Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Uncovering selection bias in case-control studies using Bayesian post-stratification

S. Geneletti[a], N. Best[b]*, M.B. Toledano[b], P. Elliott[b], S. Richardson[b]

Case-control studies are particularly prone to selection bias, which can affect odds ratio estimation. Approaches to discovering and adjusting for selection bias have been proposed in the literature using graphical and heuristic tools as well as more complex statistical methods. The approach we propose is based on a survey weighting method termed Bayesian post-stratification and follows from the conditional independences that characterise selection bias. We use our approach to perform a selection bias sensitivity analysis by using ancillary data sources that describe the target case-control population to re-weight the odds ratio estimates obtained from the study. The method is applied to two case-control studies, the first investigating the association between exposure to electro-magnetic fields and acute lymphoblastic leukaemia in children, and the second investigating the association between maternal occupational exposure to hairspray and a congenital anomaly in male babies called hypospadias. In both case-control studies, our method showed that the odds ratios were only moderately sensitive to selection bias.

Keywords: collider bias, combining data, conditional independence, epidemiology, marginal odds ratio, sensitivity analysis

1. Introduction

Sensitivity analysis is increasingly recognised as an essential part of a statistical analysis[1, 2], especially in the context of observational studies in fields such as epidemiology. A number of methods have been put forward to address problems of selection/participation bias in epidemiology, including multiple imputation[3], inverse probability weighting[4] and bias modelling[1, 5]. In this paper, we propose a novel method, based on post-stratification, to assess the sensitivity of case-control study odds ratios to selection bias. We argue that our method is conceptually simpler than other commonly advocated methods and provides an intuitive way of performing sensitivity analysis.

Selection bias can present a serious problem for valid odds ratio estimation in case-control studies as demonstrated in Mezei and Kheifhets[6] and Geneletti et al[7]. These types of study are especially sensitive to selection bias as the sampling mechanism depends on the case/control status of the participants and participation probabilities are generally hard to estimate. Selection bias arises when the exposure under investigation is associated with the selection mechanism[8]. As we typically have limited information on the distribution of the exposure other than from the study itself, we are unable to estimate the dependence between the exposure and the selection and can therefore not adjust for selection bias. However, sensitivity analysis to try and assess the *extent of the bias* is valuable. To this effect, we propose to introduce a set of variables B such that first, B separates the exposure from the selection, and second, the distribution of B can be estimated from sources of data external to the main study. In so doing, we shift the selection bias from the exposure, whose distribution cannot be estimated without bias from the study, to B, whose distribution is estimated from data external to the main study and thus is potentially unbiased. By using different sources of data to estimate the distribution of B we can investigate the sensitivity of the odds ratio to different selection processes. This approach is conceptually simple and

[a] *London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK*
[b] *Imperial College London, St Mary's Campus, Norfolk Place, London, W2 1PG, UK*
* *Correspondence to: Nicky Best, Department of Epidemiology and Biostatistics, Imperial College London, St Mary's Campus, Norfolk Place, London, W2 1PG, UK*

gives us tools to assess selection bias in a wide number of situations, as well as encouraging us to think carefully about the populations of interest and the sources of bias.

In Section 2 we discuss selection bias, then we describe our two case studies in Section 3. Section 4 presents our method for assessing the sensitivity of case-control odds ratio estimates to selection bias, while Section 5 describes the application of the method to the two case studies. We end with a discussion and some conclusions in Sections 6 and 7 respectively.

2. Selection bias

The term "selection bias" has several different usages in epidemiology. Here we follow [9] and adopt a structural classification of selection bias as any bias arising due to conditioning on a "common effect" of the disease and exposure variables. The directed acyclic graphs (DAGs) in Figure 1 describe the dependencies between variables and demonstrate how selection bias arises in this case. In DAG 1a the exposure W and the outcome Y are marginally independent in the population, represented by the absence of an edge between the them. However, amongst the individuals in the study (those for whom S, the selection indicator, is 1) they are dependent. This phenomenon is called association by conditioning on a common child, or collider bias, and characterises selection bias as defined in this paper[9, 8]. In DAG 1b the exposure and the outcome are associated in the population, however in the study this association might be distorted by conditioning on selection.

Figure 1 shows the simplest structure that can lead to selection bias. However, the association between the exposure and the selection mechanism can have a number of forms, all charcterised by the selection indicator S being a collider (i.e. a common child of W and Y). Some of these are shown in Figures 1 and 2. The exposure can directly affect the selection as in Figure 1 or via an intermediate variable B as in Figure 2. This latter is the situation we consider in this paper, since our method relies on the existence of intermediate variables B that separate the disease (Y) – exposure (W) relationship (see Section 4.1).

In case-control studies, there is, by definition, an association between disease and selection (cases are more likely to be selected than controls). Selection bias arises when the exposure under investigation is also associated with the selection mechanism. Such an association between exposure and selection often means that the study population is unrepresentative of the population of interest, i.e. has been sampled differentially with respect to key variables (e.g. B in Figure 2). In case-control studies, cases and controls must be sampled from the same population in order to obtain valid odds ratio estimates. As cases and controls are typically sampled separately (unless the case-control study is nested within a cohort study), there is no guarantee that the cases and controls are random samples from the same population.

To clarify how this is a problem and to set the stage for our method, we define the different populations involved in a case control study. Initially researchers have in mind a population of interest, which we term the *target* population. The target population can be thought of as the population the study *intends to sample from*, or the population of *substantive interest* to the researchers, i.e. the population they would like to make inference about. However, when cases are sampled, these are not necessarily a random sample from the target population but from another *source* population. This is especially true for rare diseases where it might be difficult to find cases. Once cases have been obtained, controls are sampled, often with an attempt to match them to the cases. This can result in the controls being sampled from *yet another* source population. Thus all three — target, case source and control source — could be different, albeit overlapping, populations. This process makes case-control studies of this type especially vulnerable to selection bias.

A common consequence of selection bias is that the association between exposure and outcome in those selected into the study differs from the association among those eligible. For case-control studies, this corresponds to $P(W|Y,S) \neq P(W|Y)$, and hence the naive odds ratios from such a study will not, in general, be equal to the true odds ratios in the target population. Our focus in this paper is on case-control studies in which the selection process may be biased according to the structural classification described above, and which results in biased estimates of the odds ratios in the target population.

2.1. Approaches to tackle selection bias

There are a number of approaches to tackle selection bias in both the design and analysis stage of a case-control study. Wacholder et al.[10] suggest nesting case-control within cohort studies to guarantee that source populations are the same as each other and the target population, or ensuring that any exclusion/inclusion criteria that apply to the cases are reflected in the choice of controls in non-nested studies. While the latter means case and control populations will be similar to each other and hence valid odds ratio estimates might be obtained, it does not guarantee that these estimates are representative of the target population. More recently Haneuse and colleagues [11, 12] suggest using a multi-stage approach to sampling cases and controls in order to prevent selection bias.

