Statistics
in Medicine

Research Article

Received 14 August 2012, Accepted 10 June 2013 Published online 16 July 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5894

Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data

Matthieu Resche-Rigon,[a,b,c,d*†] Ian R. White,[a] Jonathan W. Bartlett,[e] Sanne A. E. Peters,[f] Simon G. Thompson[a] and on behalf of the PROG-IMT Study Group[a‡]

A variable is 'systematically missing' if it is missing for all individuals within particular studies in an individual participant data meta-analysis. When a systematically missing variable is a potential confounder in observational epidemiology, standard methods either fail to adjust the exposure–disease association for the potential confounder or exclude studies where it is missing. We propose a new approach to adjust for systematically missing confounders based on multiple imputation by chained equations. Systematically missing data are imputed via multilevel regression models that allow for heterogeneity between studies. A simulation study compares various choices of imputation model. An illustration is given using data from eight studies estimating the association between carotid intima media thickness and subsequent risk of cardiovascular events. Results are compared with standard methods and also with an extension of a published method that exploits the relationship between fully adjusted and partially adjusted estimated effects through a multivariate random effects meta-analysis model. We conclude that multiple imputation provides a practicable approach that can handle arbitrary patterns of systematic missingness. Bias is reduced by including sufficient between-study random effects in the imputation model.

Keywords: missing data; IPD meta-analysis; multilevel model; multiple imputation: chained equations

1. Introduction

Meta-analysis of individual participant data (IPD) is increasingly used in both randomised trials and observational studies [1]. Although it is costly to implement, the use of IPD is classically viewed as the *gold standard* in systematic reviews [2]. It gives results that are less biased and less sensitive to heterogeneity between studies in reporting of results, variable definitions, inclusion criteria and analysis models than meta-analysis of published data or aggregate study-level data [1, 3]. IPD meta-analysis has the advantage of facilitating consistent definitions of outcomes, exposures and confounders and consistent analyses between studies. Of particular relevance in this paper, it enables adjustment for

[a]*MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.*
[b]*DBIM, Hôpital Saint-Louis, APHP, Paris, France*
[c]*Université Paris Diderot, Paris, France*
[d]*Inserm UMRS 717, Paris, France*
[e]*Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K.*
[f]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, PO Box 85500, 3508 GA Utrecht, The Netherlands.*
**Correspondence to: Matthieu Resche-Rigon, DBIM, Hôpital Saint-Louis, 1 avenue Claude Vellefaux, 75010 Paris, France.*
[†]*E-mail: matthieu.resche-rigon@paris7.jussieu.fr*
[‡]*Members of the PROG-IMT Collaboration are listed in the Supporting information.*

all available potential confounding factors, whereas literature-based meta-analysis must rely on the confounder adjustments used in each of the papers.

Statistical methods for IPD meta-analysis must recognise the hierarchical structure of observations within studies. Two general approaches have been described. The first approach, a one-stage analysis, combines all the data from all studies in a single analysis [2]. A simple implementation would introduce a fixed effect of study in a regression model or (in a Cox regression model) stratify by study with a common exposure–disease association, but accounting for possible heterogeneity in exposure–disease associations between studies would require more complex models using exposure by study interactions or random effects [4–7]. The second approach, a two-stage analysis, first analyses each study separately using the same model. Then, it applies standard meta-analysis methods using the summary statistics obtained at the first stage. As for the one-stage approach, univariate or multivariate random effects models are applied that allow for both within-study and between-study variation [8].

The availability of confounders typically varies between studies. We define a variable to be 'systematically missing' if it is missing for all individuals within particular studies in an IPD meta-analysis. All of the following discussion applies equally to a multicentre study, where a variable is systematically missing if it is missing for all individuals in particular centres. When a systematically missing variable is a potential confounder, two approaches are commonly used. A complete study analysis includes only studies with no systematically missing confounders: this loses information and hence precision, and is potentially biased if the systematically missing data are not missing completely at random (MCAR).Alternatively, a complete confounders analysis includes all the studies but adjusts only for the confounders available in all the studies and therefore remains potentially confounded [9]. We aim to develop a method that uses the information from all studies and adjusts for all the confounders, whatever the pattern of systematic missingness. Our approach uses multiple imputation (MI).

Multiple imputation is an attractive and effective approach for statistical analysis of incomplete data [10–12]. The main idea is to create multiple data sets that reflect the potential values of the missing data. More precisely, random draws are made from the posterior distribution of the missing values given the observed data, usually under the missing at random (MAR) assumption [13]. Estimates are combined across imputed data sets using Rubin's rules [10]. Various univariate and multivariate imputation models have been proposed, some of which rely on MCMC methods to draw from the required posterior distribution [13, 14]. When missing values occur in multiple variables, and in particular when these are a mixture of continuous and discrete variables, the method of multiple imputation by chained equations (MICE) is particularly attractive [15]. This involves specifying a separate imputation model for each incomplete variable given all the other variables and repeatedly imputing the variables in an iterated sequence [16]. As with MI in general, it is crucial that the imputation model is consistent (or congenial) with the model of interest, which will subsequently be fitted to the imputed data sets [17].

Multiple imputation methods have been proposed for multilevel data structures in the case of 'sporadically missing' covariates – that is, incomplete covariates that are observed for at least some individuals in each cluster (i.e. study in the context of meta-analysis). Arguably, the simplest approach is to impute separately within each cluster. An alternative strategy is to impute the missing data of all the clusters in a single process. The key point of both approaches is that the imputation model should allow for the hierarchical structure of the data [18, 19]. In 2002, Schafer and Yucel proposed a Gibbs sampler to generate MIs of continuous missing data from a joint multivariate linear mixed model [20]. A similar approach was proposed in an MLwiN macro by Carpenter and Goldstein [21]. More recently, Yucel modified this approach to allow imputation of missing values at any level of the multilevel structure [22]. In 2010, van Buuren proposed to impute data on a variable-by-variable basis using a MICE approach [18]. This involves specifying a random effects imputation model for each variable and using a Gibbs sampler to generate imputations.

None of these methods specifically deals with systematically missing covariates in IPD meta-analysis. MI within studies cannot handle systematically missing data, because in those studies with systematically missing variables, there is no information on which to base imputation. Therefore, we must use an imputation model that imputes missing data in all studies jointly. We could simply combine the studies and treat the data as one large study, but this would (incorrectly) ignore the dependence between subjects from the same study and also make the often unreasonable assumption that the associations between variables are homogeneous across studies. In this paper, we develop an MI method based on multilevel regression models adapted for systematically missing data in continuous variables. Importantly, our proposal allows for between-study heterogeneity not only in the intercept of the imputation models but also in the associations between variables.

The paper is organised as follows. Section 2 presents our motivating example, an IPD meta-analysis of eight studies assessing the impact of intima media thickness (IMT) on vascular risk. In Section 3, we propose our MICE approach and evaluate its performance using a small simulation study in Section 4. Section 5 illustrates these methods using the IMT data. Finally, Section 6 provides a discussion.

2. PROG-IMT data and analysis model

The PROG-IMT Collaboration aimed to explore how progression of carotid IMT predicts vascular risk and to confirm associations between IMT and vascular risk [23, 24]. At the time of our analysis, the Collaboration brought together IPD from nine prospective population-based cohort studies in five countries comprising 27 557 individuals [25–33].

Because our aim is to develop statistical methods, we focus on estimating associations between IMT and vascular risk, allowing for potential confounders. IMT was measured in various ways on two occasions an average of 4 years apart. Our measure of IMT at each occasion is the average of all the available mean common carotid artery measures including left side, right side, near wall and far wall; these are then averaged across the two occasions. The outcome of interest is time to first cardiovascular event, defined as the time to myocardial infarction, stroke or vascular death (or death if cause of death is not available) after the second IMT measure. Death from other causes was regarded as censoring. The median follow-up was 7 years.

