

Correcting for multivariate measurement error by regression calibration in meta-analyses of epidemiological studies

Fibrinogen Studies Collaboration*

SUMMARY

Within-person variability in measured values of multiple risk factors can bias their associations with disease. The multivariate regression calibration (RC) approach can correct for such measurement error and has been applied to studies in which true values or independent repeat measurements of the risk factors are observed on a subsample. We extend the multivariate RC techniques to a meta-analysis framework where multiple studies provide independent repeat measurements and information on disease outcome. We consider the cases where some or all studies have repeat measurements, and compare study-specific, averaged and empirical Bayes estimates of RC parameters. Additionally, we allow for binary covariates (e.g. smoking status) and for uncertainty and time trends in the measurement error corrections. Our methods are illustrated using a subset of individual participant data from prospective long-term studies in the Fibrinogen Studies Collaboration to assess the relationship between usual levels of plasma fibrinogen and the risk of coronary heart disease, allowing for measurement error in plasma fibrinogen and several confounders. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: measurement error; within-person variation; meta-analysis; individual participant data; regression calibration

1. INTRODUCTION

Many epidemiological studies aim to estimate the association between potential risk factors and the likelihood of disease. Because risk factors are usually measured with error and fluctuate within individuals, analyses that use only single measurements of the risk factors produce biased estimates of any aetiological association between average (or ‘usual’) risk factor levels and disease [1]. This

*The Authors/Writing Committee, Authors/Members, and Authors/Coordinating Centre of the Fibrinogen Studies Collaboration are listed at the end of this article.

Correspondence to: Angela Mary Wood, University of Cambridge, Public Health and Primary Care, Cambridge, U.K. (E-mail: amw79@medsch.cam.ac.uk).

Contract/grant sponsor: British Heart Foundation; contract/grant number: 002/02

Received 11 March 2008

Copyright © 2009 John Wiley & Sons, Ltd.

Accepted 27 November 2008

bias is caused by any or all of (i) measurement error, (ii) short-term within-person variation and (iii) longer-term within-person variation (e.g. due to behaviour changes) [2]. We assume there exist ‘usual levels’ of the risk factors which represent the true exposure of interest, and we describe any difference between measured levels and usual levels as ‘measurement error’. Measurement error in exposures leads to underestimation of exposure–disease associations (regression dilution bias) [2–4], while measurement error in confounders leads to residual confounding [5].

Various methods have been proposed to estimate the effect of measurement error in multiple covariates and to correct the disease associations estimated from single observations of the risk factors [1, 6, 7]. For continuous risk factors, the true regression coefficients in the disease model may be estimated by multiplying the vector of observed regression coefficients by the inverse of a correction matrix [6]. The correction matrix comprises coefficients from the linear regression of the true risk factors on the observed values: we call this the ‘regression calibration (RC)’ model. A second approach replaces the observed risk factors in the disease model with conditional expectations of the true values given the observed values, which are predicted from the RC model. The RC model may be estimated directly if true and observed measures are available for some subjects, or indirectly if repeat observations are available for some or all subjects. In the latter case, the RC model is the regression of an unbiased repeat measurement on the first measurement [8].

The Correction Matrix and Conditional Expectation approaches are mathematically equivalent in simple cases and have underlying assumptions that (i) the errors in different repeat measurements are independent of each other and of the true value [9] and (ii) knowledge of usual levels completely captures the risk of disease associated with the risk factors (rather than change or spikes in risk factors). Under these assumptions, RC exactly corrects for measurement error when the disease model is a linear regression or a Poisson regression; however, in general, it is an approximate method [10, 11].

Whereas RC approaches have been applied in the literature for the analysis of single studies, less work has been done in the context of individual participant data (IPD) meta-analysis of multiple studies. Earlier work [2, 12, 13] has focused on methods for correcting for univariate measurement error in meta-analyses. The Prospective Studies Collaboration estimated a usual risk factor level from a non-linear time-dependent RC model that pooled all studies [2]. The Fibrinogen Studies Collaboration (FSC) estimated study- and time-specific regression dilution ratios (RDRs) which were then combined allowing for within- and between-study heterogeneity [13]. Both ignored error in confounders. Here, we extend the multivariate RC technique to a meta-analysis framework. We consider the cases where either some or all studies have repeat measurements. First we define and compare methods using study-specific, averaged and empirical Bayes estimates of regression calibration parameters. Second, we take account of the imprecision in the measurement error corrections by adapting methods used in multiple imputation. Third, we allow for measurement error time trends by estimating time-dependent usual levels of the risk factors which can be entered into a time-dependent disease model. We also discuss the practicalities of using Correction Matrix versus Conditional Expectation approaches.

Our motivation is in providing appropriate methods for the analysis of IPD in collections of multiple epidemiological studies. By comparison with analysis of single studies, combination of data from several studies should yield greater precision and comprehensiveness. With the increasing number of such IPD collations (e.g. the Prospective Studies Collaboration [2], the Asia Pacific Cohort Studies Collaboration [14], the FSC [15], and the million-participant Emerging Risk Factors Collaboration [16]) and of purpose-designed multi-centre prospective observational studies (e.g.

the 22-centre, 520 000-person European EPIC study [17]), it is important to develop appropriate biostatistical methods for correcting for multivariate measurement errors in IPD meta-analyses.

We illustrate our proposals on a subset of IPD from prospective long-term studies in the FSC, described in Section 2. In Section 3 we introduce our model and then discuss various practicalities in Section 4. Results from the FSC data are presented in Section 5. We give a discussion in Section 6 and conclusions in Section 7.

2. THE FIBRINOGEN STUDIES COLLABORATION

2.1. Background

The FSC [15] is a meta-analysis of individual data on 154 211 adults from 31 prospective studies with information on plasma fibrinogen [18] and major disease outcomes. The FSC has previously reported moderately strong associations between plasma fibrinogen values and the risks of major vascular and non-vascular chronic disease outcomes, although the causal relevance of these associations remains uncertain [19]. As part of the FSC data collection, information was provided on repeat measures (values recorded after the initial baseline examination) of plasma fibrinogen and other measures from subsets of participants from 15 of the studies at various re-measurement times.

For illustration purposes, we will explore the association of fibrinogen, systolic blood pressure and smoking with coronary heart disease (CHD) in five representative studies in the FSC, which had repeat measures available on each of these risk factors. To remain consistent with previous FSC reports [19], baseline fibrinogen levels above 5.62 g/L (the highest 1 per cent of values) are excluded due to the potential distortions arising from assay imprecision and from acute-phase reactions, although our proposed methods would be able to address such extreme values. Individuals with a known history of previous CHD or stroke at baseline, and the few participants aged below 20 years at baseline, are also excluded.

2.2. Typical disease models

The disease models are based on Cox proportional hazards models. If we ignored measurement error, we would model the hazard for the i th individual in the s th study as

$$\log h_{si}(t|W_{si}) = \log h_{0s}(t) + \beta_s^{\text{un}} W_{si} \quad (1)$$

where $s = 1, \dots, S$, $i = 1, \dots, N_s$, and $W_{si} = (W_{si1}, W_{si2}, \dots, W_{sIP})$ is a P -vector of observed risk factors. The superscript in β^{un} indicates that no adjustment for within-person variability has been made.

In order to adjust for measurement error, we might assume that the risk factors have ‘usual levels’ $X_{si} = (X_{si1}, X_{si2}, \dots, X_{sIP})$. The Cox model then becomes

$$\log h_{si}(t|X_{si}) = \log h_{0s}(t) + \beta_s X_{si} \quad (2)$$

Alternatively, if the assumption of a constant ‘usual level’ over the life-course is not tenable [20], we might prefer to define time-dependent ‘usual levels’ $X_{si}(t)$ and the Cox model

$$\log h_{si}(t|X_{si}(t)) = \log h_{0s}(t) + \beta_s X_{si}(t) \quad (3)$$

A second-level model then links the parameters β_s in different studies. Under a ‘fixed-effect’ model, $\beta_s = \beta$, the vector of parameters β is of interest, being the adjusted log hazard ratios per unit increase in ‘usual levels’ of the risk factors. This model can be estimated in a single stage. Alternatively, we may consider a ‘random-effects’ model $\beta_s \sim \text{MVN}(\beta, Z)$. Here β is the vector of average log hazard ratios, and the variance–covariance matrix Z represents heterogeneity between studies [21]. This model is most conveniently estimated by a two-stage procedure in which the model is first estimated in each study, yielding parameter estimate $\hat{\beta}_s$ with variance–covariance matrix \hat{v}_s , and these are then combined using standard methods.

2.3. Clinical importance

In the initial analysis to determine the relationship between fibrinogen and cardiovascular disease [19], measurement error in plasma fibrinogen was corrected for using an RDR of 0.46 estimated from a linear regression of repeat measurements on baseline values of fibrinogen in each study and for each time interval, and pooled allowing for within- and between-study heterogeneity [13]. The association between plasma fibrinogen and disease was found to be moderately strong, even after adjusting for all measured confounders. However, residual confounding remains if the confounders are also measured with error. The methodological work developed in this paper was motivated by the need to correct for measurement error in multiple confounders using data from multiple studies. This paper focuses on developing the methodology rather than presenting the full applied results.

3. MULTIVARIATE RC MODELS IN META-ANALYSES

In this section we develop multivariate RC models which incorporate data from multiple studies. For ease of exposition, we ignore non-error-prone confounders, such as age and sex.

