Journal of Exposure Science and Environmental Epidemiology (2018) 28, 251–258

www.nature.com/jes

ORIGINAL ARTICLE

Berkson error adjustment and other exposure surrogates in occupational case-control studies, with application to the Canadian INTEROCC study

Tamer Oraby[1], Siva Sivaganesan[2], Joseph D. Bowman[3], Laurel Kincl[4], Lesley Richardson[5], Mary McBride[6], Jack Siemiatycki[5], Elisabeth Cardis[7] and Daniel Krewski[8,9,10] On behalf of the INTEROCC Study Group

Many epidemiological studies assessing the relationship between exposure and disease are carried out without data on individual exposures. When this barrier is encountered in occupational studies, the subject exposures are often evaluated with a job-exposure matrix (JEM), which consists of mean exposure for occupational categories measured on a comparable group of workers. One of the objectives of the seven-country case-control study of occupational exposure and brain cancer risk, INTEROCC, was to investigate the relationship of occupational exposure to electromagnetic fields (EMF) in different frequency ranges and brain cancer risk. In this paper, we use the Canadian data from INTEROCC to estimate the odds of developing brain tumours due to occupational exposure to EMF. The first step was to find the best EMF exposure surrogate among the arithmetic mean, the geometric mean, and the mean of log-normal exposure distribution for each occupation in the JEM, in comparison to Berkson error adjustments via numerical approximation of the likelihood function. Contrary to previous studies of Berkson errors in JEMs, we found that the geometric mean was the best exposure surrogate. This analysis provided no evidence that cumulative lifetime exposure to extremely low frequency magnetic fields increases brain cancer risk, a finding consistent with other recent epidemiological studies.

Journal of Exposure Science and Environmental Epidemiology (2018) **28,** 251–258; doi:10.1038/jes.2017.2; published online 29 March 2017

Keywords: arithmetic mean; Berkson error; brain cancer; electromagnetic fields; exposure surrogate; geometric mean

INTRODUCTION

A principal aim of the international occupational case-control study (INTEROCC) is to estimate the odds ratio for being diagnosed with a brain tumour in relation to occupational exposure to electromagnetic fields (EMF) in different frequency ranges. As EMF measurements for individual subjects are not available in INTEROCC, time-weighted average (TWA) magnetic field strengths are assigned to jobs using a job exposure matrix (JEM). A JEM is a tool used in occupational epidemiology to assign exposure when the investigator has collected job histories from study subjects, but not histories of exposure to specific agents. The JEM is designed to indicate the exposures that correspond to different job titles, and the job axis typically consists of a pre-existing occupational classification system such as the International Standard Classification of Occupations (ISCO).[2] By its nature a JEM produces estimates of exposure that are based on groups rather than on individuals, the group being all those workers who share the same occupation code. A JEM can be created in different ways, ranging from an expert opinion-based approach to an approach based on pre-existing measurements of exposures in the various occupations of the JEM. When a bank of measurements exists from which the JEM will be created, there are issues to be resolved of how to best summarize the data into the JEM cell entries.

In the INTEROCC study, job histories were collected from cases with brain cancer and from controls, and coded to the ISCO system. In order to assign exposure to electro-magnetic fields, a JEM was used which was adapted by Bowman *et al.*[2] for the purpose. This JEM (http://www.crealradiation.com/index.php/en/databases?id=55) includes only the extremely low frequency (ELF) band. Since many subjects worked in different jobs, and so have been exposed to different sources of EMF/ELF during their work histories, the final exposure metric used for risk estimation is individual cumulative exposure to ELF.

Berkson measurement error occurs frequently in occupational epidemiology when using group-based exposure assessments based on appropriate JEMs, which is more practical than constructing individual retrospective exposure profiles. In such studies, each group's measure of central tendency is assigned to the subjects who have the same characteristics of the group, which is the type of occupation in the underlying study. Berkson measurement error then results from random fluctuations in the (unobserved) actual exposures around the observed measurements, leading to greater variability in modeling actual exposures.[3,4]

[1]School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, Edinburg, Texas, USA; [2]Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio, USA; [3]National Institute for Occupational Safety and Health, Cincinnati, Ohio, USA; [4]College of Public Health and Human Services, Oregon State University, Corvallis, Oregon, USA; [5]School of Public Health and Hospital Research Centre, University of Montreal, Montreal, Quebec, Canada; [6]BC Cancer Agency, Vancouver, British Columbia, Canada; [7]Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; [8]McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Ottawa, Ontario, Canada; [9]Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada and [10]Risk Sciences International, Ottawa, Canada. Correspondence: Dr T Oraby, University of Texas Rio Grande Valley, School of Mathematical and Statistical Sciences, 1201 W University Dr, Edinburg, Texas 78539, USA.
Tel.: +956 665 3536. Fax: +956 665 5091.
E-mail: tamer.oraby@utrgv.edu
Received 19 November 2015; accepted 22 January 2017; published online 29 March 2017

An important question is whether Berkson measurement errors can be reduced by the use of a suitable exposure surrogate representing each occupation code in a JEM; and if so, which one of the measures of central tendency can be the best candidate. Research to date, however, has not clarified whether the arithmetic mean (AM), geometric mean (GM), or the mean true exposure given the available data (in the present context, the log-normal mean, which we call the modified geometric mean or MGM) is the best choice for the exposure surrogate in logistic regression analysis. It has been shown in both theoretical[5] and simulation[6] studies that the accuracy of risk estimates obtained by using the AM or GM depends on the shape of the underlying exposure distribution. Steenland's simulation suggested that assigning the AM to the exposure groups might lead to better estimates of the slope in logistic regression, although this result was obtained under highly simplified exposure conditions.

The measurement error adjustment approach in the ensuing analysis takes into account one source of uncertainty—Berkson error—in risk estimation. In this paper, we estimate brain tumour risk using a maximum likelihood approach with adjustment for Berkson error in ELF exposures, as well as using different exposure surrogates based on the JEM.[7] (For the Berkson error adjustment approach, we will use numerical integration[8] to integrate possible exposures of individual study subjects and estimate the resulting risks.)

