Appendix S1: Patterns of second- and third-trimester growth and discordance in twin pregnancy: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort

O. T. Stirrup, A. Khalil, F. D'Antonio, and B. Thilaganathan,
on behalf of the Southwest Thames Obstetric Research Collaborative (STORK)

1 Statistical methods

1.1 Multilevel modelling

The nested nature of the sets of measurements obtained for each fetus and pairs of twins within each pregnancy means that the analysis of fetal growth in twin pregnancies is particularly suited to the application of multilevel modelling. The term 'multilevel modelling' describes a mixed effects approach in which sets of measurements within each individual and/or individuals within groups form a nested structure. Multilevel modelling is therefore an extension of the general linear regression framework that involves the fitting of a combination of 'fixed effects' and 'random effects' to different hierarchical levels [1;2]. The fixed effects provide the expected mean value for any given combination of predictive variables, whilst the random effects are used to model the random variation associated with each level of the analysis and the covariance between different measurements obtained from the same individual or between individuals within the same group. Analysis of this nature is required in order to create adequate statistical models that take account of the dependency within data that results from obtaining serial measurements from individuals or the clustering of individuals within distinct groups.

The simplest example of a multilevel model is a two-level 'random intercept' model. In the context of analysis of longitudinal data, this could result when serial measurements are obtained from individuals and the measurement at each time point is dependent on age, with an overall mean intercept (β_0) and slope (β_1) for the relationship between the measurement and age and random components associated with both each separate measurement occasion (ϵ_{ij}) and each individual overall (u_{0j}). For the value (y_{ij}) of the i^{th} measurement occasion within the j^{th} individual at age (t_{ij}), this can be written:

$$y_{ij} = \beta_0 + u_{0j} + \beta_1 t_{ij} + \epsilon_{ij}$$

In the standard multilevel model, the random components are normally distributed and independent between levels and so:

$$\begin{aligned} Y_{ij} &= \beta_0 + U_{0j} + \beta_1 t_{ij} + E_{ij} \\ U_{0j} &\sim N(0, \tau^2) \\ E_{ij} &\sim N(0, \sigma^2) \end{aligned}$$

The randomly distributed values of u_{0j} shift the intercept for the relationship between measurements and age for each individual, hence the term 'random intercept' model. The realizations of U_{0j} specific to each individual are not estimated in the process of fitting a multilevel model, but the variance of this term (τ^2) is estimated and the resultant dependency between measurements within each individual is accounted for.

The next logical extension of the 'random intercept' model is a 'random slope' model, in which the relationship between the predictive and outcome variables is also allowed to randomly vary between individuals and/or groups. In the example outlined above, this would introduce a new random term (u_{1j}) representing variation between individuals in the relationship between the measurement and age:

$$y_{ij} = \beta_0 + u_{0j} + (\beta_1 + u_{1j}) t_{ij} + \epsilon_{ij}$$

The random terms at any given level of the model are not assumed to be independent as, for example, smaller initial size may be associated with greater subsequent growth. The possible correlation between the random slope term and the random intercept term for each individual means that they are modelled as following a multivariate normal (MVN) distribution, with covariance parameter τ_{01}:

$$\begin{aligned} Y_{ij} &= \beta_0 + U_{0j} + (\beta_1 + U_{1j}) t_{ij} + E_{ij} \\ \begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} &\sim MVN(\mathbf{0}, \boldsymbol{\Omega_u}), \text{ where } \boldsymbol{\Omega_u} = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \\ E_{ij} &\sim N(0, \sigma^2) \end{aligned}$$

Similar models can be constructed using higher order polynomial functions for the relationship between the dependent variable and age, or that take into account fixed and/or random effects associated with other variables, with a resultant increase in the number of parameters associated with the fixed and random parts of the model[3]. A greater number of hierarchical levels can also be included. In the context of modelling growth in twin pregnancies, the 'fetus' level (level 2) of the model can be considered to be nested within a 'pregnancy' level (level 3); this higher level is associated with random effects shared by the two fetuses within the pregnancy, which follow a MVN distribution independent of the other levels. This structure is illustrated in Figure 1a.