3.1. Justification of RC approach

We first justify the use of the regression coefficients of the repeat observations on the baseline observations. If model (2) is correct then approximately unbiased estimators of β_s may be obtained from a Cox regression using as covariates the conditional expectations of X_{si} given the observed baseline value W_{si0} : this is because

$$\begin{aligned}\log h_{si}(t|W_{si0}) &\approx E[\log h_{si}(t|X_{si}, W_{si0})|W_{si0}] \\ &= E[\log h_{si}(t|X_{si})|W_{si0}] \\ &= \log h_{0s}(t) + \beta_s E[X_{si}|W_{si0}]\end{aligned}\tag{4}$$

The approximation above derives from a Taylor series, valid for small cumulative incidence [11]. The first exact equality above (expressing non-differential error) depends on the correctness of model (2), since it assumes that the observation W_{si0} would add nothing to the model if X_{si} were known. Model misspecification could invalidate the corrections we propose [20]: that is, the corrected estimates could differ systematically from those obtained if we could fit model (2) directly to data on X_{si} .

Similarly, if model (3) is correct then

$$\begin{aligned}\log h_{si}(t|W_{si0}) &\approx E[\log h_{si}(t|X_{si}(t), W_{si0})|W_{si0}] \\ &= E[\log h_{si}(t|X_{si}(t))|W_{si0}] \\ &= \log h_{0s}(t) + \beta_s E[X_{si}(t)|W_{si0}]\end{aligned}\quad (5)$$

hence, the covariates in the Cox regression must be the conditional expectations of $X_{si}(t)$ given W_{si0} .

In order to estimate $E[X_{si}|W_{si0}]$ or $E[X_{si}(t)|W_{si0}]$, we make the further assumption that any repeat measurements W_{sir} are unbiased measures of the true underlying value, in the sense that $E[W_{sir}|W_{si0}, X_{si}] = X_{si}$, so that $E[X_{si}|W_{si0}] = E[W_{sir}|W_{si0}]$. We therefore focus on estimation of the RC model $E[W_{sir}|W_{si0}]$. We start assuming model (2) and describe modifications for model (3) in Section 4.3.

3.2. Regression calibration models

3.2.1. Single study with a single repeat measure. We begin with a single study with N individuals, with each individual i providing baseline measurements on P continuous risk factors, denoted by W_{i0p} for $i = 1, \dots, N$ and $p = 1, \dots, P$. We assume that $n \leq N$ individuals have exactly one repeat measure on all risk factors measured at the same time point, denoted by W_{i1p} for $i = 1, \dots, n$ and $p = 1, \dots, P$. A linear multivariate RC model can be written as

$$W_{i1p} = a_p + b_{p1}W_{i01} + b_{p2}W_{i02} + \dots + b_{pP}W_{i0P} + e_{ip} \quad (6)$$

for $i = 1, \dots, n$, $p = 1, \dots, P$, where $(e_{i1}, \dots, e_{iP})^T \sim MVN(0, \Sigma)$.

Note that repeat measurements are regressed on the baseline measurements rather than vice versa. Formulating the model in this manner is crucial for its applicability to individuals with only baseline measurements. Non-error-prone confounders, such as baseline sex and age, can easily be incorporated in equation (6). For example, if Age_{i0} denotes baseline age for individual i , the term $c_p Age_{i0}$ can be added to the right-hand side of equation (6).

3.2.2. Multiple studies with a single repeat measure. We now extend the model to allow for S studies, with study s having N_s ($s = 1, \dots, S$) individuals providing baseline measures and n_s individuals providing a single repeat measurement on P risk factors. Let the baseline measure for the p th risk factor ($p = 1, \dots, P$) from the i th individual in study s be denoted by W_{si0p} ($i = 1, \dots, N_s$) and similarly let the repeat measurement be denoted by W_{si1p} ($i = 1, \dots, n_s$). The multivariate RC model is

$$W_{si1p} = a_{sp} + b_{sp1}W_{si01} + b_{sp2}W_{si02} + \dots + b_{spP}W_{si0P} + e_{sip} \quad \text{for } i = 1, \dots, n_s \quad (7)$$

where $(e_{s1}, \dots, e_{sp})^T \sim MVN(0, \Sigma)$ and Σ is a $(P \times P)$ variance–covariance matrix. Variations on the covariance structure Σ are discussed in Section 3.4.

3.2.3. Multiple studies with multiple repeat measures. We finally extend to the case when data are available from S studies with study s having N_s individuals providing baseline measurements and n_s individuals providing up to R_s repeat measures on P risk factors. For simplicity we begin by assuming that the error in the risk factors is time-independent, but this is relaxed in Section 4.3. Let the baseline and the r th repeat measurement for the p th risk factor in study s be denoted by

W_{si0p} ($i = 1, \dots, N_s$) and W_{sirp} for ($i = 1, \dots, n_s, r = 1, \dots, R_s$), respectively. A linear multivariate RC model is

$$\begin{aligned} W_{sirp} &= a_{srp} + b_{srp1} W_{si01} + b_{srp2} W_{si02} + \dots + b_{srpP} W_{si0P} + z_{sip} + e_{sirp} \\ \text{for } s &= 1, \dots, S, \quad i = 1, \dots, n_s, \quad r = 1, \dots, R_s, \quad p = 1, \dots, P \end{aligned} \quad (8)$$

where $(e_{sir1}, \dots, e_{sirP})^T \sim \text{MVN}(0, \Sigma)$ and $(z_{si1}, \dots, z_{sip})^T \sim \text{MVN}(0, \Phi)$ for all $r = 1, \dots, R_s$, and Σ and Φ are $(P \times P)$ variance–covariance matrices denoting residual error and individual-specific variation, respectively. The latter accounts for the hierarchical structure at the individual level by allowing an individual-specific random intercept (e.g. some individuals may have consistently higher or lower repeat measurement levels of certain risk factors). Note that the RC model (7) is a special case of model (8) with $R_s = 1$. As before, non-error-prone confounders can be easily incorporated in equations (7) and (8), for example, by adding the term $c_p \text{Age}_{si0}$ or $c_{sp} \text{Age}_{si0}$ to the right-hand side. Variations on model (8) are discussed in Section 3.4.

3.3. Extracting the regression coefficients

We can extract the regression coefficients from equations (6)–(8) in three possible ways:

- (i) *Study-specific regression coefficients.* These are the set of estimated \hat{b}_{pq} obtained from fitting equation (6) separately to each study or equivalently the \hat{b}_{spq} from equation (7) ($s = 1, \dots, S$). To obtain study-specific coefficients from equation (8) we use the additional set of models:

$$b_{srpq} = b_{spq} + v_{srpq} \quad \text{for } p, q = 1, \dots, P \quad (9)$$

where arranging the random terms as a vector $v_{sr} = (v_{sr11}, v_{sr12}, \dots, v_{sr1P}, v_{sr21}, \dots, v_{srPP})$, then $v_{sr} \sim N(0, \Omega)$ where Ω is a $P^2 \times P^2$ variance–covariance matrix. Within-study heterogeneity is measured by the diagonals of Ω : positive values mean that the coefficients at different repeat measurements within studies differ more than would be expected by chance.

- (ii) *Averaged regression coefficients.* It may be appropriate to use an overall set of coefficients from equation (7) or (9) estimated using the further set of models;

$$b_{spq} = b_{pq} + u_{spq} \quad \text{for } p, q = 1, \dots, P \quad (10)$$

where the study random effect terms

$$u_s = (u_{s11}, u_{s12}, \dots, u_{s1P}, u_{s21}, \dots, u_{sPP}) \sim \text{MVN}(0, \Psi)$$

and Ψ is a $P^2 \times P^2$ variance–covariance matrix. Between-study heterogeneity is measured by the diagonals of Ψ : positive values mean that the coefficients in different studies differ more than would be expected by chance [22]. Study random effect terms could also be applied to the coefficients for non-error-prone confounders.

- (iii) *Empirical Bayes regression coefficients.* Additionally, it is possible to extract empirical Bayes regression coefficients from (9) and (10): these are obtained by combining the estimated individual-specific and/or study-specific random effects with the averaged regression

coefficients [23]. For example, the fully empirical Bayes regression coefficients which incorporate the individual and study random effects can be extracted from (8), (9) and (10) as

$$\hat{b}_{pq} + \hat{u}_{spq} + \hat{v}_{srpq} \quad \text{for } p, q = 1, \dots, P$$

where estimates \hat{u}_{spq} and \hat{v}_{srpq} are obtained as described in Goldstein [24].

These estimates are ‘optimally’ weighted averages that combine information derived from the individual and/or study with the mean for all similar groups. The advantage of these estimates is that they reflect the between-individual and/or between-study heterogeneity for individuals and studies with repeat measures and provide averaged estimates for individuals and studies without repeat measures.

Usually only a subset of studies has data available on repeat measurements, as in the FSC, and it is necessary to find ways to transfer the information to the other studies. When considerable heterogeneity exists, averaged regression coefficients may not provide an appropriate correction (although they are likely to be far better than no correction), and any factors associated with the heterogeneity should be considered. The empirical Bayes regression coefficients reflect the heterogeneity in studies with repeat measures and use the averaged regression coefficients for studies without repeat measures.