In order to determine the best exposure surrogate for use with this JEM for INTEROCC's risk analysis for all seven counties, computer simulation is used to compare risk estimates when each of the following surrogates AM, GM, and MGM, as well as the Berkson error adjustment are used. We performed two separate simulation studies, one for continuous and another for categorical exposures.[9]

MATERIALS AND METHODS

Canadian INTEROCC Study

The Canadian INTEROCC is comprised of 813 subjects, including 165 cases of brain cancer and 648 controls. Each subject is classified according to gender, education (primary-secondary, intermediate college, tertiary, or don't know), age (age < 40, $40 \le$ age < 50, $50 \le$ age < 60, and age ≥ 60), and urban centre (Montreal, Ottawa, and Vancouver). The Canadian INTEROCC data will be used in two ways in this study. The first is in the simulation study, where the subjects' job histories will be used to generate different responses at different selected odd ratios; those odds ratios are then estimated using the different exposure surrogates and the Berkson error adjustment (see below). The second involves the use of the actual Canadian INTEROCC data to estimate brain cancer risk following lifetime occupational exposure to ELF.

Job Exposure Matrix for ELF Magnetic Field Exposures

The JEM employed here consists of full-shift measurements of the TWA magnitudes of the ELF magnetic field B, in micro-tesla (μT), from multiple studies with these data grouped by their ISCO occupational codes.[2] The JEM contains entries aggregated from different exposure studies, giving the geometric mean GM, arithmetic mean AM, standard deviation SD and geometric standard deviation GSD for each of 336 ISCO job codes.

The cumulative magnetic field (MF) exposures for subject i = 1,…, S is given as

$$\mathrm{CumMF_i} = \sum_{j=1}^{J_i} \mathrm{t_{ij}\, X_{ij}}, \tag{1}$$

where J_i is the number of jobs held by subject i, $\mathrm{t_{ij}}$ is the time spent by subject i in job j in years with exposure $\mathrm{X_{ij}}$, and S is the number of subjects. Using a surrogate $\mathrm{W_j}$ for $\mathrm{X_{ij}}$ to find $\mathrm{CumMF_i}$ would result in Berkson measurement error that would lead to bias (in either direction) in the estimate of the slope of the exposure-response relationship. The surrogate $\mathrm{W_j}$ can be taken as the arithmetic mean (AM), geometric mean (GM)), which is the median of the log-normal random variable, or the mean of a log-normal random variable defined by MGM: = GM*exp(.5*log^2(GSD)). The latter surrogate very well for linear and non-linear models.[10]

In order to adjust for Berkson measurement error, individual exposures associated with each job are postulated to follow a log-normal distribution whose geometric mean and geometric standard deviation are taken from the JEM (see Supplementary Appendix 1 for further details). A log-normal random variable can be written in the form $\mathrm{X_{ij}} = \mathrm{W_j}\, \epsilon_{\mathrm{ij}}$, representing the multiplicative Berkson error model, where $\mathrm{W_j}$ is the geometric mean $\mathrm{GM_j}$ and ϵ_{ij} is distributed by the log-normal distribution $\mathrm{logNormal}\left(0,\ \sigma_j^2\right)$ and $\sigma_j^2 = \log^2 \mathrm{GSD_j}$.

Berkson Error Adjustment

As exposure can be either continuous or categorical, we will adjust for Berkson error in both cases.

(a) Continuous exposure. The likelihood function in the Berkson error model is given by integrating across the exposure error distributions given in (2). In order to adjust for Berkson error for continuous exposures, we use numerical integration of the likelihood function[9] using Gaussian quadrature.[11] Specifically, Gauss-Hermite integration of degree N[12–14] is used to approximate the likelihood function

$$L\left(y|\beta_0,\beta_1,\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2\right) = \int_x f\left(y|\beta_0,\beta_1,x\right) f\left(x|\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2\right) dx, \tag{2}$$

where $\boldsymbol{W}$ is the surrogate for x, for example GM. (Here and elsewhere, boldface symbols are used to represent vectors and matrices.)

Since the likelihood function involves an intractable integral, we approximate the likelihood function based on the following proposition, whose proof is provided in Supplementary Appendix 2.

Proposition 1. For fixed $N \ge 1$, $L\left(y|\beta_0,\beta_1,\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2\right) = \int_x f\left(y|\beta_0,\beta_1,x\right) f\left(x|\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2\right) dx$ can be approximated by

$$L_N\left(y|\beta_0,\beta_1,\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2\right) = \prod_{i=1}^{S} \sum_{\boldsymbol{k}_i \in [N]_i^{J_i}} \Pr(\boldsymbol{k}_i) \times \frac{\exp[\mathrm{y_i}\left(\beta_0 + \beta_1 \sum_{j=1}^{J_i} \mathrm{t_{ij} W_j} \exp\left(\mathrm{a_{k_j}} \times \sigma_j\right)\right)]}{1 + \exp\left(\beta_0 + \beta_1 \sum_{j=1}^{J_i} \mathrm{t_{ij} W_j} \exp\left(\mathrm{a_{k_j}} \times \sigma_j\right)\right)} \tag{3}$$

for $\beta_0 \ge 0$, $\beta_1 > 0$, where $\boldsymbol{k}_i := (k_1, k_2, \ldots, k_{J_i})$, $[N] := \{1, 2, \ldots, N\}$, J_i is the number of jobs held by subject i, N is the integral approximation degree and $[N]_i^J$ is the set of all ordered J_i-tuples of elements chosen without replacement from $[N]$. The probability $\Pr(k_1, k_2, \ldots, k_{J_i})$ is given as the product of the integration weights divided by $\sqrt{\pi}$ making their sum equal to one. The coefficients a_k are the abscissas (zeros of Hermite polynomials of degree N) times $\sqrt{2}$.