Eleven potential confounders are included in the final analysis model. Five continuous variables were measured at the two occasions: body mass index (BMI), systolic blood pressure (SBP), hemoglobin (Hb), serum creatinine (Creat) and total cholesterol (Chol). Each of these variables is included as a covariate as the mean of the two measurements. Four binary variables were measured at the two occasions: treatment for hypertension (AHT), treatment for dyslipidaemia (DTreat), smoking status (Smoke) and diabetes mellitus (Diab). These are coded in the model as 1 if the treatment, smoking or diabetes was reported at either occasion, and 0 otherwise. Lastly, the participant's sex and age in years at the second IMT measurement were included.

The analysis model is a Cox proportional hazards model with the exposure and all the confounders included as linear terms. Restricted cubic spline regressions did not indicate deviation from linearity for any of the continuous variables [34]. Proportional hazards were checked for all variables [35].

This IPD meta-analysis, like many others, was complicated by missing data. We exclude one study for which the exposure was systematically missing [33]. The remaining eight studies have both systematically and sporadically missing data for the confounders. As our purpose is to propose new methods for handling systematically missing data, here we ignore sporadically missing data: to do this, we deleted some data so that, in each study, each variable is either completely observed or completely unobserved. In each study, confounders with more than 5% of observations missing were considered as systematically missing (i.e. the observed values from that study were made missing). If less than 5% of values were missing for a variable in a study, those subjects with the variable missing were excluded from our analyses, so that the variable was then fully observed within the study. The methods that we present could easily be extended to handle both sporadic and systematically missing values simultaneously: modifications to the algorithm are given in the discussion.

The pattern of missing covariates is presented in Table I. The pattern is monotonic, which can simplify the analysis, but because other data sets are unlikely to be monotonic, we do not consider methods that exploit this feature.

3. Multiple imputation by chained equations using random effects models

The MICE is a flexible and practical approach for handling missing data. An extension of the standard MICE procedure, which handles sporadically missing values, was recently proposed by van Buuren [18]. This method uses a linear random effects imputation model and applies a Gibbs sampler, notably for the random effect, to generate MIs from the posterior distribution of the missing data. In this section, we propose an approach to multiply impute systematically missing continuous confounders using multilevel linear imputation models. Because imputation procedures should be congenial to the analysis model [36], we start by specifying the analysis model that will be applied to the imputed data.

Table I. Pattern of the missing covariates in the PROG-IMT study.

Study	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6	Study 7	Study 8
Original size	12140	3313	3120	2198	2891	3992	255	649
Analysed size	11645	2917	2865	2163	2705	3777	237	619
Hb	✓	✓	✓					
Creat	✓	✓	✓					
Chol	✓	✓	✓	✓	✓	✓		
BMI	✓	✓	✓	✓	✓	✓		
SBP	✓	✓	✓	✓	✓	✓	✓	
AHT	✓	✓	✓	✓	✓	✓	✓	✓
Smoke	✓	✓	✓	✓	✓	✓	✓	✓
Diab	✓	✓	✓	✓	✓	✓	✓	✓
DTreat	✓	✓	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓	✓	✓
Sex	✓	✓	✓	✓	✓	✓	✓	✓
IMT	✓	✓	✓	✓	✓	✓	✓	✓

A tick denotes that the variable was measured in the corresponding study. Analysed sizes correspond to the size of the studies after exclusion of observations with sporadically missing data.

The analysis model considered is a proportional hazards model with a random effect on the intercept and on some of the covariates [7, 34, 37]:

$$h_{ij}(t) = h_0(t) \exp\left(\sum_{r=1}^{k} \beta_r x_{ijr} + \sum_{r=1}^{l} u_{ir} x_{ijr} + u_{i0} \right) \tag{1}$$

where $h_{ij}(t)$ denotes the hazard function at time t for subject $j = 1, \ldots, n_i$ in study $i = 1, \ldots, n$. There are k column vectors of covariates $(x_1, \ldots, x_k)$, some of which may be systematically missing, and which are ordered such that the first $l \leqslant k$ are modelled as random effects. β_r represents the log hazard ratio (HR) associated with covariate X_r. $u_i = (u_{i0}, u_{i1}, \ldots, u_{il})$ is the l-vector of random effects from the ith study, with $u_i \sim MVN(0, \tau)$ and τ is an $l \times l$ covariance matrix.

3.1. *Imputation model*

The MICE approach proceeds by imputing each variable in turn, in each case using the latest imputed values of the other variables. We therefore consider the task of imputing the systematically missing values of one incomplete variable, say, the rth covariate (which we assume is continuous) x_{ijr}.

We anticipate heterogeneity between studies both in means of variables and in associations between variables, and so we define each imputation model as a linear mixed effects regression model, allowing both for random intercepts and for random effects on covariates. The regression coefficients used to impute systematically missing values will therefore be drawn from a distribution estimated from the coefficients in other studies.

Let x_{ir} be the $n_i \times 1$ vector of values of X_r on individuals $j = 1, \ldots, n_i$ in study i. If study i has covariate x_r observed, x_{ir} is observed; otherwise, it is completely missing. We assume that

$$x_{ir} = W_{ir}\gamma_r + Z_{ir}b_{ir} + e_{ir} \tag{2}$$

where W_{ir} is an $n_i \times p$ design matrix corresponding to the fixed effects of p covariates included in the imputation model, typically including the other covariates x_{is} $(s \neq r)$ and a function of the outcome, as discussed earlier; γ_r is a $p \times 1$ vector of fixed-effects parameters; Z_{ir} is an $n_i \times q$ design matrix, which represents q covariates, usually a subset of those in W_{ir}, on which a random effect is modelled; $b_{ir} \sim N(0, \Psi_r)$ is a $q \times 1$ vector of random effects; Ψ_r is the $q \times q$ variance–covariance matrix of the random effects; and $e_{ir} \sim N(0, \sigma_r I(n_i))$ is an $n_i \times 1$ vector of residuals, where $I(.)$ is the identity matrix and σ_r is a residual variance parameter.

Under standard noninformative priors, the posterior distribution of the parameters of a linear mixed model are not available in closed form. We therefore use a large sample approximation to them. Then, steps of the (proper) imputation procedure are as follows:

1. Obtain maximum likelihood (or restricted) maximum likelihood estimates of the imputation model parameters $\left(\hat{\gamma}_r, \hat{\Psi}_r, \hat{\sigma}_r\right)$ using studies where covariate x_r is observed.
2. Draw γ_r^* from $N\left(\hat{\gamma}_r, \text{Var}\left(\hat{\gamma}_r|\hat{\Psi}_r, \hat{\sigma}_r\right)\right)$ and $\left(\Psi_r^*, \sigma_r^*\right)$ from $N\left(\left(\hat{\Psi}_r, \hat{\sigma}_r\right), \text{Var}\left(\hat{\Psi}_r, \hat{\sigma}_r\right)\right)$.
3. For studies i with systematically missing data, draw b_{ir}^* from $N\left(0, \Psi_r^*\right)$.
4. For studies i with systematically missing data, draw x_{ijr}^* by drawing $e_{ijr}^* \sim N\left(0, \sigma_r^*\right)$ and

$$x_{ijr}^* = W_{ir}\gamma_r^* + Z_{ir}b_{ir}^* + e_{ijr}^* \tag{3}$$

Our method requires that data are MAR. In the context of systematically missing data, this means that whether data on X_r are systematically missing in a study cannot depend on the values of X_r in that study, conditional on the data observed in that study: in particular, whether data on X_r are systematically missing cannot depend on the mean of X_r or on the associations between X_r and other variables. This assumption seems plausible in general, but it would not be true if X_r was unmeasured because it was not an important confounder in a particular study.

