International Journal of Epidemiology, 2018, 1343–1354
doi: 10.1093/ije/dyy117
Advance Access Publication Date: 22 June 2018
Education Corner

Education Corner

Analysis of multicentre epidemiological studies: contrasting fixed or random effects modelling and meta-analysis

Xavier Basagaña,[1,2,3]* Marie Pedersen,[4,5] Jose Barrera-Gómez,[1,2,3] Ulrike Gehring,[4] Lise Giorgis-Allemand,[5] Gerard Hoek,[4] Massimo Stafoggia,[6,7] Mark J Nieuwenhuijsen,[1,2,3] Bert Brunekreef[4] and Rémy Slama[5]; on behalf of the ESCAPE Birth Outcomes working group

[1]ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain, [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain, [4]Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands, [5]Inserm, CNRS, University Grenoble Alpes, IAB Joint Research Center, Grenoble, La Tronche, France, [6]Department of Epidemiology, Lazio Region Health Service/ASL Roma 1, Rome, Italy and [7]Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden

*Corresponding author. ISGlobal, C/ Doctor Aiguader 88, 08003 Barcelona, Spain. E-mail: xavier.basagana@isglobal.org

Editorial decision 16 May 2018; Accepted 24 May 2018

Abstract

Multicentre studies are common in epidemiological research aiming at identifying disease risk factors. A major advantage of multicentre over single-centre studies is that, by including a larger number of participants, they allow consideration of rare outcomes and exposures. Their multicentric nature introduces some complexities at the step of data analysis, in particular when it comes to controlling for confounding by centre, which is the focus of this tutorial. Commonly, epidemiologists use one of the following options: pooling all centre-specific data and adjusting for centre using fixed effects; adjusting for centre using random effects; or fitting centre-specific models and combining the results in a meta-analysis. Here, we illustrate the similarities of and differences between these three modelling approaches, explain the reasons why they may provide different conclusions and offer advice on which model to choose depending on the characteristics of the study. Two key issues to examine during the analyses are to distinguish within-centre from between-centre associations, and the possible heterogeneity of the effects (of exposure and/or confounders) by centre. A real epidemiological study is used to illustrate a situation in which these various options yield different results. A synthetic dataset and R and Stata codes are provided to reproduce the results.

Key words: Fixed effects, random effects, multicentre study, multilevel analysis, meta-analysis

Key Messages

- Multicentre studies can be analysed in different ways to account for confounding due to differences between centres.
- Using fixed and random effects by centre in analysis of pooled data and meta-analysis of centre-specific analyses may provide different conclusions.
- Differences in within-centre and between-centre associations, and heterogeneity of the effect of confounders, can explain possibly diverging results of various statistical approaches and should be explored during the analysis.

Introduction

Epidemiological studies often include data from different towns, regions, countries or other structures that cluster study participants (e.g. schools or hospitals). Such studies are often called multicentre studies. The reasons for designing a multicentre study include the possibility of recruiting a larger number of participants in a given time, and consequently to: achieve more statistical power; have greater exposure variability; improve the generalizability of the results or at least the possibility to discuss it; or explore the geographical variation of the estimated effects. Such a multicentric design is very frequent in clinical trials,[1] nutritional (cancer) epidemiology,[2] occupational epidemiology,[3] biomarker-based environmental epidemiology[4] and air pollution epidemiology.[5]

Here, we illustrate several ways to analyse the data from a multicentre study when one is interested in the association of an exposure with an outcome, while accounting for the fact that data are collected from different centres. In particular, researchers often want to control for centre to account for the potential effect of unmeasured confounders that vary by centre. Although other techniques are available, epidemiologists often consider three main types of statistical analyses for multicentre studies: adjusting for centre using fixed effects; adjusting for centre using random effects; or analysing each centre separately and then combining the results using meta-analyses. The purpose of this paper is: to provide some guidance on which of these methods to use, depending on the study characteristics; to describe when the different methods can provide different results; and to understand the reasons why this may happen. As an illustration, we use data from an analysis of the association between exposure to ambient air pollution with fine particulate matter ($PM_{2.5}$) during pregnancy at the maternal residence, and birthweight (g), a continuous outcome, in the context of the European Study of Cohorts for Air Pollution Effects (ESCAPE).[5]

Description of models

We assume that researchers have managed to harmonize and pool data from each centre, so that all three modelling options discussed are possible. Otherwise, without pooling the data from different centres, the fixed and random effects models are no longer an option. The analyses that are conducted in a pooled dataset with all centres are sometimes called one-stage individual participant data meta-analysis,[6] and correspond to the fixed effects and random effects models discussed here. In contrast, the sometimes-called two-stage individual participant data meta-analysis,[6] which corresponds to what we simply call meta-analysis in this paper, conducts first separate analyses within each centre and then combines the results in a second step.

Fixed and random effects models

Fixed and random effects models are two different options to adjust for centre in multicentre studies. The underlying model for both of them can be written as:

$$Y_{ij} = \alpha_j + \beta X_{ij} + \gamma_1 C_{1ij} + \cdots + \gamma_k C_{Kij} + \varepsilon_{ij}, \qquad (1)$$

where Y_{ij} and X_{ij} are the continuous outcome and the exposure of interest, respectively, measured for individual i in centre j; $C_{1ij},\ldots, C_{Kij}$ are K potential confounders included in the model (the variable 'centre' is not included in this set); α_j is the centre-specific intercept in centre j; β is the regression coefficient associated with the exposure of interest; $\gamma_1,\ldots, \gamma_K$ are the regression coefficients associated with the confounders; and ε_{ij} are the residuals, which are assumed to be normally distributed with a given variance, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. The centre-specific intercepts, $(\alpha_j)_{j=1\ldots J}$, capture the remaining differences in the response variable by centre after adjusting for the exposure and confounders.[7,8]

Model (1) can be estimated using fixed effects or random effects. If the study includes J centres, a fixed effects model would include indicator variables (often called dummy variables) for centre, i.e. it would fit the model:

$$\begin{aligned} Y_{ij} = {} & \alpha_1 I(\text{centre} = 1) + \alpha_2 I(\text{centre} = 2) + \cdots \\ & + \alpha_J I(\text{centre} = J) + \beta X_{ij} + \gamma_1 C_{1ij} + \cdots \\ & + \gamma_k C_{Kij} + \varepsilon_{ij}, \end{aligned} \qquad (2)$$

where I() designates indicator variables. Such a model can be estimated with any routine for ordinary linear

regression (e.g. *regress* in Stata or *lm()* in R). If the model includes a global intercept, then one of the centres would be taken as the reference centre and only *J*-1 dummy variables would be included.