Remark 1. The case $N = 1$ is equivalent to using the GM as a surrogate where the only abscissa is equal to zero (making $\exp\left(\mathrm{a_{k_j}} \times \sigma_j\right) = \exp(0) = 1$) and the only integration weight is $\sqrt{\pi}$ (making $\Pr(\boldsymbol{k}_i) = 1$). Thus,

$$L_1\left(y|\beta_0,\beta_1,\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2\right) = \prod_{i=1}^{S} \frac{\exp[\mathrm{y_i}\left(\beta_0 + \beta_1 \sum_{j=1}^{J_i} \mathrm{t_{ij} W_j}\right)]}{1 + \exp\left(\beta_0 + \beta_1 \sum_{j=1}^{J_i} \mathrm{t_{ij} W_j}\right)}.$$

Remark 2. If $\beta_0 < 0$, $\beta_1 < 0$ the proposition still applies by switching cases and control to flip the inequality signs for the log-odds ratio. Numerically, it was found, however, that the approximation works well even when the slope and intercept are negative and without the switch.

The above proposition holds after including other covariates $\underline{\boldsymbol{Z}}$ which are not subject to measurement error, age and gender, that is $L_N\left(y|\beta_0,\beta_1,\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2\right) = \prod_{i=1}^{S} \sum_{\boldsymbol{k}_i \in [N]_i^{J_i}} \Pr(\boldsymbol{k}_i)$

$$\times \frac{\exp[\mathrm{y_i}\left(\beta_0 + \beta_1 \sum_{j=1}^{J_i} \mathrm{t_{ij} W_j} \exp\left(\mathrm{a_{k_j}} \times \sigma_j\right) + \boldsymbol{\gamma}.\underline{\boldsymbol{Z}}_i\right)]}{1 + \exp\left(\beta_0 + \beta_1 \sum_{j=1}^{J_i} \mathrm{t_{ij} W_j} \exp\left(\mathrm{a_{k_j}} \times \sigma_j\right) + \boldsymbol{\gamma}.\underline{\boldsymbol{Z}}_i\right)}.$$

(b) Categorical exposure. Categorical exposures are defined here according to specified cut-off points based on quintiles $(Qu_l; l = 1, 2, 3, 4)$ of the sampling distribution of cumulative exposurethe $\mathrm{CumMF_i} = \sum_{j=1}^{J_i} \mathrm{t_{ij} X_{ij}}$ calculated using the geometric means as surrogates (we assume that Qu_0 is 0 and Qu_5 is ∞). As a result, a 4-dimensional exposure vector $\mathbf{CatCumMF_i} = (e_{1i}, e_{2i}, e_{3i}, e_{4i})$ will be assigned to each subject i such that $e_{li} = I(\mathrm{CumMF_i} \in (Qu_l, Qu_{l+1}])$, where $I(A)$ is the indicator function that is equal to one if A is true and zero otherwise. We note that $\mathbf{CatCumMF_i} =$ (0,0,0,0) when $\mathrm{CumMF_i} \in (Qu_0, Qu_1]$. When the $\mathrm{X_{ij}}'s$ are random variables, the cumulative exposure $\mathrm{CumMF_i}$ and the 4-dimensional exposure vector, denoted by $\underline{E}_i = (E_{1i}, E_{2i}, E_{3i}, E_{4i})$, will also be random variables.

In this case, the integral in Eq. (1) has a closed form solution, and does not require the application of numerical methods. Yet, it will be essential

to find the probability $P_{li} := \Pr(E_{li} = 1) = \Pr(\text{CumMF}_i \in (Qu_l, Qu_{l+1}]) = \int_{Qu_l}^{Qu_{l+1}} g_{Cum_i}(v)dv$ for $l = 0,1,2,3,4$, where g_{Cum_i} is the probability density function of CumMF_i. Determining g_{Cum_i} is complicated; otherwise handling the continuous exposure will be a straightforward application. However P_{li} can be estimated using Monte-Carlo simulation. We give the final form of the integral in the following proposition, whose proof is included in Supplementary Appendix 4. (The proposition and its proof are not new, but are given here for completeness.) In the following formulation of the proposition, β_0 is the intercept, and β_i ($i = 1,2,3,4$) are the slopes of the four categories.

Proposition 2. The function $L(y|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2) = \int_x f(y|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,x) f(x|\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2)dx$ can be found by

$$L(y_1,\ldots,y_S|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\underline{\boldsymbol{W}},\underline{\boldsymbol{\sigma}}^2) = \prod_{i=1}^{S}\left(y_i + (-1)^{y_i}\sum_{l=0}^{4} P_{li}\frac{1}{1+exp(\beta_0+\sum_{k=1}^{4}\beta_k I(k=l))}\right), \quad (4)$$

Including other error free covariates $\underline{\boldsymbol{Z}}$, will not change the form of the likelihood function **L**.

Simulation Study

We conducted two simulation studies to compare the risk estimates obtained using each of the surrogates AM, GM, and MGM, and that obtained by performing Berkson error adjustment using a logistic regression model to describe the exposure-response relationship. When using Berkson error adjustment, we invoked Proposition 1 in the case of continuous exposure and Proposition 2 with categorical exposure. We calculated the bias and standard error of these risk estimates based on these simulations. We used the job histories of the 813 subjects in the Canadian INTEROCC study to simulate exposures using the JEM and the log-normal distribution, and then simulated the responses for the 813 subjects multiple times.

(a) Simulation with continuous exposure. An overview of the design of the simulation study based on the 813 subjects with continuous exposure is given in Figure 1. In this analysis, $N = 4$ was used as the approximation degree for the Berkson error adjustment. In each simulation run, we have performed the following calculations.

