature provided insufficient data for exact solutions. Although these semi-empiric estimates fill many of the gaps in the diverse data available, they add to the uncertainties of the exposure assessment, as shown by the simulations in Table 5. Moreover, the method we used to reliably estimate parameters from only maximum values, as proved by the relatively low bias obtained in the simulations, provides a novel approach which, to our knowledge, was lacking in the present literature. For data combinations not considered in Tables 1 and 2, which may also be found in the literature, we provided the assumptions and premise formulae needed to easily derive appropriate methods.

We provided evidence for the reliability of our methodology through both simulations and a split data set validation. While the simulated accuracies are far greater than the 25% accuracy criterion established by NIOSH for occupational exposure measurements,[31] most methods for GM had overall uncertainties of 53% or less, which we consider sufficient for retrospective epidemiology. Moreover, these accuracies are expected to improve if GSD and/or SD are extracted from the literature or larger sample sizes are used, as seen in our additional simulations and previous studies.[55] However, the impact of these larger exposure assessment errors on risk estimates should be investigated. For the methods in Table 2, a comprehensive approach for evaluating uncertainties was not found. Although some of the estimated values violated the assumptions $k_{\mathrm{over}} > 1$ and $\underline{k}_{\mathrm{under}} < 1$, one of the semi-empirical estimated parameters ($\underline{k}_{\mathrm{over}} = 1.47$) compared well with a calculated value ($k_{\mathrm{over}} = 1.41$) based on empirical monitoring data (i.e., measurements of the same location using two different ELF– magnetic field meters at a car factory in the Netherlands). However, further validations would be advisable for these correction factors as well as for the equations in Tables 1 and 2.

The influence of measurement quality on exposure and risk estimates requires a rigorous evaluation, including transparency in the way data are weighted for their actual or relative value.[28,56] Some authors[5,6,28] proposed the use of sample size or inverse variance to obtain quality-weighted exposure estimates. However, the frequent lack of this information for measurements in the EMF literature makes the use of such approaches unfeasible. We used expert confidence ratings to adjust our estimates to the quality and relevance of the pooled measurements, overcoming this difficulty. The scoring system we selected is in agreement with a recent proposal for the evaluation of exposure data quality,[28] where a method to classify measurements in four quality groups (i.e., good, moderate, poor and unacceptable) is proposed. Although we did not distinguish between poor and unacceptable measurements, those rated as low confidence (0–1) were generally excluded from the pooling. However, some low confidence measurements, for which no better data were available, were included in the SEM. On the basis of this confidence classification, sensitivity analysis may be conducted (e.g., excluding lower-quality data). This method also allows accounting for sampling characteristics, whereas other weighting approaches, such as inverse variance, only take into account the statistical uncertainty and do not consider other potentially important factors (e.g., quality of the task description and the sampling devices or focus on high exposures), which can be easily identified in the literature and may determine the quality and relevance of a measurement.[28] Thus, similarly to meta-analysis in epidemiology,[57] measurements with higher confidence have a larger contribution to the weighted mean. Finally, this approach allowed the raters to use a simple additive method to assign scores, which has been shown to be a good predictor of overall methodological quality.[58,59]

One possible weakness of the SEM was our need to include the less accurate spot and operator position dosimetries in order to provide exposure data for some of the reported sources. However, the results of the ICC showed that the overall differences between the three dosimetry types are small. Estimates obtained from operator position or spot measurements may, therefore, be reliably used as surrogates of personal exposure when this is not available. Moreover, confidence-weighted estimates were adjusted to head exposure through the confidence-weighting process. Measurements made at head location obtained higher ratings and were upgraded in the pooling. To allow the use of the SEM in studies on other locations (e.g., chest, gonads) — where different weighting approaches may be applied — the unweighted estimates were also provided. As the confidence evaluations for all eight factors are stored in the SEM database, future studies may reduce the weight given to head measurements while retaining the other seven factors affecting measurement quality.

Another weakness is the lack of use of anatomical location and distance information collected in the OEMD for spot and operator position measurements. SEM values refer, therefore, to average levels over different exposure scenarios, which provide the within-source variability inherent within each mean estimate. Pooled estimates represent different situations of exposure depending on the dosimetry type. Estimates for personal and operator position comprise measurements at different anatomical locations (e.g., head, chest or waist) while spot estimates include exposures at different distances (e.g., 30–100 cm for most ELF sources). However, as shown in Figure 3, the availability of this information may allow future modeling of exposures at specific distances and locations, useful in studies interested in other body parts.