3.4. Simplifying assumptions

Models involving (9) and/or (10) are highly multivariate and computationally complex. It may be appropriate to replace the variance–covariance matrices Σ and Φ with diagonal variance matrices, equivalent to P -independent univariate models. Such models will produce correct regression coefficients for balanced data (where all individuals provide a baseline and repeat measure on all P risk factors) but not for unbalanced data (where repeat measures for some risk factors may be missing either by design or not), although differences can often be minimal in practice. Alternatively, if sufficient repeat measures are available in all studies, then Σ and Φ may be replaced by study-specific matrices Σ_s and Φ_s reducing models (7) and (8) to S -independent multivariate models. If both simplifications are appropriate, the models can be reduced to $P \times S$ -independent univariate models.

Usually, the regression coefficient of repeat measure W_{sirp} on its corresponding baseline measure W_{si0p} (represented by b_{spp} in equation (9)) is the strongest and most important, whereas the regression coefficient of W_{sirp} on other baseline measures W_{si0q} (represented by b_{spq} in equation (9) where $q \neq p$) is relatively weak and of less importance. As a result, study-specific random effects u_{spq} or repeat-specific random effects v_{srpq} for $q \neq p$ can have very small variances and may lead to estimation problems, especially for large P . In such circumstances, we recommend only incorporating random effects on the regression coefficients b_{spp} . This simplification leads to a replacement of equations (9) and (10) by the respective models

$$\begin{aligned} b_{srpp} &= b_{spp} + v_{srpp} \quad \text{for } p = 1, \dots, P \\ b_{srpq} &= b_{spq} \quad \text{for } p, q = 1, \dots, P \text{ and } p \neq q \end{aligned} \tag{11}$$

$$\begin{aligned} b_{spp} &= b_{pp} + u_{spp} \quad \text{for } p = 1, \dots, P \\ b_{spq} &= b_{pq} \quad \text{for } p, q = 1, \dots, P \text{ and } p \neq q \end{aligned} \tag{12}$$

where $u_{spp} \sim \text{MVN}(0, \Psi)$ and $v_{srpp} \sim \text{MVN}(0, \Omega)$, and where Ψ and Ω are $(P \times P)$ variance-covariance matrices. Under these simplifications, our most general model can be defined by the equation:

$$W_{sirp} = a_{sirp} + (b_{pp} + u_{spp} + v_{srpp}) W_{si0p} + \sum_{q=1, q \neq p}^P b_{pq} W_{si0q} + z_{sip} + e_{sirp} \quad (13)$$

3.5. Estimation procedures

We here describe our proposed estimation procedures for estimating study-specific, averaged or empirical Bayes regression coefficients from the models described in Section 3.2 (and the simplifications in Section 3.4). To obtain study-specific regression coefficients from multiple studies with a single repeat measure, we use maximum likelihood estimation to fit either (i) $P \times S$ -independent univariate models defined by equation (7) or (ii) a single multivariate model defined by equation (7). To obtain averaged or empirical Bayes regression coefficients, we use restricted maximum likelihood estimation to fit either (i) P -independent univariate models defined by equations (7) and (12) or (ii) a single multivariate model defined by equations (7) and (12).

To obtain study-specific, averaged or full empirical Bayes coefficients from multiple studies with multiple repeat measures, we would ideally fit the models defined by equation (13) in a one-stage estimation approach using restricted maximum likelihood. However, this model is a cross-classified multivariate model, inducing computational difficulties. Thus, we propose two further different simplifications of the model. The first is to simplify (13) to P cross-classified univariate models, estimated by restricted maximum likelihood. Study-specific regression coefficients can be obtained from $P \times S$ cross-classified-independent univariate models defined by equations (8) and (12). Averaged or full empirical Bayes coefficients can be obtained from P cross-classified univariate models defined by equations (8), (11) and (12). The second simplification reduces the cross-classified model (13) to hierarchical univariate models or a hierarchical multivariate model by excluding either the within-study random effect term v_{srpp} or the within-subject random effect term z_{sip} from equation (13), then estimable by restricted maximum likelihood. To obtain study-specific coefficients we propose fitting either (i) $P \times S$ -independent univariate models or (ii) S multivariate models. To obtain averaged or empirical Bayes coefficients we propose fitting either (i) P -independent univariate models or (ii) a single multivariate model.

We note that (restricted) maximum likelihood estimation provides valid estimates under the missing at random assumption [25]. This is an important assumption, as generally individuals will not have complete information at all re-measurements times for all risk factors (either by design or not).

Hierarchical univariate models were fitted in STATA version 9.2 and checked using MLwiN [24], and cross-classified univariate models and hierarchical multivariate models were fitted in MLwiN [24]. An example of the STATA code to fit the hierarchical univariate models is shown in the Appendix.

4. PRACTICALITIES AND EXTENSIONS

In Section 3 we built an RC model to incorporate data from multiple studies with multiple repeat measures of multiple error-prone risk factors. Study-specific, averaged or empirical Bayes

coefficients could be extracted from the model and used to create correction matrices or conditional expectations. In this section, we define Correction Matrices and Conditional Expectations and discuss their practicalities and appropriateness. We also discuss extensions of the RC model to allow for time trends in measurement error corrections and for binary confounders. Finally, we propose a method to allow for the uncertainty from the RC model in the estimates of the risk associations of interest.

4.1. Correction matrices and conditional expectations

Two mathematically equivalent procedures use the regression coefficients from the RC model to estimate the vector of corrected risk associations $\hat{\beta}$ and its variance–covariance matrix \hat{v} . The first procedure utilizes a Correction Matrix defined for a single study by

$$\text{CM} = \begin{bmatrix} \hat{b}_{11} & \hat{b}_{12} & \dots & \hat{b}_{1P} \\ \hat{b}_{21} & \hat{b}_{22} & \dots & \hat{b}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{b}_{P1} & \hat{b}_{P2} & \dots & \hat{b}_{PP} \end{bmatrix}$$

where \hat{b}_{pq} ($p, q = 1, \dots, P$) are the estimated coefficients from, say, equation (6). Let $\hat{\beta}^{\text{un}}$ represent the vector of estimated uncorrected risk associations (e.g. log hazard ratios) with estimated variance–covariance matrix $\hat{v}^{\text{un}} = \text{var}(\hat{\beta}^{\text{un}})$. Corrected risk associations and corresponding variance–covariance matrix are estimated as $\hat{\beta} = \text{CM}^{-1}\hat{\beta}^{\text{un}}$ and $\hat{v} = \text{CM}^{-1}\hat{v}^{\text{un}}\text{CM}^{-\text{T}}$ respectively. The Correction Matrix is an extension of the ‘RDR’ used for single error-prone risk factors; the diagonal component \hat{b}_{pp} represents the estimated adjusted RDR for the p th risk factor (adjusted for all other risk factors included in the RC model). Values close to one imply small levels of measurement error and values closer to zero imply greater levels of measurement error.

The second procedure replaces observed risk factors in the disease model with the linear predictors from the RC model. These linear predictors are conditional expectations of the usual risk factor, for example, using estimated regression coefficients from equation (6):

$$\text{E}[X_{ip}|W_{i0}] = \hat{a}_p + \hat{b}_{p1}W_{i01} + \hat{b}_{p2}W_{i02} + \dots + \hat{b}_{pP}W_{i0P} \quad \text{for } i = 1, \dots, N \quad (14)$$

The associations between $\text{E}[X_{ip}|W_{i0}]$ and risk of disease directly provide estimates $\hat{\beta}$ and \hat{v} (see Section 3.1).

4.2. Practical use of correction matrices and conditional expectations

Correction Matrices and Conditional Expectations can be created using the study-specific, averaged or empirical Bayes coefficients defined in Section 3.3. Using study-specific and empirical Bayes Correction Matrices will always require a two-stage estimation approach (e.g. the first stage corrects the observed study-specific risk associations using the study-specific Correction Matrix, and the second stage combines the corrected study-specific risk associations using meta-analysis techniques). Empirical Bayes Correction Matrices that incorporate between-person heterogeneity have no practical use: instead, we propose the use of empirical Bayes Conditional Expectations. Study-specific Conditional Expectations, averaged Correction Matrices/Conditional Expectations

and empirical Bayes Conditional Expectations can be used in both single- and two-stage estimation approaches (see Section 2.2).

4.3. Time-dependent measurement error corrections

It is likely that repeat measures of risk factors are measured at different times for different studies and possibly different individuals and risk factors. Our models are generalizable to such data. However, as the time separation between baseline and repeat measures increases, the strength of the relationship may decrease [8] resulting in the need to use time-dependent corrections [2]. This can be investigated by including a time interaction with the baseline risk factor W_{si0p} in the RC model for W_{sirp} (for $r=1, \dots, R_s$). One may also wish to consider time interactions with the other baseline risk factors, although such relationships may be of less importance.

Let t_{sir} denote the time at which the r th repeat measure was made for individual i in study s . We propose the time-dependent RC model defined by:

$$\begin{aligned} W_{sirp} = & a_{srp} + a_p^{\text{time}} t_{sir} + (b_{pp} + u_{spp} + b_{pp}^{\text{time}} t_{sir}) W_{si0p} \\ & + \sum_{\substack{q=1 \\ q \neq p}}^P (b_{pq} + b_{pq}^{\text{time}} t_{sir}) W_{si0q} + z_{sip} + e_{sirp} \end{aligned} \quad (15)$$

Note that this model is incompatible with model (2) and requires the ‘current usual level’ model (3). It may also be appropriate to consider incorporating between-study random effects on the b_{pp}^{time} interaction term.