A model for the i^{th} measurement occasion of the j^{th} fetus within the k^{th} pregnancy (Y_{ijk}) could be expressed as:

$$
\begin{aligned}
&Y_{ijk} = \beta_0 + U_{0jk} + V_{0k} + (\beta_1 + U_{1jk} + V_{1k})\, t_{ijk} + E_{ijk} \qquad (1)\\
&\begin{pmatrix} V_{0k} \\ V_{1k} \end{pmatrix} \sim MVN(\mathbf{0}, \boldsymbol{\Omega_v}), \text{ where } \boldsymbol{\Omega_v} = \begin{pmatrix} \varphi_{00} & \varphi_{01} \\ \varphi_{01} & \varphi_{11} \end{pmatrix}\\
&\begin{pmatrix} U_{0jk} \\ U_{1jk} \end{pmatrix} \sim MVN(\mathbf{0}, \boldsymbol{\Omega_u}), \text{ where } \boldsymbol{\Omega_u} = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\\
&E_{ijk} \sim N(0, \sigma^2)
\end{aligned}
$$

Here, V_{0k} and V_{1k} represent the random variation in slope and intercept common to the two twins in each pregnancy, and U_{0jk} and U_{1jk} represent that specific to each individual fetus. The level 1 error term (ϵ_{ijk}) is modelled as independent between measurement occasions and between the two twins assessed at any one examination. The variance of this term can be attributed to some combination of measurement error and short-term variation of the growth trajectory of each fetus from that modelled by the polynomial growth function (including random variations in the coefficients specific to that fetus). For both of these explanations for the level 1 variance, it seems reasonable to think that the variation for the two twins at any one measurement occasion might in fact be correlated rather than independent. Within the framework of multilevel modelling, this can be investigated by considering the measurements for the two twins in each pregnancy as a multivariate response at each examination. This involves replacing level 1 with a response indicator for which twin a measurement relates to and defining level 2 instead as the measurement occasion (Figure 1b)[4].

This 'multivariate multilevel model' for growth in twin pregnancies requires the measurements obtained from 'Twin A' (y_{Ajk}) and 'Twin B' (y_{Bjk}) to be modelled as distinct outcome variables. Again considering only a linear growth function with random intercepts and slopes for simplicity, the measurements obtained at the j^{th} measurement occasion of the k^{th} pregnancy can be expressed as:

$$
\begin{aligned}
&Y_{Ajk} = \beta_0 + U_{0Ak} + (\beta_1 + U_{1Ak})\, t_{jk} + E_{Ajk} \qquad (2)\\
&Y_{Bjk} = \beta_0 + U_{0Bk} + (\beta_1 + U_{1Bk})\, t_{jk} + E_{Bjk}\\
&\begin{pmatrix} U_{0Ak} \\ U_{1Ak} \\ U_{0Bk} \\ U_{1Bk} \end{pmatrix} \sim MVN(\mathbf{0}, \boldsymbol{\Psi}), \text{ where } \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Omega_u} + \boldsymbol{\Omega_v} & \boldsymbol{\Omega_v} \\ \boldsymbol{\Omega_v} & \boldsymbol{\Omega_u} + \boldsymbol{\Omega_v} \end{pmatrix}\\
&\begin{pmatrix} E_{Ajk} \\ E_{Bjk} \end{pmatrix} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}), \text{ where } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma^2 \end{pmatrix}
\end{aligned}
$$

The structure for $\boldsymbol{\Psi}$ in this model (2) is that same as that implied by the 'nested model' (1) for the random coefficients of the polynomial growth functions for the two twins within each pregnancy[5]. However, the covariance of the residual error for the two fetuses at each examination is now included as σ_{AB}. In this model, fetus-level random effects (represented by $\boldsymbol{\Omega_u} + \boldsymbol{\Omega_v}$, inclusive of pregnancy-level variation) and