1. For each subject i, exposure during his/her work in job j is generated randomly according to $X_{ij} \sim \text{logNormal}(\log\ GM_j, \log^2 GSD_j)$, where GM_j and GSD_j are provided by the JEM. The cumulative exposure, CumMF_i, is then calculated for each subject.
2. A pre-determined intercept $\beta_0 = 1$ and slopes β_1 ranging from 0 (no relation) to 0.4, inclusively, with step length .01, were used to find the probability $p_i := \frac{1}{1+\exp(-\beta_0-\beta_1\ \text{CumMF}_i)}$ that subject i develops a brain tumour. The range of slopes corresponds to odds ratios ranging from 1 to 1.5.
3. The case status (1 = tumour; 0 = no tumour) for subject i is then generated randomly for each subject using a Bernoulli distribution with probability p_i.
4. The slope of the exposure-response curve (the logarithm of the odds ratio) is then estimated using each of the statistics AM, GM, and MGM: = GM*exp(.5*log^2(GSD)) as a proxy for the actual exposure. Berkson error adjustment based on Proposition 1 is then applied to estimate brain tumour risk.
5. Steps 3 and 4 are repeated 100 times and the median estimate is calculated for the 100 estimates of the pre-determined slope. (We chose the median as the central estimate as the distribution seems right skewed for each of the pre-determined slopes.)
6. Steps 1 to 5 are repeated 100 times, and the mean and the 2.5% and 97.5% percentiles are calculated from this distribution.

To compare between the five different approaches to risk estimation based on the different exposure surrogates considered, we calculated the bias defined by the average (over all simulation runs) risk estimate minus the pre-determined target parameter. We also calculated the root mean squared error, given by the square root of the sum of the variance estimates and the square of the bias. The variance is given by the total variance defined as the sum of two terms: the mean (average) of the conditional variances given the simulated inputs, and the variance of the conditional means given the simulated inputs.

The deviance defined by

$$D = 2 * \left[NLL\left(y|\hat{\beta}_0, 0, \underline{\boldsymbol{W}}, \underline{\boldsymbol{\sigma}}^2\right) - NLL\left(y|\hat{\beta}_0, \hat{\beta}_1, \underline{\boldsymbol{W}}, \underline{\boldsymbol{\sigma}}^2\right)\right] \quad (5)$$

and the approximated deviance

$$D_N = 2 * \left[NLL_N\left(y|\hat{\beta}_0, 0, \underline{\boldsymbol{W}}, \underline{\boldsymbol{\sigma}}^2\right) - NLL_N\left(y|\hat{\beta}_0, \hat{\beta}_1, \underline{\boldsymbol{W}}, \underline{\boldsymbol{\sigma}}^2\right)\right] \quad (6)$$

are used to test the goodness of fit of the logistic models to the data, where NLL is the negative logarithm of the likelihood function. Note that $\hat{\beta}_0$ and $\hat{\beta}_1$ depend on the adjustment method (Berkson approximation or exposure surrogate) and they will not be the same all the time. Notice also that $NLL_N\left(y|\hat{\beta}_0, 0, \underline{\boldsymbol{W}}, \underline{\boldsymbol{\sigma}}^2\right) = NLL\left(y|\hat{\beta}_0, 0, \underline{\boldsymbol{W}}, \underline{\boldsymbol{\sigma}}^2\right)$.

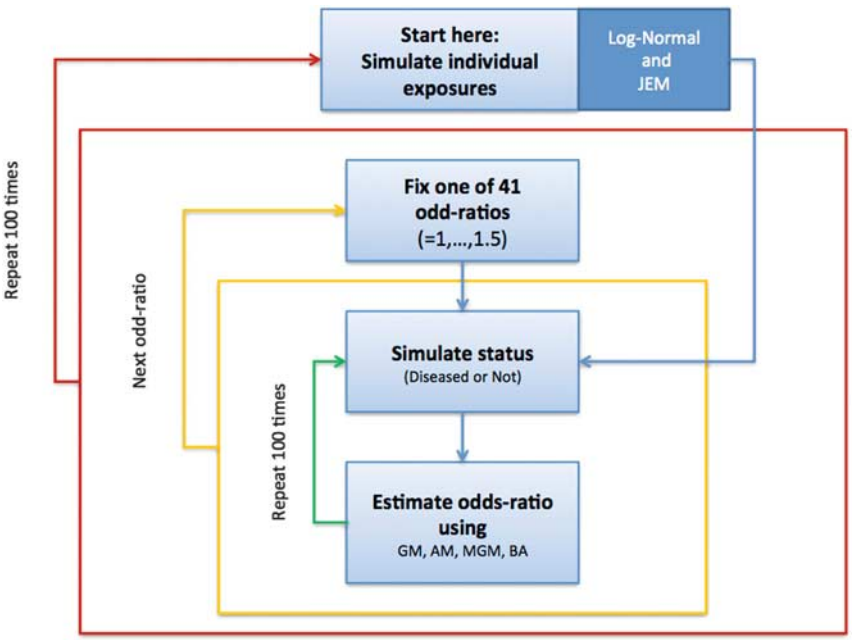

Figure 1. Schematic of simulation study of odds ratio estimation using each of the following continuous exposure metrics: AM, arithmetic mean; BA, Berkson error adjustment using numerical integration with approximation degree $N = 4$; GM, geometric mean; MGM, log-normal mean exposure.

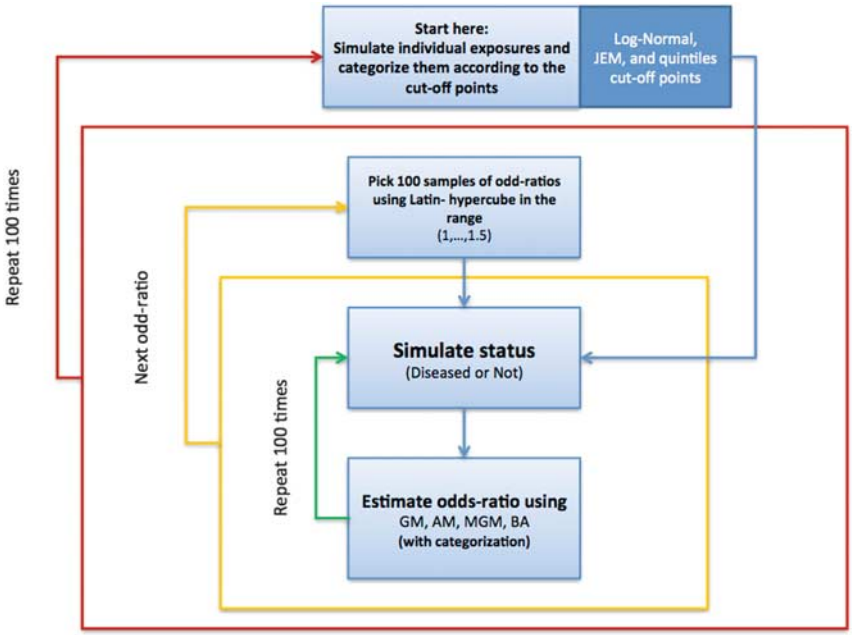

Figure 2. Schematic of the simulation study of odds ratio estimation using each of the following categorical exposure metric: AM, arithmetic mean; BA, Berkson error adjustment; GM, geometric mean; MGM, log-normal mean exposure.