3.1.1. Choice of W. For valid imputation, the matrix W_{ir} should include all the covariates and outcome of the analysis model (except of course x_{ir}) [14, 16]. In particular, when imputing missing values of analysis model covariates, the analysis model outcome must be in the imputation model [38]. One of the aims of MI is to infer plausible values for the missing data from the distribution of the observed data. This implies using the information in the outcome to fully describe the observed values. Random components are incorporated into these estimated values to reflect their uncertainty. When the analysis model is a Cox model, it is not obvious how to include the outcome: we let W_{ir} include the event indicator D and the cumulative marginal hazard $H(T)$, where T is the time to event or censoring [39]. The performance of this approach (in terms of bias and coverage) was satisfactory on the basis of a recent simulation study [40]. The cumulative marginal hazard function can be estimated by the Nelson–Aalen estimator [39]: this can be carried out either globally or for each study separately.

3.1.2. Choice of Z. We consider a range of choices of Z that allow for different aspects of heterogeneity between studies:

- Model 0: no random effects (Z empty), but a fixed study effect included in W.
- Model 1: a random intercept (Z contains a column of ones).
- Model 2: Model 1 plus a random effect on (some of) the x_{is} ($s \neq r$).
- Model 3: Model 1 plus a random effect on the event indicator D and/or the cumulative marginal hazard $H(T)$.
- Model 4: Models 2 and 3 combined.

In models 1–4, the mean of X_r varies randomly between studies. In models 2–4, associations between X_r and other variables also vary randomly between studies. Method 0 was added to compare our imputation procedure to a simplistic approach that avoids modelling between-study heterogeneity in associations between variables [41]. In this case, estimates of the imputation model depend on the choice of reference category for the cohort variable. Because systematically missing variables will be imputed as if the missing cohorts are like the reference study, draws of systematically missing values will be similar in mean to that in the reference study. This should underestimate the between-study variance in the analysis model.

3.2. Implementation

In each case, linear mixed models were fitted using the *lme* function of the *nlme* R package using REML estimation [42]. The *lme* function transforms Ψ to the logarithm of the standard deviation and Fisher's z transformation of the correlations for estimation. (Ψ^*, σ^*) was drawn on the transformed scale to improve the normality assumption. In the case of a nonpositive-definite $\text{Var}(\hat{\Psi}, \hat{\sigma})$, an improper imputation was realised by drawing b_i^* using $N(0, \hat{\Psi})$ and $e_{ij}^* \sim N(0, \hat{\sigma}^2)$. The percentages of improper imputation were lower than 5% for model 1, around 20 % for models 2 and 3 and around 50% for model 4 in our example with eight cohorts.

The overall imputation procedure was implemented using the *MICE* R package [43], with the imputation models for all incomplete covariates having the same form (e.g. all containing random effects

for the intercept, the exposure or the event indicator and using global Nelson–Aalen estimators). Ten cycles were performed for each chain.

The imputed data sets were analysed by fitting model (1) using the *coxme* R package [44]. Parameter estimates were pooled using Rubin's rules [10].

4. Monte Carlo simulation study

Although methods for MI by chained equations have been proposed for multilevel data structures and validated with simulation studies, none of them deal specifically with systematically missing data [22]. So Monte Carlo simulations were used to evaluate the performance of the MICE-RE methods with systematically missing data. We generated N independent data sets in several settings defined by the number of studies and size of the between-studies variance and analysed the data by the different MICE-RE methods and by a MICE approach using a fixed effect of studies.

4.1. Data-generating process

Data were generated according to an exponential distribution for the survival time t_{ij}:

$$t_{ij} \sim \text{Exp}(\lambda_t \exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + u_{i1} X_{ij1} + u_{i0})) \tag{4}$$

with $\begin{pmatrix} u_{i1} \\ u_{i0} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_0 \\ \rho\tau_1\tau_0 & \tau_0^2 \end{pmatrix}\right)$ and t_{ij}, X_{ij1} and X_{ij2} respectively the time to event, the exposure of interest and a confounder for individual j in study i. β_1 and β_2 represent the regression coefficients ($\log(HR)$) of the exposure and the confounder, respectively. Parameter values were suggested by fitting model (1) to the complete studies of the IMT data with IMT as exposure and (age/100) as confounder, giving $\tau_1 = 0.5$, $\tau_0 = 0.5$ and $\rho = -0.7$. We set $\beta_1 = 1$, $\beta_2 = 3$ and $\lambda_t = 1$. Noninformative censoring times were generated from an exponential distribution with rate $\lambda_c = 13.5$ to give approximately 75% censoring (as in the example). The decision not to include a random effect on X_2 was made to keep our model as simple as possible.

X_{ij2} was simulated from a univariate random effects model $X_{ij2} \sim N\left(\mu_{2i}, \sigma_2^2\right)$, with $\mu_{2i} \sim N\left(\mu_2, \tau_2^2\right)$, where μ_2 is the overall mean, τ_2^2 the between-study variance and σ_2^2 the within-study variance. X_{ij1} was simulated from the random effect model $X_{ij1} = \mu_1 + B_2 X_{ij2} + b_{i1} + b_{i2} X_{ij2} + e_{ij1}$, with $\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \sim N\left(0, \Psi = \begin{pmatrix} \psi_1^2 & \rho'\psi_1\psi_2 \\ \rho'\psi_1\psi_2 & \psi_2^2 \end{pmatrix}\right)$ and $e_{ij} \sim N\left(0, \sigma_1'\right)$. We fixed $\mu_2 = 0.63$, $\tau_2 = 0.08$, $\sigma_2 = 0.08$, $\mu_1 = -0.9$, $B_2 = 0.97$, $\psi_1 = 0.18$, $\psi_2 = 0.24$, $\rho' = -0.9$ and $\sigma_1' = 0.15$. Half the studies were chosen using an MCAR mechanism to have systematically missing data; in each of these studies, either X_1 or X_2 was chosen with equal probabilities to be systematically missing. This is the simplest nonmonotone missing data pattern, and so MICE is an appropriate procedure.

The total number of patients was fixed at 3200. The number of studies was 8 (setting A) and 32 (setting B) yielding 400 and 100 patients per study. Setting C had eight studies, and all the between-study variances and covariances were increased by a factor of 4. A total of $N = 2500$ independent data sets were generated for each setting.

4.2. Analysis methods

The data were first analysed before data deletion as a benchmark for the MI procedure. The incomplete data were then imputed $m = 5$ times using each of the imputation models 0–4 described in Section 3. Model 0 was considered as a reference model, illustrating a simple approach with a fixed study effect. Models 1–4 were designed to evaluate the impact of a random intercept and random effects on the exposure (IMT), the outcome, neither or both.

Specifically, in models 1–4, when imputing x_1, the fixed effects W_{i1} comprised X_2, $H(T)$ and D, and when imputing x_2, the fixed effects W_{i2} comprised X_1, $H(T)$ and D. In model 0, the fixed effects also included study. Models 1–4 included a random intercept for study. In models 2 and 4, random coefficients were also placed on the covariate not being imputed. In models 3 and 4, random coefficients were also placed on D but not on $H(T)$, because with low cumulative event rates, most of the event information is carried by D. Moreover, for a particular patient, $H(T)$ does not depend on covariates, and small errors in estimating $H(T)$ are known to have little impact on the imputations [39]. $H(T)$ was estimated globally using the overall data set.

The analysis model considered was a Cox model with the same structure as the data-generating model (Equation (4)):

$$h_{ij}(t) = h_0(t)\exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + u_{i1}X_{ij1} + u_{i0}). \quad (5)$$

Estimates of parameters β_1 and β_2 were pooled using Rubin's rules [10]. Confidence intervals were calculated using the t-distribution and degrees of freedom as defined by Rubin [10].