The analysis of the available measurement data for different years showed signs of a slight data quality increase over time, which is reasonable considering the improvements in industrial hygiene.[60] Exposure levels, on the contrary, showed a clear decrease pattern, which is in line with the trends shown by other technologies such as mobile phones.[61] However, as level changes are limited to one order of magnitude and OEMD data for the same source seldom span several years, we do not expect that these changes will have a strong effect on the SEM estimates.

The SEM can be used to assess EMF exposures for other occupational and residential epidemiologic studies that have collected individual information on the use of EMF sources. Such studies require questionnaires that elicit individual information about the type of EMF sources used/exposed, as well as about conditions of use (e.g., distance to the source, automation) to adjust the SEM estimates to the specific tasks and work characteristics of the individual. If the time-weighted average or cumulative exposures are desired, the questionnaire also needs to obtain information on the frequency and duration of use/exposure. In INTEROCC, industry was considered through the classification of all EMF sources into 12 occupational sections.[19] Therefore, the variability due to industrial differences is embedded within the type of source itself, which together with the aforementioned information on other exposure determinants allows a detailed estimation of a subject's level of exposure. While the means in the SEM are most useful in chronic disease studies, the EMF maxima can be applied to acute effects, such as electromagnetic interference with pacemakers and other medical devices.[62]

In conclusion, the methodology described allowed us to construct the first SEM for EMF exposure assessment, based on measurements identified in the literature, and supplemented with expert judgment estimates for sources without available measurements, whose details will be published elsewhere. These methods made use of measurement data that more conventional methods would have discarded. Although more analyses are needed on their uncertainty and validity, they may also be useful for other physical and chemical agents for which available measurement data are sparse and traditional methods are insufficient.

The SEM will be used to estimate cumulative RF and ELF exposures of the INTEROCC subjects, through algorithms that combine SEM means with individual data on exposure determinants collected by interviews. This more individualized exposure assessment will potentially increase within-job variability among subjects and reduce uncertainty because of misclassification and Berkson errors. We expect that this approach will strengthen our ability to evaluate potential health effects from EMF exposures.