Model (1) can also be estimated using random effects, in which case one further assumes that α_j follows a normal distribution centred at α with a variance σ_α^2, i.e. $\alpha_j \sim N(\alpha, \sigma_\alpha^2)$. One then needs to use a routine for linear regression with random effects (e.g. *xtmixed* in Stata or *lme()* in R). The model estimated by random effects controls for the effect of centre and returns estimates of α, β, γ_k and σ_α^2. Predictions of the different α_j for each centre are not usually of interest, but one can estimate them afterwards as a function of the estimated parameters.

Meta-analysis

Another strategy is to analyse the data from each centre separately, fitting *J* models allowing the parameters of the model (be they related to the exposure of interest or to adjustment factors) to take different values in each centre. Then, the centre-specific estimates of β (the exposure-health association) are combined using meta-analysis in order to obtain a single common estimate, together with an estimate of between-centre heterogeneity. This approach has been used in several multicentre epidemiological studies.[9–12] Meta-analysis essentially calculates a weighted average of the centre-specific estimates of β, usually using weights that are proportional to the inverse of the variance of the centre-specific estimates (Box 1).[13] Incidentally, one can combine the centre-specific estimates in a meta-analysis using the so-called fixed effects or random effects meta-analyses, depending on heterogeneity of centre-specific effect estimates. To avoid confusion, we restrict the fixed and random effects terminology to the models described in the previous section, and the results that we present for meta-analysis are applicable to any kind of meta-analysis.

Within- and between-centre effects of the exposure

The multilevel structure of the data allows the decomposition of total effects into within-centre effects and between-centre effects. Within-centre effects only use variation in Y and X which occurs within centre to obtain the regression coefficient β. The between-centre effects are obtained by examining the relationship between the centre means of the exposure and outcome. These relationships can be estimated from aggregated (ecological) data, provided the centre intercepts α_j and the centre averages of the variables in the model can be assumed to be uncorrelated.[10,14] Figure 1A represents the two types of effects in a particular hypothetical example in which they strongly differ.

Both the fixed effects model and the meta-analysis approach estimate within-centre effects, even though they do not necessarily provide identical estimates. It is important to note that when estimating within-centre effects, the differences in mean exposure between centres is not used. Many times, multicentre studies are motivated to achieve a wider exposure range. However, one should be aware that if one uses the fixed effects model or the meta-analysis approach, such between-centre variation is not used in the estimation process. This is obvious in the meta-analysis case, in which one would obtain exactly the same estimate for β using the raw data or using an X variable previously centred to a common value in all centres (thereby removing the between-centre differences in X). For the fixed effects

Box 1. Comparison of the meta-analysis approach and the fixed effects model. In meta-analysis, a different model is fitted for each centre, and each covariate can have a different effect in each centre. In the fixed effects model, a single model is fitted. The assumed equation for each centre allows a different intercept (mean average of the outcome) for each centre, but the effect of all the other covariates is assumed to be the same in all centres

Meta-analysis	Fixed effects model
Models: Several models are fitted, one for each centre	Models: A single model is fitted:
	$Y_{ij} = \alpha_j + \beta X_{ij} + \gamma_1 C_{1ij} + \cdots + \gamma_k C_{Kij}$
Implied model for each centre:	Implied model for each centre:
$Y_{i1} = \alpha_1 + \beta_1 X_{i1} + \gamma_{11} C_{1i1} + \cdots + \gamma_{K1} C_{ki1}$	$Y_{i1} = \alpha_1 + \beta X_{i1} + \gamma_1 C_{1i1} + \cdots + \gamma_k C_{Ki1}$
$Y_{i2} = \alpha_2 + \beta_2 X_{i2} + \gamma_{12} C_{1i2} + \cdots + \gamma_{K2} C_{ki2}$	$Y_{i2} = \alpha_2 + \beta X_{i2} + \gamma_1 C_{1i2} + \cdots + \gamma_k C_{Ki2}$
$Y_{i3} = \alpha_3 + \beta_3 X_{i3} + \gamma_{13} C_{1i3} + \cdots + \gamma_{K3} C_{ki3}$	$Y_{i3} = \alpha_3 + \beta X_{i3} + \gamma_1 C_{1i3} + \cdots + \gamma_k C_{Ki3}$
...	...
$Y_{iJ} = \alpha_J + \beta_J X_{iJ} + \gamma_{1J} C_{1iJ} + \cdots + \gamma_{KJ} C_{ki4}$	$Y_{iJ} = \alpha_J + \beta X_{iJ} + \gamma_1 C_{1iJ} + \cdots + \gamma_k C_{KiJ}$
Estimate for β: $\widehat{\beta} = \left(\sum_{j=1}^{J} w_j \widehat{\beta}_j\right) / \left(\sum_{j=1}^{J} w_j\right)$, for some set of weights w_j.	Estimate for β: $\widehat{\beta}$ from the common model.
Total number of parameters estimated:	Total number of parameters estimated:
J*(1 + 1+K)	J+1+K