In each data-generating setting, the performance of each method was assessed by computing the empirical mean of the parameter estimates, relative bias (Rbias), root mean square of estimated standard errors ($Se\beta$ Calc.), empirical Monte Carlo standard errors ($Se\beta$ Emp.), the coverage of nominal 95% confidence intervals (95%CI) for the $\log(HR)$ and percentages of successful runs of the MICE-RE procedure from the 2500 replications of the data set. When observed coverage is 95% from 2500 simulations, a 95%CI for true coverage is 94.1–95.9%.

4.3. Results

Results of the simulation study are given in Table II. Some bias was found in $\hat{\beta}_1$ without systematically missing data: this level of bias will be considered as our gold standard for estimates obtained with systematically missing data using MICE-RE models 0–4.

As expected, biases were greater using model 0 than using any other model under all simulation settings. Indeed, imputations with model 0 depend on the choice of reference category for the cohort variable, because systematically missing variables will be imputed as if the missing cohorts are like the reference cohort. This will also underestimate the between-study variance estimates of the final analysis. These biases seem to increase with between-study heterogeneity (setting C). Moreover, the estimated standard error was biased downwards compared with the empirical standard errors, resulting in

Table II. Simulation results for β_1 and β_2 estimates; full random effect analysis model (Equation (5)).

	$\hat{\beta}_1$						$\hat{\beta}_2$					
			Se						Se			Succ.
	Mean	Rbias	Calc.	Emp.	Cover	$\hat{\tau_1}$	Mean	Rbias	Calc.	Emp.	Cover	(%)
Setting A ($\beta_1 = 1, \tau_1 = 0.5, \beta_2 = 3$, 8 studies)												
No Missing	0.987	−0.013	0.293	0.299	0.954	0.513	3.006	0.002	0.480	0.487	0.972	100.0
Imp. Model 0	0.944	−0.056	0.336	0.410	0.882	0.477	2.864	−0.045	0.585	0.790	0.855	100.0
Imp. Model 1	0.975	−0.025	0.336	0.349	0.928	0.453	2.909	−0.030	0.607	0.607	0.946	100.0
Imp. Model 2	0.978	−0.022	0.336	0.350	0.928	0.455	2.916	−0.028	0.607	0.600	0.953	99.9
Imp. Model 3	0.983	−0.017	0.363	0.354	0.946	0.481	2.906	−0.031	0.645	0.612	0.960	99.6
Imp. Model 4	0.991	−0.009	0.372	0.357	0.943	0.494	2.901	−0.033	0.662	0.608	0.961	75.7
Setting B ($\beta_1 = 1, \tau_1 = 0.5, \beta_2 = 3$, 32 studies)												
No Missing	0.978	−0.022	0.248	0.251	0.971	0.577	3.002	0.001	0.466	0.469	0.975	100.0
Imp. Model 0	0.920	−0.080	0.300	0.331	0.920	0.524	2.832	−0.056	0.566	0.636	0.908	100.0
Imp. Model 1	0.976	−0.024	0.300	0.296	0.953	0.526	2.881	−0.040	0.582	0.576	0.950	100.0
Imp. Model 2	0.974	−0.026	0.301	0.299	0.948	0.526	2.882	−0.039	0.583	0.574	0.958	99.8
Imp. Model 3	0.975	−0.025	0.310	0.298	0.958	0.551	2.886	−0.038	0.598	0.574	0.953	98.9
Imp. Model 4	0.982	−0.018	0.310	0.300	0.950	0.556	2.872	−0.043	0.597	0.579	0.957	75.6
Setting C ($\beta_1 = 1, \tau_1 = 1, \beta_2 = 3$, 8 studies)												
No Missing	0.981	−0.019	0.403	0.447	0.948	0.935	2.998	−0.001	0.456	0.474	0.958	100.0
Imp. Model 0	0.880	−0.120	0.420	0.665	0.770	0.877	2.550	−0.150	0.522	1.216	0.614	100.0
Imp. Model 1	0.932	−0.068	0.407	0.480	0.879	0.795	2.770	−0.077	0.578	0.655	0.904	100.0
Imp. Model 2	0.936	−0.064	0.414	0.478	0.883	0.799	2.784	−0.072	0.591	0.651	0.903	100.0
Imp. Model 3	0.951	−0.049	0.461	0.493	0.900	0.839	2.752	−0.083	0.649	0.666	0.915	99.6
Imp. Model 4	0.960	−0.040	0.463	0.491	0.909	0.844	2.761	−0.080	0.663	0.662	0.928	92.3

Means, relative bias (Rbias), root mean square of estimated standard errors (Se Calc.), observed standard errors (Se Emp.), the coverage of the estimated confidence interval (Cover), root mean square of between-study standard deviations (τ_1) and the percentage of the procedure success (Succ) are given.

a coverage of less than 90% in setting A and less than 80% in the high between-study heterogeneity setting C. Increasing the number of studies improved the coverage (setting B).

Concerning $\hat{\beta}_1$. Point estimates obtained with MICE-RE are generally slightly below those obtained without any systematically missing data, especially in setting C. For models 1–4, the relative biases of the MICE-RE procedures range from 0.9% to 2.6% in settings A and B and from 4.0% to 6.8% with greater between-study heterogeneity in setting C. The relative biases generally decrease with the number of random effects and the presence of a random effect on the event indicator. Model 4 is always the least biased. In settings A and B, the calculated standard errors are smaller with models 1 and 2 than in models 3 and 4. Although models 3 and 4 overestimate standard errors compared with the empirical standard errors, they achieve close to 95% coverage. In setting C, models 3 and 4 underestimate the empirical standard errors, and coverage is about 90%.

Concerning $\hat{\beta}_2$. Point estimates obtained with MICE-RE are again always below those obtained without any missing confounders. There is no clear evidence of a link between the bias and the number of random effects in the model or the type of random effects. Standards errors are still overestimated by models 3 and 4 in settings A and B. Coverage is about 95% in settings A and B but nearer 90% in setting C.

The introduction of multiple random effects seems to improve the quality of the estimates of β_1 for all settings, but this is not the case for β_2. With few studies and many random effects, standard errors seem to be overestimated, perhaps because of overparameterisation of the imputation model. Moreover, the most complex model fails for around 25% of the simulated data sets in settings A and B, because of the difficulties of an iterative fit of *nmle* routines in the presence of low between-study heterogeneity.

Finally, simulations showed that the use of a random effects model to impute systematically missing confounders performed better than the other methods explored. A large improvement is already obtained with a simple random intercept. Although a higher number of random effects could slightly improve the estimates, we should keep in mind that the risk of failure of the imputation process increases.

5. Application to the PROG-IMT data

5.1. Standard approaches

We first estimated the unadjusted effect of IMT (corresponding to X_1 in our simulation study) on the risk of vascular death in each study. Results are displayed in Figure 1. Estimated log HRs per 1 mm increase in IMT range from 1.96 to 3.62, and all except study 8 give significant evidence of an association between IMT and the risk of vascular death. One-stage IPD meta-analysis using model (1) with IMT as the only

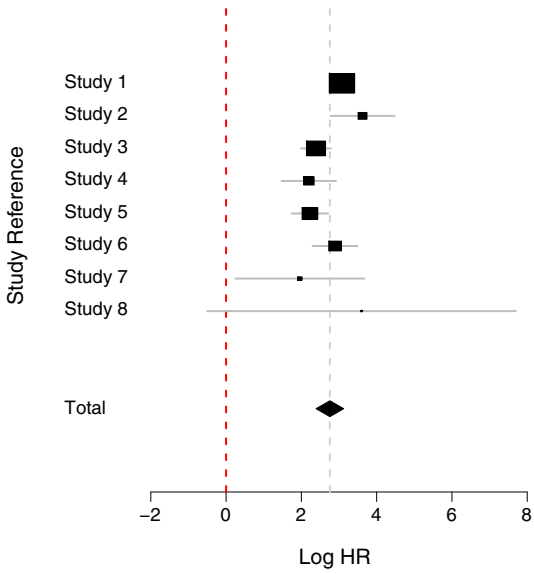

Figure 1. Unadjusted estimates of the log(hazard ratio) for vascular risk per mm increase in IMT, by study, with 95%CI.

covariate gives an estimate $\hat{\beta}_1 = 2.757$ ($Se = 0.187$), and the standard deviation of the random effect is $\hat{\tau}_1 = 0.423$ (Table IV).