(b) Simulation with categorical exposure. An overview of the simulation conducted with categorical exposure is given in Figure 2. In each run, we conducted the following calculations.

1. The probabilities $\{P_{li}; l = 0,1,2,3,4\}$ are estimated using Monte-Carlo simulation with 5 000 000 runs for each subject i.
2. For each subject i, exposure during his/her work in job j is generated randomly according to $X_{ij} \text{logNormal}(\log \text{GM}_j, \log^2 \text{GSD}_j)$, where GM_j and GSD_j are from the JEM. The cumulative exposure, $CumMF_i$, is then calculated for each subject and allocated to one of five categories identified by cut-points points that are equal to the quintiles of the sample cumulative exposure distribution of the actual control group using the geometric mean as surrogates. This step results in a four-dimensional vector, $\mathbf{CatCumMF}_i$, for each subject.
3. Using a pre-determined intercept $\beta_0 = 1$ and slopes $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ ranging from 0 to .4, inclusively, 100 combinations of the slopes were selected using Latin-hypercube sampling (see Figures 3 and 4), and used to determine the probability $p_i' := \frac{1}{1+\exp(-\beta_0 - \boldsymbol{\beta} \cdot \mathbf{CatCumMF}_i)}$ that subject I develops a brain tumour. (The range of slopes corresponds to a range of odds ratios from 1 to 1.5.
4. Case status (1 = tumour; 0 = no tumour) is generated randomly for each subject using a Bernoulli distribution with probability p_i.
5. Exposures based on each of the statistics AM, GM, and MGM were categorized using the cut-points described above, and used to estimate the slopes. Berkson error adjustment, based on Proposition 2, is then applied to estimate the risk.
6. Steps 4 and 5 are repeated 100 times and the median estimate is calculated for the 100 estimates of the pre-determined slope.
7. Steps 2 to 6 are repeated 100 times, following which the mean and the 2.5% and 97.5% percentiles are calculated.

To compare the five different approaches with categorical exposures, we calculated the bias for each of $\beta_1, \beta_2, \beta_3$, and β_4 separately. We also calculated the root mean squared error defined as above.

The Canadian Component of the INTEROCC Study

In our analysis of the Canadian INTEROCC data, we used non-stratified and stratified analyses to estimate the risk of developing brain cancer due to lifetime cumulative occupational exposure to ELF. In the non-stratified analysis, we used logistic regression[15] with and without additional covariates (age, region, gender, and education), whereas in the stratified analysis we used the conditional logistic regression.[15,16] Stratification was done using the modifiers. In all analyses, we used both categorical and continuous exposures derived from one of the exposure surrogates (AM, GM, and MGM) and performed Berkson error adjustment in the non-stratified case.

The stratified and non-stratified analyses using the AM, GM and MGM were done using SAS;[17] MatLab[18] was used for Berkson error adjustment.

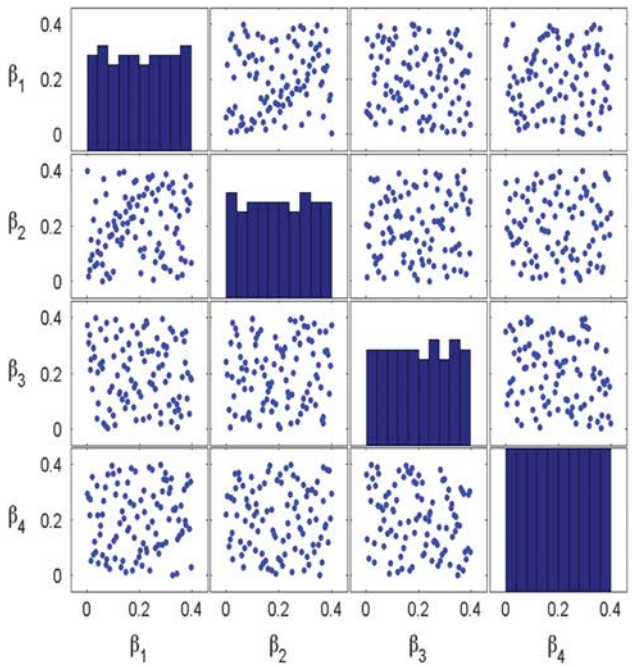

Figure 3. Scatter plots of the 100 vectors in the Latin Hypercube sample of the 4 log odds ratios (β_1, β_2, β_3, β_4) in the interval [0, 0.4].

RESULTS

Simulation Study

Simulation of the alternative exposure surrogates took almost 50 days in the case of continuous exposure and more than 3 days

with categorical exposures. The simulation was split into 5 parts in order to avoid loss of data following a computer crash (Figure 4). The simulation resulted indicated that the GM resulted in the least bias of the slope estimates and the AM resulted in the most attenuation (Figure 5a). We used also the simulated exposure data to retrieve the pre-determined assumed slopes (i.e. successfully find the estimates are equal to the pre-determined values), which was unsuccessful on average (Figure 5a and c). In almost all simulations, the intercept β_0 was successfully retrieved.

Comparisons of the root mean squared errors for the log odds ratios estimated by the five approaches considered here (Figure 6a) show that the GM is results in the lower absolute bias than the AM or MGM, and a lower standard error than the Berkson error adjustment approach (Figure 6b). See also Figure 7 for the corresponding improvements displayed by deviance.