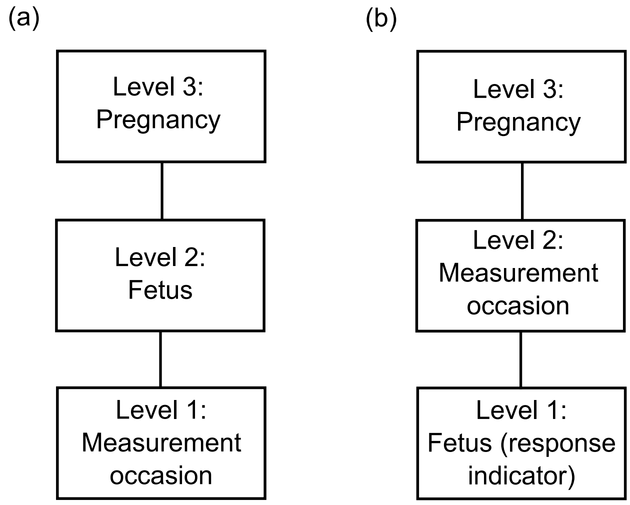

Figure 1. Diagrams showing possible ordering of hierarchy for multilevel modelling of serial measurements in twins: (a) measurement occasions are nested within individual fetuses, which are in turn nested within pregnancies; (b) measurements for the two twins in each pregnancy can instead be modelled as a multivariate response at each measurement occasion, in this situation 'level 1' is a response indicator for which twin a measurement relates to and the random variation at each examination is modelled at 'level 2'.

the within-pregnancy covariance for these (i.e. $\boldsymbol{\Omega}_v$) are both included at level 3, relating to variation in the growth trajectory for each fetus. The residual variance specific to each examination is modelled at level 2. As for any such model, the random effects defined serve to create the structure and parameterization of the marginal covariance matrix for the vector of the dependent variable as a whole.

1.2 Unbalanced and missing data

An advantage of using the multilevel modelling approach (or mixed effects models in general) to analyze longitudinal data is that it is capable of handling datasets in which the number and timing of measurement occasions vary between individuals. This is clearly important with regards to the present analysis, in which the measurements have been recorded at irregular points between 14 and 40 weeks of gestation and the number of examinations recorded per pregnancy ranges between 1 and 14. There is therefore no clear definition of what a complete set of data for any individual pregnancy should be.

The discussion of missing data in statistical modelling generally follows the terminology defined by Rubin[6], with three types of 'missingness': *missing completely at random*, where missingness is independent of the complete data; *missing at random*, where missingness is conditionally independent of the unobserved data given the observed data; and *missing not at random*, for which the missingness, conditionally given the observed data, is dependent on the unobserved data. In the present analysis, it is not safe to assume that data at later gestational ages in some pregnancies are *missing completely at random.* However, *missing at random* is a more reasonable assumption, under which inferences will be correct if maximum likelihood estimation is carried out using all available data and a correctly specified model.

1.3 Statistical software

There are numerous statistical software programs available that are capable of fitting 'mixed effects' or multilevel models; these differ in the estimation procedures employed, the model structures available and the range of additional features for assessing model fit and interpretation. MLwiN (Version 2.28; Centre for Multilevel Modelling, University of Bristol, Bristol, UK), which has been specifically developed for the purpose of multilevel modelling, was used in this analysis. This software has a number of advantageous features, including the generation of residual plots at each level of a fitted model, tools for investigating outlying data points and excluding them from further analysis, and the estimation of confidence intervals for fitted means with respect to a continuous variable (i.e. time/age).