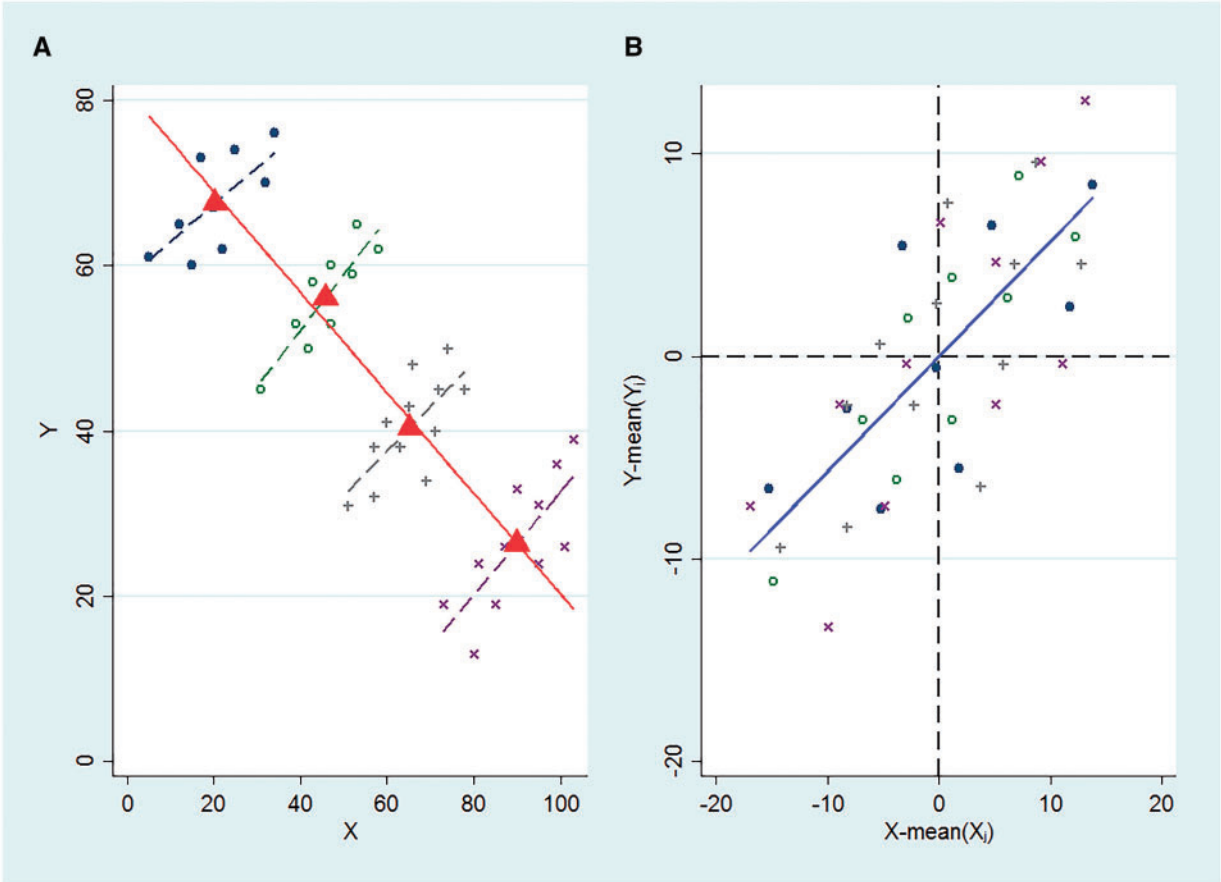

Figure 1. A: Illustration of within-centre and between-centre effects in a hypothetical example. The dots, '+' and 'x' indicate (X, Y) values in the study. The shapes of the points correspond to different centres. Triangles correspond to the means of X and Y in each centre. The solid line indicates the between-centre relationship between X and Y, and the dashed lines indicate the within-centre relationships. In this example, within- and between-centre relationships have opposite signs. B: This panel displays the original (X, Y) values in panel A after subtracting the centre means for both X and Y (denoted as X-mean(X_j) and Y-mean(Y_j) in the graph). As a result, the centre averages of (X-mean(X_j)) and (Y-men(Y_j)) are both zero. The solid line indicates the regression line resulting from fitting a no-intercept model to the data in panel B. The slope of this regression line is algebraically equivalent to the one that would be obtained by fitting a fixed effects model to the data in panel A[11].

model, one can prove that the exact same estimate of β obtained by the fixed effects model can be obtained by fitting instead a no-intercept model between the variables $\left(y_{ij}-\bar{y}_j\right)$ and $\left(x_{ij}-\bar{x}_j\right)$, where $\bar{y}_j$ and $\bar{x}_j$ are the centre averages of Y and X, respectively.[15] This is illustrated in Figure 1B, where one can see that the original between-centre differences in X and Y shown in Figure 1A are removed after the subtraction of the centre averages. It becomes then clear that, if one can obtain the same estimate of β as the fixed effects model using these new data that contain no between-centre differences, the fixed effects model must not make use of the between-centre differences in the estimation process.

The random effects model, unlike the other approaches, makes use of the between-centre variations in exposure. Under a random effects paradigm, i.e. assuming $\alpha_j \sim N(\alpha, \sigma_\alpha^2)$, one can actually fit the following model, which provides separate estimates of the within- and between-centre effects of exposure:[15]

$$Y_{ij} = \alpha_j + \beta_{within}\left(x_{ij}-\bar{x}_j\right) + \beta_{between}\bar{x}_j + \varepsilon_{ij}. \qquad (3)$$

To fit model (3), one needs to manually create the variables $\left(x_{ij}-\bar{x}_j\right)$ and $\bar{x}_j$ and include both of them in the model. Note that, in the specific case in which the within- and between-centre effects are the same, model (3) reduces to model (1). This remark importantly shows that, when one fits model (1) using random effects, one is assuming that the within-centre relationship between Y and X is the same as the between-centre relationship between Y and X, and that these two relationships can be combined into a single estimate of β.