## ABBREVIATIONS

AM, Arithmetic mean; B-field, Magnetic flux density, in μT (low-frequency fields); CVD, Chemical vapor deposition; E-field, Electric field strength, in V/m; ELF, Extremely low frequency (3–3000 Hz); EMF, Electromagnetic fields; GM, Geometric mean; GSD, Geometric standard deviation; H-field, Magnetic field strength, in A/m (high-frequency fields); HVTL, High-voltage transmission lines; IF, Intermediate frequency (3 kHz–10 MHz); Max, Maximum; Min, Minimum; N, sample size; ODR, Outside dynamic range (The range between an EMF instrument's overload input and its minimum input with acceptable accuracy); PD, Power density, in watts per square meter (W/m$^2$); RF, Radiofrequency (10 MHz–300 GHz); SD, Standard deviation; SMF, Static Magnetic Fields, in microTesla (μT), 0 Hz; TIG, Tungsten inert gas; $z_{Max}$, Standard normal quantile of a data set's maximum value.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank Dave Conover (deceased), Ed Mantiply and Leeka Kheifets (USA); Dave McLean (New Zealand); Hans Kromhout (the Netherlands); Paolo Vecchia (Italy); Louis Nadon (Canada); Wout Joseph (Belgium); Martie van Tongeren, Simon Mann, Myron Maslanyj, Cristian Goiceanu and Carolina Calderon (UK), Peter Gajšek (Slovenia) and Tommi Alanko, Maila Hietanen and Maria Tiikkaja (Finland) for providing and/or assessing measurements. Jérôme Lavoué (Canada) and Stanley Shulman (USA) contributed to the development of the SEM methodology. We also thank Professor Pere Puig (Autonomous University of Barcelona) for his input on the history of estimation. This work was funded by the National Institutes for Health (NIH) Grant No. 1R01CA124759-01. Coding of the French occupational data was in part funded by AFSSET (Convention N° ST-2005-004). The INTERPHONE study was supported by funding from the European Fifth Framework Program, "Quality of Life and Management of Living Resources" (contract 100 QLK4-CT-1999901563) and the International Union against Cancer (UICC). The UICC received funds for this purpose from the Mobile Manufacturers' Forum and GSM Association. Provision of funds to the INTERPHONE study investigators via the UICC was governed by agreements that guaranteed INTERPHONE's complete scientific independence (http://interphone.iarc.fr/interphone_funding.php). In Australia, funding was received from the Australian National Health and Medical Research 5 Council (EME Grant 219129), with funds originally derived from mobile phone service licence fees; a University of Sydney Medical Foundation Program; the Cancer Council NSW and The Cancer Council Victoria. In Montreal, Canada, funding was received from the Canadian Institutes of Health Research (project MOP-42525); the Canada Research Chair programme; the Guzzo-CRS Chair in Environment and Cancer; the Fonds de la recherche en sante du Quebec; the Société de recherché sur le cancer; in Ottawa and Vancouver, Canada, from the Canadian Institutes of Health Research (CIHR), the latter including partial support from the Canadian Wireless Telecommunications Association; the NSERC/SSHRC/McLaughlin Chair in Population Health Risk Assessment at the University of Ottawa. In France, funding was received by l'Association pour la Recherche sur le Cancer (ARC; Contrat N85142) and three network operators (Orange, SFR, Bouygues Telecom). In Germany, funding was received from the German Mobile Phone Research Program (Deutsches Mobilfunkforschungsprogramm) of the German Federal Ministry for the Environment, Nuclear 45 Safety, and Nature Protection; the Ministry for the Environment and Traffic of the state of Baden — Wurttemberg; the Ministry for the Environment of the state of North Rhine-Westphalia; the MAIFOR Program (Mainzer Forschungsforderungsprogramm) of the University of Mainz. In New Zealand, funding was provided by the Health Research Council, Hawkes Bay Medical Research Foundation, the Wellington Medical Research Foundation, the Waikato Medical Research Foundation and the Cancer Society of New Zealand. Additional funding for the UK study was received from the Mobile Telecommunications, Health and Research (MTHR) program, funding from the Health and Safety Executive, the Department of Health, the UK Network Operators (O2, Orange, T-Mobile, Vodafone, '3") and the Scottish Executive. All industry funding was governed by contracts guaranteeing the complete scientific independence of the investigators.

## DISCLAIMER

The findings and conclusions in this paper have not been formally disseminated by the National Institute for Occupational Safety and Health and should not be construed to represent any agency determination or policy.

## INTEROCC STUDY GROUP MEMBERS: INTERNATIONAL COORDINATION

Elisabeth Cardis[9], Laurel Kincl[10], Lesley Richardson[11], Geza Benke[12], Jérôme Lavoué[13] and Jack Siemiatycki[13], Daniel Krewski[14], Marie-Elise Parent[15], Martine Hours[16], Brigitte Schlehofer[17] and Klaus Schlaefer[17], Joachim Schüz[18], Maria Blettner[19], Siegal Sadetzki[20], Dave McLean[21], Sarah Fleming[22], Martie van Tongeren[23], Joseph D Bowman[24]

[9]CREAL, Spain; [10]now at Oregon State University, USA; [11]now at University of Montreal Hospital Research Centre, Canada; [12]Monash University, Australia; [13]University of Montreal Hospital Research Centre, Canada; [14]University of Ottawa, Canada; [15]INRS-Institut Armand-Frappier, France; [16]IFSTTAR, Germany; [17]DKFZ, Germany; [18]now at IARC, France; [19]Universitätsmedizin Mainz, Germany; [20]Gertner Institute, Chaim Sheba Medical Center and Tel Aviv University, Israel; [21]Massey University, New Zealand; [22]University of Leeds, UK; [23]Institute of Occupational Medicine, UK; [24]NIOSH, USA.