Time-dependent Correction Matrices or Conditional Expectations can simply be extracted as previously described for various blocks of follow-up time (e.g. 0–5, 5–10, 10–15 years, etc.) [2, 8]. It is then possible either to apply the time-dependent Correction Matrices to the corresponding risk associations estimated in each of these time blocks or to estimate the risk associations in each time block directly using the corresponding Conditional Expectations. However, time-dependent Conditional Expectations can be used more flexibly, and time need not be categorized into such blocks. Usual levels of the error-prone risk factors at time t can be estimated by replacing t_{sir} by t in (15) and entered into the Cox model (3).

4.4. Error-prone binary risk factors

Our RC models have been built on a normality assumption for continuous variables. Measurement error can also exist in binary variables (such as smoking status), but the problem is that errors are usually correlated with the true value, invalidating one of the underlying assumptions of regression calibration (e.g. if the true value of a binary variable is 0 then the error is 0 or 1, while if the true value is 1, then the error is 0 or -1). However, correction for measurement error in binary *confounders* (whose coefficients are not of interest) need not address this problem [26], under the further assumption that the true *exposure* of interest (e.g. in our case, usual plasma fibrinogen) is uncorrelated with the error in the binary confounder.

For our illustration, binary smoking status is used in the RC models as if it were continuous and treated as an error-prone binary confounder. In a single study, such corrections can give valid risk associations in the disease model for the exposures, but over-correct the risk associations for the binary confounder (although a further correction factor can be applied [26]). We also

investigated using Conditional Expectations estimated from a univariate logistic RC model; this produced similar results and is not reported here.

4.5. Allowing for the uncertainty in the RC model

It is uncommon in practice to allow for the uncertainty in the RC model in the corrected risk associations. Suggested approaches have focused on using bootstrapping [27] which is fairly computer intensive. We propose that instead of extracting a best estimate of study-specific, averaged, or empirical Bayes regression coefficients, one draws a set of M plausible regression coefficients which add random noise to the estimated regression coefficients and incorporate the residual and appropriate random effects error. The set of regression coefficients can be used to provide a set of Correction Matrices or Conditional Expectations. For each set, corrected risk associations, say Q_k , and corresponding variances, say W_k (for $k=1, \dots, M$), are obtained using standard procedures. We regard this as a form of multiple imputation [28] although in our case the ‘missing data’ are the true regression coefficients, not the values of the true exposure variables. The risk associations are then pooled using rules derived by Rubin [28]:

$$\text{Pooled risk association } Q^* = \frac{1}{M} \sum_{k=1}^M Q_k$$

with

$$\text{Var}(Q^*) = Z + \left(1 + \frac{1}{M}\right) B$$

where

$$Z = \frac{1}{M} \sum_{k=1}^M Z_k \quad \text{‘within-imputations variance’}$$

and

$$B = \frac{1}{M-1} \sum_{k=1}^M (Q_k - Q^*)^2 \quad \text{‘between-imputations variance’}$$

The between-imputations variance inflates the variance of the risk associations to account for the uncertainty in the RC model. In our application we use $M=5$ [28].

5. APPLICATION TO FSC

5.1. Description of the studies, risk factors and repeat measures

Our analyses are restricted to data from 27 779 individuals from five studies with repeat measurements on fibrinogen, systolic blood pressure and smoking (Table I). In total 16 529 fibrinogen, 35 629 systolic blood pressure and 33 512 smoking status repeat measures were available from a total of 12 926 individuals at various time intervals spanning roughly 15 years in the 5 studies (Figure 1 and Table I). Three studies provided multiple repeat measures and for such studies we identified each measurement as belonging to repeat 1, repeat 2, etc. according to cut-off times

Table I. Five prospective studies of plasma fibrinogen and coronary heart disease (CHD) in general populations: characteristics of studies and individuals with repeat measures.

Study	N	Baseline characteristics				No. of repeat measures			
		No. of CHD cases	Per cent Male	Median (IQR) Age in years	Mean (SD) Fibrinogen (g/L)	Systolic Blood Pressure (mmHg)	Per cent Current Smokers	Fibrinogen	Systolic blood pressure
Cardiovascular Health Study	3961	418	39	71 (68–76)	3.2 (0.6)	136 (21)	13	3197	21750
Northwick Park Heart Study 1	2367	174	73	48 (41–55)	3.0 (0.6)	139 (22)	47	1693	1876
Prospective CV Munster Study	9749	150	79	47 (41–53)	2.6 (0.5)	131 (18)	35	3541	3503
StrongHeart Study	3821	205	40	54 (49–61)	3.0 (0.7)	127 (19)	35	5882	6068
Whitehall II Study	7881	200	68	48 (45–52)	2.5 (0.6)	121 (14)	23	2216	2394
Total	27779	1155					16529	35629	33512

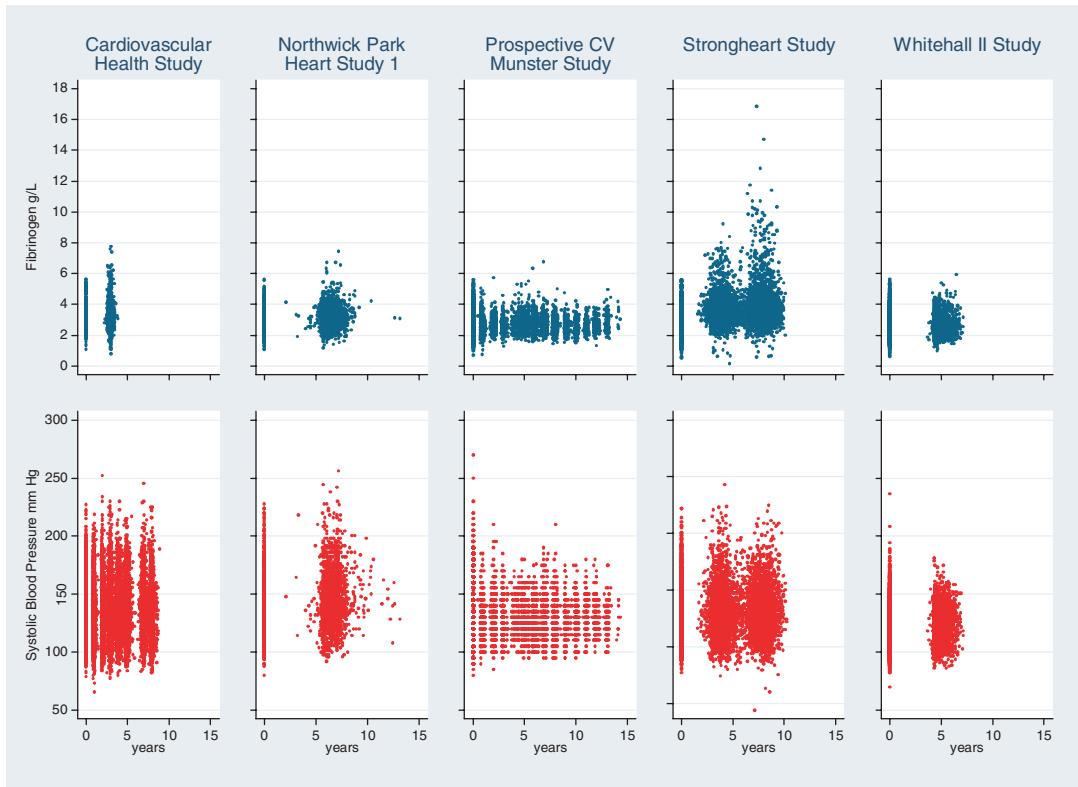


Figure 1. Levels and timings of repeat measures of fibrinogen and systolic blood pressure by study. Each point represents the observed measurements per person-visit. Repeat measurements for smoking status not shown, but follow similar timings to systolic blood pressure.

selected by inspection of Figure 1. Repeat measurements were available from surviving individuals who were not lost to follow-up. Individuals with repeat measures were generally younger, and somewhat more likely to be women and non-smokers than individuals without repeat measures [13].

The analysis involved a random-effects Cox proportional hazards model (see Section 2.2), stratified by sex. The median follow-up time was 8 years (inter-quartile range 7–10 years). Adjusted hazard ratios and 95 per cent confidence intervals between CHD and the baseline risk factors, fibrinogen, systolic blood pressure, smoking and age are provided in Table II (row 1).

5.2. Measurement error correction using a single repeat measure

Figure 2 presents the estimated study-specific RC coefficients using the first repeat measures of fibrinogen, systolic blood pressure and smoking (except in the Cardiovascular Health Study, when the third repeat measures of systolic blood pressure and smoking were used to correspond to the first repeat measure of fibrinogen). The meta-analysis combined adjusted RDRs for fibrinogen, systolic blood pressure and smoking status are 0.52 (95 per cent CI: 0.46, 0.59), 0.56 (95 per cent CI: 0.53–0.58) and 0.76 (95 per cent CI: 0.70–0.81) respectively, shown as diamonds in the main

Table II. Adjusted hazard ratios (95 per cent confidence intervals), stratified by cohort and sex, for coronary heart disease with various measurement error corrections using conditional expectations estimated from 5 studies using a single repeat measure.

