

# Multivariate multilevel spline models for parallel growth processes: application to weight and mean arterial pressure in pregnancy

Corrie Macdonald-Wallis,<sup>a,\*†</sup> Debbie A. Lawlor,<sup>a</sup> Tom Palmer<sup>a</sup>  
and Kate Tilling<sup>b</sup>

Growth models are commonly used in life course epidemiology to describe growth trajectories and their determinants or to relate particular patterns of change to later health outcomes. However, methods to analyse relationships between two or more change processes occurring in parallel, in particular to assess evidence for causal influences of change in one variable on subsequent changes in another, are less developed. We discuss linear spline multilevel models with a multivariate response and show how these can be used to relate rates of change in a particular time period in one variable to later rates of change in another variable by using the variances and covariances of individual-level random effects for each of the splines. We describe how regression coefficients can be calculated for these associations and how these can be adjusted for other parameters such as random effect variables relating to baseline values or rates of change in earlier time periods, and compare different methods for calculating the standard errors of these regression coefficients. We also show that these models can equivalently be fitted in the structural equation modelling framework and apply each method to weight and mean arterial pressure changes during pregnancy, obtaining similar results for multilevel and structural equation models. This method improves on the multivariate linear growth models, which have been used previously to model parallel processes because it enables nonlinear patterns of change to be modelled and the temporal sequence of multivariate changes to be determined, with adjustment for change in earlier time periods. Copyright © 2012 John Wiley & Sons, Ltd.

**Keywords:** linear spline; multivariate; multilevel model; structural equation model; longitudinal; change

## 1. Introduction

Growth models are often used in life course epidemiology to describe patterns of growth or change, to assess the influences of rates and timing of change on later health outcomes and to investigate the effects of exposures on patterns of change [1–6]. In some cases two or more change processes occur in parallel, such as gains in height and weight during childhood, and it may be of interest to assess relationships between such processes, or to adjust for the effects of one change process when assessing the associations of another with a later health outcome. It may also be hypothesised that change in one process causes a change in another, or that an unmeasured or latent construct causes change in both of the processes simultaneously. Bivariate multilevel models (MLMs) for repeated measurements with linear relationships of both outcome variables with time have been used to demonstrate the correlation between values of the two variables at baseline and rates of change in the variables over time [7–9]. However, this does not demonstrate whether change in one variable temporally precedes change in another and is thus unable to provide evidence towards a causal effect.

<sup>a</sup>MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, U.K.

<sup>b</sup>School of Social and Community Medicine, University of Bristol, U.K.

\*Correspondence to: Corrie Macdonald-Wallis, MRC CAiTE Centre, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, U.K.

†E-mail: c.macdonald-wallis@bristol.ac.uk

Linear spline models have recently been proposed as a method of representing growth [5] or change [10], which reduces the dimensionality of repeated measurements. The shape of the trajectory of change is assumed to be piecewise linear, with knot points defining changes in the magnitude or direction of association of the response variable with time. The selection of the number and location of knot points may be determined by the data or by prior knowledge and therefore the linear splines represent the shape of the change trajectory in a meaningful way, and the linear periods of change provide coefficients that are more interpretable than for polynomial curves. By first fitting splines to each change process and then modelling these processes in parallel, we are able to model associations between observed rates of change in one variable in one period of time and observed rates of change in another variable in a subsequent period of time, with time periods defined by the splines, and thus we may identify whether changes in one variable precede changes in another. By examining the temporal sequence of events there is the possibility of improving causal inferences for pathways between variables. Furthermore, prior knowledge or hypotheses about which pathways exist may be tested by constraining covariances between individual-level random effects for different periods of change to be zero and comparing the model fit to models where these covariances are freely estimated.

We will use the example of weight and mean arterial pressure (MAP) during pregnancy to demonstrate these models. In normal pregnancy there is a decrease in blood pressure in early pregnancy followed by a rise in late pregnancy [11]. Hypertensive disorders of pregnancy (HDP), defined by high blood pressure in late pregnancy (after 20 weeks' gestation), are associated with risk of adverse health outcomes for both the mother and offspring [12–14]. and gestational weight gain (GWG) has been found to be positively associated with the risk of developing an HDP [15–17]. However, it is not clear whether changes in weight during pregnancy precede changes in blood pressure, or whether increases in blood pressure precede GWG because of increases in oedema and plasma volume expansion [18]. MAP is the average pressure in an artery over a complete cycle of one heart beat. It is estimated by combining systolic and diastolic blood pressure and allowing for the lower pressure during the diastolic phase of the cardiac cycle (see below for description of how this is calculated). In this paper we have used MAP so that we have one blood pressure variable that we can relate to weight change in pregnancy and that takes account of blood pressure in both systole and diastole. We show how MLMs and structural equation models (SEMs) with linear splines for gestational age can be applied to weight and MAP to learn whether an increase in weight precedes a rise in MAP, and in which periods of pregnancy associations are strongest. The paper is structured as follows: in Section 2 we describe the data. In Section 3 we define the multilevel multivariate response model and the derivation of regression coefficients from the variances and covariances of random effects. In Section 4 we show that the model can be equivalently defined in SEM form. In Section 5 we provide an application of the model to weight and MAP changes in pregnancy, and compare results from fitting the model in MLM software (MLWIN) (Rasbash J, Browne W, Healy M, Cameron B, Charlton C, Centre for Multilevel Modelling, University of Bristol, UK) and SEM software (MPLUS) (Muthen and Muthen, Los Angeles, California). Section 6 is a discussion of the methods described.

## 2. Data

The data are from the Avon Longitudinal Study of Parents and Children (ALSPAC), which is described in full elsewhere [19] and on the website <http://www.bris.ac.uk/alspac>. A total of 14,541 pregnant women were recruited, who were living in a defined area of Avon including the city of Bristol during their pregnancy and had an expected delivery date between 1 April 1991 and 31 December 1992. Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and from the National Health Service local ethics committee. We restrict analysis to singleton live term births ( $\geq 37$  weeks' gestation) with no evidence of pre-eclampsia (defined according to the International Society for the Study of Hypertension in Pregnancy criteria [20]) or a previous diagnosis of hypertension ( $N = 11,650$ ). All weight and blood pressure measurements, which were taken routinely as part of antenatal care by midwives or obstetricians, were abstracted from obstetric records by six trained research midwives. This resulted in a median of 14 (interquartile range 11 to 16) blood pressure measurements and median of 12 (interquartile range 10 to 14) weight measurements per woman. MAP was calculated as  $1/3 \times$  systolic blood pressure  $+ 2/3 \times$  diastolic blood pressure. Confounding variables considered here are maternal height, age, parity, education, smoking and offspring sex. Maternal age and offspring sex were obtained from obstetric records and maternal height, parity, highest educational qualification and smoking status were obtained from questionnaires administered during pregnancy. Smoking status was classified as

**Table I.** Characteristics of the dataset ( $N = 11,650$ ) and the subset with complete data on all covariates ( $N = 9429$ ).

Maternal characteristic	$N$ with data	Mean (SD) or %	Mean (SD) or % in complete case
<b>Height (cm)</b>	10,278	164.0 (6.69)	164.1 (6.65)
<b>Age (yrs)</b>	11,650		
< 20		4.84	3.45
20 – 24		19.30	16.66
25 – 29		38.66	39.56
30 – 34		27.60	29.95
35+		9.60	10.38
<b>Parity</b>	10,830		
Nulliparous		44.56	44.63
Multiparous		55.44	55.37
<b>Smoking in pregnancy</b>	10,945		
Never		66.20	68.69
Prepregnancy/first trimester		13.85	13.41
Throughout		19.95	17.90
<b>Highest educational qualification</b>	10,492		
CSE/vocational		29.89	28.10
O level		34.51	34.99
A level		22.57	23.35
Degree		13.03	13.55
<b>Sex of offspring</b>	11,650		
Male		51.12	50.79
Female		48.88	49.21

‘never’, ‘immediately prepregnancy or first trimester only’, or ‘throughout pregnancy’ for women who continued to smoke after the first trimester. Table I shows the characteristics of the women included in the analysis. Linear spline random effects models have previously been fitted to weight measurements in pregnancy, to model changes in weight with gestational age [21]. A similar method was used to fit a linear spline random effects model to MAP (see Section 5).