At the analysis stage, a number of researchers in epidemiology [13, 14, 15] derived a simple and intuitive formula to show how selection bias could be adjusted for when selection probabilities were known or estimable. These were in fact identical to Horvitz-Thompson estimators [16] commonly used in survey statistics. Scott and Wild [17, 18] adopt a similar approach. Lin and Paik look at the problem in matched case-control studies where they are able to model the selection probabilities. [19].

In more complex analyses of selection bias, three models are typically considered: one for the question of interest, another for the the selection process, and a third for the relation between the observed and the missing variables. The different methods tend to make different assumptions and/or model these three components differently. Indeed, these three models are implicit in all analytic methods of handling selection bias, but in the simplest analyses, only the outcome-exposure model is considered (typically the standard logistic regression used for case-control studies). This corresponds to making an implicit assumption about ignorability of the selection process, which may or may not be reasonable.

Inverse probability weighted methods (IPW),[4] which are extensions of Horvitz-Thompson estimators, have been developed in the context of drop-out from AIDs trials. Typically IPW requires some knowledge of the drop-out mechanism and is therefore not as easily applicable to case-control studies where the selection probabilities are usually unknown. Advanced IPW methods often model all three components mentioned above. While this is an advantage of IPW methods, as each aspect of the problem has to be dealt with explicitly, it is typically complex and often computationally intensive.

In the context of case-control studies, we can think of the non-participating individuals who are part of the target population as missing, and so missing data methods such as multiple imputation[3] can be used to deal with selection problems. Typically, *missing at random* (MAR)[20, 3] assumptions are made that allow the researchers to ignore the selection process model. However, if selection depends on unobserved values of the exposure or outcome of non-participants in a case-control study then the data are *missing not at random*(MNAR). There is less guidance in the literature about how to tackle MNAR analyses, and results can be sensitive to the specification of the selection process model needed in this case. (Note that if data are MNAR this does not necessarily imply selection bias [21]).

In the context of sensitivity analyses, multiple bias approaches have been advocated to tackle selection bias. These are Bayesian models, or approximations to Bayesian models, that estimate *bias parameters*[5, 1] to quantify different sources of bias. These approaches involve formulating informative prior distributions on the bias parameters as usually there are no direct data available to estimate them. Results can be sensitive to choice of bias parameter priors.

We focus here on another technique with roots in the survey literature, termed post-stratification[22, 18, 23], which we use to assess the sensitivity of the case-control odds ratio to selection bias. In fact the H-T and post-stratification formulas are the same in the context of case-control studies (see web-based support material). The principal differences between them are how the weights are interpreted and thus estimated, and crucially, what information is available to the researcher. As mentioned above, H-T and IPW weights rely on information on the selection being available before the study is conducted whereas typically post-stratification weights are derived after the study although this is not necessarily the case [23]. Our method is based on two assumptions: the existence of variables (that we term *selection influence variables*, SIV) that *inform the selection and obey a specific conditional independence assumption*; and the availability of data on these variables that is *representative of the target population* of the case-control study. As we discuss in Section 6, these assumptions can be met in a wide range of situations since large administrative and population-based databases such as census, household surveys and cohort studies, are becoming publicly available.

The conditional independence assumption we make in our approach enables us to avoid dealing with models for the selection process and for imputing the missing data, and in this sense is analogous to making MAR assumptions conditional on the SIV's. A similar strategy is often used with multiple imputation methods, whereby fully observed auxiliary variables are included in the imputation model in an attempt to convert an MNAR process into one that can be assumed to be MAR. However, unlike multiple imputation methods, we do not need an imputation model, but instead use weights that can be obtained empirically from external population-based data sources. Thus, we focus on whether there are variables and data that conform with the assumptions above and not about modelling (although the weights can be modelled if required). The advantage is that thinking about SIV's and considering potential data sets to represent the target population is more straight-forward and accessible to non-statisticians such as epidemiologists than more complex modelling.

Our method does not rely on individual level data being available to inform the selection process – although partial information on individuals in the study can be used when available.

We describe the assumptions of our method in Section 4 after a description of the case studies we use to motivate our methods.

3. Case studies

3.1. Extremely low- Electromagnetic fields and childhood leukaemia case-control study (EE-ALL study)

Extremely low-frequency electromagnetic fields (EL-EMF) have been designated as possibly carcinogenic by the International Agency for Research on Cancer based on epidemiologic studies in children[24]. Also, EL-EMF studies have been much researched in the context of bias estimation [1, 25]. We consider here a US based study investigating the link between EL-EMF exposure from power lines and childhood acute lymphoblastic leukaemia (ALL)[26], which found little evidence of an association (odds ratio for ALL of 1.24, 95% confidence interval (0.86,1.79) at exposures of 0.2μT or greater as compared with less than 0.065μT). In a later analysis, Hatch et al[27] suggested that there might be selection bias due to differential participation rates in different socio-economic strata.

Cases were contacted by the Children's Cancer Group and controls were selected by random digit dialling and matched to cases according to the first eight digits of their phone numbers, age and race. Demographic details were collected over the telephone. EL-EMF measurements inside the residence of those who had completed the telephone interview were attempted by technicians blinded to the case/control status. For details see Linet et al.[26]. We term participants for whom indoor measurements were made, *full* participants, whilst those for whom indoor measurements were not made, either because of refusal or because the family had moved etc., are termed *partial* participants.

We concentrate on two socio-economic indicators (SEI): annual household income (income) and whether the family lived in a city or otherwise (urban). These seemed the most important SEIs likely to be related to exposure and participation (selection) in the study, particularly urban, as individuals living in the countryside tend to be exposed to more EL-EMF via power lines than those living in the city.

Table 1 shows the numbers (%) of individuals in the SEI groups by full and partial participant status. Full and partial participants are significantly different: half of full participants versus one third of the partial participants are in the highest income bracket (chi-squared test of association p-value = 1.2×10^{-14}) whilst two thirds of the full participants were city dwellers compared to only one quarter of the partial participants (chi-squared test of association p-value = 2.2×10^{-16}). If EL-EMF exposure is associated with the SEI of the participants then this can result in selection bias. Finally, whilst the case-control split amongst full participants is fairly even, a slightly larger proportion of the partial participants are controls (chi-squared test of association p-value = 0.0035). This leads us to further question whether the full and partial participants are drawn from the same (target) population.

3.2. Hairspray exposure and hypospadias case-control study (H-H study)

We also consider a case-control study investigating the association between hypospadias, a congenital anomaly of baby boys, and occupational exposure to hairspray. Ormond et al[28] estimated that maternal exposure to hairspray was associated with increased risk of giving birth to a baby boy with the anomaly (odds ratio = 2.4, 95% confidence interval (1.40,4.17), adjusting for income and smoking).