Measurement error correction method	Fibrinogen per 1-g/L increase	Systolic blood pressure per 10- mmHg increase	Smoking status	
			0=non- ex-smoker	1= current smoker
Uncorrected	1.47 (1.29, 1.68)	1.18 (1.13, 1.22)	1.67 (1.41, 1.98)	2.05 (1.69, 2.50)
Univariate RC model				
Study specific*	1.96 (1.48, 2.59) (1.48, 2.60)	1.31 (1.22, 1.39) (1.22, 1.39)	1.89 (1.43, 2.49) (1.43, 2.49)	1.85 (1.57, 2.18) (1.56, 2.18)
allowing for uncertainty in RC coefficients				
Averaged†	1.93 (1.50, 2.49) (1.47, 2.55)	1.30 (1.22, 1.39) (1.21, 1.39)	1.88 (1.49, 2.38) (1.47, 2.40)	1.84 (1.50, 2.25) (1.50, 2.25)
allowing for uncertainty in RC coefficients				
Empirical Bayes‡	1.94 (1.48, 2.53) (1.47, 2.55)	1.30 (1.22, 1.40) (1.21, 1.40)	1.90 (1.49, 2.42) (1.48, 2.44)	1.84 (1.51, 2.26) (1.51, 2.26)
allowing for uncertainty in RC coefficients				
Multivariate RC model				
Study specific‡	1.98 (1.50, 2.62) (1.49, 2.63)	1.30 (1.22, 1.39) (1.22, 1.40)	1.89 (1.43, 2.49) (1.43, 2.50)	1.84 (1.56, 2.17) (1.55, 2.18)
allowing for uncertainty in RC coefficients				
Averaged§	1.93 (1.50, 2.48) (1.47, 2.53)	1.30 (1.22, 1.39) (1.21, 1.40)	1.88 (1.49, 2.38) (1.47, 2.41)	1.84 (1.50, 2.25) (1.50, 2.25)
allowing for uncertainty in RC coefficients				
Empirical Bayes§	1.93 (1.48, 2.52) (1.47, 2.54)	1.30 (1.22, 1.39) (1.22, 1.40)	1.90 (1.49, 2.42) (1.48, 2.43)	1.84 (1.51, 2.25) (1.51, 2.26)
allowing for uncertainty in RC coefficients				

Conditional expectations estimated from the following models:

*3 × 5 independent univariate RC models defined by equation (7).

†3 independent univariate RC models defined by equations (7) and (12).

‡a single multivariate RC model defined by equation (7).

§a single multivariate RC model defined by equations (7) and (12).

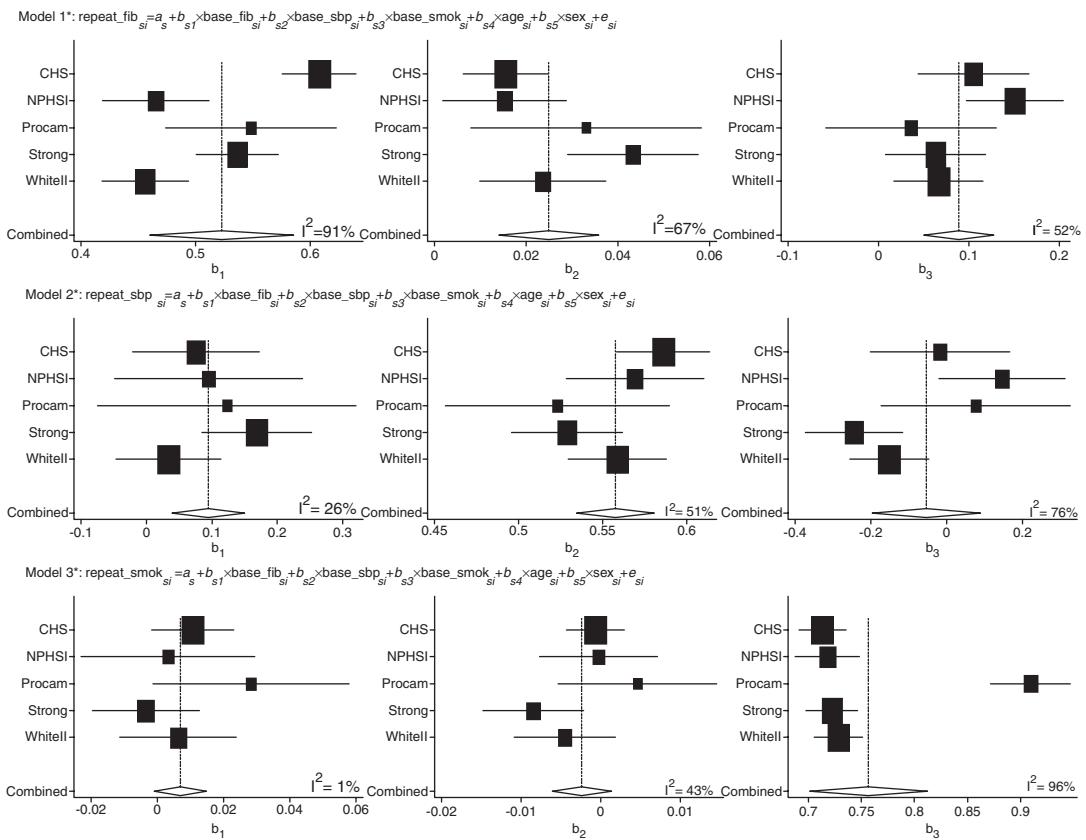


Figure 2. Meta-analysis of the study-specific regression coefficients estimated from independent univariate RC models using a single repeat measure. repeat_var and base_var represent the repeat and baseline measurements, respectively, for variables fibrinogen (fib), systolic blood pressure tenths (sbp) and smoking status (smok). The RC models are also adjusted for baseline age and sex. CHS = Cardiovascular Health Study, NPHSI = Northwick Heart Park Study 1, Procam = Prospective CV Munster Study, Strong = StrongHeart Study and WhiteII = Whitehall II Study.

diagonal of Figure 2. Between-study heterogeneity is assessed in terms of I^2 , the percentage of variance in the estimated regression coefficients from each study that is attributable to between-study variation as opposed to sampling variation [29]. Values of I^2 close to 0 per cent correspond to lack of heterogeneity. Substantial between-study heterogeneity exists between the adjusted RDRs for fibrinogen and smoking status (Figure 2). In particular, the adjusted RDR for smoking differs greatly between the Prospective CV Munster study and other studies; a random-effects model may be inappropriate here but is fitted for illustrative purposes. Regression coefficients of repeat exposures on baseline confounders are comparatively small, although some heterogeneity between studies exists. These findings support our proposed simplifications to the model described in Section 3.4. Similar estimates for the averaged RDRs were obtained from a single multivariate model defined by equation (7).

Table II displays the adjusted hazard ratios for CHD using Conditional Expectations constructed from various RC models and a random-effects Cox proportional hazards disease model. The hazard

ratios for fibrinogen, systolic blood pressure and smoking status increase after correcting for measurement error, whereas the hazard ratio for age decreases slightly. The results across the different RC models are similar. Slightly greater hazard ratios for fibrinogen are obtained from the study-specific corrections. Accounting for the uncertainty in the RC model had very little effect on the confidence intervals for the hazard ratios. Results from a fixed-effect Cox proportional hazards disease model were similar (not shown).

5.3. Measurement error corrections using multiple repeat measures

Table III displays adjusted hazard ratios for CHD using Conditional Expectations constructed from various RC models, using all repeat measures from the 5 studies. The corrected hazard ratios for fibrinogen, systolic blood pressure and smoking are generally higher than those shown in Table II, which used only a single repeat, measured within approximately seven years of follow-up, to estimate the Conditional Expectations. This is because the estimated RDRs were generally lower when later repeat information was used due to increased variability in the risk factors over time (see Figure 3). The effect of accounting for uncertainty in these models was not considered as we expected an even smaller effect than that shown in Table II due to increased sample sizes.

The cross-classified models produced an average RDR for fibrinogen of 0.48 (95 per cent CI: 0.43, 0.53). The corresponding between-study heterogeneity standard deviations were 0.02 (95 per cent CI: 0, 0.05) and the corresponding within-study heterogeneity standard deviations were 0.07 (95 per cent CI: 0.03, 0.09). Ignoring the within-study heterogeneity gave similar findings. For example, the univariate model produced an averaged RDR for fibrinogen of 0.50 (95 per cent CI: 0.44, 0.55) with corresponding between-study heterogeneity standard deviation of 0.06 (95 per cent CI: 0.01, 0.10). The different models also produced similar RDR results for systolic blood pressure and smoking.

Between-study heterogeneity is reflected in the different hazard ratios produced by study-specific, averaged and empirical Bayes measurement error corrections. A similar pattern exists in the corrected hazard ratios across the three general RC models: hazard ratios corrected using empirical Bayes Conditional Expectations tend to be higher for fibrinogen, but lower for systolic blood pressure and smoking status.

Similar results were observed from univariate RC models which included between-study random effects on b_{spq} terms in equation (8) (i.e. replacing equation (12) with equation (10)), and by replacing the within-subject random effects with within-study random effects (i.e. excluding the w_{sip} term in equation (8) but including the v_{srpp} term in equation (11)).

5.4. Allowing for time trends

Figure 3 suggests declines in the RDRs over time for fibrinogen, systolic blood pressure and smoking status, although most RDRs estimated after 5 years of follow-up were dominated by the Prospective CV Munster study. The averaged RDRs for fibrinogen decreased by 0.025 (95 per cent CI: 0.017, 0.032) per year, thus from these fitted values, the mean averaged RDR for fibrinogen was 0.60 at 1 year, declining to 0.38 at 10 years. Similarly, the averaged RDRs for systolic blood pressure and smoking decreased by 0.025 (95 per cent CI: 0.019, 0.028) and 0.039 (95 per cent CI: 0.037, 0.41) per year respectively.