MLwiN has been developed particularly for hierarchically structured data, and as such fitting a model with nested levels as illustrated by Figure 1a (Page 4) is relatively straightforward. Accounting for the potential for correlation between the lowest level random terms for each fetus at each measurement occasion is more problematic, but can be achieved by defining a multivariate multilevel model as shown by Figure 1b. The way in which this modelling approach requires a restructuring of the dataset, largely performed automatically by the MLwiN software. For this scenario, as two distinct outcome variables are effectively being defined, it is necessary to apply parameter constraints to specify that the fixed effects for 'Twin A' are equal to those for 'Twin B' in each pregnancy and that, for the random effects (as described on Page 3), '$\boldsymbol{\Omega_u}+\boldsymbol{\Omega_v}$' is equal for the two twins in each pregnancy and $\boldsymbol{\Omega_v}$ is symmetric. Parameter constraints are included in the estimation procedure, including adjustment of standard errors, as described by Goldstein[7]. Although we have chosen to use the notation '$\boldsymbol{\Omega_u}+\boldsymbol{\Omega_v}$' for the covariance matrix of fetus-level random effects, because conceptually this is inclusive of pregnancy-level variation, the components of this matrix are estimated directly in the model described.

To ensure that any possible systematic difference would not interfere with the estimation procedure, the labelling of 'Twin A' and 'Twin B' in each pregnancy was randomized using the 'experiment' package in R (R Foundation, Vienna, Austria). It would also be possible to include the desired covariance term specific to each measurement occasion through the use of a cross-classified (with respect to the nested structure) random effect, but this is not readily implemented in MLwiN.

Unlike standard linear regression models, the parameter estimates when fitting multilevel models do not have a 'closed form' solution that can be easily calculated. Instead, an iterative procedure is required that attempts to find the optimal solution through repetition of an algorithm until convergence is achieved. It is important to note that convergence is not always reached, particularly if a model is misspecified or too complex for the data available. In MLwiN, the procedure used is iterative generalized least squares (IGLS) estimation[1], which provides maximum likelihood estimates for normal response models.

Maximum likelihood estimates for mixed effects models can be seriously biased for small sample sizes, but this is not a problem in the present analysis given the large size of the dataset. For smaller datasets the use of restricted maximum likelihood (REML) procedures provides less biased estimates, and this can be implemented in MLwiN using restricted iterative generalized least squares (RIGLS) estimation. However, for REML procedures estimation is based on a likelihood function that is independent of the fixed effects in the model, meaning that likelihood ratio tests cannot be used to compare models that differ with respect to the fixed effects included. Because of this, IGLS estimation will be used to allow statistical comparison of models with differences in both the random and fixed effects components.

All statistical graphics were generated using the 'ggplot2' package in R.

1.4 Model fitting procedure

For each model, the polynomial growth function in terms of gestational age was determined first. Gestational age was expressed in weeks and was centred at 14 weeks, meaning that intercept terms can be interpreted as representing variation at this point in pregnancy. The process began with a 'random intercept' model for linear growth, with fetus-level random terms for the intercept that were correlated within each pregnancy, and a 'random slopes' model was fitted next with additional random effect terms for the slope of the relationship between the variable under consideration and gestational age. Following this, additional fixed effect and fetus-level random effects terms (with the latter again correlated within each pregnancy) were added simultaneously to the model to create incrementally higher order polynomial growth functions. It would have been possible to add higher order terms only to the fixed effects part of the model, but this would constrain the curvature to be identical across individuals, which has been described as "... a constraint that seems antithetical to the model-building exercise..."[8].

A maximum likelihood-based estimation procedure (i.e. IGLS) was used in order to allow likelihood ratio tests of statistical significance for the simultaneous addition of fixed effect and random effects terms to a model. The likelihood ratio statistic is twice the difference in log-likelihood ($\Delta 2\ell$) between the models being compared, and is compared to a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters (Δp) between the two models. Following the parameterization for a 'multivariate multilevel model' for twins (as described on Page 3), and with parameter constraints used to achieve this, extending a model to the n^{th} level of a polynomial growth function involves the addition of a single fixed effect parameter and $2(n{+}1)$ variance and covariance parameters for the random effects at each stage.