Indeed, the random effects model estimate of β in model (1) is actually a weighted average of the within- and

between-centre effects, $\widehat{\beta}_{\text{random}} = (1-\lambda)\widehat{\beta}_{\text{within}} + \lambda\widehat{\beta}_{\text{between}}$ (note that β_{random} does not refer to a model including a random slope for the exposure variable but to model (1), in which there are only random intercepts).[15] The value of λ depends both on the intraclass correlation of the response and on the slope of the regression of x_{ij} on $\bar{x}_j$. For exposures assessed at the centre level (i.e. without within-centre variation, as occurs in so-called ecological studies), $\lambda = 1$ and, as a consequence, $\widehat{\beta}_{\text{random}}$ estimates a between-centre effect. If all centres have the same mean of exposure, $\lambda = 0$ and, as a consequence, $\widehat{\beta}_{\text{random}}$ estimates a within-centre effect. In the remaining cases, $\widehat{\beta}_{\text{random}}$ estimates a weighted average of the two effects, giving more weight to the between-centre effect as the within-centre variation of exposure decreases. Other authors have argued that it is better to fit the following model instead of model (3):[16]

$$Y_{ij} = \alpha_j + \beta_{within} x_{ij} + \beta_{contextual} \bar{x}_j + \varepsilon_{ij}. \quad (4)$$

Both models (3) and (4) provide identical estimates of β_{within}. The estimation of $\beta_{contextual}$ from model (4) can be interpreted as a contextual effect (i.e. the effect of living in a centre with value $\bar{x}_j$, regardless of the individual value of x_{ij}), whereas $\beta_{between}$ from model (3) is an ecological effect, capturing both individual and contextual effects.[16,17] Note that $\beta_{between} = \beta_{within} + \beta_{contextual}$.

Comparison of methods

Fixed vs random effects

The previous section showed that the fixed effects model estimates within-centre effects. whereas the random effects model estimates a mixture of within-centre and between-centre effects, thus showing a first difference between the two methods. In fact, for centre-level variables, random effects models are the only possibility, as within-centre effects cannot be estimated. Here, we evaluate some other properties of the methods. First, we outline two situations in which the estimates of the association between exposure and health from fixed and random effects models tend to be the same, in which case the choice of one method over the other becomes less important. The similarity/convergence between the two methods occurs: either when (i) the intraclass correlation of the response is close to 1, i.e. most of the variation in response Y is between centres; or when (ii) there is a large number of observations within each centre, in which case within-centre associations convey most of the weight in the random effect estimate; or when (iii) the intraclass correlation of exposure is close to zero, i.e. there is no between-centre variation in exposure, or equivalently all centres have a similar average of exposure.[18]

Two important properties that are usually evaluated when comparing estimators are bias and variance. Bias quantifies the average difference between the true value of the parameter and the estimated quantity over hypothetical repetitions of the study. It is desirable that the bias is zero, in which case the estimator is called unbiased, or as close to zero as possible. Variance quantifies the variability in the estimates we would obtain over hypothetical repetitions of the study. Ideally the variability should also be as small as possible. Since, when comparing two estimators, one can be better in terms of bias and the other in terms of variability, estimators are often compared in terms of mean squared error (MSE), which combines squared bias and variance into a single index by summing them. Estimators with small MSE are preferred.

We first discuss the issue of bias. The fixed effects model provides unbiased estimates of β (provided model (1) is correct), because using dummy variables provides unbiased control of the centre effects.[17] The random effects model will provide biased estimates of β if the centre effects α_j are correlated with some of the variables included in the model.[18] Correlation between α_j and X, for example, likely indicates correlation between an unobserved variable (say Z) and X. This unobserved variable Z will be an omitted confounder, and therefore will induce bias in associations if both Z and X are associated with Y. There would be no correlation between α_j and the variables in the model, and therefore no bias, if there were no differences in the average of the response variable between centres, or if there were no differences in the averages of the exposure and covariates between centres. In the context of observational studies, there will almost always be some correlation between α_j and $\bar{x}_j$, and therefore there is potential for the random effects estimate to be biased. The greater this correlation, the greater the bias. Comparing the fixed effects and the random effects estimates may provide insights on the potential presence of bias in the random effects estimate.

In terms of variance, the fixed effects model can be suboptimal if the study has few observations per centre, in which case the centre effects would be estimated with low precision, or if exposure varies little within centres as compared with between centres.[18] In those cases, the random effects models can work better. The improvement comes from borrowing information between centres; i.e. the estimates for the centre intercepts in the random effects model are not only based on data from that centre, but also on the entire sample, hence providing more stable estimates.

Two recent simulation studies compared the performance of the fixed and random effects models in terms of MSE.[18,19] Those studies only examined scenarios with up to a total of 5000 participants (up to 100 centres, up to

50 individuals per centre). This choice of scenarios does not cover the situation observed in many epidemiological studies, with hundreds of participants per centre. However, their main conclusions are likely to hold as well in such scenarios. In particular, the authors concluded that when studies become large (say, more than 500 observations in total), fixed effects models should be preferred. Random effects models were only preferred mainly in two situations, both with a small sample size (roughly, less than 500 observations in total): (i) the regression model explained a small proportion of the variance of the outcome (roughly, less than 50%); and (ii) the within-centre variation of exposure was small (roughly, less than 20%). At high correlations between α_j and $\bar{x}_j$ (e.g. greater than 0.3), the random effects model became much worse than the fixed effects model, in terms of MSE, except in the two exceptions explained above.

Fixed effects vs meta-analysis

We have mentioned above that both adjusting for centre using a fixed effects model and the meta-analysis approach estimate within-centre effects of exposure. However, when both approaches are applied to the same dataset, they can provide different results, especially in the presence of confounders. The reason is that in the meta-analysis approach many more parameters are used, giving more flexibility to the model, as illustrated in Box 1. In particular, the meta-analysis approach allows the effect of confounders to be different in each centre, whereas this is not the case in the fixed effects model. If such centre by covariate interactions exist and are not included in the fixed effects model, the model is misspecified. If those interaction terms are associated with exposure, they act as omitted confounders and the resulting estimate of the parameter associated with exposure from the misspecified fixed effects model can be biased.