The use of categorical exposures produced similar results. The GM showed less bias than the Berkson error adjustment approach, demonstrating near zero bias with wide credible intervals (see Supplementary Figure SI1 in the Supplementary Information). The AM and MGM showed almost the same bias as the true exposures. Berkson error adjustment was also seen to result in less precise risk estimates of than the GM (see Supplementary Figure SI2 in the Supplementary Information).

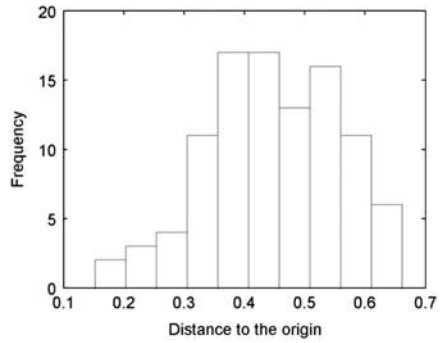

Figure 4. Histogram for the Euclidean distances of the 100 vectors $\beta = (\beta_1,\beta_2,\beta_3,\beta_4)$ to the origin in the Latin-hypercube sample of the 4 log odds ratios $(\beta_1,\beta_2,\beta_3,\beta_4)$ in the interval [0, 0.4].

Analysis of the Canadian INTEROCC Data

Analysis of the actual Canadian INTEROCC data was conducted using each of the exposure estimates described above. The results showed non-significant odds ratios in all analyses using all exposure metrics considered (see Tables 1, 2, and 3), except in the highest exposure category with categorical exposure after adjusting for Berkson error in the non-stratified approach. Differences between estimates of odds ratios in the stratified and non-stratified analyses are negligible in the case of continuous exposures (Table 1) and very small with categorical exposure (the maximum absolute difference between estimates was 0.18.) (Tables 2 and 3), as is generally expected (ref. 15, p. 205).

DISCUSSION

Measurement error can bias the association between exposure and outcome towards the null hypothesis of no effect and overestimate the precision of the corresponding risk estimate,[19] as well as reduce the power of the analyses.[20] Thus, exposures measured with error and/or uncertainty can give rise to

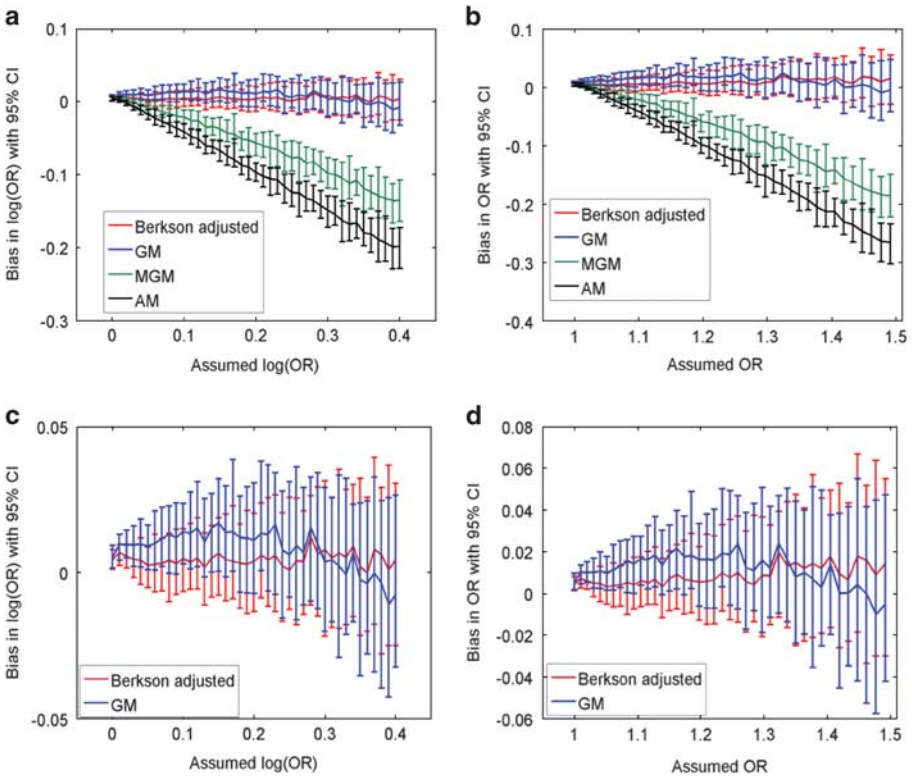

Figure 5. Bias in the estimates of (**a**) the log odds ratios, (**b**) the odds ratio using the five exposure metric: GM = geometric mean, AM = arithmetic mean, MGM = log-normal mean, and Berkson error adjustment using numerical integration. The same outputs are shown only for the two best approaches: the GM = geometric mean and Berkson error adjustment using numerical integration for (**c**) the log odds ratios (**d**) the odds ratio.

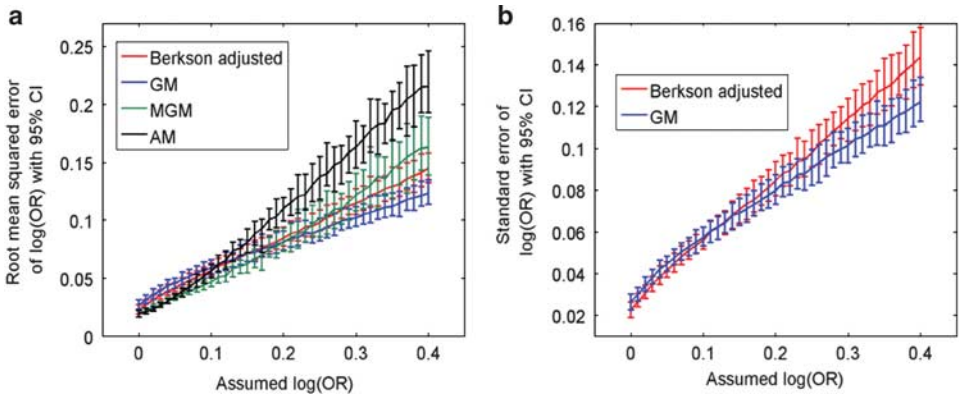

Figure 6. (**a**) Root mean squared error for the estimates of the log odds ratios using the five exposure metrics: AM, arithmetic mean; GM, geometric mean; MGM, log-normal mean; and Berkson error adjustment using numerical integration. (**b**) Standard error for the estimates using the GM and Berkson error adjustment.