### 3. Multilevel spline model with a multivariate response

#### 3.1. Multivariate response linear spline random effects model

Multilevel models for repeated measurements have been described by Goldstein [22] and others. We will first consider a univariate MLM for repeated measurements of a response variable  $y$ , measured on occasions,  $j = 1, \dots, r_k$  for each individual  $k = 1, \dots, n$ , with two levels: *measurement occasion* $_j$  within *individual* $_k$ . The response variable is modelled as a linear function of time and the intercept and linear slopes are allowed to vary at the individual level by including individual-level random effects for these parameters as follows:

$$y_{jk} = \beta_{0k} + (\beta_{1k} + u_{1k})t_{jk} + u_{0k} + \varepsilon_{0jk} \quad (1)$$

where  $y_{jk}$  is the value of the response variable,  $y$ , at the  $j^{\text{th}}$  measurement for the  $k^{\text{th}}$  individual,  $t_{jk}$  is the time of the  $j^{\text{th}}$  measurement for the  $k^{\text{th}}$  individual,  $u_{0k}$  and  $u_{1k}$  are the individual level (level-2) residuals for the intercept and slope respectively and the  $\varepsilon_{0jk}$ s are the measurement occasion level (level-1) residuals. These parameters have the following assumptions:

$$\begin{aligned} \begin{bmatrix} u_{0k} \\ u_{1k} \end{bmatrix} &\sim N(0, \Omega_u), & \Omega_u &= \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix} \\ \varepsilon_{0jk} &\sim N(0, \Omega_\varepsilon), & \Omega_\varepsilon &= \sigma_{\varepsilon0}^2 \end{aligned} \quad (2)$$

A bivariate response model to investigate linear change in two variables  $y^{(i)}$  for  $i = 1, 2$  over time has three levels: *responses* $^{(i)}$  within *measurement occasion* $_j$  within *individual* $_k$  and thus a model with

a random intercept and random slope at the individual level is defined by the following formula and assumptions [22]:

$$y_{jk}^{(1)} = \beta_0^{(1)} + (\beta_1^{(1)} + u_{1k}^{(1)})t_{jk} + u_{0k}^{(1)} + \varepsilon_{0jk}^{(1)}$$

$$y_{jk}^{(2)} = \beta_0^{(2)} + (\beta_1^{(2)} + u_{1k}^{(2)})t_{jk} + u_{0k}^{(2)} + \varepsilon_{0jk}^{(2)}$$

$$\begin{bmatrix} u_{0k}^{(1)} \\ u_{1k}^{(1)} \\ u_{0k}^{(2)} \\ u_{1k}^{(2)} \end{bmatrix} \sim N(0, \Omega_u), \quad \Omega_u = \begin{bmatrix} \sigma_{u0}^{2(1)} & & & \\ \sigma_{u01}^{(1)} & \sigma_{u1}^{2(1)} & & \\ \sigma_{u00}^{(1)(2)} & \sigma_{u10}^{(1)(2)} & \sigma_{u0}^{2(2)} & \\ \sigma_{u01}^{(1)(2)} & \sigma_{u11}^{(1)(2)} & \sigma_{u01}^{(2)} & \sigma_{u1}^{2(2)} \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} \varepsilon_{0jk}^{(1)} \\ \varepsilon_{0jk}^{(2)} \end{bmatrix} \sim N(0, \Omega_\varepsilon), \quad \Omega_\varepsilon = \begin{bmatrix} \sigma_{\varepsilon 0}^{2(1)} & \\ \sigma_{\varepsilon 00}^{(1)(2)} & \sigma_{\varepsilon 0}^{2(2)} \end{bmatrix}$$

To extend this to a model where time is represented by a set of  $m^{(i)}$  linear splines for response ( $i$ ) where, for  $m^{(i)}$  splines there are  $m^{(i)} + 1$  time points,  $t_0^{(i)} < t_1^{(i)} < \dots < t_{m^{(i)}}^{(i)}$  then

$$s_{ljk}^{(i)} = 0 \quad \text{if } t_{jk} \leq t_{l-1}^{(i)}$$

$$s_{ljk}^{(i)} = (t_{jk} - t_{l-1}^{(i)}) \quad \text{if } t_{l-1}^{(i)} < t_{jk} \leq t_l^{(i)} \quad \text{for } l = 1, \dots, m^{(i)}; i = 1, 2 \quad (4)$$

$$s_{ljk}^{(i)} = (t_l^{(i)} - t_{l-1}^{(i)}) \quad \text{if } t_{jk} > t_l^{(i)}$$

Hence, a linear spline model with a bivariate response has the form

$$y_{jk}^{(i)} = \beta_0^{(i)} + \sum_{l=1}^{m^{(i)}} (\beta_l^{(i)} + u_{lk}^{(i)}) s_{ljk}^{(i)} + u_{0k}^{(i)} + \varepsilon_{0jk}^{(i)} \quad \text{for } i = 1, 2$$

$$\begin{bmatrix} u_{0k}^{(1)} \\ \vdots \\ \vdots \\ \vdots \\ u_{m^{(2)}k}^{(2)} \end{bmatrix}_{(m^{(1)}+m^{(2)}+2) \times 1} \sim N(0, \Omega_u), \quad \Omega_u = \begin{bmatrix} \sigma_{u0}^{2(1)} & & & & \\ \sigma_{u01}^{(1)} & \sigma_{u1}^{2(1)} & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \sigma_{u0m^{(2)}}^{(1)(2)} & \dots & \dots & \dots & \sigma_{um^{(2)}}^{2(2)} \end{bmatrix}_{(m^{(1)}+m^{(2)}+2) \times (m^{(1)}+m^{(2)}+2)}$$

$$\begin{bmatrix} \varepsilon_{0jk}^{(1)} \\ \varepsilon_{0jk}^{(2)} \end{bmatrix} \sim N(0, \Omega_\varepsilon), \quad \Omega_\varepsilon = \begin{bmatrix} \sigma_{\varepsilon 0}^{2(1)} & \\ \sigma_{\varepsilon 00}^{(1)(2)} & \sigma_{\varepsilon 0}^{2(2)} \end{bmatrix} \quad (5)$$

The individual-level random effects give each individual's deviance from the average intercept ( $u_{0k}$ ) and from the average slope of each of the splines ( $u_{1k}, \dots, u_{m^{(i)}k}$ ) and the individual-level covariance matrix contains the pair wise covariances between each set of these random effects for the intercept and slopes within each of the response variables and between the response variables. For simplicity we have defined a bivariate response model, but this can be extended to more than two response variables. In this case the covariance matrix at the individual level would include all variances and covariances between individual-level residuals and the covariance matrix at the occasion level would include all variances and covariances between occasion-level residuals.

Of note, the model estimated variances and covariances of individual-level random effects,  $\hat{\sigma}_{u_{l_1}^{(i_1)} u_{l_2}^{(i_2)}}^{2(i_1)}$ , are not the same as the variances and pairwise covariances of the estimated individual-level residuals produced by MLMs,  $\hat{u}_{l_1 k}^{(i_1)}, \hat{u}_{l_2 k}^{(i_2)}$ , because the residuals are shrunken towards the population mean by varying degrees, reflecting the level of uncertainty with which each individual's residual is

estimated [23]. Because the individual-level variance–covariance matrix estimated by the model provides estimates of the population variances and covariances of the random effects this should be used to assess associations between the intercepts and slopes in the population rather than the shrunken residuals.

### 3.2. Testing hypotheses about temporal relationships between response variables

The model defined in Equation (5) allows the individual-level random effects relating to the intercepts and each of the slopes of the response variables in different time periods to be correlated with each other. Specific hypotheses about time periods in which changes in each of the variables are associated may be applied by setting the covariances between the corresponding pairs of the random effects to be zero, thus simplifying the model. The hypotheses may be tested by comparing the simplified models with the full unrestricted model. The fit of the models may be compared by using a likelihood ratio test, with the difference in degrees of freedom between the models being the number of constrained covariances.

For simplicity of explanation, in the following example hypotheses it will be assumed that the same splines are used for both response variables; however, these definitions can easily be adapted to take into account time periods with different endpoints.

#### **Hypothesis 1: Changes in each of the variables are only associated in the same or adjacent time periods**

This hypothesis can be tested by fitting a model with covariances between individual-level random effects relating to change in nonadjacent time periods to be zero:

$$\sigma_{ul_1l_2}^{(i_1)(i_2)} = 0 \quad \text{for all } |l_1 - l_2| > 1; \quad l_1, l_2 \neq 0,$$

and comparing the fit of this model with the full model with no constraints. The hypothesis can be used to test whether change in one variable in one time period has long-term associations with change in the other variable, or whether it is only associated with change in the immediately subsequent time period.

#### **Hypothesis 2: Changes in variable 1 are associated with subsequent changes in variable 2, but changes in variable 2 are not associated with subsequent changes in variable 1**

For this hypothesis, in the restricted model, all covariances between random effects relating to change in variable 2 and random effects relating to change in variable 1 in a later time period are set to zero:

$$\sigma_{ul_1l_2}^{(i_1)(i_2)} = 0 \quad \text{for all } l_1 > l_2; \quad l_1, l_2 \neq 0$$

This can be used to test for evidence that changes in variable 1 precede changes in variable 2, rather than changes in variable 2 preceding change in variable 1.