Adjustment for confounders changes these results substantially. Figure 2 displays results obtained using the 'complete confounders' analysis, which adjusts only for the confounders available in all studies (Table III). The estimated log HRs range from 0.57 to 2.21. One-stage estimates are $\hat{\beta}_1 = 1.270$ ($Se = 0.155$) and $\hat{\tau}_1 = 0.245$ (Table IV). A 'complete studies' analysis, which includes all the desired confounders but only three studies, gives estimates $\hat{\beta}_1 = 1.212$ ($Se = 0.140$) and $\hat{\tau}_1 = 0.041$ (Table IV). Intermediate between complete confounders and complete studies approaches is a partially adjusted model, which adjusts for the covariates except Hb and Creat (Table I) and can therefore use six studies without systematically missing data. Estimates are $\hat{\beta}_1 = 0.997$ ($Se = 0.160$) and $\hat{\tau}_1 = 0.243$ (Tables III and IV).

These results confirm the impact of confounders on the estimation of the IMT effect and also that improved strategies to deal with missing data could importantly change results. Only the complete studies analysis directly estimates the quantity of interest – the impact of IMT adjusted for the complete set of confounders – but with this approach, the number of studies is low and the random effects variance seems to be underestimated.

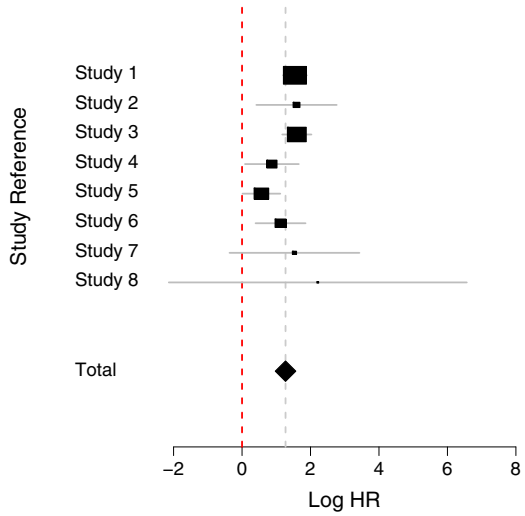

Figure 2. Partially adjusted estimates of the log(hazard ratio) for vascular risk per mm increase in IMT, by study, with 95%CI.

Table III. Estimates of the log(hazard ratio) for vascular risk per mm increase in IMT.

	Confounders adjusted for					
	All[1]		Partly complete[2]		Complete[3]	
Study	$\hat{\beta}_1^f$	(Se)	$\hat{\beta}_1^{p2}$	(Se)	$\hat{\beta}_1^{p1}$	(Se)
1	1.11	(0.18)	1.17	(0.18)	1.54	(0.18)
2	1.35	(0.63)	1.31	(0.63)	1.59	(0.60)
3	1.40	(0.22)	1.40	(0.22)	1.60	(0.22)
4			0.82	(0.41)	0.87	(0.40)
5			0.58	(0.29)	0.57	(0.28)
6			0.51	(0.38)	1.13	(0.37)
7					1.53	(0.97)
8					2.21	(2.22)

[1]All confounders listed in Table I.
[2]Confounders observed in studies 1–6 in Table I.
[3]Confounders observed in all studies in Table I.

Table IV. Estimates of the log hazard ratio for vascular risk per mm increase in IMT: various methods.

Type of analysis	Adjustment	No. of studies	$\hat{\beta}_1$	(Se)	$\hat{\tau}_1$
Unadjusted	None	8	2.757	(0.187)	0.423
Complete studies	Full	3	1.212	(0.140)	0.041
Complete confounders	Partial	8	1.270	(0.155)	0.245
Intermediate*	Partial	6	0.997	(0.160)	0.243
MICE model 0	Full	8	1.058	(0.148)	0.203
MICE-RE model 1	Full	8	1.017	(0.159)	0.245
MICE-RE model 2	Full	8	1.023	(0.158)	0.237
MICE-RE model 3	Full	8	1.037	(0.163)	0.225
MICE-RE model 4	Full	8	1.019	(0.155)	0.234
FPAMA, bivariate	Full	8	0.952	(0.234)	0.351
FPAMA, trivariate	Full	8	1.021	(0.155)	0.204

*See text

5.2. *MI approaches*

We increased the number of imputed data sets to 30 to reduce Monte Carlo error [16]. We first imputed systematically missing data using only a fixed study effect in the imputation model (model 0). Estimates are shown in Table IV. We then imputed missing data using the MICE-RE random effects imputation models. We considered a random effect on the single intercept (model 1), on the intercept and the exposure (IMT) (model 2), on the intercept and the outcome (model 3) or on the intercept, the outcome and the exposure (model 4). Point estimates of the log HR were similar, ranging from 1.017 to 1.037 (Table IV). Estimates using within-study estimation instead of a global estimation of the cumulative hazard function were similar (results not shown).

As expected from the simulation study, the standard error was smaller in the fixed effect imputation model. Standard errors were similar using the MICE-RE procedure to those obtained with six or eight studies with the standard approaches. Changes in the standard errors according to the number of random effects in the imputation model were very small.

5.3. *Comparison with multivariate two-stage meta-analysis*

Recently, to overcome concerns about systematically missing covariates, Jackson and White proposed a method that exploits the relationship between fully and partially adjusted analyses [9]. Considering a simple pattern in which the systematically missing variables are the same across the incomplete studies, they proposed a two-stage bivariate random effects meta-analysis for the fully and partially adjusted estimates of effect [8, 45]. The studies without missing confounders are used to provide information on fully adjusted estimates, partially adjusted estimates and their association, whereas the studies with systematically missing confounders are used only to obtain partially adjusted estimates. A joint likelihood function is then used to fit the bivariate random effects model and make inferences about the fully adjusted effect using all the available studies. We call this method the fully and partially adjusted meta-analysis (FPAMA). One of the main limitations of that approach is that the incomplete studies must all have the same subset of systematically missing confounders. In the absence of such a pattern, some simplification is required, for example, ignoring some confounders even if they are available in some complete studies. We generalised this approach by considering several partially adjusted estimates (each adjusted for different subsets of confounders) and a multivariate random effects model (Appendix A). Estimates were obtained using REML.

Because only eight studies were available, we considered only a bivariate model as in [9] and a trivariate model. We defined β_1^{p1} as the coefficient of IMT adjusted for the fully observed covariates $\{AHT, Smoke, Diab, DTreat, Age, Sex\}$ (Table I); β_1^{p1} was estimated in all studies. We defined β_1^{p2} as the coefficient of IMT adjusted for the larger set of covariates $\{Chol, BMI, SBP, AHT, Smoke, Diab, DTreat, Age, Sex\}$; β_1^{p2} was estimated in studies 1–6. Fully adjusted estimates are denoted β_1^f

and were estimated in studies 1–3. The bivariate model used β_1^{p1} and β_1^f and so used all the data except Chol, BMI and SBP in the incomplete studies 4–7. The trivariate model used β_1^{p1}, β_1^{p2} and β_1^f and so used all the data except SBP in study 7.

Estimates $\hat{\beta}_{1i}^f$, $\hat{\beta}_{1i}^{p1}$ and $\hat{\beta}_{1i}^{p2}$ of the effect of IMT in the three models, and their standard errors, for each study i are shown in Table III. The bivariate FPAMA (using the $\hat{\beta}_{1i}^f$ and $\hat{\beta}_{1i}^{p1}$ estimates) gave $\hat{\beta}_1^f = 0.952$, the smallest of all methods (Table IV). The trivariate FPAMA (using all three sets of estimates) gave $\hat{\beta}_1^f = 1.021$ with a standard error of 0.155 and $\hat{\tau}_1 = 0.204$.