Table 1. Odds ratios for brain cancer risk in relation to cumulative lifetime occupational exposure to ELF in the Canadian INTEROCC study based on four continuous exposure metrics.

		Berkson adjusted		*AM*		*GM*		*MGM*	
Analysis	*Covariates*	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
Non-stratified	Excluded	0.999	0.980, 1.019	0.999	0.981, 1.018	1	0.978, 1.022	1	0.982, 1.018
	Included	0.996	0.996, 0.9661	0.992	0.961, 1.023	0.998	0.972, 1.023	0.996	0.972, 1.020
Stratified	NA[a]	NA	NA	0.997	0.976, 1.018	1	0.979, 1.021	0.999	0.982, 1.017

[a]NA = not applicable.

Table 2. Odds ratios for brain cancer risk in relation to cumulative lifetime occupational exposure to ELF in the Canadian INTEROCC study, based on four categorical exposure metrics (non-stratified analysis including covariates).

	Berkson adjusted		*AM*		*GM*		*MGM*	
Exposure Category (uT)	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
< 1.48[a]	1.00	NA[b]	1.00	NA	1.00	NA	1.00	NA
1.48–2.45	1.508	0.863, 2.634	0.876	0.448, 1.714	1.057	0.600, 1.862	0.475	0.230, 0.980
2.45–3.30	2.398	0.299, 19.201	1.146	0.588, 2.236	0.710	0.374, 1.349	0.954	0.487, 1.869
3.30–4.42	0.668	0.122, 3.661	0.616	0.311, 1.223	0.694	0.358, 1.348	0.576	0.290, 1.144
> 4.42	2.567	2.394, 2.752	0.793	0.423, 1.487	0.995	0.525, 1.888	0.737	0.392, 1.387

[a]Reference category. [b]NA = not applicable.

Table 3. Odds ratios for brain cancer risk in relation to cumulative lifetime occupational exposure to ELF in the Canadian INTEROCC study, based on four categorical exposure metrics (stratified analysis excluding covariates).

	AM		*GM*		*MGM*	
Category ... vs. category 1 (< 1.48 uT)	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
1.48–2.45 uT	0.829	0.409, 1.681	0.959	0.534, 1.722	0.447	0.209, 0.955
2.45–3.30 uT	0.963	0.482, 1.923	0.648	0.335, 1.256	0.827	0.413, 1.657
3.30–4.42 uT	0.554	0.273, 1.124	0.708	0.363, 1.382	0.539	0.265, 1.095
> 4.42 uT	0.788	0.414, 1.502	1.142	0.589, 2.213	0.776	0.408, 1.478

misleading epidemiological results if risk estimates are not adjusted for those errors. Such adjustments, however, generally increase the standard errors of the estimates, and hence the uncertainty about the actual level of risk.[7]

In this paper, we developed a structural approach to adjust for Berkson error in measures of occupational exposure to EMF. A multiplicative Berkson error model is employed to this study based on the exposure studies from which the JEM was mostly

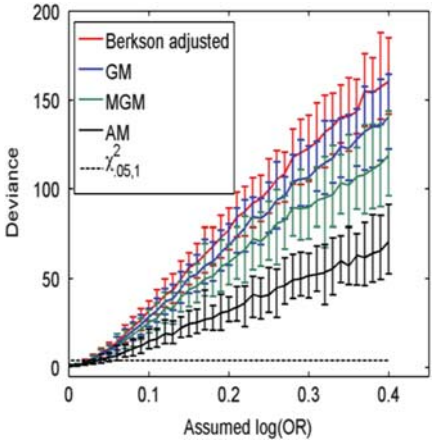

Figure 7. Deviance estimated using the four exposure metrics: AM, arithmetic mean; GM, geometric mean; MGM, log-normal mean, as well as the approximated deviance based on Eq. (6) for Berkson error adjustment using numerical integration. Deviances are shown relative to the critical point $\chi^2_{.05,1} = 3.8415$, indicated by the dashed horizontal line.

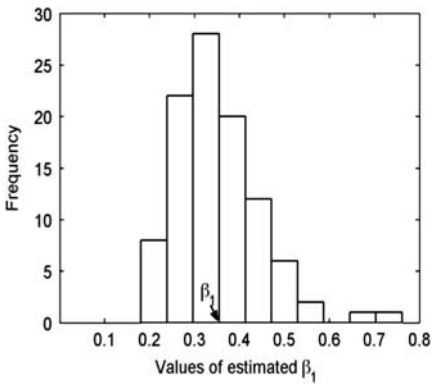

Figure 8. Histogram of estimates of β_1 corresponding to the true value of $\beta_1 = 0.35$ at one set of simulated continuous exposures with the disease status of each subject simulated 100 times.

formulated. In those studies, logarithm of field exposures are modeled by two-way random models (see Supplementary Appendix I). Additive models are not realistic here since they might lead to negative exposures of some ISCO when those exposures have small mean and large variability.

Using both computer simulation and analyses of the actual Canadian INTEROCC data, we compared the odds ratios estimated using three exposure surrogates: the arithmetic mean (AM), geometric mean (GM) and the mean of log-normal (which we call modified geometric mean, MGM) with estimates based on the likelihood integration approach to adjust for the Berkson measurement error. The estimates of the log-odds ratio using the GM and the Berkson error adjustment showed very small bias in our computer simulations, in contrast to those estimates based on the AM and MGM. One possible reason for this is that the GM is a first-degree approximation of the multiple integrals comprising the structural form of the likelihood function. As the degree of approximation increases, the bias decreases on average. We found also that the larger the log-odds ratio, the wider the corresponding credible interval.