#### **Hypothesis 3: There is a lag time between change in one variable and change in another**

To test this hypothesis, in the restricted model the covariance between random effects relating to change in one variable in a particular time period and random effects relating to change in the other variable in subsequent time periods is set to zero for any time period where it is hypothesised that there may be a lag, but covariances with random effects relating to change in time periods after the hypothesised lag time remain unconstrained

$$\sigma_{ul_1l_2}^{(i_1)(i_2)} = 0 \quad \text{for } l_2 - l_1 < \delta_s; \quad l_1, l_2 \neq 0,$$

where  $\delta_s$  is the number of splines/time periods that make up the hypothesised lag time.

The choice of hypotheses to be tested should be guided by the research question, which the analysis is designed to answer, and prior knowledge about the plausible relationships between the variables studied.

### 3.3. Deriving regression coefficients from the random effects variance–covariance matrix

From the covariance matrix of the individual-level random effects relating to baseline values of the response variable or rates of change in these variables in each time period described by the splines it is possible to derive regression coefficients to describe the associations between baseline values and rates

of change in each of the variables in different time periods. The regression coefficient for the change in random effect  $u_{l_1}^{(i_1)}$  associated with a unit change in random effect  $u_{l_2}^{(i_2)}$  is given by [24]

$$\beta_{l_1 l_2}^{(i_1)(i_2)} = \frac{\sigma_{ul_1 l_2}^{(i_1)(i_2)}}{\sigma_{ul_2}^{2(i_2)}} \quad i_1, i_2 = 1, 2; \quad l_1 = 0, 1, \dots, m^{(i_1)}; \quad l_2 = 0, 1, \dots, m^{(i_2)} \quad (6)$$

This formula can be extended to give regression coefficients for the relationship between two variables when adjusting for others. Fisher described how to adjust regression coefficients for additional variables [25] and expressions for regression coefficients when adjusting for up to four covariates are given in the online supplemental material 1.

In the more general case, to regress random effect  $u_{l_0}^{(i_0)}$  on  $p$  independent random effect variables:  $u_{l_1}^{(i_1)}, u_{l_2}^{(i_2)}, \dots, u_{l_p}^{(i_p)}$ , let  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$  be the  $p \times 1$  vector of regression coefficients relating to  $u_{l_1}^{(i_1)}, u_{l_2}^{(i_2)}, \dots, u_{l_p}^{(i_p)}$ , respectively, adjusting for each of the other variables. Also, let  $\sigma_0 = (\sigma_{ul_0 l_1}^{(i_0)(i_1)}, \sigma_{ul_0 l_2}^{(i_0)(i_2)}, \dots, \sigma_{ul_0 l_p}^{(i_0)(i_p)})'$  be the  $p \times 1$  vector of covariances between  $u_{l_0}^{(i_0)}$  and each of the independent variables and

$$\Sigma = \begin{bmatrix} \sigma_{ul_1}^{2(i_1)} & \sigma_{ul_1 l_2}^{(i_1)(i_2)} & \dots & \sigma_{ul_1 l_p}^{(i_1)(i_p)} \\ & \ddots & & \vdots \\ & & \ddots & \vdots \\ & & & \sigma_{ul_p}^{2(i_p)} \end{bmatrix}$$

be the  $p \times p$  matrix of variances and pairwise covariances of the independent variables. Then  $\Sigma\beta = \sigma_0$ , meaning that the mutually adjusted regressions can be found by solving the equation:

$$\beta = \Sigma^{-1}\sigma_0 \quad (7)$$

### 3.4. Standard errors of regression coefficients derived from the random effects variance–covariance matrix

We consider a number of methods for calculating the standard errors of the regression coefficients derived from the random effects variance–covariance matrix and compare the performance of these methods on the example dataset in Section 5. The first method estimates the standard errors as if we had sampled  $n$  observations of the individual-level random effects from an underlying population (where  $n$  is the number of individuals in our MLM), whereas Methods 2 and 3 estimate the standard errors by treating the regression coefficients as nonlinear combinations of the estimates of the variances and covariances of the random effects, but do this in different ways.

**Method 1:** If we treat our estimates of the variances and covariances of the random effects as if they were obtained from a sample of the individual-level random effects, then the standard errors of the regression coefficients can also be derived using formulae described by Fisher [25]. For the general case with  $p$  independent variables, where  $\beta$ ,  $\Sigma$  and  $\sigma_0$  are as in (7) above, let  $d$  be the  $p \times 1$  vector containing the diagonal elements of the inverse of  $\Sigma$ ,  $d = (\Sigma^{-1}[1, 1], \Sigma^{-1}[2, 2], \dots, \Sigma^{-1}[p, p])'$ . Then the  $p \times 1$  vector of standard errors,  $SE_\beta = (SE(\beta_1), SE(\beta_2), \dots, SE(\beta_p))'$  is given by

$$SE_\beta = \sqrt{\frac{(\sigma_{l_0}^{2(i_0)} - \beta'\sigma_0) d}{n - p - 1}} \quad (8)$$

**Method 2:** The second approach is to use the delta method to approximate the standard errors [26, 27]. This can be implemented in STATA (StataCorp LP, College Station, Texas) using the command nlcom and by explicitly defining the formulae for the regression coefficients from the estimated variances and covariances of random effects (see supporting information for formulae)<sup>‡</sup>. We note that some programs

<sup>‡</sup>Supporting information may be found in the online version of this article.

estimate variances and covariances on transformed scales [28], and implementations of the delta method are also available in R (The R Foundation for Statistical Computing, Vienna, Austria) [29]. This method is practical when adjusting for up to three or even four variables; however, because there is no convenient expression for the inverse of a square matrix of order greater than four, a method that solves the inverse of  $\Sigma$  numerically (see Method 3) is more appropriate when adjusting for more variables.

**Method 3:** The third approach is to generate a large number,  $q$ , of realisations of a  $(m^{(1)} + m^{(2)} + 2) \times 1$  vector, which we shall call  $\tilde{\sigma}$ , containing all of the variances and covariances of the random effects from the multivariate multilevel spline model (as described in Equation (5)),  $\tilde{\sigma}_h = (\tilde{\sigma}_{u0h}^{2(1)}, \tilde{\sigma}_{u01h}^{(1)}, \tilde{\sigma}_{u1h}^{2(1)}, \dots, \tilde{\sigma}_{u0m_h^{(2)}}^{(1)(2)}, \dots, \tilde{\sigma}_{um_h^{(2)}}^{2(2)})'$  for  $h = 1, 2, \dots, q$ , using a multivariate normal distribution,  $\tilde{\sigma} \sim N(\mathbf{v}, \Gamma)$ . The mean vector,  $\mathbf{v}$ , of the multivariate normal distribution is the  $(m^{(1)} + m^{(2)} + 2) \times 1$  vector of estimates of the variances and covariances of the random effects from the MLM,  $\mathbf{v} = (\hat{\sigma}_{u0}^{2(1)}, \hat{\sigma}_{u01}^{(1)}, \hat{\sigma}_{u1}^{2(1)}, \dots, \hat{\sigma}_{u0m^{(2)}}^{(1)(2)}, \dots, \hat{\sigma}_{m^{(2)}}^{2(2)})'$ , and the  $(m^{(1)} + m^{(2)} + 2) \times (m^{(1)} + m^{(2)} + 2)$  variance-covariance matrix,  $\Gamma$ , is the matrix of the variances and covariances of these estimates,

$$\Gamma = \begin{bmatrix} \text{var}(\hat{\sigma}_{u0}^{2(1)}) & & & & \\ \text{cov}(\hat{\sigma}_{u0}^{2(1)}, \hat{\sigma}_{u01}^{(1)}) & \text{var}(\hat{\sigma}_{u01}^{(1)}) & & & \\ \vdots & & \ddots & & \\ \text{cov}(\hat{\sigma}_{u0}^{2(1)}, \hat{\sigma}_{um^{(2)}}^{2(2)}) & \dots & \dots & \text{var}(\hat{\sigma}_{um^{(2)}}^{2(2)}) & \end{bmatrix}.$$

After producing  $q$  realisations of  $\tilde{\sigma}$ , the coefficients relating to the regression of one of the random effects on a subset  $p$  of the random effects from the MLM are derived for each of the realisations of  $\tilde{\sigma}$  separately. The  $h^{\text{th}}$  realisations of  $\Sigma$ ,  $\tilde{\Sigma}_h$ , and  $\sigma_0$ ,  $\tilde{\sigma}_{0h}$ , defined in (7) are formed from the estimates in  $\tilde{\sigma}_h$  and the equation  $\tilde{\beta}_h = \tilde{\Sigma}_h^{-1} \tilde{\sigma}_{0h}$  is solved for each of the  $q$  realisations. The overall estimate of the vector of regression coefficients  $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_p)$  is produced by taking the mean of the  $q$  realisations,  $\tilde{\beta}_h = (\tilde{\beta}_{1h}, \tilde{\beta}_{2h}, \dots, \tilde{\beta}_{ph})'$ , of each of these regression coefficients,  $\tilde{\beta} = \sum_{h=1}^q (\tilde{\beta}_h/q)$ . 95% confidence intervals for each of the elements of  $\tilde{\beta} : \tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_p$  are then produced by using the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of their distributions over the  $q$  realisations as the lower and upper limits, respectively.