Approaches including the maximum of information (trivariate FPAMA and MICE-RE) gave comparable point estimates and standard errors. This illustrates the disadvantage of the simplified missing data pattern used in the bivariate FPAMA. In cases of complex systematic missingness patterns, a feasible (low dimension) FPAMA would have to ignore some confounders in some studies, and a MICE-RE approach would appear preferable.

6. Discussion

Our aim in this paper was to propose a method to deal with systematically missing covariates in IPD meta-analysis. We developed an approach based on MI, which is one of the most popular methods to handle complex missing data patterns [46,47]. Classical MI is sometimes criticised as 'making up data', and here, we appear to be 'making up data' for a whole study. To refute this criticism, we need to argue that MI does not create information but only allows for the uncertainty about the missing data by creating several different plausible imputed data sets [11,47]. When imputing a systematically missing covariate, it is therefore essential to show that we have allowed not only for the usual sources of uncertainty (parameter uncertainty and random error) but also for potential heterogeneity in imputation models between the different studies. Because the variable being imputed is a confounder in the analysis model of interest, it seems most important that the imputation model should allow for heterogeneity in its associations with the exposure and the outcome in the analysis model of interest. We have made this possible by including both random intercepts and random slopes in our imputation model (Equation (2)).

Several methods have previously been proposed to impute missing data in a multilevel framework, but most of them were developed using a joint multivariate linear mixed model [20, 21]. We instead followed van Buuren and chose the chained equations framework, adding random effects terms in each imputation model. Indeed, MICE has been already proposed by van Buuren for multilevel data, but only in the case of sporadic missing values and not for systematically missing data [18]. His method uses an MCMC algorithm to draw imputed values at each step of the chained equations algorithm. Results of his simulations showed an improvement over standard practice (complete case analysis, MI ignoring any clustering structure and MI treating cluster as fixed factor): biases were reduced and coverage of 95%CIs was improved. Our aim was firstly to conserve the flexibility of the MICE approach by using a random effects regression model and secondly to use REML estimators and a classical proper imputation algorithm instead of van Buuren's Bayesian proper imputation.

Our simulation studies showed broadly good performance for the MICE methods but small bias towards the null. It is interesting to note that small biases seem present even before the missingness is imposed. Such fixed effects estimates slightly biased towards zero have been reported by Hirsch using R packages for random effects Cox models and notably *coxme* [48]. Concerning results obtained in presence of missing data, similar biases have been found elsewhere after imputing the covariates of a (single level) survival analysis model. White and Royston found relative biases between 0 and −10% in the regression coefficients after using a MICE algorithm to impute a binary missing covariate with 50% missing values [39]. Marshall *et al.* also showed underestimation of the absolute value of the regression coefficients [40]. Underestimation seemed to increase with the percentage of missing values and was worse for a skewed variable (which was not our case). The relatively good results that we obtained with MICE-RE were also observed using a stratified analysis model (results not shown), suggesting that however the analysis model handles between-study heterogeneity, allowing random effects in the imputation model improves the results. A possible advance in the imputation model would be to estimate the cumulative hazard separately by study, but when we did this in our simulation study, results showed no notable improvement but an increase in the computational failure rate in some settings. Finally, the coverage in our simulation seemed to be good for small between-study heterogeneity, but it decreased slightly with higher between-study heterogeneity and seemed to increase with the number of random effects.

From a practical point of view, MICE-RE is implemented as an extra routine, available upon request, for use with the *MICE* R package [43]. Imputation steps use *nlme* R package [42] and are significantly slowed by increasing study size and number of random effects. Nevertheless, the increase of computation time for an imputation model with a single random effect compared with no random effect is slight.

A large number of random effects in the imputation model can cause computational problems, illustrated by the percentage of failures in model 4 and the increase of the standard error for models 3 and 4 in setting A. The choice of the number and type of random effects should be carefully made in the light of the number of studies, the size of the studies and the expected heterogeneity. In our simulation study, results using a simple random intercept imputation model (model 1) seemed adequate and very similar to the results obtained with a larger number of random effects (models 2–4). The between-cluster heterogeneity seems to be largely captured by the random intercept. This model might be a good choice in the presence of few complete studies. If one random effect is to be added, then it should probably be placed on the event indicator, and a second random effect could be placed on the exposure of interest. If a fuller random slopes model were required but computationally difficult, one might consider assuming independent random effects, to reduce the number of parameters of the imputation model.

The models used in our example do not take into account interaction terms between covariates, but our methods allow such interactions to be included in the analysis model and/or in the imputation models. However, the presence of interactions in the analysis model places demands on the choice of imputation model, which can be difficult to satisfy [16, 49].

The models used in our example are linear, but our algorithm could be extended to impute binary or categorical systematically missing variables using generalised linear random effects models. Nevertheless, methods to fit general linear model with random effects are less robust than methods for linear models. Thus, the performance and stability of such imputation model should be further explored. Our algorithm could also be extended to allow for sporadic as well as systematically missing data. In this case, in the nonsystematically missing cohorts, after fitting model 2, we would obtain $b_{ir}|x_{ir} \sim N(m(x_{ir}), v(x_{ir}))$ with $v(x_{ir}) = \Psi_r - \Psi_r Z_{ir}^T \left(Z_{ir}\Psi_r Z_{ir}^T + \Sigma_{ir}\right)^{-1} Z_{ir}\Psi_r$ and $m(x_{ir}) = \Psi_r Z_{ir}^T \left(Z_{ir}\Psi_r Z_{ir}^T + \Sigma_{ir}\right)^{-1} (x_{ir} - X_{ir}\beta_r)$ [50, 51]. After step 2 of the single imputation procedure, we would obtain $v(x_{ir})^*$ and $m(x_{ir})^*$ by replacing $(\gamma_r, \Psi_r, \sigma_r)$ with $\left(\gamma_r^*, \Psi_r^*, \sigma_r^*\right)$. Then, we would draw b_{ir}^* from $N\left(m\left(x_{ir}\right)^*, v\left(x_{ir}\right)^*\right)$ in step 3. Finally, x_{ijr}^* would be drawn using Equation (3). All these extensions are under development.

In the example, we compared our method with the FPAMA method that estimates the fully adjusted parameter using its association with the partly adjusted parameter [9]. This method has the advantage over MICE-RE that it does not require any model for the distribution of the incomplete covariates, and it could be directly applied for missing categorical covariates. However, FPAMA requires some studies with complete data and is best with a simple pattern of missingness. In the presence of a complex pattern of missingness, a small number of available studies will constrain the dimension of the multivariate random effects model used, forcing the choice of a simplified pattern of missingness, which could importantly modify the results, as in our example where bivariate and trivariate analyses give different estimates (which is a matter of concern). Finally, FPAMA does not handle sporadically missing values: these could be multiply imputed on a study-by-study basis, with FPAMA applied to the results of applying Rubin's rules within each study.

Our method was developed for systematically missing covariates in an IPD meta-analysis but could be easily applied to other types of multilevel data. In particular, large cohort studies involving several centres could have systematically missing data because some examination was unavailable in some centres. In such a case, our approach would allow the final analysis to include all the individual participants under a MAR assumption. Our approach could also be applied to handle a systematically missing exposure or a systematically missing outcome, although we would not expect this to yield much gain in precision.

The MI and FPAMA approaches are valid under a MAR mechanism, which allows the probability that a variable is systematically missing in a given study to depend on the observed variables in that study. However, systematic missingness is a decision made at the study design stage and is more likely to be independent of all observed and unobserved data – that is, a MCAR mechanism, which is a special case of a MAR mechanism. The similarity of the 'complete studies' and MICE-RE estimates in Table IV supports the view that the PROG-IMT data are MCAR. We cannot, however, exclude a missing not at random (MNAR) mechanism, which would arise if, for example, a covariate were not collected in a given study because it was believed not to be a confounder in that population: under MNAR, the MI and

FPAMA approaches can give biased estimates. Adapting the imputation procedure to allow for MNAR is a topic for future research.