As in the EE-ALL study, there were full and partial participants in this study. The full participants completed a telephone questionnaire whereas partial participants only responded to an initial mailing. Again, there are good reasons to suspect that both exposure and participation in the study are related to socio-economic factors, which may give rise to selection bias. We obtained a measure of socio-economic status — the 1991 Carstairs score (an area-level deprivation index[29]) — for each full and partial participant with a residential postcode that could be linked to an electoral ward (for details, see Section 1.1 of the Web-based supporting material). The Carstairs deprivation score was discretised into three categories, high, medium and low, using tertiles. Table 2 shows the differential distribution of deprivation category between full and partial participants (chi-squared test of association p-value = 0.003).

Maternal age is also associated with occupational exposure to hairspray[30, 31], and is another socio-economic indicator that may be differentially associated with participation of cases and controls. Maternal age is only available for full participants, so we cannot examine empirically whether it is associated with selection into the study. However, amongst full participants, there is a slightly lower proportion of controls (9%) than cases (13%) in the youngest age group ($<$ 25 years) and a correspondingly higher proportion (27% of controls versus 22% of cases) in the oldest age group ($\geq$ 35 years) (chi-squared test of association p-value = 0.06) which may be at least partially due to differential participation rates amongst young and old mothers of cases versus controls. Table 2 also shows a strong imbalance between the case:control ratio for full versus partial participants (chi-squared test of association p-value = 1.1×10^{-8}), suggesting that full and partial participants may not be drawn from the same population.

4. Sensitivity analysis using post-stratification

In this section we describe the method we propose to investigate sensitivity of the marginal odds ratio (MOR) (also termed the causal odds ratio) to selection bias. While the MOR is not always the reported parameter in case-control studies (conditional odds ratios are usually reported as they are easily calculated using logistic regression), it is often of interest as it gives population estimates. We discuss how marginal odds ratios can differ from conditional odds ratios in Section 6.

The MOR can be expressed as follows:

$$\begin{aligned} MOR &= \frac{p(Y=1|W=1)p(Y=0|W=0)}{p(Y=0|W=1)p(Y=1|W=0)} \\ &= \frac{p(W=1|Y=1)p(W=0|Y=0)}{p(W=0|Y=1)p(W=1|Y=0)}, \end{aligned} \tag{1}$$

where Y is the outcome and W the exposure [14]. If there is selection bias then, as shown in Figure 2, the exposure is associated with the selection. This means that the MOR cannot be estimated from the study, rather, we estimate a biased odds ratio:

$$SBOR = \frac{p(W=1|Y=1,\mathbf{S}=\mathbf{1})p(W=0|Y=0,\mathbf{S}=\mathbf{1})}{p(W=0|Y=1,\mathbf{S}=\mathbf{1})p(W=1|Y=0,\mathbf{S}=\mathbf{1})} \tag{2}$$

where S takes value 1 if an individual is selected into a study and 0 otherwise. If W and S are dependent, the selection biased odds ratio is not necessarily equal to the MOR.

4.1. *Identifying the selection influence variables*

We follow Geneletti et al. [7] and attempt to assess selection bias by assuming that we can find a set of variables B, termed the *selection influence variables* (SIV) (bias breaking variables in [7]), that obey two assumptions:

- A1 they separate the disease – exposure association of interest from the nuisance selection mechanism, thereby transferring the selection bias from the exposure to the selection variable;
- A2 they are such that data on the SIV are readily available for the subjects in the study (e.g. from baseline measurements collected on all eligible participants) and do not suffer from selection bias, or that external sources of data on the SIV are available for a representative sample from the target population.

In addition we assume that the SIV are either categorical or can be categorised for the purposes of this analysis.

We formally express assumption (A1) in terms of conditional independences

$$W \perp\!\!\!\perp S|(Y,B), \tag{3}$$

where B represents the SIV. Conditional independence (3) says that if the values of the outcome Y and the SIV B are known, then the value of the exposure W does not depend on the value of S – it is the *same* in the study and the target population for a given value of Y and B. If conditional independence (3) holds then no additional difficulties are encountered if B is also a confounder for the effect of W on Y. This relationship is depicted graphically in the DAGs in Figure 2. In causal DAG terms, B and Y block all paths from W to S.

We can use conditional independence (3) to decompose the probability of exposure W given case/control status as follows:

$$p(W=1|Y=y) = \sum_B \underbrace{p(W=1|Y=y,B)}_{(a)} \times \underbrace{p(B|Y=y)}_{(b)}. \tag{4}$$

(a) is estimated directly using data from those for whom $S=1$, as conditional independence (3) allows us to replace it with $p(W=1|Y=y,B,S=1)$, the B-stratum-specific exposure probability for the case/control participants. Equation (4) is an example of a post-stratification equation[22] and (b) is a post-stratification weight, denoted $\mathbf{p}_b$.

4.2. *Confounding*

Conditional independence (3) can be extended to (5) if, in addition to B there are confounders C, *not implicated* in the selection process.

$$(W,C) \perp\!\!\!\perp S|(Y,B). \tag{5}$$

Conditional independence (5) implies that, given the values of Y and B, both W and C do not depend on the selection status S. Conditional independence (5) holds in a number of contexts, some of which are exemplified in the three DAGs in Figure 3: when B and C are marginally independent (3a); when they are associated (3b); and when B is also a confounder for the relationship between W and Y (3c). We can use conditional independence (5) to decompose the probability of exposure W given case/control status, as follows:

$$p(W=1|Y=y) = \sum_B \sum_C \underbrace{p(W=1|Y=y,B,C)}_{(a)} \times \overbrace{p(C|Y=y,B)}^{(b)} \times \underbrace{p(B|Y=y)}_{(c)} \quad (6)$$

In equation (6), (a) is the B-stratum-specific probability of exposure given Y adjusted for confounding, (b) is a new term that represents the weight associated with a confounding adjustment and (c) is the post-stratification weight $\mathbf{p}_b$. As before, we can estimate (a) and (b) using data from the the full participants (i.e. those for whom $S=1$) as conditional independence (5) allows us to replace these expressions with $p(W=1|Y=y,B,C,S=1)$ and $p(C|Y=y,B,S=1)$ respectively.

In the context of the post-stratification approach we are proposing, the difference between confounders and SIV is that, while we deal with confounding by adjusting for it using the full participant data (the equivalent of adding it as a covariate in a regression), the distribution of the SIV B needs to be estimated from additional data.

We focus on the simplest case without additional confounders in our analysis below. An extension where we tackle additional confounding can be found in the Web-based supporting material Section 1.3.

4.3. Estimating the post-stratification weights, $\mathbf{p}_b$: Sensitivity analysis

How do we estimate $\mathbf{p}_b$ as required by assumption (A2)? Figure 2 shows that S is a collider between Y and B. Thus from the study we estimate $p(B|Y,\mathbf{S}=\mathbf{1})$ not $p(B|Y)=\mathbf{p}_b$ as required. Our solution is to estimate $\mathbf{p}_b$ by using data *outside* the full participant data.