Allowing for study-specific time trends in the error corrections (see Section 4.3) reduced the hazard ratio for current usual fibrinogen to 1.64 (1.09, 2.48) (Table IV). This inappropriate estimate is the result of an estimated study-specific decline of 0.16 (95 per cent CI: -0.14, 0.46) per year

Table III. Adjusted hazard ratios (95 per cent confidence intervals), stratified by cohort and sex, for coronary heart disease with various measurement error corrections using conditional expectations estimated from 5 studies using multiple repeat measures.

Measurement error correction method	Fibrinogen per 1-g/L increase	Systolic blood pressure per 10-mmHg increase	Smoking status 0 = non- or ex-smoker 1 = current smoker	Age per 10 year increase
<i>Univariate RC model: cross-classified</i>				
Study-specific*	2.06 (1.52, 2.79)	1.32 (1.23, 1.42)	1.87 (1.41, 1.49)	1.85 (1.57, 2.18)
Averaged†	2.05 (1.57, 2.69)	1.32 (1.23, 1.41)	1.82 (1.43, 2.33)	1.86 (1.53, 2.27)
Full empirical Bayes†	2.13 (1.67, 2.70)	1.27 (1.16, 1.38)	1.62 (1.25, 2.10)	1.90 (1.57, 2.31)
<i>Univariate RC model: excluding within-study random effects</i>				
Study-specific‡	2.09 (1.56, 2.80)	1.31 (1.22, 1.41)	1.89 (1.42, 2.52)	1.84 (1.56, 2.17)
Averaged§	2.05 (1.57, 2.69)	1.32 (1.23, 1.41)	1.82 (1.43, 2.33)	1.86 (1.53, 2.27)
Full empirical Bayes§	2.12 (1.64, 2.74)	1.27 (1.17, 1.38)	1.64 (1.27, 2.12)	1.90 (1.57, 2.31)
<i>Multivariate RC model: excluding within-study random effects</i>				
Study-specific¶	2.16 (1.55, 3.00)	1.31 (1.22, 1.40)	1.84 (1.39, 2.45)	1.84 (1.56, 2.17)
Averaged	2.05 (1.57, 2.69)	1.32 (1.23, 1.41)	1.82 (1.43, 2.33)	1.86 (1.53, 2.27)
Full empirical Bayes	2.16 (1.70, 2.75)	1.24 (1.16, 1.33)	1.53 (1.22, 1.91)	1.93 (1.61, 2.31)

Conditional expectations estimated from the following models:

* 3×5 cross-classified-independent univariate RC models defined by equations (8) and (11).

† 3×5 cross-classified-independent univariate RC models defined by equations (8), (11) and (12).

‡ 3×5 independent univariate RC models defined by equations (8) and (11) ignoring the v_{srpp} term.

§ 3×5 independent univariate RC models defined by equations (8), (11) and (12) ignoring the v_{sypp} term.

¶ 5 multivariate RC models defined by equations (8) and (11) ignoring the v_{srpp} term.

|| a single multivariate RC model defined by equation (13) ignoring the v_{srpp} term.

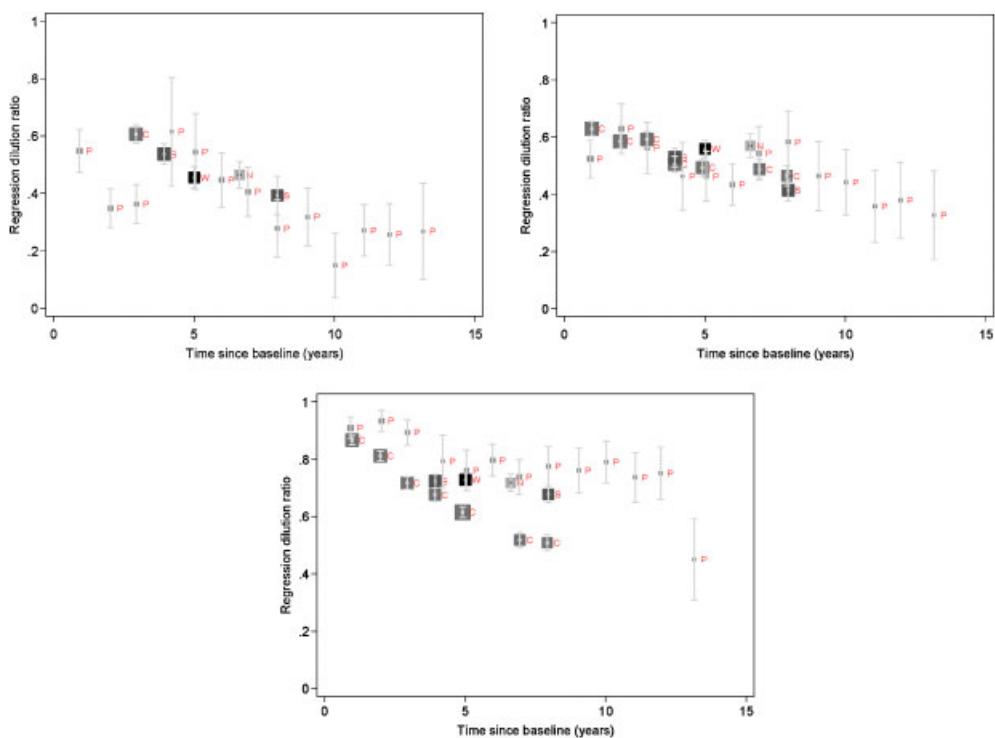


Figure 3. Regression dilution ratios for: (a) fibrinogen; (b) systolic blood pressure; and (c) smoking status with 95 per cent CI, estimated independently for each study and repeat measurement time. C=Cardiovascular Health Study, N=Northwick Heart Park Study 1, P=Prospective CV Munster Study, S=StrongHeart Study and W=Whitehall II Study. Each block represents the estimated study- and time-specific RDR estimated from the model defined by equation (6), with baseline fibrinogen, systolic blood pressure, smoking, age and sex included in each model. Each repeat is represented by the mean time since baseline. The relative sizes of the blocks are proportional to the inverse of the RDR standard error.

in RDRs for fibrinogen within the Cardiovascular Health Study, calculated from a single repeat measured over a relatively short time period (Figure 1). Using averaged or empirical Bayes measurement error corrections appropriately weights the study-specific declines, providing estimates calculated across the full follow-up period and accounting for between-study heterogeneity. These hazard ratios relate to the usual current values of the risk factors, and are similar to those shown in Table III. Incorporating between-study random effects on the b_{pp}^{time} interaction term produced similar results.

6. DISCUSSION

The major complications arising in meta-analyses of observational studies are heterogeneity between studies and measurement error in exposures and confounders. We have addressed these

Table IV. Adjusted hazard ratios (stratified by cohort and sex) for coronary heart disease estimated from a time-dependent Cox model using time-dependent univariate conditional expectations estimated from 5 studies with multiple repeat measures.

Measurement error correction method	Fibrinogen per 1-g/L increase	Systolic blood pressure per 10-mmHG increase	Smoking status 0 = non-smoker or ex-smoker 1 = current smoker	Age per 10 year increase
<i>Univariate RC model: excluding within-study random effects</i>				
Study-specific*	1.64 (1.09, 2.48)	1.27 (1.20, 1.33)	1.90 (1.40, 2.58)	1.91 (1.60, 2.27)
Averaged†	2.09 (1.57, 2.79)	1.30 (1.23, 1.39)	1.79 (1.34, 2.39)	1.88 (1.53, 2.30)
Full empirical Bayes†	2.23 (1.72, 2.87)	1.23 (1.16, 1.31)	1.51 (1.18, 1.94)	1.93 (1.61, 2.32)

* 3×5 independent univariate RC models defined by a modified equation (15), replacing terms $b_{pp} + u_{spp}$ with b_{spp} .

†3 independent univariate RC models defined by equation (15).

complications together. We have described and compared various approaches for correcting for multivariate measurement error in the setting of IPD meta-analyses. The approaches are illustrated on repeat measures from 12 926 participants in 5 prospective studies from the FSC. The main findings from these data are: (i) there can be substantial between-study heterogeneity in RDRs, as seen by comparing RC model coefficients, (ii) there is likely to be little advantage of using a multivariate RC model over separate univariate models for each error-prone variable, (iii) RC model uncertainty has only a small impact and (iv) time trends are likely to exist in the RC models, but subsequently allowing for such trends has relatively little effect on the disease associations. We discuss these findings below in turn.

Heterogeneity: As shown in our example, there can be considerable heterogeneity in the RDRs between different studies, and ignoring such differences can lead to biased estimates [12]. Such biases may remain even after allowing for between-study heterogeneity in the disease model; hence, it is appropriate to allow for heterogeneity in the RDR estimates. We allowed for between-study heterogeneity in the RC model by using a random-effects model; alternatives include allowing one study to have a completely different RC model. In our example, incorporating further study-specific random effects on other predictors in the RC models had negligible effect on the corrected hazard ratios and was practically difficult to implement. Allowance for within-study heterogeneity was less important, although this may be of greater relevance for a meta-analysis of studies with larger numbers of repeat measures.

Multivariate vs univariate RC models: We compared corrections from multivariate and univariate RC models to deal with multiple error-prone risk factors. While we were unable to fit the full proposed cross-classified multivariate model (equation (13)), we did explore various simplifications, all of which produced similar results suggesting a robustness to different model assumptions.