Collaborative projects to collate data of studies from different countries are increasingly popular. The studies may be epidemiological cohorts or randomised clinical trials, and the aim may be to gain power, to measure more precisely the effect of risk factors or treatment or to explore subgroup effects. In such international collaborations, data recording frequently differs between countries. For example, in studying kidney failure, creatinine is usually recorded in Europe, but blood urea level is more often recorded in North America and some studies record both, so adjustment for kidney failure may be impossible without a method to handle systematically missing values. Given the increasing popularity of IPD meta-analysis, and the wide availability of methods and software to perform such analysis, our method for handling systematically missing covariates should be increasingly important in facilitating appropriate adjustment for confounding and making best use of the available information.

Appendix A

In this Appendix, we generalise the two-stage approach proposed by D. Jackson and I. White considering a fully adjusted and several partially adjusted estimates.

Within-cohorts model

Let n be the number of cohorts. In a particular cohort $i \in \{1, \ldots, n\}$, let $(X_1, \ldots, X_k)$ be k column vectors of covariates. We assume that X_1 is complete and is the exposure of interest. We assume that some of $X_2, \ldots, X_k$ are systematically missing and represent potential confounders.

We assume that at least two cohorts have all the covariates available. In those cohorts, we assume a full proportional hazards model.

$$h(t) = h_0^f(t) \exp\left(\beta_1^f X_1 + \sum_{j=2}^{k} \beta_j^f X_j\right)$$

where β_1^f is the fully adjusted log HR that we want to estimate and $h(t)$ the hazard function.

Let $\{E_2, \ldots, E_l\}$ be $l-1$ different subsets of $\{2, \ldots, k\}$. We assume the following partially adjusted models:

$$h(t) = h_0^{p2}(t) \exp\left(\beta_1^{p2} X_1 + \sum_{j \in E_2} \beta_j^{p2} X_j\right) \tag{A1}$$

$$h(t) = h_0^{p3}(t) \exp\left(\beta_1^{p3} X_1 + \sum_{j \in E_3} \beta_j^{p3} X_j\right) \tag{A2}$$

.... (A3)

$$h(t) = h_0^{pl}(t) \exp\left(\beta_1^{pl} X_1 + \sum_{j \in E_l} \beta_j^{pl} X_j\right) \tag{A4}$$

where $\left\{\beta_1^{p2}, \beta_1^{p3}, \ldots, \beta_1^{pl}\right\}$ represent the partially adjusted log HR corresponding respectively to the quantities obtained with the subsets of covariates $\{E_2, \ldots, E_l\}$. The superscript f refers to fully adjusted quantities, whereas superscripts $p2$, $p3, \ldots,$ pl refer to partially adjusted quantities. All the models cannot be simultaneously true [8], but they nevertheless bring useful information in reflecting the relative impact of the exposure and provide adequate descriptions of the data [52].

In each cohort i, some or all of estimates $\hat{\beta}^f$, $\hat{\beta}_1^{p2}$, $\hat{\beta}_1^{p3}, \ldots$, or $\hat{\beta}_1^{pl}$ are obtained by fitting the corresponding proportional hazards models on the subsets E_j available for the cohort i.

Between-cohorts model

We consider now a multivariate meta-analysis framework [8], where the different effects $\left\{\beta_1^f, \beta_1^{p2}, \beta_1^{p3}, \ldots, \beta_1^{pl}\right\}$ are considered as outcomes.

We assume for any given cohort the following:

$$\begin{pmatrix} \hat{\beta}_1^f \\ \hat{\beta}_1^{p2} \\ \dots \\ \hat{\beta}_1^{pl} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1^f \\ \beta_1^{p2} \\ \dots \\ \beta_1^{pl} \end{pmatrix}, S \right), \quad S = \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \dots & \rho_{1,l}\sigma_1\sigma_l \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \dots & \rho_{2,l}\sigma_2\sigma_l \\ \dots & \dots & \dots & \dots \\ \rho_{1,l}\sigma_1\sigma_l & \rho_{2,l}\sigma_2\sigma_l & \dots & \sigma_l^2 \end{pmatrix} \tag{A5}$$

where S is the within-cohort variance–covariance matrix considered here as fixed [53]. Studies with systematically missing covariates have missing values for $\hat{\beta}_1^f$ and possibly some of the $\hat{\beta}_1^{pk}$: for such studies, model (A5) is adapted to describe just the observed $\hat{\beta}_1^{pk}$'s. We assume that $\left(\beta_1^f, \beta_1^{p2}, \dots, \beta_1^{pl}\right)$ may vary from cohort to cohort. We model this variation by the following:

$$\begin{pmatrix} \beta_1^f \\ \beta_1^{p2} \\ \dots \\ \beta_1^{pl} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta^f \\ \beta^{p2} \\ \dots \\ \beta^{pl} \end{pmatrix}, \Sigma, \right) \quad \Sigma = \begin{pmatrix} \tau_1^2 & \kappa_{1,2}\tau_1\tau_2 & \dots & \kappa_{1,l}\tau_1\tau_l \\ \kappa_{1,2}\tau_1\tau_2 & \tau_2^2 & \dots & \kappa_{2,l}\tau_2\tau_l \\ \dots & \dots & \dots & \dots \\ \kappa_{1,l}\tau_1\tau_l & \kappa_{2,l}\tau_2\tau_l & \dots & \tau_l^2 \end{pmatrix} \tag{A6}$$

where Σ represents the within-cohort variance–covariance matrix. We obtain the following:

$$\begin{pmatrix} \beta_1^f \\ \beta_1^{p2} \\ \dots \\ \beta_1^{pl} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta^f \\ \beta^{p2} \\ \dots \\ \beta^{pl} \end{pmatrix}, S + \Sigma \right) \tag{A7}$$

In practice, the σ_j are estimated within each cohort using the variances estimated with each Cox model, and the $\rho_{j,k}$ are estimated by a bootstrap procedure [9].

Log-likelihood

In practice, the number l of subsets E_j should be small notably because the number of parameters to be estimated in Equation (A7) is $(l^2/2 + 3l/2)$ and increases rapidly with l. Let $\beta = (\beta^f, \beta^{p2}, \dots, \beta^{pl})$ and

$$L(\beta, \tau, \kappa) = \sum_{i=1}^{n} \log f_i(\hat{\beta}) \tag{A8}$$

where $f_i(\hat{\beta})$ represents the multivariate density from model (A5) corresponding to the subset of β_1^f, β_1^{p2}, $\beta_1^{p3}, \dots, \beta_1^{pl}$ obtained within cohort i. All the parameters of the different multivariate densities are taken from Equation (A7).

Maximising the log-likelihood gives an estimate of β_1^f, which exploits all the information in the data, including that from cohorts that do not directly estimate β_1^f. In the PROG-IMT data, with only eight studies, we reduced the number of partial estimations to 2 or 3 using a bivariate or a trivariate model in Equation (A7).

Implementation

The FPAMA method was implemented using R software [54]. Cox models were fitted using the *Coxph* function of the *survival* package. Estimation of the $\rho_{j,k}$ involved a nonparametric bootstrap procedure with 2000 replications. β_1^f and its confidence interval were estimated using maximum partial likelihood [16]. The *optim* function of R was used using the Broyden, Fletcher, Goldfarb, and Shannon method to fit model (A7). The standard errors were directly obtained by a numerically differentiated Hessian matrix given by *optim*.