The choice between the meta-analysis approach and the fixed effects model represents a trade-off between bias and variance. Indeed, the meta-analysis approach, using a larger number of parameters, provides better control for confounding and therefore reduces bias at the price of increasing variance (the results may be more unstable). The fixed effects (or the random effects) model provides less control for confounding, which can introduce bias, but results may be less variable. In studies with small numbers of participants per centre, it is preferable to use fixed effects (or random effects) models instead of meta-analysis, as the centre-specific coefficients are estimated with poor precision.

Although this is not often done in practice, one could allow the effects of confounders to be different in each centre in the fixed effects model by including the interactions of centre with all confounders, or only with those whose effects are expected to vary with centre. For example, one could allow for the effect of socioeconomic covariates on health to vary between countries, because of differences between the social systems across countries. This may result in the fixed effects model and the meta-analysis approach providing more similar results. One could also do that with random effects models, by allowing for random slopes (see Supplementary material, available as Supplementary data at *IJE* online, section 3).[20,21]

There are other reasons why the fixed effect model and meta-analysis can differ.[6] For example, in the fixed effects model, the residual variance is often assumed to be the same in each centre, whereas in meta-analysis one estimates a different residual variance in each centre. Another difference between the two methods is that, whereas the fixed effects model accounts for all correlations between variables in the model, the meta-analysis approach does not unless a multivariate meta-analysis is used (i.e. centre-specific coefficients are meta-analysed jointly and not one by one).

Table 1 provides a summary of main characteristics of the models described so far.

Worked example

We use data from an analysis of the association between ambient air pollution exposure during pregnancy (using $PM_{2.5}$ as an air pollution marker) and birthweight, in the context of the European Study of Cohorts for Air Pollution Effects (ESCAPE).[5] The study included data from $J = 16$ European mother-child or birth cohorts. In order to better illustrate the differences between fixed and random effects, one of the largest cohorts included in the original study (MoBA, Norway) was not included in the worked example. The example is only provided for illustrative purposes. Those interested in the effects of ambient air pollution on birthweight should refer to the original publication.[5] The code used in the example, in both Stata and R statistical software, is provided in the Supplementary material, available as Supplementary data at *IJE* online. In addition, we provide a synthetic dataset that provides similar results to the ones presented here and that can be used to run the statistical code. This synthetic dataset is entirely simulated based on the relationships between variables in the original data using the synthpop R package.[22] Even though all data in the dataset were randomly generated and therefore no real individuals were included, we checked that the simulated dataset contained no rows that were identical to rows in the real dataset. This dataset is only useful for educational purposes and no valid scientific conclusions can be derived from it.

Table 1. Advantages and disadvantages of the different models

Method	Advantages	Disadvantages
Fixed effects	Estimates within-centre effects Complete control of confounding by centre Can control for centre by covariate interactions by including them in the model	Cannot estimate between-centre effects (e.g. it cannot be used to estimate the effects of centre-level variables) Between-centre variation in exposure is not used, reducing effective exposure range Can provide less efficient estimates than random effects models in settings with small sample size and: (i) low R^2 of the model; or (ii) small exposure variation within centres
Random effects	Estimates a mixture of within- and between-centre effects, making use of between-centre variation in exposure Simple generalizations allow separation of within-centre and between-centre effects Can provide more efficient estimates than fixed effects models in settings with small sample size and: (i) low R^2 of the model; or (ii) small exposure variation within centres Can control for centre by covariate interactions by including random slopes	Can provide biased estimates when the centre effects are correlated with variables in the model
Meta-analysis	Estimates within-centre effects Provides full control of centre by covariate interactions, preventing potential bias	Cannot estimate between-centre effects (e.g. it cannot be used to estimate the effects of centre-level variables) Between-centre variation in exposure is not used, reducing effective exposure range Estimates can be highly variable in settings with centres with small sample sizes Estimates a large number of parameters

Table 2. ESCAPE birthweight study. Number of participants in each centre. Note that one centre was excluded, compared with the original data

Centre	Country	Number of participants	Percentage
BAMSE, Jarfalla	Sweden	709	1.7
BAMSE, Solna	Sweden	608	1.5
BAMSE, Sundbyberg	Sweden	377	0.9
BAMSE, Stockholm	Sweden	685	1.7
DNBC	Denmark	15 625	38.1
KANC	Lithuania	3873	9.4
ABCD	Netherlands	7124	17.4
GENERATION R	Netherlands	5592	13.6
PIAMA, North	Netherlands	1010	2.5
PIAMA, West	Netherlands	854	2.1
PIAMA, Middle	Netherlands	1345	3.3
DUISBURG	Germany	194	0.5
APREG	Hungary	1095	2.7
GASPII	Italy	643	1.6
INMA, Sabadell	Spain	542	1.3
RHEA	Greece	761	1.9
Total		41 037	100.0

Table 2 summarizes the number of participants in each centre. There was a large heterogeneity in centre sample size, with the smallest centre having 194 participants and the largest having 15 625, with an average size of 2565. Figure 2 illustrates the distribution of birthweight by centre. Small differences, with a Northern Europe to Southern Europe gradient (with higher birthweights in Northern Europe), were observed, but most of the variation in the distribution of birthweight was observed within centres (only 5% of the total birthweight variation was due to between-centre differences). Figure 3 illustrates the distribution of ambient air pollution ($PM_{2.5}$) across centres. In this case, 92% of the total variation was due to variation between centres.