Multi-level multivariate RC models offer added flexibility by allowing different risk factor measurement error variances to be correlated within studies, and allowing within-subject variances to be correlated across risk factors. These are plausible assumptions, as some studies may employ more rigorous methods to reduce measurement errors and other sources of within-person variation, and some individuals may have consistently higher (or lower) levels of correlated risk factors (e.g. higher systolic blood pressure and higher body mass index). However, our example has shown that allowing for this can have little effect on corrected hazard ratios. This may not always be true in practice, especially for highly correlated risk factors (e.g. high density lipoprotein and low density lipoprotein cholesterol), and while multi-level univariate RC models are easier to implement in standard software (e.g. SAS, STATA), some checking and comparisons with the multivariate approach may be necessary.

Uncertainty in the RC model: We have shown that accounting for the uncertainty in the estimates from the RC models has minimal effect on the confidence intervals for the corrected hazard ratios. Allowing for the uncertainty was expected to make most difference in the study-specific corrections, because the standard errors from the RC models tended to be larger. However, the uncertainty in the RC models is small in comparison with the uncertainty in the estimated coefficients in the disease model.

Time trends: Our proposed correction methods assume that disease risk depends either on a ‘usual level’ or on a ‘current usual level’ of exposure and confounders. In our example, despite there being significant time trends in the RDRs, such trends had relatively little effect on the estimated hazard ratios. This may be because CHD was a relatively rare outcome (typically 10 per cent) and because the distribution of repeat measures across the study period was similar

to that of events. Time trends in the RDRs are also important when considering hazard ratios within subgroups which have different durations of follow-up, such as those formed by age at risk [2].

Limitations: In this paper we focused primarily on the additive RC model, which has an underlying assumption about homogeneity of variance with respect to the usual values. It is not uncommon for risk factors to have a measurement error variance that increases with level [13, 30] which would lead to a non-linear relationship between repeat measures. Taking log-transformations or including suitable interactions or quadratic terms in the RC model may be required. RC corrections in linear regression are valid without any assumption of linearity between repeat measures [8], but similar results hold only approximately for non-linear regression [11].

Similarly, we have assumed linear associations between exposure (and confounders) and disease, but in some cases the appropriate disease models may include interactions or non-linear terms. If such terms are known, or can be assumed, then conditional expectations can be estimated from appropriate RC models. However, assessing the shape of an exposure–disease relationship in the presence of measurement error is not straightforward. The measurement error can make the relationship appear more linear and standard methods do not allow for this [31]. Further, we have not considered the possibility that disease risk may depend on the past history of the exposure or confounders, rather than their current levels. If, for example, the risk of disease depends on the temporal rate of change in the exposure, then RC corrections are known to typically overcorrect [20]; life course methods would be more appropriate, although they have greater data requirements [32].

Multivariate measurement error correction, as undertaken in this paper, aims to estimate more closely the aetiological association of risk factors with disease. However, such corrections cannot correct for unmeasured confounders, and the potential for residual confounding will usually remain in practice. An entirely different but complementary approach to estimating aetiological associations is Mendelian randomization [33], but this may also have limitations in practice [33–35].

7. CONCLUSIONS

The methods described in this paper have general applicability to other IPD meta-analyses. Preliminary data checks should assess the quality of the information on repeat measures (e.g. comparing baseline measures between studies with and without repeat measures and between individuals with and without repeat measures) and the assumptions underlying regression calibration (RC) methods. We recommend the use of empirical Bayes conditional expectations extracted from an RC model to encompass between-study heterogeneity. Checks should be performed to justify any model simplifications, such as ignoring the within-study heterogeneity or ignoring the multivariate structure. While the multivariate regression RC model is preferable to a set of univariate RC models, the latter is more computationally convenient. Even when repeat measures are available from all studies (as in our illustration), the use of study-specific corrections may result in bias from smaller or outlying studies. Combining information across studies strengthens the reliability and precision of the estimated RC coefficients.

APPENDIX: EXAMPLE STATA CODE TO FIT HIERARCHICAL UNIVARIATE MODEL

Suppose that the data are formatted in the following way:

Study	id	repeat	repeat_fib	base_fib	base_sbp	base_smok	Age	Sex
1	1	1	3.7	3.3	123	0	76	0
1	1	2	3.3	3.3	123	0	76	0
1	1	3	3.6	3.3	123	0	76	0
1	2	1	2.8	2.3	142	1	48	0
1	2	2	3.1	2.3	142	1	48	0
1	3	1	1.9	2.9	147	0	52	1
1	4	1	3.0	2.8	128	1	66	1
1	4	2	3.2	2.8	128	1	66	1
:	:	:	:	:	:	:	:	:
2	1	1	5.4	4.1	110	1	67	1
2	2	1	3.6	3.2	115	0	70	1
2	2	2	3.8	3.2	115	0	70	1
:	:	:	:	:	:	:	:	:

and we wish to fit the following model:

$$\text{repeat_fib}_{si} = a_{sr} + (b_{11} + u_{s11})\text{base_fib}_{si0} + b_{12}\text{base_sbp}_{si0} + b_{13}\text{base_smok}_{si0} \\ + c_1\text{age}_{si} + c_2\text{sex}_{si} + z_{si} + \epsilon_{si}$$

This is a simplified version of equation (13) including non-error-prone confounders. The STATA code and results are shown below:

```
xi: xtmixed repeat_fib i.study*i.repeat base_fib base_sbp  
base_smok age sex ||  
study: base_fib, noconstant || id:
```

```
i.study      _Istudy_1--5    (naturally coded; _Istudy_1 omitted)  
i.repeat      _Irepeat_1--13   (naturally coded; _Irepeat_1 omitted)  
Mixed-effects REML regression          Number of obs=16529
```

Group Variable	No. of Groups	Minimum	Observations per Group	Average	Maximum
Study	5	1693	3305.8	5882	
id	11787	1	1.4	13	

```
Wald chi2(24)=1151.88  
Log restricted-likelihood=-17858.492          Prob>chi2=0.0000
```

repeat_fibCoefficient	Std. Err.	z	P> z	[95 per cent Confidence Interval]
_Istudy_2 0.8041103	0.2102337	3.82	0.000	0.3920598 1.216161
_Istudy_3 1.271994	0.1977429	6.43	0.000	0.8844249 1.659563

_Istudy_4	0.257616	0.2087866	1.23	0.217	-0.1515982	0.6668303
_Istudy_5	0.3751409	0.205475	1.83	0.068	-0.0275827	0.7778645
_Irepeat_2	-0.5561327	0.404178	-1.38	0.169	-1.348307	0.2360417
_Irepeat_3	-0.2681136	0.4038413	-0.66	0.507	-1.059628	0.5234007
_Irepeat_4	-0.1557177	0.36144	-0.43	0.667	-0.8641271	0.5526918
_Irepeat_5	-0.0534753	0.4074761	-0.13	0.896	-0.8521138	0.7451632
_Irepeat_6	-0.1237561	0.4066654	-0.30	0.761	-0.9208056	0.6732934
_Irepeat_7	-0.1187755	0.4066415	-0.29	0.770	-0.9157782	0.6782271
_Irepeat_8	-0.224244	0.4072799	-0.55	0.582	-1.022498	0.5740099
_Irepeat_9	-0.2488897	0.4077776	-0.61	0.542	-1.048119	0.5503396
_Irepeat_10	-0.3015269	0.4078587	-0.74	0.460	-1.100915	0.4978615
_Irepeat_11	-0.2141387	0.4080348	-0.52	0.600	-1.013872	0.5855948
_Irepeat_12	-0.198489	0.4083559	-0.49	0.627	-0.9988519	0.6018739
_Irepeat_13	-0.0237129	0.4095358	-0.06	0.954	-0.8263883	0.7789624
_IstuXre_4_2	0.3080618	0.0478986	6.43	0.000	0.2141823	0.4019412
_IstuXre_4_3	-0.128406	0.4089363	-0.31	0.754	-0.9299064	0.6730945
base_fib	0.497483	0.0271787	18.30	0.000	0.4442137	0.5507522
base_sbp	0.0021802	0.0003385	6.44	0.000	0.0015167	0.0028437
base_smok	0.1248417	0.0142951	8.73	0.000	0.0968237	0.1528596
age	0.0056616	0.0009317	6.08	0.000	0.0038354	0.0074878
sex	0.1040671	0.0132093	7.88	0.000	0.0781773	0.1299569
_cons	0.7814809	0.3736456	2.09	0.036	0.0491489	1.513813

Random-effects Parameters	[95 per cent			
	Estimate	Std. Err.	Conf. Interval]	
study: Identity sd(base_fib)	.0553705	.0230456	.0244908	.1251855
id: Identity sd(_cons)	.3813449	.0087969	.3644872	.3989822
sd(Residual)	.6135454	.0053735	.6031034	.6241682

The overall RDR for fibrinogen from this model, given by the coefficient of base_fib, is 0.50 (95 per cent CI 0.44, 0.55). Between-study heterogeneity in the RDR, given by the standard deviation sd(base_fib), is 0.06 (95 per cent CI 0.02, 0.13), suggesting that the RDR in different studies differ more than would be expected by chance. The amount of individual-specific variation and residual error are represented by their standard deviation estimates 0.38 (95 per cent CI 0.36, 0.40) and 0.61 (95 per cent CI 0.60, 0.62), respectively.