We have uploaded STATA commands that implement Methods 2 and 3, named `reffadjust4nlcom` and `reffadjustsim`, respectively, to the Statistical Software Components archive [30].

#### 4. Comparison with a structural equation model

Structural equation models contain a measurement equation and a structural equation [31, 32]. In the measurement equation, each of the measurements  $y_j^{(i)}$  loads on the latent variables  $f_l^{(i)}$  where  $l = 0, 1, \dots, m^{(i)}$ . The intercept term for variable  $i$  is  $f_0^{(i)}$ , and the loadings for each of the measurements  $y_j^{(i)}$  on this latent variable are 1. The other  $m^{(i)}$  factors represent the slopes for each of the periods defined by the splines. Using the same  $m^{(i)} + 1$  time points,  $t_0^{(i)} < t_1^{(i)} < \dots < t_m^{(i)}$  as above, the measurement equation is

$$y_j^{(i)} = f_0^{(i)} + \sum_{l: t_l^{(i)} \leq t_j} (t_l^{(i)} - t_{l-1}^{(i)}) f_l^{(i)} + \sum_{l: t_{l-1}^{(i)} < t_j < t_l^{(i)}} (t_j - t_{l-1}^{(i)}) f_l^{(i)} + \sum_{l: t_{l-1}^{(i)} \geq t_j} 0 f_l^{(i)} + \varepsilon_j^{(i)} \quad \text{for } i = 1, 2. \tag{9}$$

Any rescaling of this equation proportional to these loadings can also be used. To see that this is equivalent to Equation (5), an extra subscript representing individuals,  $k$ , is added and  $(\beta_l^{(i)} + u_{lk}^{(i)})$  is substituted for each of the  $f_l^{(i)}$ . This gives

$$\begin{aligned}
 y_{jk}^{(i)} &= (\beta_0^{(i)} + u_{lk}^{(i)}) + \sum_{l: t_l^{(i)} \leq t_{jk}} (t_l^{(i)} - t_{l-1}^{(i)}) (\beta_l^{(i)} + u_{lk}^{(i)}) + \sum_{l: t_{l-1}^{(i)} < t_{jk} < t_l^{(i)}} (t_{jk} - t_{l-1}^{(i)}) (\beta_l^{(i)} + u_{lk}^{(i)}) \\
 &+ \sum_{l: t_{l-1}^{(i)} \geq t_{jk}} 0 (\beta_l^{(i)} + u_{lk}^{(i)}) + \varepsilon_{0jk}^{(i)} \\
 &= (\beta_0^{(i)} + u_{lk}^{(i)}) + \sum_{l=1}^{m^{(i)}} s_{ljk}^{(i)} (\beta_l^{(i)} + u_{lk}^{(i)}) + \varepsilon_{0jk}^{(i)} \quad \text{for } i = 1, 2,
 \end{aligned}
 \tag{10}$$

where  $s_{ljk}^{(i)}$  is defined as in Equation (4).

The structural equation of the SEM defines the mutual relationships between the factors. It can be shown that the regression coefficient for the univariate relationship between two factors,  $f_{l_1}^{(i_1)}$  regressed on  $f_{l_2}^{(i_2)}$  is equivalent to that between two MLM random effects  $u_{l_1}^{(i_1)}$  and  $u_{l_2}^{(i_2)}$  (given in (6)) by substituting  $(\beta_l^{(i)} + u_l^{(i)})$  for each of the  $f_l^{(i)}$  as follows:

$$\begin{aligned}
 \beta_{l_1 l_2}^{(i_1)(i_2)} &= \frac{\text{cov}(f_{l_1}^{(i_1)}, f_{l_2}^{(i_2)})}{\text{var}(f_{l_2}^{(i_2)})} = \frac{\text{cov}((\beta_{l_1}^{(i_1)} + u_{l_1}^{(i_1)}), (\beta_{l_2}^{(i_2)} + u_{l_2}^{(i_2)}))}{\text{var}(\beta_{l_2}^{(i_2)} + u_{l_2}^{(i_2)})} \\
 &= \frac{\text{cov}(u_{l_1}^{(i_1)}, u_{l_2}^{(i_2)})}{\text{var}(u_{l_2}^{(i_2)})} = \frac{\sigma_{u_{l_1 l_2}^{(i_1)(i_2)}}}{\sigma_{u_{l_2}^{(i_2)}}^2}.
 \end{aligned}
 \tag{11}$$

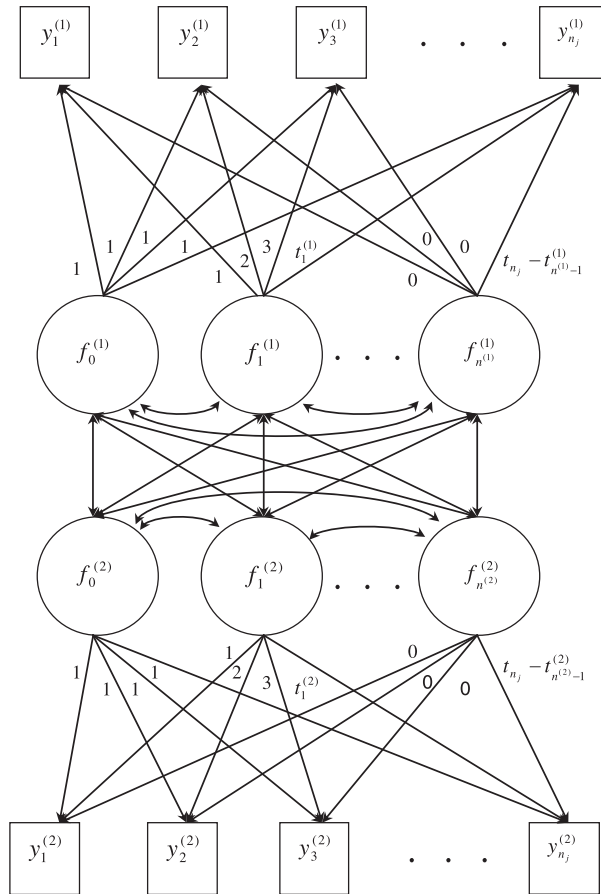
Adjusted regression coefficients for the relationships between factors and their standard errors can be shown to be equivalent to those between random effects in a similar way.

The MLM described in Equation (5) is equivalent to the SEM shown in the path diagram in Figure 1. There have been a number of papers demonstrating the equivalence of linear MLMs and SEMs when each individual has the same number of measurements occurring at the same time points meaning that the data are balanced [31–33]. It has also been shown that SEMs can be estimated equivalently to MLMs when data are unbalanced, by treating the design as balanced with missing data and using full information maximum likelihood estimation [33]. However, nonlinear MLMs and generalized linear MLMs do not always have equivalent parametrisations within the SEM framework [7, 33]. For our weight and MAP example, where individuals have different numbers of measurements and these were taken at different times for each individual, this can be modelled in an SEM by including all times when measurements were taken on any individual (rounded to whole week periods) and treating a particular individual's measurements as missing for any time points on which they were not measured.

## 5. Example: bivariate model of weight and MAP change in pregnancy

Where there was more than one blood pressure or weight measurement in a 2-week period, one was selected at random for the analysis to prevent any individual from having too high an influence on the models. This left a median of 10 and range of 1 to 18 measurements per woman for analysis. Linear spline models were fitted to weight and MAP separately to describe the shapes of the average trajectories of weight and MAP with gestational age [11, 21]. Full details of this process are provided in the online supporting information. The model that minimised the differences between predicted and actual values over pregnancy for weight had two knot points, at 18 and 28 weeks' gestation, and for MAP had three knot points, at 18, 30 and 36 weeks' gestation. For simplicity, because the first two knot point locations were similar for weight and MAP and altering the position of the second knot point by 1 week had little effect on the fit of these models, we have used knot points of 18 and 29 weeks for weight and 18, 29 and 36 weeks for MAP. Baseline was set at 8 weeks' gestation for both weight and MAP because there were few measurements prior to this and model predictions earlier in pregnancy than this may be unreliable. Gestational age was rounded to the nearest whole week to make the MLM comparable