All analyses were adjusted for gestational age, gestational age squared, child sex, maternal age, maternal education, parity, maternal active smoking during the second trimester, maternal pre-pregnancy weight ('broken-stick' model with a knot at 60 kg), maternal height and season of conception; in total, confounders accounted for $k = 15$ parameters. Table 3 shows the results of the different analyses. Focusing on the first three rows, we see that the three main methods described in the paper provided different

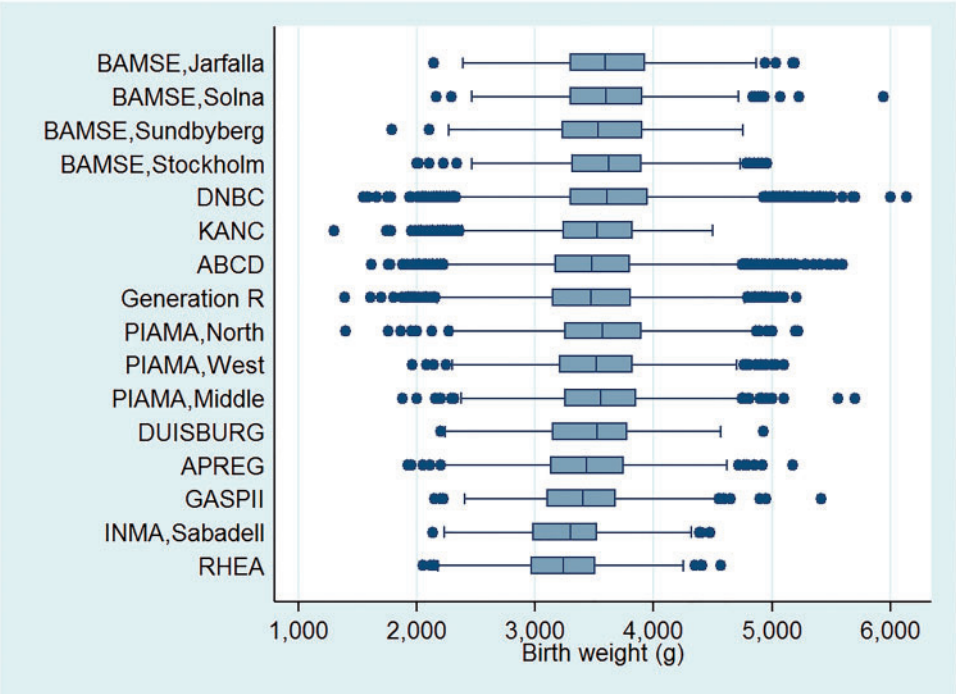

Figure 2. Boxplots of the distribution of birthweight (g) in each centre.

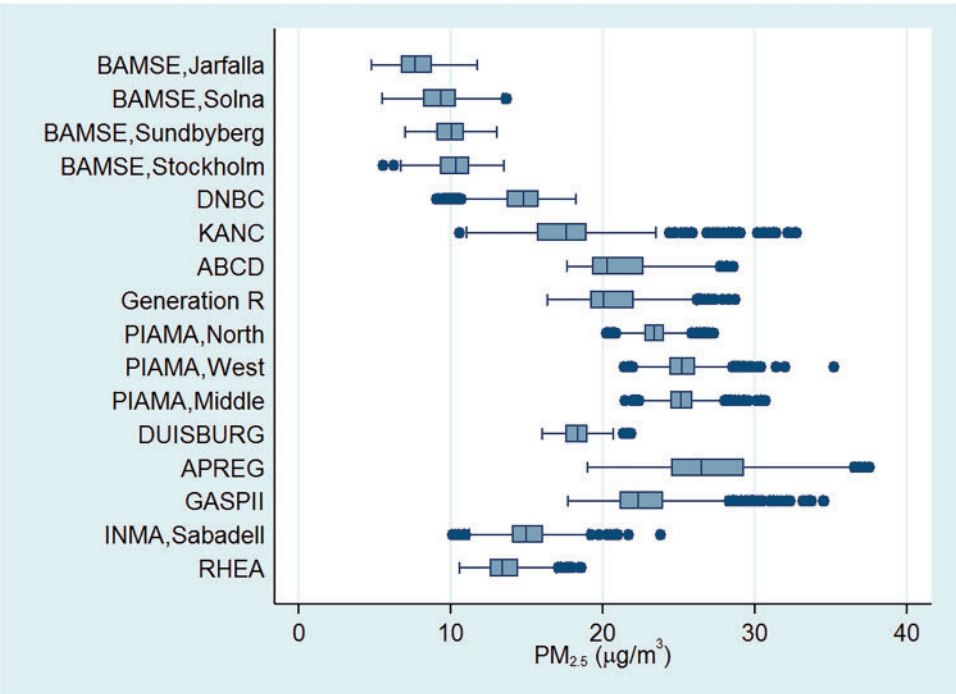

Figure 3. Boxplots of the distribution of $PM_{2.5}$ ($\mu g/m^3$) in each centre.

estimates, leading to different conclusions (e.g. in terms of statistical significance at 5% level). The random effects model estimated a statistically significant 27-g [standard error (s.e.) 10.8] reduction in birthweight associated with a 10-$\mu g/m^3$ increase in $PM_{2.5}$ (P-value = 0.012). The fixed effects model estimated a smaller reduction of 18 g (s.e. 11.5), with a P-value of 0.111. Finally, the meta-analysis estimated a clearly non-significant 6-g (s.e. 12.3) reduction in birthweight (P-value = 0.632). With such results, the researcher is forced to make a decision on which estimates to

Table 3. Estimated effect of a 10-μg/m^3 increase in residential exposure to ambient air pollution with $PM_{2.5}$ during the full pregnancy on offspring birthweight (g) obtained from different analyses

Model	Estimate (95% CI)	s.e.	*P*-value
Fixed effects (pooled)	−18 (−41, 4)	11.5	0.111
Random effects (pooled)	−27 (−48, −6)	10.8	0.012
Meta-analysis	−6 (−30, 18)	12.3	0.632
Random effects (pooled):			
Within-centre (model (2) or model (3))	−19 (−41, 4)	11.6	0.110
Between-centre (model (2), Neuhaus and Kalbfleisch[11])	−85 (−137, −34)	26.4	0.001
Contextual (model (3), Begg and Parides[12])	−67 (−123, −11)	28.8	0.020
Fixed effects (pooled), centre by confounder interactions (if heterogeneity present)	−8 (−34, 16)	12.4	0.512
Fixed effects (pooled), all centre by confounder interactions	−5 (−30, 19)	12.4	0.682