For non-nested case-control studies the source populations of cases and controls are not necessarily the same as each other or the target population, as covered in Section 2. However, we know what the target population is and although we typically do not have the whole target population at our disposal, we often have data sets that can be thought of as representative or random samples from the target population. We can use these to estimate the weights and the MOR and thus assess the sensitivity of the MOR to selection bias. It is difficult to be prescriptive about what constitutes robust or highly sensitive results, but if the MOR hardly varies, e.g. if the point estimates for each sensitivity analysis do not differ by more than, say, 10% or 20% of the width of the 95% uncertainty interval, then we can probably report any of the estimates with confidence. However, if the MOR is highly variable, e.g. if the point estimates for some sensitivity analyses fall outside the 95% uncertainty intervals for other sensitivity analyses, then we would almost certainly want to re-asses the study conclusions and validity.

4.4. Bayesian estimation of the post-stratified MOR

4.4.1. Bayesian post-stratification Bayesian methods have become increasingly popular in the epidemiology literature[32, 33, 34]. Bayesian post-stratification has been used in a number of contexts[35, 22] in the survey literature. Geneletti et al.[7] applied non-Bayesian post-stratification to the H-H case-control study. We extend these approaches here.

In the current context, using a Bayesian framework has several advantages over a classical analysis. It makes estimating the variance of the adjusted MOR simple as it can be derived as the variance of the posterior distribution of the adjusted MOR obtained by using Markov Chain Monte Carlo (MCMC) methods. By contrast, a cumbersome approximation derived using the delta method is needed to obtain standard errors for the adjusted MOR estimated using frequentist techniques[7]. Also, Bayesian methods give us the flexibility to jointly model the post-stratification weights as well as the exposure probabilities, if required. This is useful when data used to estimate the post-stratification weights is sparse or must be constrained in some way. In such situations it is possible to use prior information to augment and stabilise the raw frequencies. However, in the current analysis it was not necessary to model the post-stratification weights as the datasets were sufficiently large. Finally, a Bayesian approach to modelling the MOR can lead to more reliable estimates than several non-Bayesian alternatives for case-control studies with small samples, as we demonstrate by a simulation study reported in the web-based supporting material.

4.4.2. Modelling the MOR In a Bayesian analysis, we cannot in general use the prospective/retrospective equivalence of the logistic regression model to make inference about retrospective exposure probabilities as can be done in the frequentist set-up[36]. Instead the design of case-control studies allows us to directly estimate the vector of probabilities $p(W, B|Y = 1)$ and $p(W, B|Y = 0)$[36] for all combinations of W and B. We further stratify by B so that we estimate $p(W|B, Y = 1) = \boldsymbol{\gamma}$ and $p(W|B, Y = 0) = \boldsymbol{\phi}$ for all the categories of W and B which we can then plug into Equation (4)(a). If additional confounders are used we can directly estimate $p(W|B, C, Y)$ as needed in equation (6)(a). The discussion that follows can be extended to this case.

In the examples we consider, we have a binary exposure for the EE-ALL study and a three valued exposure for the H-H study as hairspray takes on values "no exposure" (0), "exposure" (1) and "unsure" (2). For the former we use a Beta-Binomial model, for the latter we use a Dirichlet-Multinomial model. We discuss the Dirichlet-Multinomial model in some detail below; the arguments extend with minor variations to the Beta-Binomial model. See the Web-based supporting material for further discussion of the latter.

We want to estimate $\boldsymbol{\gamma} = p(W|B, Y = 1)$ and $\boldsymbol{\phi} = p(W|B, Y = 0)$ where W and B are categorical variables. In the most general case, we have an exposure W taking on Q values, and R categorical/ordinal variables B (and confounders C if any are included), excluding the exposure and the case/control status. If each of these R variables take on r_k values for $k \in \{1, \ldots, R\}$, then we have

$$\prod_{k}^{R} r_k = T$$

multinomial models, each with Q cells, for each of cases and controls. Let case multinomial cell ij contain n_{ij} individuals and control multinomial cell ij contain m_{ij} individuals for $i \in \{1, \ldots, T\}$, $j \in \{1, \ldots, Q\}$. If $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{iQ})$ are the case exposure probabilities and $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{iQ})$ are the control exposure probabilities for each strata, then the retrospective likelihood is given by

$$\prod_{i=1}^{T} l(\boldsymbol{\gamma}_i, \boldsymbol{\phi}_i, \mathbf{n_i}, \mathbf{m_i}) \propto \prod_{i=1}^{T} \prod_{j=1}^{Q} (\gamma_{ij}^{n_{ij}} \times \phi_{ij}^{m_{ij}}), \tag{7}$$

where $\mathbf{n}_i = (n_{i1}, \ldots, n_{iQ})$ and similarly for $\mathbf{m}_i$. A convenient choice of prior is then

$$\boldsymbol{\gamma}_i \sim Dir(a_{i1}, \ldots, a_{iQ}) \tag{8}$$

$$\boldsymbol{\phi}_i \sim Dir(b_{i1}, \ldots, b_{iQ}), \tag{9}$$

where Dir represents the Dirichlet distribution.

There are two points to note relative to the model above. The first is that this model implicitly assumes that the cases and controls are independent of one another a-priori. This is easily seen by noting that the prior parameters a and b are not correlated. The second point is that we can interpret the Dirichlet prior parameters a_{ij} and b_{ij} as prior "sample sizes" for stratum ij for cases and controls respectively. This is because, by Dirichlet-Multinomial conjugacy, the posterior distribution of the probabilities is another Dirichlet distribution with parameters $a_{ij} + n_{ij}$ for cases and $b_{ij} + m_{ij}$ for controls. In order to express our relatively weak prior beliefs in the current context, we set all our Dirichlet parameters to 0.2 for both cases and controls. We checked robustness to this choice in the two case studies by also trying other weak priors, e.g. setting the Dirichlet parameters to 0.5 or 1.

We use WinBUGS to model and estimate the posterior Dirichlet parameters. The WinBUGS model code for these models is given in the Web-based supporting material. We used WinBUGS rather than performing the conjugate analysis analytically because we wanted estimates of the posterior distributions of the odds ratios which required additional manipulation of the Dirichlet posterior distributions involving the post-stratified weights as shown in Equation (4).

5. Applying the sensitivity analysis to the case studies

5.1. Identifying the selection influence variables

Table 1 shows that in the EE-ALL study, the SEI (urban status and income) have different distributions for full and partial participants indicating that selection and the SEI are associated. Also, it is plausible that EL-EMF exposure is associated with the SEI. This implies that selection and exposure to EL-EMF are marginally associated thus potentially giving rise to selection bias (Figure 2). If we are willing to assume that EL-EMF (W) and selection (S) are only associated through the SEI (B) and case/control status (Y), i.e. conditional independence (3) holds, then the SEI are potential SIVs.

Using a similar argument we assume that the Carstairs deprivation score and maternal age are the SIV in the H-H study (Table 2).

5.2. *Estimating the post-stratification weights,* $\mathbf{p}_b$

In our proposed sensitivity analysis strategy, we can vary two factors. The first is the *source of data* we use to obtain estimates of $\mathbf{p}_b$. These are sources of data that are potential representative samples from the target population. The second is the nature of the estimate of $\mathbf{p}_b$, which can be either *conditional* on, or *marginal* to, the case/control status.