## REFERENCES

1 Sauvé J-F, Beaudry C, Bégin D, Dion C, Gérin M, Lavoué J. Statistical modeling of crystalline silica exposure by trade in the construction industry using a database compiled from the literature. *J Environ Monit JEM* 2012; **14**: 2512–2520.
2 Sauvé J-F, Beaudry C, Bégin D, Dion C, Gérin M, Lavoué J. Silica exposure during construction activities: statistical modeling of task-based measurements from the literature. *Ann Occup Hyg* 2013; **57**: 432–443.
3 Koh D-H, Nam J-M, Graubard BI, Chen Y-C, Locke SJ, Friesen MC. Evaluating temporal trends from occupational lead exposure data reported in the published literature using meta-regression. *Ann Occup Hyg* 2014; **58**: 1111–1125.
4 Koh D-H, Locke SJ, Chen Y-C, 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**: 605–616.
5 Hein MJ, Waters MA, van Wijngaarden E, Deddens JA, Stewart PA. Issues when modeling benzene, toluene, and xylene exposures using a literature database. *J Occup Environ Hyg* 2008; **5**: 36–47.
6 Hein MJ, Waters MA, Ruder AM, Stenzel MR, Blair A, Stewart PA. Statistical modeling of occupational chlorinated solvent exposures for case-control studies using a literature-based database. *Ann Occup Hyg* 2010; **54**: 459–472.
7 Lavoué J, Bégin D, Beaudry C, Gérin M. Monte Carlo simulation to reconstruct formaldehyde exposure levels from summary parameters reported in the literature. *Ann Occup Hyg* 2007; **51**: 161–172.
8 Park D, Stewart PA, Coble JB. Determinants of exposure to metalworking fluid aerosols: a literature review and analysis of reported measurements. *Ann Occup Hyg* 2009; **53**: 271–288.
9 Burau KD, Huang B, Whitehead LW, Delclos GM, Downs TD. A system linking occupation history questionnaire data and magnetic field monitoring data. *J Expo Anal Environ Epidemiol* 1998; **8**: 231–252.
10 Forssén UM, Mezei G, Nise G, Feychting M. Occupational magnetic field exposure among women in Stockholm County, Sweden. *Occup Environ Med* 2004; **61**: 594–602.
11 Bowman JD, Touchstone JA, Yost MG. A population-based job exposure matrix for power-frequency magnetic fields. *J Occup Environ Hyg* 2007; **4**: 715–728.
12 Gobba F, Bravo G, Rossi P, Contessa GM, Scaringi M. Occupational and environmental exposure to extremely low frequency-magnetic fields: a personal monitoring study in a large group of workers in Italy. *J Expo Sci Environ Epidemiol* 2011; **21**: 634–645.
13 Huss A, Vermeulen R, Bowman JD, Kheifets L, Kromhout H. Electric shocks at work in Europe: development of a job exposure matrix. *Occup Environ Med* 2013; **70**: 261–267.
14 Vergara XP, Fischer HJ, Yost M, Silva M, Lombardi DA, Kheifets L. Job exposure matrix for electric shock risks with their uncertainties. *Int J Environ Res Public Health* 2015; **12**: 3889–3902.
15 Kelsh MA, Kheifets L, Smith R. The impact of work environment, utility, and sampling design on occupational magnetic field exposure summaries. *AIHAJ J Sci Occup Environ Health Saf* 2000; **61**: 174–182.
16 Kheifets L, Bowman JD, Checkoway H, Feychting M, Harrington JM, Kavet R *et al*. Future needs of occupational epidemiology of extremely low frequency electric and magnetic fields: review and recommendations. *Occup Environ Med* 2009; **66**: 72–80.
17 Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998; **55**: 651–656.
18 Greenland S, Fischer HJ, Kheifets L. Methods to explore uncertainty and bias introduced by job exposure matrices. *Risk Anal Off Publ Soc Risk Anal* 2015; **36**: 74–82.
19 Vila J, Bowman JD, Richardson L, Kincl L, Conover DL, McLean D *et al*. A source-based measurement database for occupational exposure assessment of electromagnetic fields in the INTEROCC study: a literature review approach. *Ann Occup Hyg* 2016; **60**: 184–204.
20 Bowman JD, Kelsh MA, Kaune WT. *Manual for Measuring Occupational Electric and Magnetic Field Exposures*. DHHS, CDC, National Institute for Occupational Safety and Health (NIOSH): Cincinnati, OH, USA, 1998: http://www.cdc.gov/niosh/docs/98-154/pdfs/98-154.pdf.
21 Hitchcock RT, Patterson RM. *Radio-Frequency and ELF Electromagnetic Energies: A Handbook for Health Professionals*. Van Nostrand Reinhold: New York, 1995.
22 Rappaport S, Kupper L. *Quantitative Exposure Assessment*. Stephen Rappaport: El Cerrito, CA, USA, 2008.
23 Roosli M(ed.) *Epidemiology of Electromagnetic Fields*. CRC Press: Boca Raton, 2014.
24 Aitchison J, Brown JAC *The Lognormal Distribution*. Cambridge University Press: Cambridge, UK, 1963.
25 Royston JP. Algorithm AS 177: expected normal order statistics (exact and approximate). *J R Stat Soc* 1982; **31**: 161–165.
26 Zwillinger D, Kokoska S, Order statistics. In: *Standard Probability and Statistics Tables and Formulae*. Chapman and Hall. CRC Press: Boca Raton, USA 1999: https://www.crcpress.com/CRC-Standard-Probability-and-Statistics-Tables-and-Formulae/Zwillinger-Kokoska/9781584880592. Accessed 7 December 2015.