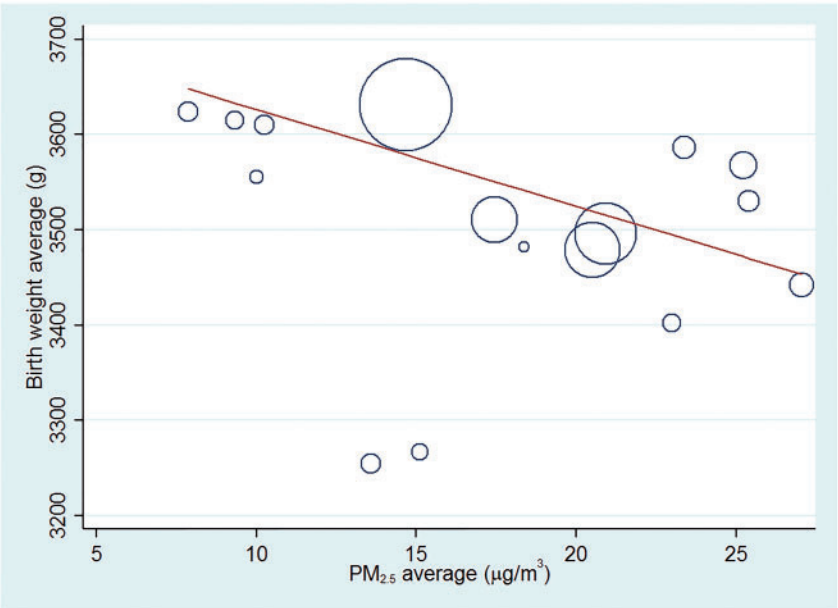

Figure 4. Crude relationship between average birthweight (g) at each centre and average $PM_{2.5}$ (μg/m^3) levels at each centre. Points are proportional to the number of participants in each centre. The solid line is the estimated regression line.

trust. We will use the results described in the first part of the present paper to guide this decision.

The first discrepancy between methods was the difference between the fixed effects and the random effects models. In our example, the total sample size was large, in which case, according to the simulation studies,[18,19] we should favour the fixed effects estimate regardless of other parameters. In addition, the estimated correlation between the estimated centre-specific intercepts α_j from the random effects model and the average $PM_{2.5}$ by centre was high in absolute terms (−0.48), although it had a wide 95% confidence interval (−0.79, 0.02). This large correlation may introduce bias in the random effects estimate which is not compensated by a reduction in variance.[17] All of this suggests that the results from the fixed effects should be trusted more than the results from the random effects model. However, the percentage of within-centre variation in exposure (8%) and the explanatory power of the model (R^2=28%) were low, two conditions that favour the random effects models in small samples.

Another useful way to understand the discrepancy between the fixed and random effects models is to fit model (3), which separates within-centre and between-centre effects. In doing so, we obtained a within-centre estimate that was almost identical to the one obtained in the fixed effects model (Table 3). The between-centre estimate, using the Neuhaus and Kalbfleisch method (model 2),[15] or the contextual estimate, using the Begg and Parides[16] method (model 4), provided much larger estimates (in absolute value), indicating that centres with higher $PM_{2.5}$ levels tended to have lower mean birthweights, even after accounting for the different variables in the model (e.g. maternal height). Figure 4 plots the crude relationship

between $PM_{2.5}$ and birthweight at centre level and illustrates this point. The fact that the within-centre and between-centre effects differ is the reason behind the discrepancy between the random effects model, which estimates an average of the two effects, and the fixed effects model, which estimates solely the within-centre effect. Between-centre effects can provide useful information, but one should keep in mind that information on between-centre effects is of ecological nature, albeit adjusted for individual-level variables. In particular, the between-centre estimate can be biased if a centre-level confounder is omitted, whereas the within-centre estimate is not. For centre-level variables, however, between-centre effects, and consequently random effects models, are the only option.

From the above discussions, the fixed effects model is seen to be likely to provide a better estimate than the random effects model, even though it only makes use of 8% of the variability in exposure. If one wants to make use of the remaining 92%, it is recommended to provide separate estimates of within-centre and between-centre effects when the two differ, instead of a weighted average of the two as provided by the random effects model. Besides, one should interpret between-centre estimates with caution.

The remaining method, meta-analysis, also provides different estimates than does the fixed effects model. In this case, both methods are estimating a within-centre effect (Table 3). As described above, the meta-analysis approach allows for different effects of confounders in each centre, which can explain the discrepancy in results. In fact, when the meta-analysis approach was applied to all terms in the model, significant heterogeneity (P-values <0.05) in the effects of several variables were detected, as shown in Table 4, suggesting that the effects of some of the confounders differed by centre. When we introduced the interactions between centre and all the confounders that had significant heterogeneity in the meta-analysis into the fixed effects model, or when we included all possible interactions between centre and confounders, the results of the fixed effects model became much closer to the results of the meta-analysis (Table 3). The meta-analysis approach, or the fixed effects model that includes all interactions with centre, estimated 272 parameters (17 in each centre), whereas the fixed effects model estimated only 32. Given the size of the dataset, none of these numbers is overwhelmingly high and the results of the meta-analysis should not be discarded on this basis. Even in the smallest centre, the sample size would fit the rule of thumb of having at least 10 participants per variable in the model.[23] Thus, in this particular example, the results from the meta-analysis, or from a fixed effects model that includes interactions of confounders and centre, should be preferred; there is no obvious argument to prefer one over the other, which is not an issue given their similar conclusions.