5.2.1. EE-ALL study In the EE-ALL study we consider two sources of data: (i) *External* data from the Current Population Survey (CPS), a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics in the US which include income and urban status; (ii) *Internal* data in the form of the *combined* full and partial participant data for which the SIV's, income and urban status, are known. We refer to these as external and internal data sources to highlight the fact that the external data are from sources unrelated to the study whereas the internal partial participant data were collected as part of the study but were excluded in a full participant analysis.

We argue that these two data sources are plausible representative samples from the target population. With respect to the external data, most children with leukaemia are eventually identified, so it makes sense to assume that the target population of the cases is indeed the population in which the study took place, i.e. the nine US states in which the study was conducted. This extends to children who have passed away from the disease. As the CPS data are representative of this population, they provide a valid estimate of the distribution of income and urban status.

The study organisers went to great lengths to correctly identify the target population and sample controls from it, thus the internal sample of combined full and partial participants that were initially recruited provides another plausible data source.

The external data source, the CPS, does not contain information on the case/control status of the sample. It is therefore only possible to obtain estimates of $p(B)$ marginalised over the case/control status, and not $p(B|Y)$. However it is plausible that the distribution of B in controls is represented by the CPS data. We therefore use these data only to re-weight the controls as these are tend to suffer more from selection bias than cases and thus be less well represented in the study.

In the internal data, it is possible to obtain both marginal estimates of $\mathbf{p}_b$ (combining the case and control data) and estimates conditional on the case/control status.

The top half of Table 3 lists the different sensitivity analyses we performed for the EE-ALL study, based on various combinations of data to estimate the post-stratification weights. We re-weight cases and controls differently to reflect the fact that they are likely to suffer from selection bias to different degrees.

5.2.2. H-H study Similar arguments hold for the H-H study where the SIV B are Carstairs deprivation score and maternal age. In this study we considered three data sources for estimating the post stratification weights: the combined full and partial study data, the 2001 UK Census data (which provides information on the population distribution of the Carstairs deprivation score) and the Millennium Cohort Study (MCS)[37]. The MCS is a nationally representative longitudinal birth cohort study following the families of children born between 2000 and 2001 in the UK, and contains information on maternal age and deprivation. We used these three datasets to estimate $p(Depriv, MA|Y)$ where our selection variable was $B = \{Depriv, MA\}$ and $Depriv$ and MA stand for deprivation category and maternal age, respectively. The external post-stratification weighted estimates require the use of both the Census data and the MCS data (for details see Web-based supporting Material Section 1.3). The various sensitivity analyses performed for the H-H study based on different data sources to estimate the post-stratification weights are listed in the bottom half of Table 3.

5.3. *Results*

5.3.1. EE-ALL study Table 4 shows the posterior medians and credible intervals for the standard MOR estimate, together with a number of plausible post-stratification weighted estimates. Although the differences between the standard estimate and the adjusted estimates are modest and do not affect the epidemiologic conclusions of the study, i.e. that EL-EMF exposure is not associated with an increase in incidence of ALL, there is some evidence of the odds ratio's sensitivity to different post-stratification weights and consequently to different assumptions about the target population. In particular (and not surprisingly) using the external CPS data to re-weight controls (SA5, SA6) has a bigger impact on the OR estimate than using the internal full and partial participant data. This suggests that the original sample of full and partial participants is not completely representative of the CPS population which is a likely candidate for the target population.

For reference, Table 4 also shows the conditional odds ratio estimate from a frequentist logistic regression with B as explanatory variables (i.e. ignoring selection bias), as this is commonly reported for case-control studies. There is no substantial difference between this and the post-stratified estimates. Table 4 also shows the corresponding re-weighted estimates using the raw frequencies for the case-control probabilities instead of modelling them using Bayesian techniques. Again results are similar across the scenarios although raw variances are slightly smaller. The Bayesian approach takes into account modelling uncertainty, further the variances obtained are based on posterior sampling and are thus theoretically grounded. This is in contrast to those derived for the raw variances which are based on the delta method[7]. Changing the

values of the Dirichlet prior parameters in the Bayesian post-stratification models also had negligible impact on the results (not shown).

5.3.2. H-H study The results for the post-stratification weighted estimates are given in Table 5 and are very similar across data sources. The only difference relates to SA5 and SA6, the estimates involving the external data sources, where the MOR is slighty smaller although this difference is negligible with respect to the width of the credible intervals. These two scenarios also produce slightly wider 95% credible intervals. We found modest sensitivity of the MOR estimates to the choice of Dirichlet prior parameters in this application, since some of the multinomial cell counts were very sparse. For example, setting the all the Dirichlet prior parameters to 1 rather than 0.2 reduced the posterior median (95% credible interval) for the MOR under the standard scenario to 2.23 (1.39, 3.61). Differences across the different post-stratification scenarios was robust however, since all scenarios were similarly affected by the change in Dirichlet prior setting.

The conditional odds ratio estimate from a frequentist logistic regression with B as explanatory variables is slightly higher than any of the Bayesian post-stratified estimates, although the substantive conclusions do not change. This difference partly reflects the sensitivity of the Bayesian MOR estimates to the choice of prior setting for the Dirichlet parameters in this example. We do not include the raw estimates for this analysis as analogous results appear in [7] and are similar in size and variability to the Bayesian estimates shown here.

6. Discussion

In both studies we observe some sensitivity to different models and weights. However the estimates and the statistical significance of the results does not change in either study, suggesting that the original study estimates are robust and unlikely to be affected by substantial selection bias. This conclusion is reassuring to the epidemiologists involved in the studies. By having performed this quantitative sensitivity analysis using additional data sources, we can say with some confidence that there is no substantial impact of selection bias mediated through the SIV we chose. Studies with highly variable MORs and large differences between the full participant and additional data distributions of the SIV will be less robust.

We only show results where the SIV are assumed to be the same for cases and controls. We have have also investigated how to adjust cases and controls using different SIV. For example, in the H-H study we adjusted cases using only deprivation and controls using maternal age and deprivation (not shown). This did not significantly change the results.

Odds ratios conditional on particular characteristics are often of interest. If these are conditional on the SIV (e.g. the odds ratio for urban dwellers in the EE-ALL study) then these are not subject to selection bias (assuming urban status is the only SIV). If we are interested in conditional odds ratios that are not conditional on SIV (e.g. the odds ratio for boys in the EE-ALL study) then these may still be subject to selection bias and the method described above can be applied within strata of the variable we are interested in conditioning on.

When suitable data to perform post-stratification cannot be found, sensitivity analysis can be performed by proposing plausible distributions for the SIV B in the same spirit as Greenland [1], using prior knowledge to elicit and constrain these distributions. Similar arguments hold when the data on B are sparse.