Acknowledgements

M. Resche-Rigon was supported by Fondation de France grant Engt 2010-12449. I. R. White was supported by the Medical Research Council (Unit Programme number U105260558). J. W. Bartlett was supported by a grant from the ESRC Follow-On Funding scheme (RES-189-25-0103) and MRC grant G0900724. The authors thank the PROG-IMT Study Group (http://www.prog-imt.org) for the permission to use their data and all the PROG-IMT Study Group members (Supporting information[§]).

References

1. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *British Medical Journal* 2010; **340**(7745):521–525.
2. Simmonds MC, Higgins J, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2005; **2**(3):209–217.
3. Stewart LA, Parmar M. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993; **341**(8842):418–422.
4. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**(3):177–188.
5. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2000; **19**(24):3417–3432.
6. Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Statistics in Medicine* 2005; **24**(9):1307–1319.
7. Katsahian S, Latouche A, Mary JY, Chevret S, Porcher R. Practical methodology of meta-analysis of individual patient data using a survival outcome. *Contemporary Clinical Trials* 2008; **29**(2):220–230.
8. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Statistics in Medicine* 2011; **30**:2481–2498. DOI: 10.1002/sim.4172.
9. The Fibrinogen Studies Collaboration (writing committee: Jackson D, White I). Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Statistics in Medicine* 2009; **28**(8):1218–1237.
10. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc.: Hoboken, New Jersey, 2004.
11. Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical Research* 1999; **8**(1):3.
12. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004; **5**(3):445–464.
13. Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*. John Wiley & Sons Inc: Hoboken, New Jersey, 2007.
14. Schafer JL. *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC: Boca Raton, Florida, 1997.
15. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999; **18**(6):681–694.
16. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* 2011; **30**(4):377–399.
17. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological Methods* 2002; **7**(2):147.
18. van Buuren S. Multiple imputation of multilevel data. In *The Handbook of Advanced Multilevel Analysis*, Hox JJ, Roberts JK (eds). Taylor & Francis, Ltd: New York, 2010; 173–196.
19. Andridge RR. Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biometrical Journal* 2011; **53**(1):57–74.
20. Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* 2002; **11**(2):437–457.
21. Carpenter J, Goldstein H. Multiple imputation using MLwiN. *Multilevel Modelling Newsletter* 2004; **16**(2):9–18.
22. Yucel RM. Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Royal Society of London Philosophical Transactions Series A* 2008; **366**:2389–2403.
23. Lorenz MW, Bickel H, Bots ML, Breteler M, Catapano AL, Desvarieux M, Hedblad B, Iglseder B, Johnsen SH, Juraska M, *et al.* Individual progression of carotid intima media thickness as a surrogate for vascular risk (PROG-IMT): rationale and design of a meta-analysis project. *American Heart Journal* 2010; **159**(5):730–736.
24. Lorenz MW, Polak JF, Kavousi M, Mathiesen EB, Vlzke H, Tuomainen T, Sander D, Plichart M, Catapano AL, Robertson CM, *et al.* Carotid intima-media thickness progression to predict cardiovascular events in the general population (the PROG-IMT collaborative project): a meta-analysis of individual participant data. *Lancet* 2012; **379**(9831):2053–2062.
25. Nambi V, Chambless L, Folsom AR, He M, Hu Y, Mosley T, Volcik K, Boerwinkle E, Ballantyne CM. Carotid intima-media thickness and presence or absence of plaque improves prediction of coronary heart disease risk: the ARIC (Atherosclerosis Risk In Communities) study. *Journal of the American College of Cardiology* 2010; **55**(15): 1600–1607.
26. Chambless LE, Folsom AR, Clegg LX, Sharrett AR, Shahar E, Nieto FJ, Rosamond WD, Evans G. Carotid wall thickness is predictive of incident clinical stroke: the Atherosclerosis Risk in Communities (ARIC) study. *American Journal of Epidemiology* 2000; **151**(5):478.
27. Lorenz MW, Schaefer C, Steinmetz H, Sitzer M. Is carotid intima media thickness useful for individual prediction of cardiovascular risk? Ten-year results from the Carotid Atherosclerosis Progression Study (CAPS). *European Heart Journal* 2010; **31**(16):2041.

[§] *Supporting information may be found in the online version of this article.*

28. Cao JJ, Arnold AM, Manolio TA, Polak JF, Psaty BM, Hirsch CH, Kuller LH, Cushman M. Association of carotid artery intima-media thickness, plaques, and C-reactive protein with future cardiovascular disease and all-cause mortality: the Cardiovascular Health Study. *Circulation* 2007; **116**(1):32.
29. Sander K, Bickel H, Förstl H, Etgen T, Briesenick C, Poppert H, Sander D. Carotid-intima media thickness is independently associated with cognitive decline. The INVADE study. *International Journal of Geriatric Psychiatry* 2010; **25**(4):389–394.
30. Prabhakaran S, Singh R, Zhou X, Ramas R, Sacco RL, Rundek T. Presence of calcified carotid plaque predicts vascular events: the Northern Manhattan Study. *Atherosclerosis* 2007; **195**(1):e197–e201.
31. Bots ML, Grobbee DE, Hofman A, Witteman J. Common carotid intima-media thickness and risk of acute myocardial infarction: the role of lumen diameter. *Stroke* 2005; **36**(4):762–767.
32. Mathiesen EB, Johnsen SH, Wilsgaard T, Bonaa KH, Lochen ML, Njolstad I. Carotid plaque area and intima-media thickness in prediction of first-ever ischemic stroke: a 10-year follow-up of 6584 men and women: The Tromso study. *Stroke* 2011; **42**(4):972–978.
33. Bonora E, Willeit J, Kiechl S, Oberhollenzer F, Egger G, Bonadonna R, Muggeo M. Relationship between insulin and carotid atherosclerosis in the general population: the Bruneck study. *Stroke* 1997; **28**(6):1147–1152.
34. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer Verlag: New York, 2000.
35. Grambsch PM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**(3):515–526.
36. Meng XL. Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* 1994; **9**(4):538–558.
37. Vaida F, Xu R. Proportional hazards model with random effects. *Statistics in Medicine* 2000; **19**(24):3309–3324.
38. Moons KG, Donders RA, Stijnen T, Harrell F, Jr. Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology* 2006; **59**(10):1092.
39. White IR, Royston P. Imputing missing covariate values for the Cox model. *Statistics in Medicine* 2009; **28**(15):1982–1998.
40. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Medical Research Methodology* 2010; **10**(1):7.
41. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, *et al.* Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Soc Cardiology.* 2013; **34**:19, 1404–1413.
42. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team. nlme: Linear and nonlinear mixed effects models, 2011. R package version 3.1-102.
43. Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2010; **45**(3):1–68.
44. Therneau T. coxme: Mixed effects Cox models. URL urlhttp://CRAN.R-project.org/package=coxme,r package version 2.1-3, 2011.
45. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**(4):589–624.
46. Sterne J, White I, Carlin J, Spratt M, Royston P, Kenward M, Wood A, Carpenter J. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ (Clinical research ed.)* 2009; **338**:b2393, 157–160.
47. Rubin D. Multiple imputation after 18+ years. *Journal of the American Statistical Association* 1996; **91**(434):473–489.
48. Hirsch K, Wienke A. Software for semiparametric shared gamma and log-normal frailty models: an overview. *Computer methods and programs in biomedicine* 2012; **107**(3):582–597.
49. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Medical Research Methodology* 2012; **12**(1):46.
50. West BT, Welch KB, Galecki AT. *Linear Mixed Models: A Practical Guide Using Statistical Software*. CRC Press: Boca Raton, Florida, 2007.
51. Lee Y, Nelder JA, Pawitan Y. *Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood.* CRC Press: Boca Raton, Florida, 2006.
52. Ford I, Norrie J, Ahmadi S. Model inconsistency, illustrated by the Cox proportional hazards model. *Statistics in Medicine* 1995; **14**(8):735–746.
53. Hardy R, Thompson S. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629.
54. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2010. URL http://www.R-project.org/.