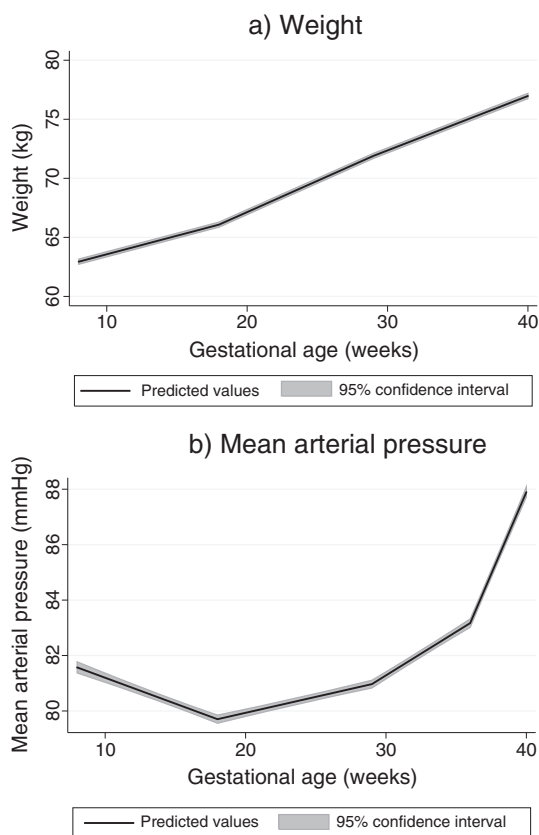
**Figure 1.** Path diagram for bivariate response spline model in structural equation format. Path diagram assumes that measurements  $y_j^{(i)}$  occur at equally spaced time points.

with the SEM. The formulae for the separate multilevel spline models for weight and MAP, without any confounders included in models were

$$\begin{aligned}
 Weight_{jk} &= 62.943 + \left(0.314 + u_{1k}^{weight}\right) s_{1jk}^{weight} + \left(0.529 + u_{2k}^{weight}\right) s_{2jk}^{weight} + \left(0.461 + u_{3k}^{weight}\right) s_{3jk}^{weight} \\
 &\quad + u_{0k}^{weight} + \varepsilon_{0jk}^{weight} \\
 MAP_{jk} &= 81.572 + \left(-0.187 + u_{1k}^{MAP}\right) s_{1jk}^{MAP} + \left(0.115 + u_{2k}^{MAP}\right) s_{2jk}^{MAP} + \left(0.314 + u_{3k}^{MAP}\right) s_{3jk}^{MAP} \\
 &\quad + \left(1.182 + u_{4k}^{MAP}\right) s_{4jk}^{MAP} + u_{0k}^{MAP} + \varepsilon_{0jk}^{MAP},
 \end{aligned}$$

where  $s_1^{weight}$  is the first gestational age spline for weight, up to 18 weeks' gestation,  $s_2^{weight}$  is the second spline, from 18 – 29 weeks, and  $s_3^{weight}$  is the third spline, from 29 weeks onwards. For MAP  $s_1^{MAP}$  is the first gestational age spline, up to 18 weeks' gestation,  $s_2^{MAP}$  is the second spline, from 18 – 29 weeks,  $s_3^{MAP}$  is the third spline, from 29 – 36 weeks and  $s_4^{MAP}$  is the fourth spline, from 36 weeks onwards. The shape of the average weight and MAP trajectories across gestation predicted from these models are shown in Figure 2, and Figure S1 shows example predicted trajectories of weight and MAP for five individual women.

We combined both of these separate spline models into one bivariate spline model, with two responses: weight and MAP. The distributions of weight and MAP at 8 weeks and rates of change in weight and MAP in each period of pregnancy predicted by this unadjusted bivariate model are shown in Table SI. We then included the potential confounders: maternal height (continuous variable); age (< 20, 20 – 24, 25 – 29, 30 – 34, 35 + years); parity (nulliparous, multiparous); smoking (never, prepregnancy/first trimester, throughout pregnancy); education (CSE/vocational, O level, A level, degree) and offspring sex (male, female). Maternal height was centred about the mean (164 cm) and the reference categories for



**Figure 2.** Average trajectories of (a) weight and (b) mean arterial pressure across pregnancy predicted by univariate multilevel linear spline models ( $N = 11,650$ ).

the other variables were: age 25 – 29 years, nulliparous, never smoked, O level qualification and male offspring. We completed analyses using STATA version 11.2 and MLWIN version 2.24 with runmlwin [34] for the MLMs and MPLUS version 5 for the SEMs.

The level 2 (individual-level) covariance matrix and correlations between individual-level random effects for this model, with all covariances between random effects freely estimated, are shown in Table II. We will call this MLM Model 1. We then considered several restricted models and tested different hypotheses about the relationships between changes in weight and MAP by comparing these restricted models to the full model (MLM Model 1). MLM Model 2 was constrained so that changes in weight and MAP could only be correlated with weight and MAP change in the same and adjacent periods, but weight and MAP at baseline were allowed to correlate with changes in weight and MAP in all periods. In MLM Model 3 MAP change was not allowed to correlate with weight change in subsequent periods of pregnancy, but weight change was allowed to correlate with later MAP changes, and in MLM Model 4 weight change was not allowed to correlate with MAP change in subsequent periods of pregnancy, but MAP change was allowed to correlate with later weight changes. These models (MLM Models 3 and 4) were used to test whether changes in weight precede changes in MAP or vice versa. In MLM Model 5, MAP change up to 18 weeks was not allowed to correlate with weight at baseline or weight changes in any period, to test whether baseline weight and changes in weight were only associated with increases in MAP after 18 weeks, but not associated with MAP change prior to 18 weeks, when it is decreasing and less likely to be influenced by HDP. Table III shows the deviances of each of the models tested. The fits of the constrained models were all much poorer than the full model (MLM Model 1) so this was selected as the model that best represented the relationships in the data and all of the alternative hypotheses were rejected in favour of this unrestricted model.

We also fitted an SEM to the data as in Figure 1 with each time point representing a week of pregnancy up to 44 weeks. The earliest week of gestation for which we had data was week 2, producing 43 time points. We included four splines of gestational age for MAP with the same knot points as in the MLM and adjusted for the same set of confounders. All intercept and slope latent variable variances

**Table II.** Covariances (standard errors) (upper right triangle and diagonal) and correlations (lower left triangle) between individual-level random effects in the full multilevel model (MLM Model 1;  $N = 9429$ ). \*

	Weight at 8 weeks	Weight change 8 – 18 weeks	Weight change 18 – 29 weeks	Weight change 29 weeks onwards	MAP at 8 weeks	MAP change 8 – 18 weeks	MAP change 18 – 29 weeks	MAP change 29 – 36 weeks	MAP change 36 weeks onwards
Weight at 8 weeks	119.611 (1.786)				25.574 (1.244)	0.387 (0.142)	-0.530 (0.093)	0.196 (0.143)	-0.435 (0.276)
Weight change 8 – 18 weeks	-0.24	0.050 (0.001)			-0.113 (0.031)	0.001 (0.004)	0.012 (0.003)	-0.012 (0.004)	0.003 (0.008)
Weight change 18 – 29 weeks	-0.16	0.28	0.033 (0.001)		-0.124 (0.023)	0.006 (0.003)	0.011 (0.002)	0.000 (0.003)	0.006 (0.005)
Weight change 29 weeks onwards	0.02	0.10	0.47	0.044 (0.001)	-0.018 (0.026)	0.006 (0.003)	0.006 (0.002)	0.021 (0.003)	0.027 (0.006)
MAP at 8 weeks	0.39	-0.08	-0.11	-0.01	36.579 (1.444)	-1.071 (0.144)	-0.278 (0.087)	0.127 (0.135)	-0.176 (0.268)
MAP change 8 – 18 weeks	0.08	0.01	0.07	0.07	-0.42	0.180 (0.020)	-0.035 (0.011)	-0.029 (0.016)	0.015 (0.031)
MAP change 18 – 29 weeks	-0.14	0.16	0.18	0.08	-0.14	-0.24	0.113 (0.010)	-0.042 (0.012)	-0.012 (0.020)
MAP change 29 – 36 weeks	0.03	-0.09	0.00	0.17	0.04	-0.12	-0.21	0.352 (0.023)	-0.108 (0.034)
MAP change 36 weeks onwards	-0.03	0.01	0.03	0.11	-0.03	0.03	-0.03	-0.16	1.314 (0.077)

\*Model is adjusted for maternal height, age, parity, smoking in pregnancy, education and offspring sex.

**Table III.** Model fit comparisons for full and restricted multilevel models adjusted for confounders ( $N = 9429$ ).

Model	Deviance	Difference in df (compared with model 1)	$\chi^2$ $p$ -value
MLM Model 1: Full model	984211.981	—	—
MLM Model 2: Changes in MAP or weight are only associated in adjacent periods of gestation	984277.574	8	< 0.001
MLM Model 3: Weight change precedes MAP change	984241.507	3	< 0.001
MLM Model 4: MAP change precedes weight change	984235.532	5	< 0.001
MLM Model 5: Weight at baseline and changes in weight are only associated with MAP changes after 18 weeks	984228.987	4	0.002

Note:

MLM Model 1: Full model with no restrictions.

MLM Model 2: Individual-level random effects for periods of MAP and weight change are correlated only with MAP and weight changes in concurrent and immediately subsequent periods, but not with later periods (although MAP and weight at 8 weeks (intercept) may correlate with MAP and weight changes in any time period).

MLM Model 3: Individual-level random effects for periods of MAP change are not correlated with weight change in later periods of gestation; weight change can correlate with later MAP changes.