6.1. Differences between estimates

There are some differences between re-weighted estimates based on different data sources. For the EE-ALL data, the re-weighted estimates tend to estimate the effect of exposure to EL-EMF as lower than the standard estimate, in particular for SA5 and SA6 which are based on the external CPS data. The post-stratified estimates overlap with the logistic regression estimate.

In the H-H study, the post stratified estimates were very stable when using internal data but resulted in slightly lower MOR for SA5 and SA6 which used the external Census and MCS data.

The differences between the logistic regression odds ratio and the adjusted or marginal odds ratios (in particular for the H-H study) could be due to a number of reasons. We have already noted that the MOR estimates in the H-H study were somewhat sensitive to the choice of Dirichlet prior parameters. Other explanations are that the logistic regression estimates an odds ratio conditional on the SIV having fixed values, rather than a marginal odds ratio; there could be an interaction between the exposure and the SIV; or it could be a product of the non-collapsibility of the odds ratio[38]. We found no evidence of effect modification. If the differences are due to non-collapsibility or the conditional nature of the logistic regression odds ratio, they are not significant enough to change any of the conclusions. It is, however, worth bearing in mind when comparing the logistic regression estimate with others obtained using post-stratification, and we recommend that the main comparison be with a standard MOR estimated under the assumption of no selection bias.

6.2. *Assumptions underlying our approach*

The sensitivity analysis rests on assumptions (A1) and (A2) in Section 4.1 being plausible. While in our case studies we have argued that the assumptions are realistic, is this generally the case? When B is a set of socio-economic indicators, as it often is in case-control studies where control participation (like case participation) is voluntary, it is usually possible to find data on these indicators from routinely gathered data sources such as census or population surveys that are carried out by most national statistics offices. When B is not a socio-economic indicator then finding suitable datasets requires more thought. Disease registers, such as cancer registries, can often provide information about symptoms of disease and patient characteristics that can be used to post-stratify when the selection bias comes about through the cases rather than the controls, e.g. due to susceptibility bias.

7. Conclusions

Our post-stratification based method provides a simple empirically and theoretically grounded approach to selection bias sensitivity analysis. In particular, if potential SIV can be identified and plausible data sources can be found to estimate their distribution, then no additional modelling is required. We assume that given conditional independence (3) holds, the data we have on the SIV already embodies the relationships between the exposure and the outcome and there is no need for additional modelling of this relationship nor the selection mechanism. Any variability in the odds ratio can be clearly attributed to the changes in the population that the researcher has chosen to use for re-weighting, and not the the specification of models of missingness or bias parameter priors. The only modelling issue in our method is the choice of prior values for the Dirichlet (or Beta) distributions. We recommend using weak priors and checking sensitivity to different prior specifications, particularly when the case or control data in some of the multinomial cells are very sparse. If the data on the SIV used to calculate the post-stratification weights are sparse, we also recommend carrying out some modelling of these data to produce more stable weights.

Our method is computationally simple (see WinBUGS code in the Web-based supporting material), in particular when the weights are estimated using raw frequencies[7]. Indeed, using WinBUGS/MCMC for estimation is not strictly necessary (but is computationally convenient), since all that is needed is to generate samples from the appropriate Dirichlet posterior distribution (see section 4.4.2), which is available in closed-form, and then re-weight these samples using the relevant post-stratification weights. Our method can also easily be extended if models are needed to estimate the post-stratification weights, and to contexts outside case-control studies where selection bias is present. Also, further structure can be included in order to model case/control probabilities or weights.

While our method is accessible and requires minimal modelling, it is not as generic as multiple imputation and bias modelling as it relies on the existence of SIV and the appropriateness of the associated conditional independence assumptions, and also on the availability of suitable data on the SIV. In fact, the main aim of our approach is to explore potential for selection bias rather than to propose a "quick fix" adjustment. In particular, using our approach helps to think carefully about how case and control populations differ; what our target population is; what the sources of bias are and what variables can be used to assess the sensitivity of the results to these biases.

In this paper we have focussed on using Bayesian post-stratification to assess sensitivity of case-control study results to selection bias. As this was our explicit aim, we tried to find data sources that best represented the target populations of the studies. These were the CPS data for the EE-ALL study and the census and MCS study for the H-H study. However, as suggested by a referee, our method could also be used to assess how the exposure-outcome association varies across *different target populations* of interest. Consider for example the EE-ALL study: this was conducted in 9 US states and we pooled the CPS data to obain weights for the analysis. We could instead have looked at the effect on the MOR of using the data from the 9 states individually. If these states differed considerably with respect to urban/rural composition and income distribution then we might well have seen differences in the MOR. This might be of interest to policy makers.

Increasingly, epidemiologists and medical statisticians are incorporating sensitivity analyses in their research[39]. They recognise that, when there are a number of potential sources of bias in their data, conclusions based on a single "best" model do not account sufficiently for uncertainty in the results. The challenge for the researcher is then to identify appropriate methods and data sources to perform suitable sensitivity analyses in their specific context.

If selection bias is a potential problem, we encourage using all available data sources and exploring various SIV that are in line with subject matter knowledge. If selection bias appears to be present then it is up to the researcher to decide whether the post-stratified adjusted estimates should be reported as describing the range of possible values of the odds ratio, or whether a more complex analysis is required.

Acknowledgements

The authors would like to thank Dr. Martha Linet who kindly provided data files and coding manuals for the National Cancer Institute - Children's Cancer Group case-control study of residential exposure to magnetic fields and acute lymphoblastic leukaemia in children. We would also like to than two reviewers and an associate editor for making comments that improved the paper. S. Geneletti, S. Richardson and N. Best would like to acknowledge financial support from ESRC (grant RES-576-25-0015). M. Toledano and P. Elliott would like to acknowledge financial support from UK Department of Health (grant 12167262) for the hypospadias study. We would also like to acknowledge the reviewers and associate editor that refereed the paper.