27 Baker S, Driver J, McCallum D (eds.) *Residential Exposure Assessment*. Springer: Boston, MA, USA, 2001: http://link.springer.com/10.1007/978-1-4615-1279-0. Accessed 11 July 2016.
28 Tielemans E, Marquart H, De Cock J, Groenewold M, Van Hemmen J. A proposal for evaluation of exposure data. *Ann Occup Hyg* 2002; **46**: 287–297.
29 Harrell FE Jr, Hmisc: Harrell Miscellaneous. R package version 3.17-0 2015: http://CRAN.R-project.org/package; http://www.inside-r.org/packages/cran/hmisc/docs/wtd.stats = Hmisc.
30 Teschke K, Olshan AF, Daniels JL, De Roos AJ, Parks CG, Schulz M *et al*. Occupational exposure assessment in case-control studies: opportunities for improvement. *Occup Environ Med* 2002; **59**: 575–593 discussion 594.
31 NIOSH. Manual of Analytical Methods *National Institute for Occupational Safety and Health*, 4th edn. US Dept of Health and Human Services (NIOSH). 1994: http://www.cdc.gov/niosh/docs/2003-154/default.html. Accessed 25 October 2016.
32 British Standards Institution, Workplace atmospheres - General requirements for the performance of procedures for the measurement of chemical agents, European Standard BS EN 482:1994, ISBN 0580236447.
33 R Core Team. R: a language and environment for statistical computing *R Foundation for Statistical Computing*, URL http://www.R-project.org/, 2014.
34 Benke G, Sim M, Fritschi L, Aldred G. Beyond the job exposure matrix (JEM): the task exposure matrix (TEM). *Ann Occup Hyg* 2000; **44**: 475–482.
35 Benke G, Sim M, Fritschi L, Aldred G. A task exposure database for use in the alumina and primary aluminium industry. *Appl Occup Environ Hyg* 2001; **16**: 149–153.
36 Dick FD, Semple SE, van Tongeren M, Miller BG, Ritchie P, Sherriff D *et al*. Development of a task-exposure matrix (TEM) for pesticide use (TEMPEST). *Ann Occup Hyg* 2010; **54**: 443–452.
37 Hyland RA, Yates DH, Benke G, Sim M, Johnson AR. Occupational exposure to asbestos in New South Wales, Australia (1970-1989): development of an asbestos task exposure matrix. *Occup Environ Med* 2010; **67**: 201–206.
38 Semple S, Cherrie JW. Factors influencing personal magnetic field exposure: preliminary results for power utility and office workers. *Ann Occup Hyg* 1998; **42**: 167–171.
39 Coble JB, Dosemeci M, Stewart PA, Blair A, Bowman J, Fine HA *et al*. Occupational exposure to magnetic fields and the risk of brain tumors. *Neuro Oncol* 2009; **11**: 242–249.
40 Friesen MC, Coble JB, Lu W, Shu X-O, Ji B-T, Xue S *et al*. Combining a job-exposure matrix with exposure measurements to assess occupational exposure to benzene in a population cohort in shanghai, china. *Ann Occup Hyg* 2012; **56**: 80–91.
41 Koh D-H, Bhatti P, Coble JB, Stewart PA, Lu W, Shu X-O *et al*. 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**: 9–16.
42 Seixas NS, Robins TG, Moulton LH. The use of geometric and arithmetic mean exposures in occupational epidemiology. *Am J Ind Med* 1988; **14**: 465–477.
43 Crump KS. On summarizing group exposures in risk assessment: is an arithmetic mean or a geometric mean more appropriate? *Risk Anal* 1998; **18**: 293–297.
44 EPA. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. US Environmental Protection Agency (USEPA): Washington, DC, USA, 1992.
45 Steenland K, Deddens JA, Zhao S. Biases in estimating the effect of cumulative exposure in log-linear models when estimated exposure levels are assigned. *Scand J Work Environ Health* 2000; **26**: 37–43.
46 Deng Q, Wang X, Wang M, Lan Y. Exposure-response relationship between chrysotile exposure and mortality from lung cancer and asbestosis. *Occup Environ Med* 2012; **69**: 81–86.
47 Pronk A, Preller L, Raulf-Heimsoth M, Jonkers ICL, Lammers J-W, Wouters IM *et al*. Respiratory symptoms, sensitization, and exposure response relationships in spray painters exposed to isocyanates. *Am J Respir Crit Care Med* 2007; **176**: 1090–1097.
48 Lippmann M. The search for non-linear exposure-response relationships at ambient levels in environmental epidemiology. *Nonlinearity Biol Toxicol Med* 2005; **3**: 125–144.
49 Wallace ME, Grantz KL, Liu D, Zhu Y, Kim SS, Mendola P. Exposure to ambient air pollution and premature rupture of membranes. *Am J Epidemiol* 2016; **183**: 1114–1121.
50 Bateson TF, Wright JM. Regression calibration for classical exposure measurement error in environmental epidemiology studies using multiple local surrogate exposures. *Am J Epidemiol* 2010; **172**: 344–352.
51 Simon SL, Hoffman FO, Hofer E. The two-dimensional Monte Carlo: a new methodologic paradigm for dose reconstruction for epidemiological studies. *Radiat Res* 2015; **183**: 27–41.