ACKNOWLEDGEMENTS

The Fibrinogen Studies Collaboration (FSC) is supported by Special Project Grant 002/02 from the British Heart Foundation. A variety of sources have supported recruitment, follow-up, and laboratory measurements in the 31 cohorts contributing to the FSC. Investigators from several of these studies have contributed to a list naming some of these funding sources, which can be found at <http://www.phpc.cam.ac.uk/MEU/FSC/Studies.html>.

Authors/Writing Committee of the Fibrinogen Studies Collaboration:

A. M. Wood, I. R. White, S. G. Thompson.

Authors/Members of the Fibrinogen Studies Collaboration:

Aspirin Myocardial Infarction Study: J. B. Kostis, A. C. Wilson; *Atherosclerosis Risk in Communities Study*: K. Wu; *Bezafibrate Infarction Prevention Study*: M. Benderly, U. Goldbourt; *Bruneck Study*: J. Willeit, S. Kiechl; *Caerphilly Study*: J. W. G. Yarnell, P. M. Sweetnam, P. C. Elwood; *Cardiovascular Health Study*: M. Cushman, R. P. Tracy (see <http://chs-nhlbi.org> for acknowledgments); *Copenhagen City Heart Study*: A. Tybjærg-Hansen; *European Concerted Action on Thrombosis and Disabilities (ECAT) Angina Pectoris Study*: F. Haverkate, S. G. Thompson; *Edinburgh Artery Study and Edinburgh Claudication Study*: A. J. Lee, F. B. Smith; *Finnish National Risk Factor Survey 1992, Hemostasis Study*: V. Salomaa, K. Harald, V. Rasi, P. Jousilahti, J. Pekkanen; *Framingham Study*: R. D'Agostino, P. W. F. Wilson, G. Tofler, D. Levy; *GISSI-Prevenzione Trial*: R. Marchioli, F. Valagussa (deceased); *Göteborg 1913 and Göteborg 1933 studies*: A. Rosengren, G. Lappas, H. Eriksson; *Göttingen Risk Incidence and Prevalence Study*: P. Cremer, D. Nagel; *Honolulu Heart Program*: J. D. Curb, B. Rodriguez, K. Yano; *Kuopio Ischaemic Heart Disease Study*: J. T. Salonen, K. Nyysönen, T.-P. Tuomainen; *Malmö Study*: B. Hedblad, G. Engström, G. Berglund; *MONICA/KORA Augsburg Study*: H. Loewel, H. W. Hense; *Northwick Park Heart Study I*: T. W. Meade, J. A. Cooper, B. De Stavola, C. Knottenbelt; *Northwick Park Heart Study II*: G. J. Miller (deceased), J. A. Cooper, K. A. Bauer, R. D. Rosenberg; *Osaka Study*: S. Sato, A. Kitamura, Y. Naito, H. Iso; *Platelet Activation and Inflammation Study*: V. Salomaa, K. Harald, V. Rasi, E. Vahtera, P. Jousilahti, T. Palosuo; *Prospective Epidemiological Study of Myocardial Infarction*: P. Ducimetiere, P. Amouyel, D. Arveiler, A. E. Evans, J. Ferrieres, I. Juhan-Vague, A. Bingham; *Prospective Cardiovascular Münster Study*: H. Schulte, G. Assmann; *Quebec Cardiovascular Study*: B. Cantin, B. Lamarche, J.-P. Després, G. R. Dagenais; *Scottish Heart Health Study*: H. Tunstall-Pedoe, G. D. O. Lowe, M. Woodward; *Speedwell Study*: Y. Ben-Shlomo, G. Davey Smith; *Strong Heart Study*: V. Palmieri, J. L. Yeh; *Thrombosis Prevention Trial*: T. W. Meade, P. Brennan, C. Knottenbelt, J. A. Cooper; *Physicians' Health Study*: P. Ridker; *Vicenza Thrombophilia and Atherosclerosis Project*: F. Rodeghiero, A. Tosetto; *West of Scotland Coronary Prevention Study*: J. Shepherd, G. D. O. Lowe, I. Ford, M. Robertson; *Whitehall II Study*: E. Brunner, M. Shipley; *Zutphen Elderly Study*: E. J. M. Feskens, D. Kromhout.

Authors/Coordinating Centre of the Fibrinogen Studies Collaboration:

E. Di Angelantonio, S. Kaptoge, S. Lewington, G. D. O. Lowe, N. Sarwar, S. G. Thompson, M. Walker, S. Watson, I. R. White, A. M. Wood, J. Danesh (coordinator).

Conflict of interest: none declared.

The following FSC investigators contributed data to the current study but did not participate as co-authors: A. R. Folsom, L. Chambless, B. M. Psaty, M. P. M. de Maat, F. G. R. Fowkes, E. Vahtera, W. B. Kannel, L. Wilhelmsen, W. Koenig, A. Rudnicka.

REFERENCES

- Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models*. Chapman & Hall: London, 1995.
- Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; **360**:1903–1913.

3. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, Abbott R, Godwin J, Dyer A, Stamler J. Blood pressure, stroke, and coronary heart disease. Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; **335**:765–774.
4. Law M, Wald N, Wu T, Hackshaw A, Bailey A. Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: data from the BUPA study. *British Medical Journal* 1994; **308**:363–366.
5. Phillips AN, Davey Smith G. How independent are independent effects? Relative risk estimation when correlated exposures are measured imprecisely. *Journal of Clinical Epidemiology* 1991; **44**:1223–1231.
6. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology* 1990; **132**:734–745.
7. Rosner B, Spiegelman D, Willett W. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American Journal of Epidemiology* 1992; **136**:1400–1413.
8. Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, Peto R. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *American Journal of Epidemiology* 1999; **150**:341–353.
9. Wong MY, Day NE, Bashir SA, Duffy SW. Measurement error in epidemiology: the design of validation studies. I: univariate situation. *Statistics in Medicine* 1999; **18**:2815–2829.
10. Rosner B, Willett W, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* 1989; **8**:1051–1069.
11. Hughes MD. Regression dilution in the proportional hazards model. *Biometrics* 1993; **49**:1056–1066.
12. Carroll RJ, Stefanski LA. Measurement error, instrumental variables and corrections for attenuation with applications to meta-analysis. *Statistics in Medicine* 1994; **13**:1265–1282.
13. Fibrinogen Studies Collaboration. Regression dilution methods for meta-analysis: assessing long-term variability in plasma fibrinogen among 27 247 adults in 15 prospective studies. *International Journal of Epidemiology* 2006; **35**:1570–1578.
14. Woodward M, Barzi F, Martiniuk A, Fang X, Gu D, Imai Y, Lam T, Pan W, Rodgers A, Suh I, Jee SH, Ueshima H, Huxley R. Cohort profile: the Asia Pacific Cohort Studies Collaboration. *International Journal of Epidemiology* 2006; **35**:1412–1416.
15. The Fibrinogen Studies Collaboration. Collaborative meta-analysis of prospective studies of plasma fibrinogen and cardiovascular disease. *European Journal of Cardiovascular Prevention and Rehabilitation* 2004; **11**:9–17.
16. The Emerging Risk Factors Collaboration. Analysis of individual data on lipid, inflammatory and other markers in over 1.1 million participants in 104 prospective studies of cardiovascular diseases. *European Journal of Epidemiology* 2007; **22**(12):839–869.
17. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel-Chapelon F, Thiébaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, de Mesquita HBB, Peeters PHM, Lund E, Engeset D, González CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci R. European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. *Public Health Nutrition* 2002; **5**:1113–1124.
18. Lowe GDO, Rumley A, Mackie IJ. Plasma fibrinogen. *Annals of Clinical Biochemistry* 2004; **41**:430–440.
19. The Fibrinogen Studies Collaboration. Plasma fibrinogen and the risk of major cardiovascular diseases and non-vascular mortality: an individual participant meta analysis. *JAMA* 2005; **294**:1799–1809.
20. Frost C, White IR. The effect of measurement error in risk factors that change over time in cohort studies: do simple methods over-correct for regression-dilution? *International Journal of Epidemiology* 2005; **34**:1359–1368.
21. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624.
22. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal* 1994; **309**:1351–1355.
23. Normand SL. Meta-analysis: formulating, evaluating, combining and reporting. *Statistics in Medicine* 1999; **18**:321–359.
24. Goldstein H. *Multilevel Statistical Models* (3rd edn). Edward Arnold: London, 2003.
25. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). Wiley: Hoboken, NJ, 2002.
26. White IR, Frost C, Tokunaga S. Correcting for measurement error in binary and continuous variables using replicates. *Statistics in Medicine* 2001; **20**:3441–3457.
27. Carpenter J. Test inversion bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B* 1991; **61**:159–172.

28. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
29. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *British Medical Journal* 2003; **327**:557–560.
30. Thompson SG, Pocock SJ. The variability of serum cholesterol measurements: implications for screening and monitoring. *Journal of Clinical Epidemiology* 1990; **43**:783–789.
31. Frost C, Thompson SG. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society, Series A* 2000; **163**:173–189.
32. De Stavola BL, Nitsch D, dos Santos Silva I, McCormack V, Hardy R, Mann V, Cole T, Morton S, Leon D. Statistical issues in life course epidemiology. *American Journal of Epidemiology* 2006; **163**:84–96.
33. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environment in complex diseases. *Lancet* 2001; **358**:1356–1360.
34. Smith GD, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**:1–22.
35. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *American Journal of Epidemiology* 2006; **163**:397–403.