References

1. Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc Ser A Stat Soc* 2005; **168**(2):267–306.
2. Saltelli A, Tarantola S, Campolongo F. Sensitivity analysis as an ingredient of modeling. *Statistical Science* NOV 2000; **15**(4):377–395.
3. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Statistical Methods in Medical Research* 2007; **16**(3):199–218, doi:{10.1177/0962280206075304}. Workshop on Multiple Imputation, Cambridge, ENGLAND, APR 12-13, 2005.
4. Robins J, Hernan M, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* SEP 2000; **11**(5):550–560.
5. Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society Series A-Statistics in Society* 2009; **172**(Part 1):21–47, doi:{10.1111/j.1467-985X.2008.00547.x}.
6. Mezei G, Kheifets L. Selection bias and its implications for case-control studies: a case study of magnetic field exposure and childhood leukaemia. *International Journal of Epidemiology* 2006; **35**:397–406.
7. Geneletti S, Richardson S, Best N. Adjusting for selection bias in retrospective, case-control studies. *Biostatistics* 2009; **10**(1):17–31.
8. AAAI ((ed.)). *Controlling Selection Bias in Causal Inference*, 2011.
9. Hernan M, Hernandez-Diaz S, Robins J. A structural approach to selection bias. *Epidemiology* 2004; **15**(5):615–625.
10. Wacholder S, Silverman D, McLaughlin J, Mandel J. Selection of controls in case-control studies .2. Types of controls. *American Journal of Epidemiology* MAY 1 1992; **135**(9):1029–1041.
11. Haneuse S, Schildcrout J, Crane P, Sonnen J, Breitner J, Larson E. Adjustment for Selection Bias in Observational Studies with Application to the Analysis of Autopsy Data. *Neuroepidemiology* 2009; **32**(3):229–239, doi:{10.1159/000197389}.
12. Haneuse S, Chen J. A Multiphase Design Strategy for Dealing with Participation Bias. *BIOMETRICS* MAR 2011; **67**(1):309–318, doi:{10.1111/j.1541-0420.2010.01419.x}.
13. Kleinbaum D, Morgenstern H, Kupper L. Selection bias in epidemiologic studies. *American Journal of Epidemiology* 1981; **113**(4):452–463.
14. Austin M, Criqui M, Barret-Connor E, Holdbrook M. The Effect of response bias on the odds ratio. *American Journal od Epidemiology* 1981; **114**(1):137–143.
15. Greenland S, Neutra R. An Analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. *Journal of Chronic Diseases* 1981; **34**(9-10):433–438, doi:{10.1016/0021-9681(81)90002-3}.
16. Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 1952; (47):663685.
17. Scott A, Wild C. Fitting logistic models under case-control or choice based sampling. *Journal of the Royal Statistical Society, Series B* 1986; (48):170–182.
18. Scott A, Wild C. Fitting logistic regression models in stratified case-control studies. *Biometrics* 1991; (47):497–510.
19. Lin I, Paik M. Matched case-control data analysis with selection bias. *Biometrics* DEC 2001; **57**(4):1106–1112, doi:{10.1111/j.0006-341X.2001.01106.x}.
20. Heitjan D. What can be done about missing data? Approaches to imputation. *Americal Journal of Public Health* APR 1997; **87**(4):548–550, doi:{10.2105/AJPH.87.4.548}.
21. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. *STATISTICAL METHODS IN MEDICAL RESEARCH* JUN 2012; **21**(3, SI):243–256, doi:{10.1177/0962280210394469}.
22. Gelman A. Struggles with survey weighting and regression modelling. *Stat Sci* 2007; **22**(2):153–164.
23. Lu H, Gelman A. A method for estimating sampling variances for surveys with weighting, poststratification and raking. *Journal of Official Statistics* 2003; **19**(2):133–151.
24. *Non-ionizing radiation, Part 1: Static and Extremely low-frequency Electric and magnetic fields*, *IARC Monographs on the evaluation of Carcinogenic Risks to Humans*, vol. 80, IARC Press, Lyon, France, 2002.
25. Greenland S, Kheifets L. Leukemia attributable to residential magnetic fields: Results from analyses allowing for study biases. *Risk Anal* 2006; **26**(2).
26. Linet M, Hatch E, Kleinerman R, LL R, Kaune W, Friedman D, Severson R, Haines C, Hartstock C, Niwa S, *et al.*. Residential exposure to magnetic fields and acute lymphoblastic leukaemia in children. *New Engl J Med* 1997; **337**(1):1–7.
27. Hatch E, Kleinerman R, Linet M, Tarone R, Kaune W, Anssi A, Dasul B, Robison L, Wacholder S. Do confounding or selection factors of residential wire codings and magnetic fields distort findings of electromagnetic field studies? *Epidemiology* 2000; **11**(2):189–198.
28. Ormond G, Nieuwenhuijsen M, Nelson P, Toledano M, Izatt N, Geneletti S, Elliott P. Endocrine disruptors in the workplace, hair spray, folate supplementation, and risk of hypospadias: Case-control study. *Environ Health Perspect* 2009; **117**(2):303–307.
29. Carstairs V, Morris R. *Deprivation and Health in Scotland*. Aberdeen University Press: Aberdeen, 1991.
30. Office of National Statistics. *The National Statistics Socio-economic Classification: User Manual* 2005.
31. Takkouche B, Regueira-Mendez C, Montes-Martinez A. Risk of cancer among hairdressers and related workers: a meta-analysis. *International Journal of Epidemiology* DEC 2009; **38**(6):1512–1531, doi:{10.1093/ije/dyp283}.
32. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology* JUN 2006; **35**(3):765–775, doi:{10.1093/ije.dyi312}.

33. Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *International Journal of Epidemiology* FEB 2007; **36**(1):195–202, doi:{10.1093/ije/dyl289}.
34. Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *International Journal of Epidemiology* DEC 2009; **38**(6):1662–1673, doi:{10.1093/ije/dyp278}.
35. Gelman A, Carlin B. Poststratification and weighting adjustments. *Survey Nonresponse*, Groves R, Dillman D, Eltinge J, Little R (eds.), New York: Wiley., 2001; 289–302.
36. Seaman S, Richardson S. Bayesian analysis of case-control studies with categorical covariates. *Biometrika* 2001; **88**(4):1073–1088.
37. University of London IoE. Centre for Longitudinal Studies, Millennium Cohort Study: First Survey, 2001-2003 [computer file]. Colchester, Essex: UK Data Archive [distributor] March 2007. SN: 4683.
38. Greenland S, Robins J, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science* FEB 1999; **14**(1):29–46.
39. Geneletti S, Best N, Mason A. Adjusting for selection effects in epidemiologic studies, why sensitivity analysis is the only "solution". *Epidemiology* 2011; **22**(1):36–39.

Table 1. Numbers (percentages) of individuals in each SEI category and case/control category, by full and partial participant status in the EE-ALL dataset. There are a total of 1092 (72%) full and 424 (28%) partial participants.

	Full	Partial
income		
$\leq$ \$20k	134(12)	119(27)
\$20k-\$39k	381(35)	156(37)
>\$39k	577(53)	152(36)
urban		
city	733(67)	117(27)
rural	359(33)	310(73)
case/control		
cases	576(53)	189(44)
controls	516(47)	238(56)

Table 2. Numbers (percentages) of individuals in each deprivation category and case/control category, by full and partial participant status in the H-H dataset. There are a total of 857 (70%) full and 359 (30%) partial participants.

	Full	Partial
deprivation category		
low	300(35)	100(28)
medium	290(34)	112(31)
high	267(31)	147(41)
case/control		
cases	422(49)	112(31)
controls	435(51)	247(69)

Table 4. EE-ALL study: medians and 95% credible intervals for the posterior distribution of the Bayesian post-stratified MOR for different sensitivity analyses as well as the standard estimate; corresponding point estimates and 95% confidence intervals of the post-stratified MOR estimated using raw case and control frequencies; and logistic regression estimate and 95% confidence interval for the conditional odds ratio adjusted for the SIV (income and urban status).