52 Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol* 2006; **21**: 871–876.
53 Mahajan S. *Street-Fighting Mathematics: The Art of Educated Guessing and Opportunistic Problem Solving*. The MIT Press: Cambridge, Massachusetts, USA, and London, England, UK, 2010: http://mitpress.mit.edu/books/street-fighting-mathematics.
54 Weinstein L, Adam JA. *Guesstimation: Solving the World's Problems on the Back of a Cocktail Napkin*. Princeton University Press: Princeton, New Jersey, USA and Woodstock, Oxfordshire, UK, 2008.
55 Bartley DL, Shulman SS, Schlecht PC. Measurement uncertainty and NIOSH method accuracy range. In: *National Institute for Occupational Safety and Health, NIOSH Manual of Analytical Methods*. US Dept of Health and Human Services (NIOSH), 4th edn, Chapter P, pp 208-227, 2003. http://www.cdc.gov/niosh/docs/2003-154/pdfs/chapter-p.pdf. Accessed 25 October 2016.
56 Money CD, Margary SA. Improved use of workplace exposure data in the regulatory risk assessment of chemicals within Europe. *Ann Occup Hyg* 2002; **46**: 279–285.
57 Higgins JPT, Green S. Analysing data and undertaking meta-analyses. In: *Cochrane Handbook for Systematic Reviews of Interventions* 2009: http://handbook.cochrane.org.
58 Dawes RM. The robust beauty of improper linear models in decision making. *Am Psychol* 1979; **34**: 571–582.
59 Radin DI, Ferrari DC. Effects of consciousness on the fall of dice: a meta-analysis. *J Sci Explor* **5**: 61–83.
60 Rose VE, Cohrssen B (Ed). *Patty's Industrial Hygiene*, 6th edn. John Wiley and Sons, Inc: New York, 2010.
61 Kelsh MA, Shum M, Sheppard AR, McNeely M, Kuster N, Lau E *et al.* Measured radiofrequency exposure during various mobile-phone use scenarios. *J Expo Sci Environ Epidemiol* 2011; **21**: 343–354.
62 Bowman JD, Calvert GM, Gerard G, Witters DM. *Managing the Potential Hazards from Electromagnetic Interference (EMI) with Personal Medical Electronic Devices in Workplaces*, Abstract CS-111-01. American Industrial Hygiene Conference and Exposition: Baltimore, Maryland, USA, 2016. http://healthandenvironment.org/uploads/EMI_Risk_Management_-_AIHce_abstract_for_NIOSH_approval.pdf. Accessed 25 October 2016.

Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (http://www.nature.com/jes)