MLM Model 4: Individual-level random effects for periods of weight change are not correlated with MAP change in later periods of gestation; MAP change can correlate with later weight changes.

MLM Model 5: Individual-level random effects for weight at 8 weeks and periods of weight change are not correlated with MAP change between 8 and 18 weeks (but may correlate with later MAP changes).

and covariances were freely estimated, but the residual variances for weight in each time period were constrained to be equal, the residual variances for MAP in each time period were constrained to be equal and a common residual covariance between weight and MAP at each time point was also estimated, to make equivalence with MLM Model 1. Table IV shows the covariances and correlations between the intercept and slope latent variables for each of the splines for weight and MAP estimated from the SEM in MPLUS. The covariances and correlations are very similar to the full MLM Model (MLM Model 1). We also compared the estimates of mean weight and MAP at 8 weeks' gestation (baseline) and the mean weight and MAP slopes in each period of gestation between multilevel and structural equation models that were not adjusted for confounders and these were equivalent in the MLM and SEM.

Table V shows the associations of weight at 8 weeks and GWG in each period of pregnancy with MAP changes in the same and subsequent periods of gestation, with and without adjustment for weight and MAP at baseline and earlier changes in these variables. These regression coefficients were derived using the variances and covariances of random effects from MLM Model 1, as set out in Equation (7). The three different methods for producing the standard errors described in Section 3.4 are compared in Table V. We did not use the delta method (Method 2) for regression models that adjusted for more than three variables because this method involved specifying an explicit formula for each of the regression coefficients and these quickly became extremely complex when adjusting for more variables, and impractical to implement. Confidence intervals using Method 1, which assumed the regression coefficients had been obtained from a sample of observations of the individual-level random effects, were much narrower than for Methods 2 and 3, which treated the regression coefficients as nonlinear transformations of the estimates of variances and covariances of the random effects, and these two methods produced very similar confidence intervals.

Table V shows that weight at 8 weeks' gestation was positively associated with MAP at 8 weeks and with MAP change between 8 and 18 weeks, but negatively associated with MAP change between 18 and 29 weeks. GWG between 8 and 18 weeks was not strongly associated with MAP in the same period but was positively associated with MAP change between 18 and 29 weeks and negatively associated with MAP change between 29 and 36 weeks, and these associations remained similar when adjusting

**Table IV.** Covariances (standard errors) (upper right triangle and diagonal) and correlations (lower left triangle) of intercept and slope latent variables in the full SEM ( $N = 9429$ ).

	Weight at 8 weeks	Weight change 8 – 18 weeks	Weight change 18 – 29 weeks	Weight change 29 weeks onwards	MAP at 8 weeks	MAP change 8 – 18 weeks	MAP change 8 – 29 weeks	MAP change 29 – 36 weeks	MAP change 36 weeks onwards
Weight at 8 weeks	119.620 (1.788)				25.582 (1.232)	0.386 (0.141)	-0.530 (0.092)	0.197 (0.143)	-0.436 (0.271)
Weight change 8 – 18 weeks	-0.24 (0.001)	0.050 (0.001)	0.011 (0.001)	0.005 (0.001)	-0.114 (0.032)	0.001 (0.004)	0.012 (0.003)	-0.012 (0.004)	0.002 (0.008)
Weight change 18 – 29 weeks	-0.16	0.27	0.033 (0.001)	0.018 (0.001)	-0.123 (0.023)	0.005 (0.003)	0.011 (0.002)	0.000 (0.003)	0.006 (0.005)
Weight change 29 weeks onwards	0.02	0.11	0.47	0.044 (0.001)	-0.018 (0.026)	0.006 (0.003)	0.006 (0.002)	0.021 (0.003)	0.027 (0.006)
MAP at 8 weeks	0.39	-0.08	-0.11	-0.01	36.582 (1.180)	-1.072 (0.149)	-0.278 (0.087)	0.127 (0.135)	-0.176 (0.264)
MAP change 8 – 18 weeks	0.08	0.01	0.06	0.07	-0.42	0.180 (0.020)	-0.035 (0.011)	-0.029 (0.016)	0.015 (0.031)
MAP change 18 – 29 weeks	-0.14	0.16	0.18	0.09	-0.14	-0.25	0.113 (0.010)	-0.042 (0.012)	-0.012 (0.020)
MAP change 29 – 36 weeks	0.03	-0.09	0.00	0.17	0.04	-0.12	-0.21	0.352 (0.023)	-0.108 (0.034)
MAP change 36 weeks onwards	-0.03	0.01	0.03	0.11	-0.03	0.03	-0.03	-0.16	1.314 (0.081)

\* Model is adjusted for maternal height, age, parity, smoking in pregnancy, education and offspring sex.

**Table V.** Mean differences in mean arterial pressure at 8 weeks and average changes in mean arterial pressure in each period associated with a 10 kg increase in weight at 8 weeks or 400 g/week increase in weight gain in the same or earlier periods of gestation, using variances and covariances of the random effects from MLM Model 1 ( $N = 9429$ ).

Weight variable	Average MAP change (mmHg/week)														
	MAP at 8 weeks (mmHg)			8 – 18 weeks			18 – 29 weeks			29 – 36 weeks			36+ weeks		
	Mean diff.	95% CI	Mean diff.	95% CI	Mean diff.	95% CI	Mean diff.	95% CI	Mean diff.	95% CI	Mean diff.	95% CI			
<b>Weight at 8 weeks (per 10 kg)</b>															
Method 1	2.138	(2.035, 2.241)	0.032	(0.025, 0.040)	-0.044	(-0.050, -0.038)	0.016	(0.005, 0.027)	-0.036	(-0.057, -0.015)					
Method 2	2.138	(1.944, 2.332)	0.032	(0.009, 0.056)	-0.044	(-0.059, -0.029)	0.016	(-0.007, 0.040)	-0.036	(-0.082, 0.009)					
Method 3	2.139	(1.946, 2.334)	0.032	(0.009, 0.056)	-0.044	(-0.060, -0.029)	0.017	(-0.007, 0.040)	-0.036	(-0.082, 0.009)					
<b>Average weight change (400 g/wk):</b>															
Method 1			0.011	(-0.004, 0.026)	0.095	(0.083, 0.106)	-0.094	(-0.116, -0.073)	0.020	(-0.021, 0.061)					
Method 2			0.011	(-0.046, 0.068)	0.095	(0.055, 0.134)	-0.094	(-0.156, -0.033)	0.020	(-0.101, 0.141)					
Method 3			0.011	(-0.046, 0.068)	0.094	(0.054, 0.134)	-0.095	(-0.157, -0.033)	0.020	(-0.101, 0.140)					
<b>8 – 18 weeks</b>															
Method 1			0.031	(0.017, 0.045)	0.079	(0.067, 0.091)	-0.092	(-0.114, -0.070)	0.004	(-0.039, 0.046)					
Method 2			0.031	(-0.015, 0.077)	0.079	(0.038, 0.120)	-0.092	(-0.155, -0.029)	0.004	(-0.120, 0.128)					
Method 3			0.031	(-0.015, 0.076)	0.079	(0.038, 0.120)	-0.092	(-0.156, -0.029)	0.004	(-0.122, 0.126)					
<b>18 – 29 weeks</b>															
Method 1			0.134	(0.119, 0.148)	0.134	(0.119, 0.148)	-0.004	(-0.030, 0.023)	0.075	(0.025, 0.126)					
Method 2			0.134	(0.092, 0.176)	0.134	(0.092, 0.176)	-0.004	(-0.069, 0.061)	0.075	(-0.050, 0.201)					
Method 3			0.134	(0.092, 0.176)	0.134	(0.092, 0.176)	-0.004	(-0.068, 0.063)	0.075	(-0.052, 0.198)					
<b>29+ weeks</b>															
Method 1			0.113	(0.099, 0.128)	0.113	(0.099, 0.128)	0.044	(0.017, 0.072)	0.060	(0.006, 0.113)					
Method 2			0.113	(0.072, 0.154)	0.113	(0.072, 0.154)	0.045	(-0.033, 0.128)	0.060	(-0.091, 0.207)					
Method 3			0.195	(0.172, 0.217)	0.195	(0.172, 0.217)	0.246	(0.172, 0.217)	0.246	(0.202, 0.290)					
Method 1			0.195	(0.140, 0.249)	0.195	(0.140, 0.249)	0.246	(0.140, 0.249)	0.246	(0.142, 0.350)					
Method 2			0.195	(0.140, 0.250)	0.195	(0.140, 0.250)	0.246	(0.140, 0.250)	0.246	(0.144, 0.350)					
Method 3			0.265	(0.240, 0.289)	0.265	(0.240, 0.289)	0.288	(0.240, 0.289)	0.288	(0.238, 0.338)					
Method 1			0.265	(0.200, 0.331)	0.265	(0.200, 0.331)	0.289	(0.200, 0.331)	0.289	(0.151, 0.427)					

All coefficients are adjusted for maternal height, age, parity, smoking, education and offspring sex (by including these in the bivariate multilevel model); shaded cells are also adjusted for weight and MAP at 8 weeks and weight and MAP changes in periods prior to the exposure period (by including these random effects in the regression).  
 Method 1: Regression coefficients and standard errors calculated as if we had a sample of individual-level random effects, using Equations (7) and (8).  
 Method 2: Regression coefficients and standard errors produced using the delta method, implemented by nlcom in Stata.  
 Method 3: Regression coefficients produced by averaging over 10,000 realisations of the variance-covariance matrix of random effects, with 95% confidence intervals formed from the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the distribution of the regression coefficients over the 10,000 generated matrices.

for weight and MAP at baseline. GWG between 18 and 29 weeks was positively associated with MAP changes in the same period but not in subsequent periods and GWG from 29 weeks onwards was positively associated with MAP changes between 29 and 36 weeks and from 36 weeks onwards, in models with and without adjustment for weight and MAP at baseline and earlier changes in these variables.