Table 4. *P*-value for heterogeneity in the meta-analysis of the estimated centre-specific effects

Covariate	*P*-value for between-centre heterogeneity
$PM_{2.5}$	0.297
Gestational age	0.088
Gestational age squared	0.079
Sex	0.192
Maternal age	0.001
Maternal education (middle)	0.266
Maternal education (high)	0.067
Parity: 1	<0.001
Parity: ≥2	<0.001
Maternal smoking	0.006
Maternal weight: lower spline	0.019
Maternal weight: upper spline	0.136
Maternal height	0.314
Season of conception (April-June)	0.033
Season of conception (July-September)	0.213
Season of conception (October-December)	0.061

Further reading

We described some reasons why the different methods used to analyse multicentre studies can provide different estimates, but there are others. Many of them are described in a recent paper, with a strong focus on clinical trials.[6] We did not touch on models with random slopes, which allow variation in the effects of exposure or confounders by centre in a context of random effects models. Interested readers are referred to introductory texts on random effects models.[20,21] In addition, we have restricted the paper to linear models. Many of the issues described apply as well to generalized linear models. However, the generalized linear model carries extra difficulties, such as the difference between conditional and marginal models (e.g. generalized estimating equations; GEE).[24–26] In addition, we did not discuss issues related to the definition of centre, which may not be obvious in some studies, or to whether centres can be assumed to be independent or the data present other levels of nesting.[8,20]

Conclusions

We have discussed three commonly used methods to analyse multicentre epidemiological studies, namely fixed effects models, random effects models, and centre-specific models combined through meta-analysis. Although in some

situations the three approaches may provide similar solutions, in others they may all provide different results. We illustrated some of the reasons why this can happen. Two key issues are: the difference between within-centre and between-centre associations, within-centre associations being often seen as the true parameter of interest;[27] and the heterogeneity of the effects of confounders between centres. We recommend that these two important aspects are always explored in multicentre analyses. Accounting for these factors, along with considerations on sample size, within- and between-centre variability of exposure and outcome, correlation between random effects and variables in the model and variance explained by the model, can allow us to better gauge the conclusions of the analysis.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

No specific funding was received for this work. The ESCAPE study has received funding from the European Union Seventh Framework Program (grant 211250). Marie Pedersen holds a grant from the Danish Council for Independent Research (grant DFF-4004-00179).

Acknowledgements

ESCAPE Birth Outcomes working group: Ferran Ballester, Giulia Cesaroni, Marie-Aline Charles, Leda Chatzi, Asta Danileviciute, Kees de Hoogh, Audrius Dedele, Marisa Estarlich, Ana Fernández-Somoano, Francesco Forastiere, Regina Grazuleviciene, Olena Gruzieva, Barbara Heude, Carmen Iñiguez, Vincent W V Jaddoe, Johanna Lepeule, Aitana Lerchundi, Anne-Marie Nybo Andersen, Bente Oftedal, Daniela Porta, Ole Raaschou-Nielsen, Peter Rudnai, Tamara Schikowski, Per Schwarze, Mette Sørensen, Jordi Sunyer, Goran Pershagen, Manolis Kogevinas.

Conflict of interest: None declared.

References

1. Breedveld FC, Weisman MH, Kavanaugh AF *et al*. The PREMIER study: a multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis Rheum* 2006;**54**:26–37.
2. Riboli E, Hunt K, Slimani N *et al*. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;**5**:1113.
3. Cassidy A, Mannetje A, Tongeren MV *et al*. Occupational exposure to crystalline silica and risk of lung cancer: a multicenter case-control study in Europe. *Epidemiology* 2007;**18**:36–43.
4. Govarts E, Nieuwenhuijsen M, Schoeters G *et al*. Birth weight and prenatal exposure to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE): a meta-analysis within 12 European birth cohorts. *Environ Health Perspect* 2011;**120**: 162–70.
5. Pedersen M, Gehring U, Beelen R *et al*. Elemental constituents of particulate matter and newborn's size in eight European cohorts. *Environ Health Perspect* 2016;**24**:141–50.
6. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med* 2017;**36**:855–75.
7. Localio AR, Berlin JA, Ten Have TR. Confounding due to cluster in multicenter studies—causes and cures. *Health Serv Outcomes Res Methodol* 2002;**3**:195–210.
8. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001;**135**:112–23.
9. Cesaroni G, Forastiere F, Stafoggia M *et al*. Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. *BMJ* 2014;**348**:f7412.
10. Chinn S, Jarvis D, Burney P; European Community Respiratory Health Survey. Relation of bronchial responsiveness to body mass index in the ECRHS. European Community Respiratory Health Survey. *Thorax* 2002;**57**:1028–33.
11. Nagel G, Weinmayr G, Flohr C, Kleiner A, Strachan DP; ISAAC Phase Two Study Group. Association of pertussis and measles infections and immunizations with asthma and allergic sensitization in ISAAC Phase Two. *Pediatr Allergy Immunol* 2012;**23**: 736–46.
12. InterAct Consortium. Dietary fibre and incidence of type 2 diabetes in eight European countries: the EPIC-InterAct Study and a meta-analysis of prospective studies. *Diabetologia* 2015;**58**: 1394–408.
13. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010;**1**:97–111.
14. StataCorp. *Stata Longitudinal Data/Panel Data Reference Manual, Release 14*. College Station, TX: Stata Press, 2015.
15. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998;**54**: 638.
16. Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Stat Med* 2003;**22**:2591–602.
17. Schempf AH, Kaufman JS. Accounting for context in studies of health inequalities: a review and comparison of analytic approaches. *Ann Epidemiol* 2012;**22**:683–90.
18. Clark TS, Linzer DA. Should I use fixed or random effects? *Polit Sci Res Methods* 2015;**3**:399–408.
19. Dieleman JL, Templin T. Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: a simulation study. *PLoS One* 2014;**9**:e110257.
20. Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000;**21**:171–92.
21. Snijders TAB, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd edn. Los Angeles, CA: Sage, 2012.

22. Nowok B, Raab GM, Dibben C. synthpop: Bespoke creation of synthetic data in R. *J Stat Softw* 2016;**74.** doi: http://hdl.handle.net/10.18637/jss.v074.i11.
23. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer, 2001.
24. Betensky RA, Williams PL, Lederman HM. A comparison of models for clustered binary outcomes: analysis of a designed immunology experiment. *J R Stat Soc C* 2001;**50**:43–61.
25. Muff S, Held L, Keller LF. Marginal or conditional regression models for correlated non-normal data? *Methods Ecol Evol* 2016;**7**:1514–24.
26. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat Med* 2009;**28**: 221–39.
27. Sheppard L. Insights on bias and information in group-level studies. *Biostatistics* 2003;**4**:265–78.