For tests of additional random effects terms involving covariance parameters, it can be shown that the likelihood ratio test is conservative as it does not take account of the bounded parameter space of any covariance terms. An adjustment to the standard likelihood ratio test to ensure optimal power is available when only random effects terms are being considered[9], but it is not possible to apply this when also considering an additional fixed effect term in the same comparison. However, in the present analysis this minor loss of power to identify more complex models could be safely accepted as, given the large dataset, any terms of borderline statistical significance are likely to be of no practical importance.

Once an optimal polynomial growth model had been fitted, the fit of the model was assessed through diagnostic standardized residual plots for each level. A function for the measurement occasion-specific variance in terms of gestational age was considered if this appeared to be non-constant. Following this, residual plots were again reviewed and for any extreme outlying points (i.e. $z \geq |6|$) the data for that pregnancy were excluded before refitting the model.

2 Structure of fitted models

In both DCDA and MCDA twin pregnancies, following the initial 'random intercept' model, highly statistically significant improvements in model fit were achieved by a 'random slope' model and subsequently a quadratic model for the relationship with gestational age. The latter model included both a quadratic 'fixed effects' function for the mean and quadratic fetus-level 'random effects' in terms of gestational age in each case, with these correlated between Twin A and Twin B in each pregnancy. Model-fitting was attempted including cubic components for both the fixed and random effects, but convergence was not achieved.

For fitted quadratic models, plotting of the examination-specific standardized residuals against the fixed part prediction (wholly dependent on gestational age) revealed a clear increase in the absolute magnitude of the residuals with increasing predicted mean. As such, the examination-specific variance was modelled as a linear function of gestational age, achieving a significant improvement in model fit.

The structure of the fitted model can be expressed in the following form, in which Y_{Ajk} and Y_{Bjk} represent the measurements for 'Twin A' and 'Twin B' in the k$^{\text{th}}$ pregnancy at j$^{\text{th}}$ time point t_{jk} (in weeks centred at 14), the β-coefficients represent the fixed effects estimates for the overall relationship with gestational age, $\boldsymbol{\Omega_u} + \boldsymbol{\Omega_v}$ represents the fetus-level covariance matrix of random effects (inclusive of pregnancy-level variation), $\boldsymbol{\Omega_v}$ represents the between-twin covariance matrix (interpretable as equivalent to a pregnancy-level random effects matrix) and $\boldsymbol{\Sigma}_t$ is the covariance matrix for examination-specific random effects:

$$\begin{aligned} Y_{Ajk} &= \beta_0 + U_{0Ak} + (\beta_1 + U_{1Ak})\, t_{jk} + (\beta_2 + U_{2Ak})\, t_{jk}^2 + E_{Ajk} \\ Y_{Bjk} &= \beta_0 + U_{0Bk} + (\beta_1 + U_{1Bk})\, t_{jk} + (\beta_2 + U_{2Bk})\, t_{jk}^2 + E_{Bjk} \end{aligned} \tag{3}$$

$$\begin{pmatrix} U_{0Ak} \\ U_{1Ak} \\ U_{2Ak} \\ U_{0Bk} \\ U_{1Bk} \\ U_{2Bk} \end{pmatrix} \sim MVN(\mathbf{0}, \mathbf{\Psi}), \text{ where } \mathbf{\Psi} = \begin{pmatrix} \mathbf{\Omega_u} + \mathbf{\Omega_v} & \mathbf{\Omega_v} \\ \mathbf{\Omega_v} & \mathbf{\Omega_u} + \mathbf{\Omega_v} \end{pmatrix}$$

$$\begin{pmatrix} E_{Ajk} \\ E_{Bjk} \end{pmatrix} \sim MVN(\mathbf{0}, \mathbf{\Sigma}_t), \text{ where } \mathbf{\Sigma_t} = \begin{pmatrix} \sigma_t^2 & \sigma_{AB\_t} \\ \sigma_{AB\_t} & \sigma_t^2 \end{pmatrix}$$