We also considered associations of MAP at 8 weeks and MAP changes with subsequent changes in weight, to examine whether there was evidence that MAP increases may influence greater weight gain for example through increased oedema (Table SII). In models that were unadjusted for baseline weight and earlier GWG and MAP changes, MAP at 8 weeks was negatively associated with GWG in all three periods of gestation and MAP changes between 8 and 18 weeks, and between 18 and 29 weeks were positively associated with GWG in all subsequent periods. However, after adjustment for weight at 8 weeks, there was little evidence that MAP at 8 weeks was associated with changes in weight across pregnancy, except for a weak association with GWG between 18 and 29 weeks. MAP change between 8 and 18 weeks was not associated with GWG between 18 and 29 weeks but was weakly positively associated with GWG from 29 weeks onwards in models adjusted for weight at 8 weeks and GWG and MAP changes in earlier periods of gestation. MAP change between 18 and 29 weeks was not associated with subsequent GWG in fully adjusted models.

## 6. Discussion

We have shown that linear splines for two or more change processes can be used in a multivariate response MLM to investigate associations between rates of change in adjacent periods and to determine whether change in one time period in one variable is associated with change in another variable in the next time period, with time periods defined by the splines. It is also possible to adjust for the value of one or both of the variables at baseline or changes in the variables in earlier time periods.

We compared several methods of deriving standard errors for the regression coefficients for the associations between changes in different time periods and found that standard errors using Method 1, which acted as though we had a sample of individual-level random effects rather than estimates of their variances and covariances, produced confidence intervals that were too narrow and overestimated the evidence against the null hypothesis. The delta method is an appropriate way of estimating the variance of linear and nonlinear transformations of parameter estimates [27], and incorporated the variance of the estimates of the variances and covariances of random effects in the standard errors of the regression coefficients derived from these. Using percentiles of the distribution of the regression coefficients to combine estimates over 10,000 simulated datasets produced comparable confidence intervals for these regression coefficients to those produced using the delta method. However, it should be noted that these methods (Methods 2 and 3) both assume that the distributions of the variances and covariances of the random effects have a multivariate normal distribution, and so should only be used for large samples where this approximation is justified.

We presented regression coefficients for the associations of pregnancy weight and MAP at baseline and rates of change in these variables with subsequent rates of change in the other variable, both with and without adjustment for weight and MAP at baseline and changes in earlier periods of pregnancy for comparison. This was to assess whether associations between pregnancy weight and MAP change were confounded by earlier changes or baseline values. In general, careful consideration is needed to identify which variables to adjust for to tease out evidence for causal relationships and this decision should be motivated by a particular hypothesis to be tested or by prior knowledge of the relationships between the variables studied.

Previous studies have used covariation in the random effects of two variables in MLMs, either measured at a single point in time [35] or longitudinally by examining the covariance between random intercept and random slope terms in linear growth models [7, 8] to investigate relationships between two variables or rates of change in these variables measured in the same individuals. However, the advantage of using linear splines is that the timings of associations can be identified. For this example analysis we were able to assess whether changes in weight or MAP preceded changes in the other variable, and whether this occurred in all periods of pregnancy or only in early or late pregnancy. This is not possible with a bivariate linear growth model. The linear splines also have the advantage of representing nonlinear patterns of change in an interpretable way, so that the correlations between periods of linear change are more meaningful than correlations between random effects for squared or cubic terms in polynomial models would be. We could also extend the model to include categorical outcomes, provided that the

assumption of normally distributed random effects was not violated by including these variables. It is also possible to include different sets of splines for each of the variables if patterns of change in the variables are different.

Cross-lagged SEMs have been used effectively with panel data to demonstrate possible causal relationships between two or more variables measured at several points in time [36,37] and Granger investigated causal relationships between two time series by assessing whether earlier values of each of the variables predicted later observations of the other variable [38]. The synchronous common factor model may be used when a latent variable is thought to influence the two measured processes simultaneously [39]. These models have the flexibility to include prior information by adding constraints and removing unlikely paths from the models. However, the cross-lagged panel design relies on observations being made at the same point in time for each individual and for each measure, and these models are restricted to researcher-determined timing of measurements, rather than time points that may be more biologically meaningful.

Linear MLMs can be equivalently estimated using the SEM framework with both balanced and unbalanced data structures [31–33] but many nonlinear and generalised linear MLMs cannot be equivalently estimated as SEMs [7,33]. The MLM approach is a more efficient method of fitting linear models with unbalanced data because within the MLM framework it is not necessary for data to be measured for all individuals at all time points, provided that missingness is at random, whereas in SEMs missing data must be dealt with in some way [31] such as by full information maximum likelihood estimation. Also, the data management and code become complicated in an SEM when individuals have different numbers of measurements and these occur at different times for each individual, and the models are computationally intensive if there are many possible time points. Furthermore, it is possible to fit a wider range of hierarchical models within the MLM framework [31]. However, a particular advantage to the SEM framework is its high flexibility to specify complex covariance structures, which may not fit into standard MLMs, and the ability to include latent outcome or predictor variables, which are measured by a set of indicator variables [32]. It is also possible to investigate mediation by estimating direct and indirect effects within the SEM framework and obtain tests of overall model fit, which are not available in MLM software [33]. Curran suggested that there was no benefit to using an SEM instead of a standard MLM, but that SEMs could be used to incorporate additional model features that could not be included in an MLM [32]. The multivariate spline MLMs described here could be extended within the SEM framework to produce growth mixture models from the multivariate splines, where patterns of change in two or more variables are used as indicators of latent class membership to extract subgroups of the population with similar change trajectories.

The MLMs and maximum likelihood procedure used in the SEM models to deal with missing data both assume that the data are missing at random, meaning that the missingness in the data can be explained by observed variables [40]. If there is reason to suspect that the missing data mechanism is nonignorable, there are methods available to test the sensitivity of inferences to data being not missing at random for multivariate longitudinal data [41].

We have demonstrated that these models can be fitted in both MLM and SEM frameworks. The SEM took longer to converge than the MLM, because of the low coverage of the data. All of the women had far fewer weight and MAP observations than the number of time periods used and the SEM treated each time period in which the individual did not have a measurement as a missing value that needed to be estimated using maximum likelihood estimation. However, despite the difficulties with estimation in the SEM, the variance–covariance matrix of the latent intercept and slope factors produced by this model was very similar to that of the individual-level random effects produced by the MLM.

Convergence can also be an issue in MLMs, especially if there are many splines and therefore many variances and covariances of individual level random effects to be estimated. If the model does not converge or calculates negative variances for some of the random effects, it is not possible to derive regression coefficients from these. The choice of the number of splines to use for each variable should depend on the pattern of change in the variable over time, and should be sufficient to achieve an acceptable fit to the observations. It is possible to include a greater number of splines if there is a larger sample size and a high number of measurements per individual, and convergence may be easier if the variables are measured at the same time points for each individual and fewer covariates are included in the model. Difficulties in achieving convergence could be overcome by adding one random effect at a time to the model and reestimating using initial values from the previous model, using fewer splines, or by reducing the accuracy of the timings of the observations, for example rounding gestational age to whole weeks rather than the nearest day, because this reduces the variability in measurement times between



individuals. The complexity of the model fitted in our example appeared to be close to the limit of what is possible with this dataset, as in some of our constrained models convergence became difficult.

In summary, we found that GWG up to 18 weeks' gestation was associated with a greater increase in MAP in the next period of pregnancy (18 – 29 weeks), but associated with a smaller rise in MAP later on (29 – 36 weeks). GWG between 18 and 29 weeks was positively associated with concurrent increases in MAP but not with MAP in later periods, and GWG from 29 weeks onwards was also associated with concurrent MAP. There was some evidence that MAP change was also related to later GWG, with a weak positive association between MAP change up to 18 weeks and GWG from 29 weeks onwards. This method could be applied to many other areas where it is of interest whether changes in one variable precede changes in another, and in what time periods these relationships exist and are strongest. For example it could be used to investigate whether changes in weight are related to changes in respiratory function or bone mass or whether changes in socioeconomic position over time precede changes in a wide range of health and social outcomes.