	Bayesian posterior estimates of the MOR		**Raw estimates of the MOR**	
Scenario	*median*	*95% credible interval*	*point est.*	*95% confidence interval*
Standard	1.15	(0.81,1.62)	1.14	(0.85,1.52)
SA1	1.12	(0.80,1.58)	1.13	(0.84,1.50)
SA2	1.18	(0.84,1.67)	1.15	(0.86,1.54)
SA3	1.08	(0.77,1.53)	1.13	(0.85,1.50)
SA4	1.14	(0.81,1.62)	1.10	(0.83,1.48)
SA5	0.93	(0.64,1.35)	0.98	(0.75,1.29)
SA6	0.98	(0.68,1.42)	1.00	(0.77,1.31)
	Frequentist logistic regression estimate of the conditional OR			
	point est.		*95% confidence interval*	
Frq Logist	1.16		(0.79,1.71)	

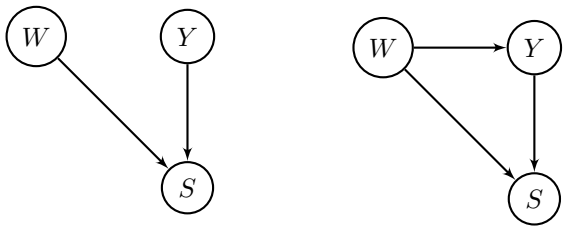
EE-ALL study				
	Cases		*Controls*	
Scenario	Type	Data Source	Type	Data Source
Standard	Conditional	Internal: full participant	Conditional	Internal: full participant
SA1	Conditional	Internal: full participant	Conditional	Internal: combined full and partial participant
SA2	Conditional	Internal: combined full and partial participant	Conditional	Internal: combined full and partial participant
SA3	Conditional	Internal: combined full and partial participant	Marginal	Internal: combined full and partial participant
SA4	Conditional	Internal: full participant	Marginal	Internal: combined full and partial participant
SA5	Conditional	Internal: full participant	Marginal	External: CPS
SA6	Conditional	Internal: combined full and partial participant	Marginal	External: CPS

H-H study				
	Cases		*Controls*	
Scenario	Type	Data Source	Type	Data Source
Standard	Conditional	Internal: full participant	Conditional	Internal: full participant
SA1	Conditional	Internal: full participant	Conditional	Internal: combined full and partial participant for $Depriv$, full participant for MA
SA2	Conditional	Internal:combined full and partial participant for $Depriv$, full participant for MA	Conditional	Internal: combined full and partial participant for $Depriv$, full participant for MA
SA3	Conditional	Internal:combined full and partial participant for $Depriv$, full participant for MA	Marginal	Internal: combined full and partial participant for $Depriv$, full participant for MA
SA4	Conditional	Internal: full participant	Marginal	Internal: combined full and partial participant for $Depriv$, full participant for MA
SA5	Conditional	Internal: full participant	Marginal	External:Census for $Depriv$ and MCS for MA
SA6	Conditional	Internal: combined full and partial participant for $Depriv$, full participant for MA	Marginal	External:Census for $Depriv$ and MCS for MA

Table 3. Table of the weights used in the sensitivity analyses. SA stands for sensitivity analysis. MCS is the Millennium Cohort Study. CPS is Current Population Survey. MA stands for Maternal Age and $Depriv$ stands for Carstairs deprivation score. For details on how to combine MA and $Depriv$ estimates in the H-H study see Web-based supporting Material Section 1.2.

Table 5. H-H study: medians and 95% credible intervals for the posterior distribution of the Bayesian post-stratified MOR for different sensitivity analyses as well as the standard estimate, plus logistic regression estimate and 95% confidence interval for the conditional odds ratio adjusted for the SIV (deprivation and maternal age).

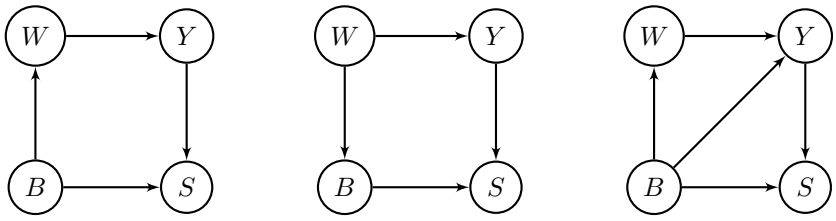
	Bayesian posterior estimates of the MOR	
Scenario	*median*	*95% credible interval*
Standard	2.51	(1.53,4.26)
SA1	2.51	(1.53,4.25)
SA2	2.51	(1.54,4.28)
SA3	2.50	(1.52,4.29)
SA4	2.50	(1.52,4.30)
SA5	2.42	(1.34,4.45)
SA6	2.43	(1.34,4.45)
	Frequentist logistic regression estimate of the conditional OR	
	point est.	*95% confidence interval*
Frq Logist	2.62	(1.51,4.54)



(a) W and Y are marginally independent

(b) W and Y are associated

Figure 1. DAGs demonstrating selection bias. W is the exposure under investigation, Y is the outcome and S the selection indicator. The absence of arrows indicates marginal independence. In DAG (a), W and Y are marginally independent but associated in the study and due to conditioning on S. In DAG (b), W and Y are marginally associated but this association is distorted by S when estimated in the study.

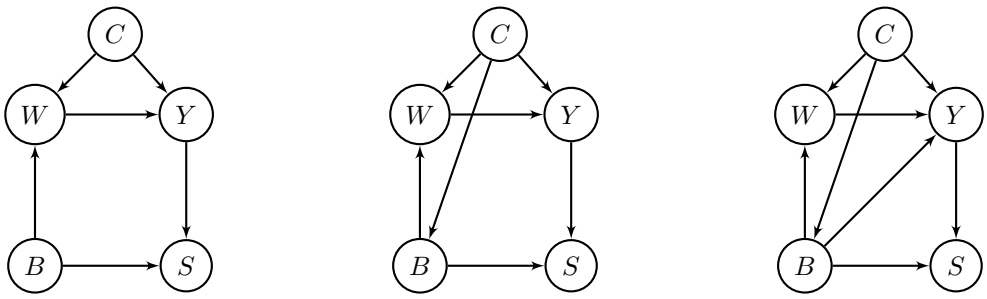


(a) B is a "parent" of W

(b) B is a "child" of W

(c) B is a "parent" of W and Y

Figure 2. These DAGs represent the assumed relationship between the variables in the problem. In the EE-ALL study, W is EL-EMF exposure, Y the case/control status, S is a participation/selection indicator, B is urban status and income. B and Y separate W and S and conditional independence (3) holds in all three DAGs. The main difference between the DAGs is in the interpretation of B: In (a) B can be though of as a influencing both exposure and participation; In (b) B is influenced by the exposure and in turn influences S. In (c) B is also a confounder for the relationship between W and Y.



(a) Including a confounder C

(b) Confounder associated with B

(c) B is also a confounder

Figure 3. In these DAGs, a confounder C is introduced as in conditional independence (5). The three DAGs correspond to different scenarios relating B (the SIV) to C (the confounder).