Standard errors of the estimates were smallest using the GM, followed by the Berkson error adjustment. However, when the log-odds ratio is large, Berkson error adjustment produces smaller standard errors than the GM (Figure 7). This finding is in line with the fact that measurement error adjustment generally increases the uncertainty in the estimates of risk.[7] The log-likelihood ratio goodness of fit test demonstrated that the full likelihood is well approximated by numerical integration.

Contrary to other simulation studies like,[6] we simulated full lifetime exposure profiles for all individuals in the Canadian INTEROCC study; repeated simulations were employed to achieve stable estimates (see Figures 1 and 2). As indicated in Figure 8, this expanded simulation strategy demonstrated a high probability that the estimated slopes (i.e., the estimated log-odds ratios) will be far from the true slope.

To illustrate the application of our methods, we conducted both stratified and non-stratified analyses of the actual data from the Canadian component of the INTEROCC study. Conditional logistic regression including relevant covariates was conducted using both continuous and categorical exposures. This analysis provided no evidence that cumulative lifetime occupational exposure to ELF is associated with brain tumour risk. This was seen using the AM, GM, and MGM as exposure surrogates, as well as when exposure was adjusted for Berkson error. These findings are in accordance with those recently reported by[2] for cumulative lifetime exposure in the full seven country INTEROCC study, as well as those in a prospective cohort study in the Netherlands conducted by.[21] Finally, we note that the use of the geometric mean as an exposure surrogate by[2]—a judgmental choice made in advance of the results of the present study—is consistent with our simulation results comparing different exposure surrogates.

CONFLICT OF INTEREST

Dr Cardis is the principal investigator of the INTEROCC study, which is funded by the NIH. The authors declare no other conflicts of interest.

ACKNOWLEDGEMENTS

This research was conducted in part while the first author was visiting the Centre for Research in Environmental Epidemiology (CREAL) at the University of Barcelona in the fall of 2010. The authors are grateful to Jordi Figuerola at CREAL for his assistance in preparing the data files used in this paper. D. Krewski is the Natural Sciences and Engineering Research Chair in Risk Science at the University of Ottawa. J Siemiatycki holds the Guzzo-Cancer Research Society Chair in Environment and Cancer, at the Université de Montréal. The INTEROCC study was funded by the National Institutes for Health (NIH) Grant No. 1R01CA124759 (PI E. Cardis).

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.

REFERENCES

1 Turner MC, Benke G, Bowman JD, Figuerola J, Fleming S, Hours M *et al.* Occupational exposure to extremely low-frequency magnetic fields and brain tumor risks in the INTEROCC study. *Cancer Epidemiol Biomarkers Prev* 2014; **23**: 1863–1872.
2 Bowman J, Touchstone J, Yost M. A population-based job exposure matrix for power-frequency magnetic fields. *J Occup Environ Hyg* 2007; **4**: 715–728.
3 Berkson J. Are there two regressions. *Journal of the American Statistical Association* 1950; **45**: 164–180.
4 Heid I, Kuchenhoff H, Miles J, Kreienbrock L, Wichmann H. Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. *Journal of Exposure Analysis and Environmental Epidemiology* 2004; **14**: 365–377.
5 Kim HM, Burstyn I. Bayesian method for improving logistic regression estimates under group-based exposure assessment with additive measurement errors. *Archives of Environmental & Occupational Health* 2009; **64**: 261–265.

6 Steenland K, Deddens JA, Zhao S. Biases in estimating the effect of cumulative exposure in log-linear models when estimated exposure levels are assigned. *Scandinavian Journal of Work, Environment & Health* 2000; **26**: 37–43.
7 Carroll R, Ruppert D, Tefanski L, Crainiceanu C *Measurement error in nonlinear models. A modern perspective*. Chapman & Hall/CRC. Monographs on Statistics and Applied Probability: Boca Raton, FL, 2006.
8 Stroud AH. *Approximate Calculation of Multiple Integrals*. Prentice-Hall: Englewood Cliffs: N.J., 1971.
9 Spiegelman D, Rosner B, Logan R. estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association* 2000; **95**: 51–61.
10 Thomas D, Stram D, Dwyer J. Exposure measurement error: Influence on exposure-disease relationships and methods of correction. *Annual Review of Public Health* 1993; **14**: 69–93.
11 Golub GH, Welsch JH. Calculation of Gauss quadrature rules. *Mathematics of Computation* 1969; **23**: 221–230.
12 Higdon R, Schafer DW. Maximum likelihood computations for regression with measurement error. *Computational Statistics & Data Analysis* 2001; **35**: 283–299.
13 Crouch EAC, Spiegelman D. The evaluation of integrals of the form S f(t) exp ($-t^2$) dt: Application to logistic-normal models. *Journal of the American Statistical Association* 1990; **85**: 464–469.
14 Gonzalez J, Tuerlinckx F, De Boeck P, Cools R. Numerical integration in logistic-normal models. *Computational Statistics & Data Analysis* 2006; **51**: 1535–1548.
15 Breslow NE, Day NE. *The Analysis of Case-Control Studies*. IARC Scientific Publications, International Agency for Research on Cancer: Lyon, France, 1980, pp 32.
16 Cox DR, Hinkley DV. *Theoretical Statistics*. Chapman & Hall: London, UK, 1974.
17 SAS Institute Inc. The SAS system for Windows. Release 9.2, Cary, NC, USA 2011.
18 MathWorks Inc. *MatLab (r2011b)*. Natick: MA, USA, 2011.
19 Fuller WA. *Measurement Error Models*. Wiley: New York, USA, 1987.
20 Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998; **55**: 651–656.
21 Koeman T, van den Brandt PA, Slottje P, Schouten LJ, Goldbohm RA, Kromhout H *et al.* Occupational extremely low-frequency magnetic field exposure and selected cancer outcomes in a prospective Dutch cohort. *Cancer Causes Control* 2014; **25**: 203–214.

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