3 Parameter estimates for fitted models

The table below give a summary of parameter estimates for the final fitted models for DCDA and MCDA pregnancies. Following earlier notation:

$$\mathbf{\Omega_v} = \begin{pmatrix} \varphi_{00} & \varphi_{01} & \varphi_{02} \\ \varphi_{01} & \varphi_{11} & \varphi_{12} \\ \varphi_{02} & \varphi_{12} & \varphi_{22} \end{pmatrix} \qquad \mathbf{\Omega_u} + \mathbf{\Omega_v} = \begin{pmatrix} \tau_{00}+\varphi_{00} & \tau_{01}+\varphi_{01} & \tau_{02}+\varphi_{02} \\ \tau_{01}+\varphi_{01} & \tau_{11}+\varphi_{11} & \tau_{12}+\varphi_{12} \\ \tau_{02}+\varphi_{02} & \tau_{12}+\varphi_{12} & \tau_{22}+\varphi_{22} \end{pmatrix}$$

$$\sigma_t^2 = \sigma_{00} + 2 * \sigma_{01} * t \qquad \sigma_{AB\_t} = \sigma_{AB\_00} + 2 * \sigma_{AB\_01} * t$$

Although we have retained the notation $\mathbf{\Omega_u} + \mathbf{\Omega_v}$, because conceptually this represents the sum of fetus-level and pregnancy-level variation, the terms that comprise this matrix have been estimated directly by MLwiN. Estimated standard errors are given in parentheses.

3.1 AC model parameters

	DCDA	**MCDA**
Fixed		
β_0	77.025 (0.286)	76.661 (0.438)
β_1	12.562 (0.054)	12.191 (0.098)
β_2	-0.093 (0.002)	-0.081 (0.005)
Random		
Level 3		
$\tau_{00}+\varphi_{00}$	29.911 (4.438)	41.044 (5.896)
$\tau_{01}+\varphi_{01}$	-2.039 (0.776)	-2.958 (1.085)
$\tau_{11}+\varphi_{11}$	1.058 (0.171)	1.324 (0.265)
$\tau_{02}+\varphi_{02}$	0.095 (0.032)	0.173 (0.050)
$\tau_{12}+\varphi_{12}$	-0.040 (0.007)	-0.056 (0.012)
$\tau_{22}+\varphi_{22}$	0.002 (0.000)	0.003 (0.001)
φ_{00}	8.670 (4.124)	6.638 (5.675)
φ_{01}	-0.119 (0.751)	-0.328 (1.062)
φ_{11}	0.405 (0.17)	0.610 (0.264)
φ_{02}	0.029 (0.032)	0.007 (0.050)
φ_{12}	-0.018 (0.007)	-0.024 (0.012)
φ_{22}	0.001 (0.000)	0.001 (0.001)
Level 2		
σ_{00}	11.575 (1.385)	14.648 (1.858)
σ_{AB_00}	9.625 (0.731)	7.424 (1.085)
σ_{01}	1.359 (0.056)	1.102 (0.093)
σ_{AB_01}	0 (0)	0 (0)

References

[1] Goldstein H. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73:43–56, 1986.

[2] Snijders TAB and Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling; Second Edition*, chapter 5 The Hierarchical Linear Model. Sage, 2011.

[3] Singer JD and Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, chapter 5 Treating TIME more flexibly. OUP USA, 2003.

[4] Rasbash J, Steele F, Browne WJ, and Goldstein H. *A User's Guide to MLwiN, v2.26*, chapter 14 Multivariate Response Models. Centre for Multilevel Modelling, University of Bristol, 2012.

[5] Raudenbush SW. *Handbook of Multilevel Analysis [Eds: de Leeuw J and Meijer E]*, chapter 5 Many Small Groups, page 217. Springer, 2008.

[6] Rubin DB. Inference and missing data. *Biometrika*, 63:581–592, 1976.

[7] Goldstein H. *Multilevel Statistical Models*, chapter 3.3 Parameter Constraints. Arnold, 1999.

[8] Singer JD and Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, chapter 6 Modeling Discontinuous and Nonlinear Change, page 222. OUP USA, 2003.

[9] Stram DO and Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50:1171–1177, 1994.