## Acknowledgements

We are extremely grateful to all of the families who took part in this study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The authors thank the editor and anonymous reviewers for their useful suggestions, which helped to considerably improve the original manuscript. This work was supported by the UK Wellcome Trust [grant number WT087997MA], which funds the salary of C M-W, and the UK Medical Research Council [grant number G1000726]. Core support for ALSPAC is provided by the UK Medical Research Council, the Wellcome Trust and the University of Bristol. The UK Medical Research Council provides funding for the MRC Centre for Causal Analyses in Translational Epidemiology [grant number G0600705], including the salary for TP.

## References

1. Pan H, Goldstein H. Multi-level repeated measures growth modelling using extended spline functions. *Statistics in Medicine* 1998; **17**(23):2755–2770.
2. McCarthy A, Hughes R, Tilling K, Davies D, Davey SG, Ben-Shlomo Y. Birth weight; postnatal, infant, and childhood growth; and obesity in young adulthood: evidence from the Barry Caerphilly growth study. *American Journal of Clinical Nutrition* 2007; **86**(4):907–913.
3. Fraser A, Hughes R, McCarthy A, Tilling K, Davies D, Rumley A, Lowe GDO, Davey SG, Ben-Shlomo Y. Early life growth and hemostatic factors. *American Journal of Epidemiology* 2008; **168**(2):179–187.
4. Ben-Shlomo Y, McCarthy A, Hughes R, Tilling K, Davies D, Davey SG. Immediate postnatal growth is associated with blood pressure in young adulthood - The Barry Caerphilly growth study. *Hypertension* 2008; **52**(4):638–644.
5. Howe LD, Tilling K, Galobardes B, Davey SG, Gunnell D, Lawlor DA. Socioeconomic differences in childhood growth trajectories: at what age do height inequalities emerge? *Journal of Epidemiology and Community Health* 2012; **66**(2):143–148.
6. Howe LD, Tilling K, Galobardes B, Davey SG, Ness AR, Lawlor DA. Socioeconomic disparities in trajectories of adiposity across childhood. *International Journal of Pediatric Obesity* 2010; **6**(2-2):e144–53.
7. MacCallum RC, Kim C, Malarkey WB, KiecoltGlaser JK. Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research* 1997; **32**(3):215–253.
8. Suvak MK, Walling SM, Iverson KM, Taft CT, Resick PA. Multilevel regression analyses to investigate the relationship between two variables over time: Examining the longitudinal association between intrusion and avoidance. *Journal of Traumatic Stress* 2009; **22**(6):622–631.
9. Pickles A, Croudace T. Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research* 2010; **19**(3):271–289.
10. Fraser A, Tilling K, Macdonald-Wallis C, Sattar N, Brion MJ, Benfield L, Ness A, Deanfield J, Hingorani A, Nelson SM, Davey SG, Lawlor DA. Association of maternal weight gain in pregnancy with offspring obesity and metabolic and vascular traits in childhood. *Circulation* 2010; **121**(23):2557–2U48.
11. Macdonald-Wallis C, Tilling K, Fraser A, Nelson SM, Lawlor DA. Established pre-eclampsia risk factors are related to patterns of blood pressure change in normal term pregnancy: Findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Journal of Hypertension* 2011; **29**(9):1703–1711.
12. Steegers EA, von DP, Duvekot JJ, Pijnenborg R. Pre-eclampsia. *Lancet* 2010; **376**(9741):631–644.
13. Lewis G (ed.). *The Confidential Enquiry into Maternal and Child Health, CEMACH. Saving Mothers' Lives: Reviewing Maternal Deaths to Make Motherhood Safer-2003-2005. The Seventh Report on Confidential Enquiries into Maternal Deaths in the United Kingdom*. Confidential Enquiry into Maternal and Child Health: London, United Kingdom, 2007.
14. Duley L. The global impact of pre-eclampsia and eclampsia. *Seminars in Perinatology* 2009; **33**(3):130–137.
15. Gaillard R, Steegers EAP, Hofman A, Jaddoe VWV. Associations of maternal obesity with blood pressure and the risks of gestational hypertensive disorders. The Generation R Study. *Journal of Hypertension* 2011; **29**(5):937–944.

16. Beyerlein A, Schiessl B, Lack N, von Kries R. Associations of gestational weight loss with birth-related outcome: a retrospective cohort study. *British Journal of Obstetrics and Gynaecology* 2011; **118**(1):55–61.
17. Viswanathan M, Siega-Riz AM, Moos MK, Deierlein A, Mumford S, Knaack J, Thieda P, Lux LJ, Lohr KN. Outcomes of maternal weight gain. *Evidence Report/Technology Assessment Number 168. AHRQ pub no: 08-E009*, Agency for Healthcare Research and Quality, 2008.
18. IOM (Institute of Medicine) and NRC (National Research Council). *Weight Gain During Pregnancy: Reexamining the Guidelines*. The National Academies Press: Washington, DC, 2009.
19. Golding J, Pembrey M, Jones R, ALSPAC-The Avon Longitudinal Study of Parents and Children - I. Study methodology. *Paediatric and Perinatal Epidemiology* 2001; **15**(1):74–87.
20. Brown MA, Lindheimer MD, de Swiet M, Van Assche A, Moutquin JM. The classification and diagnosis of the hypertensive disorders of pregnancy: Statement from the International Society for the Study of Hypertension in Pregnancy (ISSHP). *Hypertension in Pregnancy* 2001; **20**(1):9–14.
21. Fraser A, Tilling K, Macdonald-Wallis C, Hughes R, Sattar N, Nelson SM, Lawlor DA. Associations of gestational weight gain with maternal body mass index, waist circumference, and blood pressure measured 16 y after pregnancy: The Avon Longitudinal Study of Parents and Children. *American Journal of Clinical Nutrition* 2011; **93**(6):1285–1292.
22. Goldstein H. *Multilevel Statistical Models*, (2nd edn). Edward Arnold: London, 1995.
23. Robinson GK. That BLUP is a good thing: the estimation of random effects. *Statistical Science* 1991; **6**(1):15–32.
24. Tilling K, Sterne JAC, Wolfe CDA. Multilevel growth curve models with covariate effects: application to recovery after stroke. *Statistics in Medicine* 2001; **20**(5):685–704.
25. Fisher RA. Chapter 5: Tests of significance of means, differences of means, and regression coefficients. In *Statistical methods for research workers*. Oliver and Boyd: Edinburgh, 1925; 101–137.
26. Oehlert GW. A note on the Delta method. *American Statistician* 1992; **46**(1):27–29.
27. Greene WH. *Econometric Analysis*, (5th edn). Prentice Hall: Upper Saddle River, New Jersey, 2003.
28. Buis ML. Stata tip 97: Getting at rho's and sigma's. *The Stata Journal* 2011; **11**(2):315–317.
29. Jackson CH. Multi-state models for panel data: The msm package for R. *Journal of Statistical Software* 2011; **38**(8):1–28.
30. Palmer T, Macdonald-Wallis C. REFFADJUST: Stata module to estimate adjusted regression coefficients for the association between two random effects variables, 2012. (Available from: <http://ideas.repec.org/c/boc/bocode/s457403.html>). Accessed 1st March 2012.
31. Steele F. Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2008; **171**:5–19.
32. Curran PJ. Have multilevel models been structural equation models all along? *Multivariate Behavioral Research* 2003; **38**(4):529–568.
33. Bauer DJ. Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics* 2003; **28**(2):135–167.
34. Leckie G, Charlton C. runmlwin: Stata module for fitting multilevel models in the MLwiN software package, Centre for Multilevel Modelling, University of Bristol, 2011.
35. Van Minh H, Huong D, Wall S, Chuc N, Byass P. Multilevel analysis of covariation in socioeconomic predictors of physical functioning and psychological well-being among older people in rural Vietnam. *BMC Geriatrics* 2010; **10**(7):1–7.
36. Burkholder GJ, Harlow LL. An illustration of a longitudinal cross-lagged design for larger structural equation models. *Structural Equation Modeling* 2003; **10**(3):465–486.
37. Liu H, Powers DA. Growth curve models for zero-inflated count data: An application to smoking behavior. *Structural Equation Modeling* 2007; **14**(2):247–279.
38. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 1969; **37**(3):424–438.
39. Dormann C. Modeling unmeasured third variables in longitudinal studies. *Structural Equation Modeling* 2001; **8**(4):575–598.
40. Little R, Rubin D. *Statistical Analysis with Missing Data*. Wiley: New York, 2002.
41. Mahabadi SE, Ganjali M. An index of local sensitivity to non-ignorability for multivariate longitudinal mixed data with potential non-random dropout. *Statistics in Medicine* 2010; **29**(17):1779–1792.